University of Cincinnati

Date: 2/26/2015

I, Jennifer Chubinski, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Regional Development Planning.

It is entitled:
Relationships Between Neighborhoods, Housing, and Health Outcomes: A Multilevel Analysis of a Midwestern County

Student’s name: Jennifer Chubinski

This work and its defense approved by:

Committee chair: Rainer vom Hofe, Ph.D.

Committee member: Christopher Auffrey, Ph.D.

Committee member: Eric Rademacher, Ph.D.
Relationships Between Neighborhoods, Housing, and Health Outcomes: A Multilevel Analysis of a Midwestern County

A dissertation submitted to the Graduate School of the University of Cincinnati in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Planning of the College of Design, Architecture, Art, and Planning 2015

by Jennifer Chubinski

M.S.P.P.M. Carnegie Mellon University

Committee Chair: Rainer vom Hofe, Ph.D.
Abstract

Being healthy is a critical piece of having a high quality of life. Improving quality of life has long been a cornerstone of the fields of both regional planning and public health. While overall life expectancies have increased over recent decades, progress toward improved health and quality of life seems elusive in U.S. communities. Many U.S. communities have seen skyrocketing rates of obesity, increasing regional sprawl, and persistent disparities in health and environmental outcomes. Over the last decade, there has been exponential growth in the research that investigates the relationship between the built environment and health outcomes, but there is still much work to do to understand the link between health and the built environment (Frank, Sallis, Conway, Chapman, Saelens, & Bachman, 2006; Dannenberg, Frumkin, & Jackson, 2011). This increased interest in how our built environment influences our health stems from the assumption that improvements to the built environment could, with limited individual effort, improve the health of a large number of people (see discussion of Health Impact Pyramid in Section 1.3).

The focus of this study is on whether modeling housing quality proves to have additive explanatory power on self-reported health after controlling for other powerful drivers of health (poverty, unemployment, chronic conditions, health behaviors, etc.). This study separated the individual versus the community-level characteristics and their unique contributions to self-reported health outcomes, with a particular focus on housing variables. Housing quality was modeled using the neighborhood variable percent renter, and socioeconomic conditions of the neighborhood were modeled using mobility (percent of homeowners who moved to the neighborhood in the last five years).

The results of this study are in line with much of the literature on what contributes to improved self-reported health. The Level-1 factors (individual health characteristics) are more influential on the model (94.5% of the influence compared to 5.5% for neighborhood-level
Relationships Between Neighborhoods, Housing, Health

factors). This study found that chronic health conditions, race, smoking, physical activity, social support, employment status, and poverty status contribute to changes in self-reported health. At the neighborhood level, percent mobility and percent renter proved to be significant variables. Percent renter is one of the model’s housing variables suggesting that housing conditions in a neighborhood influence self-reported health. Percent mobility was also significant and was one of the model’s socioeconomic environment variables suggesting that socioeconomic conditions in the neighborhood influence self-reported health; percent mobility, however, did not influence self-reported health in the way anticipated. While Model 3 (including neighborhood variables) is not significantly better than Model 2 (with only individual variables), the individual significance of the parameter estimates ($\beta$) suggests that percent mobility and percent renter are important environmental characteristics to consider when trying to model self-reported health.

The results from this study suggest that health policy changes that target changes to the built environment in order to influence health improvements should expect only small improvements in individual self-reported health at a point-in-time. However, it is not known how the small point-in-time improvements could cumulatively lead to larger improvements in self-reported health over time.
Acknowledgements

I would like to thank my dissertation committee for their advice, assistance and encouragement through various stages of this study. Thank you for this unique opportunity to grow as a researcher. In particular, I would like to thank Rainer vom Hofe, the chair of this committee; a scholar whose drive for high-quality research and thinking is admirable. This study is much better because of your influence. I am particularly thankful for your personal support throughout my whole degree program; you never once doubted that a mother of three young children with a full-time job could successfully complete a doctoral program. I am equally thankful to Eric Rademacher, colleague and mentor; your support has been tremendous. Thank you for your help in navigating the complicated process that is a doctoral dissertation. My sincere thanks to Chris Auffrey, who through sharing his research interests helped me to find my own.

As with any project of this magnitude, I found inspiration and support from many people. Adam Carle and Aline Sayer, two experts in multilevel modeling, and two of the most positive, optimistic people I have ever met. Their tough methodological questions were always followed with encouragement and support for the study. Mark Carrozza, colleague, friend, and stellar applied methodologist, thank you for letting me pick your brain and borrow extensively from your library. Beth Walter Honadle, who helped me see that I needed a doctoral degree to continue to progress in the profession that I love. And Gail Myers whose editing support was invaluable.

Special thanks to the many friends and colleagues, in particular at Interact for Health, who have filled my sails with wind when I needed it most. Two friends and colleagues, in particular, stand out for their encouragement and support at critical moments: Sarah Walsh and Judy Harmony.

To my children Alexander, Elisabetta, and Giulia, you helped me keep my priorities straight in all the right moments. It was not time to write, but carve pumpkins, or collect flowers
or go for a bike ride. Thank you for the infinite joy you have brought to my life. And of course Sergio, my husband and biggest supporter – probably the only person more impatient than me about finishing this degree. Thank you for lifting me up when I was down and for leaving me alone when I was being productive. Hopefully you will remember what life is like when you have a wife again.
Table of Contents

List of Figures ........................................................................................................... ix
List of Tables ........................................................................................................... x

Chapter 1. Introduction

1.1 Introduction to the study .................................................................................. 1
1.2 Definition of terms and concepts used ............................................................ 2
1.3 Research objectives ......................................................................................... 4
  1.3.1 Theoretical construct ............................................................................... 4
1.4 Review of the study’s contributions ............................................................... 8
1.5 Assumptions and limitations of the study ....................................................... 9
1.6 Dissertation organization ............................................................................. 10

Chapter 2. Literature Review

2.1 The intersection of the fields of planning and public health ......................... 11
2.2 Relationships between neighborhoods, poverty, housing and health outcomes .................................................................................................................. 16
2.3 Self-reported health ....................................................................................... 23
2.4 Health measurement in a multilevel model ................................................... 28

Chapter 3. Data and Measures

3.1 Description of individual (Level-1) health data: The Greater Cincinnati Community Health Status Survey .......................................................................................... 32
3.2 Description of community data ..................................................................... 35
3.3 Determining geographic boundaries .............................................................. 37
3.4 Measures ........................................................................................................... 41
  3.4.1 Outcome variable: self-reported health status ........................................ 41
  3.4.2 Individual (Level-1) predictors ............................................................... 42
  3.4.3 Neighborhood (Level-2) predictors ........................................................ 48

Chapter 4. Methods

4.1 Introduction to multilevel modeling ............................................................... 51
4.2 Description of a general multilevel model ..................................................... 55
4.3 Centering ......................................................................................................... 59
4.4 Parameter estimation ...................................................................................... 60
  4.4.1 Level-1 coefficients, $\beta_{qj}$ .................................................................. 60
  4.4.2 Level-2 coefficients, $\gamma_{qs}$ ................................................................ 62
  4.4.3 Maximum likelihood estimates of variance and covariance components, $\sigma^2$ (Level-1) and $T$ (Level-2) .................................................... 65
4.5 Analytical plan ................................................................................................. 67
## Chapter 5. Findings and Discussion

5.1 Methodological Summary ................................................................. 69
  5.1.1 Model 1: Unconditional model .................................................. 69
  5.1.2 Model 2: Level-1 model/One-way ANCOVA with random effects .......... 71
  5.1.3 Model 3: Final model/Extension of the random-effects ANCOVA with Level-2 covariates ................................................................. 74
5.2 Model assumptions ........................................................................... 77
  5.2.1 Adequacy of sample sizes and missing data .................................... 77
  5.2.2 Normality of distributions at Level-1 and Level-2 ........................... 78
  5.2.3 Absence of multicollinearity ......................................................... 80
  5.2.4 Independence of errors ............................................................... 80
5.3 Results ...................................................................................... 81
5.4 Interpretation and discussion of results ............................................. 83
  5.4.1 Interpretation of Model 1 results ................................................ 84
  5.4.2 Interpretation of model coefficients based on variable type ............... 85
  5.4.3 Interpretation of Model 2 results ................................................ 88
  5.4.4 Interpretation of Model 3 results ................................................ 90
  5.4.5 Comparing the influence of each model variable .......................... 94
  5.4.6 Comparing Models 2 and 3 ......................................................... 97

## Chapter 6. Conclusions and Future Research

6.1 Conclusions ........................................................................ 98
6.2 Relationship of results to theory .................................................... 101
6.3 Policy implications ................................................................... 108
6.4 Research contributions ............................................................... 113
6.5 Research limitations and future research ...................................... 113

References .................................................................................. 120

Appendices

Appendix A – Detailed list of geographic boundaries ............................... 137
Appendix B – An analysis of histograms for select variables ................... 142
List of Figures

Figure 1.1 The Health Impact Pyramid .................................................................6
Figure 1.2 Relationships of interest for this study ..................................................7
Figure 3.1 Greater Cincinnati Community Health Status Survey Region.................33
Figure 3.2 Geographic aggregation of census tracts..............................................40
Figure 4.1 Nested data model – Adults clustered within communities ....................52
Figure 5.1 Q-Q plot of Chi-square distribution values by Mahalanobis distance .......79
Figure 5.2 Comparing a low and high mobility neighborhood ..............................93
Figure 6.1 The Health Impact Pyramid .................................................................102
Figure 6.2 Reflections on the relationships of interest for this study ......................106
List of Tables

Table 2.1 Models for measuring internal housing quality ........................................19
Table 3.1 Response counts for study sample ..........................................................35
Table 3.2 Source for community level data .........................................................36
Table 3.3 Community-level measures of socioeconomic conditions and housing environment .................................................................36
Table 3.4 Major steps in geographic boundary determination ................................38
Table 3.5 Descriptive statistics for continuous variables in CHSS data ................42
Table 3.6 Descriptive statistics for neighborhood predictors ...............................49
Table 4.1 Variables in multilevel model ...............................................................53
Table 4.2 Types of multilevel models used in this study ......................................67
Table 5.1 Results ..................................................................................................82
Table 5.2 Variable type .......................................................................................85
Table 5.3 Percent change in self-reported health for a one-point change in each variable .................................................................95
1.1 Introduction to the study

Being healthy is a critical piece of having a high quality of life. Improving quality of life for citizens has long been a cornerstone of both the fields of regional planning and public health. While overall life expectancies have increased over recent decades, progress toward improved health and quality of life seems elusive in U.S. communities. Many U.S. communities have seen skyrocketing rates of obesity, increasing regional sprawl, and persistently wide disparities in health and environmental outcomes between the “haves” and the “have nots.” Over the last decade, there has been exponential growth in the research that investigates the relationship between the built environment and health outcomes, but there is still much work to do to understand the link between health and the built environment (Curtis, Cave, & Coutts, 2002; Doyle et al., 2006; Frank et al., 2006; Dannenberg, Frumkin, & Jackson, 2011). This increased interest in how our built environment influences our health stems from the assumption that improvements to the built environment could, with limited individual effort, improve the health of a large number of people (see discussion of Health Impact Pyramid in section 1.3).

While the link between the built environment and health outcomes is still a relatively new field of research; the link between poorer health and higher poverty status has been well established (Wilkinson & Marmot, 2003; Wen, Browning, & Cagney, 2003; Patel, Eschback, Rudkin, Peek, & Markides, 2003; Galea, Ahern, Rudenstine, Wallace, & Vlahov, 2005). This includes a clear link between individual poverty and poorer health outcomes and between living in a community with very low median income and poorer health (Wilkinson & Marmot, 2003; Wen, Browning, & Cagney, 2003; Patel et al., 2003; and Galea et al., 2005). Poverty and poor self-reported health are strongly correlated; however, relatively little research has been done on the added influence of the housing environment on health status. This study is particularly interested in whether modeling housing quality proves to have additive explanatory power on
self-reported health after controlling for other powerful individual and community drivers of health (poverty, unemployment, chronic conditions, health behaviors, etc.).

1.2 Definitions of terms and concepts used

Before proceeding, there are a number of key concepts to define: health, public health, regional planning, the built environment, the social determinants of health, and self-reported health. For the purposes of this study, it is critical to articulate a complete working definition of each.

**Health:** According to the Constitution of the World Health Organization (WHO):

*Health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity. The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being, without distinction of race, religion, political belief, economic or social condition* (2006).

This definition has not changed since 1948. WHO’s definition is important, because it recognizes that health goes far beyond what happens within medical facilities; that health starts in people’s communities, workplaces, and homes. And, most important, that health encompasses mental, emotional, and social health.

**Public Health:** According to the American Public Health Association (APHA) public health is, “...the practice of preventing disease and promoting good health within groups of people, from small communities to entire countries” (“What is Public Health,” n.d.). Populations can be large, like a whole country, or small, like a small village, but one of the primary aims of public health is to promote healthy behaviors, healthy communities, and healthy environments.

**Regional Planning:** According to the American Planning Association (APA) planning “...works to improve the welfare of people and their communities by creating more convenient, equitable, healthful, efficient, and attractive places for present and future generations” (“What is Planning,” n.d.). Planning is the study of efficient land use and creation of the built environment. The built environment includes: parks, homes, schools, sidewalks, and roads,
Relationships Between Neighborhoods, Housing, Health

among many other things (Sallis, 2009). Regional planning studies broad systematic changes in larger land areas, for example multiple cities, villages, or counties.

The fields of public health and regional planning share in common the goal to promote healthy environments and improve citizens' quality of life.

It is also important to include a definition for the social determinants of health. This is an area of interest that has grown out of public health research, but which is important for regional planning.

Social Determinants of Health: According to WHO, the social determinants of health are:

...the conditions in which people are born, grow, live, work, and age, including the health system. These circumstances are shaped by the distribution of money, power, and resources at global, national, and local levels. The social determinants of health are mostly responsible for health inequities – the unfair and avoidable differences in health status seen within and between countries. (Social determinants of health, n.d.)

The social determinants of health is the name that the field of public health has given to the study of inequality, but the concept of working to resolve inequality is evident in regional planning as well. In their book on the social determinants of health, Wilkinson and Marmot discuss that improving the social and economic conditions that make people ill is more important for population-level health improvement than increasing access to medical care (2003).

Self-reported health is the answer to the question: “In general how would you rate your health, excellent, very good, good, fair, poor?” and is the outcome variable for the model for this study. Self-reported health has a very robust and consistent research history of reliably capturing morbidity and mortality across the world and among many different demographic groups (Brunner, 2006; Dowd & Zajacova, 2007; John & Montgomery, 2012; McFadden, Luben, Bingham, Wareham, Kinmouth, & Khaw, 2009; Regidor, Guallar-Castillón, Gutiérrez-Fisac, Banegas, & Rodríguez-Artalejo, 2010; Singh-Manoux, Guéguen, Martikainen, Ferrie, Marmot, & Shipley, 2007; Eriksson, Undén & Elofsson, 2001; and Undén & Elofsson, 2006). Research
suggests that respondents consider mental, physical, and emotional health. The question is simple to understand and easy to implement.

While it is clear that the fields of public health and regional planning have overlapping goals, improving quality of life in the communities they serve, the two fields have not always worked in coordinated ways. This will be discussed further in the following chapter.

### 1.3 Research objectives

#### 1.3.1 Theoretical construct

“Life expectancy is shorter and most diseases are more common further down the social ladder in each society. Health policy must tackle the social and economic determinants of health” (Wilkinson & Marmot, 2003, p. 10). Wilkinson and Marmot forcefully point out in their 2003 book for the WHO that the disparity in life expectancy between the poor and the rich, “...has drawn attention to the remarkable sensitivity of health to the social environment” (p. 1).

This thinking from Wilkinson and Marmot, along with Frieden’s public health impact pyramid (referred to in the Introduction), discussed below, both identify understanding the socioeconomic influences of environment on health outcomes as critical work to continuing to improve the health of the general population, in particular those that are most vulnerable.

Part of the theoretical construct for this study is Tom Frieden’s Health Impact Pyramid (2010). Dr. Frieden, the current head of the Centers for Disease Control and Prevention (CDC), presented this pyramid in a 2010 article in the *American Journal of Public Health*. It has been widely adopted in the field of public health as a useful way to discuss where and how public health interventions fit and what their goals should be. The pyramid (see Figure 1.1) presents a framework for organizing and explaining public health interventions from high population impact/
low individual effort needed (at the bottom of the pyramid), to a relatively small population affected and high individual effort needed (at the top of the pyramid).

Efforts that focus on “socioeconomic factors,” like poverty, education, transportation, or economic opportunities, according to Frieden, have the "greatest potential impact" and would be at the bottom of the pyramid. Moving up the pyramid, the second layer is called, “changing the context to make individuals’ default decisions healthy.” These are changes to the environment, like building communities to encourage walking rather than driving a car, increasing air or water quality, providing high-quality housing to everyone, and passing smoke-free laws. The third level of the pyramid, “long-lasting protective interventions,” are interventions that are effective but do not require ongoing clinical contact. Frieden’s examples include immunizations, colonoscopies, and smoking cessation programs. The fourth level of the pyramid is “clinical interventions”; this is where most traditional medical interventions enter the pyramid. Finally, the tip of the pyramid is “counseling and education.” Examples include counseling patients to change their diet or exercise more. In Frieden’s words, counseling and education, “is generally less effective than other interventions; successfully inducing individual behavioral change is the exception rather than the rule” (2010, n.p.). However, he points out that implementing programs at each level of the pyramid would achieve the maximum possible sustained public health benefit.
Using the language of the pyramid as a guide, this study will examine, while taking into consideration many socioeconomic factors of both the individual and the community (Level One of the pyramid), the relative influence of housing quality (Level Two of the pyramid) on self-reported health. We know that individual characteristics are very important (genes, health condition, health behaviors), but it is also very difficult to influence individual change (at the top of the pyramid). While environmental characteristics have less influence on health outcomes, these changes are often easier to implement and can have an effect on many people simultaneously. Not only are the bottom two strata of the pyramid of interest to public health researchers; they also are issues often of concern to planners.
The fundamental theoretical construct for this study is expressed in Figure 1.2.

**Figure 1.2: Relationships of interest for this study**

A long history of research literature has established a link between the social determinants of health (this includes the built environment) and actual health status. However, this link is heavily influenced by individual behavior and individual health condition. Individual characteristics typically drive a large portion of health outcomes as compared to environmental variables (Sayer, 2013; Picket & Pearl, 2001; Snijders & Bosker, 1999). If researchers do not separate the individual from the environmental, the effects of individual characteristics are likely to statistically overpower the environmental effects (Sayer, 2013). Multilevel modeling (discussed in Chapter 2) allows for this separation.
1.4 Review of study’s contributions

This study offers several contributions to the field.

First and foremost, very little multilevel modeling has been done in the field of planning. There are a few exceptions (Doyle et al., 2006; Van Dyck, Cardon, Deforche, Sallis, Owen, & De Bourdeaudhuij, 2010; and Sharma, 2008). As a doctoral dissertation in the field of Regional Planning, this study, using multilevel methods, benefits from a deep and long tradition of use of multilevel methods in other fields, for the benefit of the field of planning.

This study adds to the scholarly discussion that the fields of Planning and Public Health, which developed together in the 19th century, and which have since drifted from each other, should work more closely together. Combining efforts would benefit the constituents of both fields.

In research to date, geographic level has varied tremendously in similar studies of health ranging from national level-analyses (Kennedy, Kawachi, Prothrow-Stith, Lochner, & Gupta, 1998) to counties (Patel et al., 2003) to large cities (Prince et al., 2012; Doyle et al., 2006). Many comparable studies focused on health have been done outside the United States (Prince et al., 2012; Cummins et al., 2005; Stafford, 2001; and Drukker & van Os, 2003). This study, similar to Prince et al., 2012, will use ecologically meaningful neighborhood clusters within a larger administrative boundary. It is the only known model of this type within a mid-sized city in a Midwestern county. This small level of geography is unusual, and the ability to have smaller neighborhood clusters will allow for more specificity around neighborhood variables. The relatively low-level geography is a strength of the study, along with the fact that the geographic boundaries were selected to coincide with groupings of neighborhoods that are meaningful to the community. The neighborhood groupings were chosen specifically to group communities that are relatively homogenous neighborhood characteristics (poverty status, racial makeup, and home ownership rate).
Existing studies use data sources at each level of the model that are from multiple years, some with data collection dates that range over a period of ten years or more. The preponderance of data for this study all comes from the same year (2010). The exception to that are a few data points that come from the 2006-2010 American Community Survey (ACS). In Diez, Roux, and Mair, researchers postulate that the influence of the built environment has an ongoing long-term influence on people, and this suggests that this longer-term measurement of neighborhood environment before the measurement of individual health characteristics is within acceptable bounds (2010).

1.5 Assumptions and limitations of the study

A series of methodological limitations also need to be discussed. This study, like most multilevel models that look at the area effects on health, relies on cross-sectional data rather than longitudinal data. Cross-sectional data provide a snapshot at the time of the data collection rather than a more powerful examination of individual characteristics over a longer period of time and long-term observation of each respondent’s environment as it changes (or not).

This study will use administrative boundaries, similar to nearly all multilevel model studies looking at the area effects on health. Riva, Gauvin, and Barnett point out that there is a weakness in using fixed administrative boundaries (e.g., census tracts, postal codes, governmental boundaries), “their potential lack of intrinsic meaning in relation to health” (2006, p. 857). This study will use locally defined community boundaries, at as fine a geographic level as the power of the data allowed. But these boundaries are still arbitrary and might not match the area of environmental influence or neighborhood boundaries precisely for each respondent. Arbitrary boundary selection also influences the ability to make causal links. It is hard to separate whether it is the neighborhood that is influencing individual health or an intrinsic set of
characteristics of individuals who self-select into poorer neighborhoods. This is a weakness of most of these types of analysis: “…most neighborhood effects analyses are vulnerable to the criticism that significant neighborhood level coefficients may be capturing characteristics of individuals that select into more disadvantaged neighborhoods” (Browning & Cagney, 2002, p. 395).

This study includes variables on housing conditions, and it is important to note that while research shows that poor housing environments and employment conditions can lead to poor health outcomes, there is less work that shows the effects of improvements to deprived urban conditions. Part of this is due to the vast array of site specific (i.e., not generalizable research) on the link (Curtis, Cave & Coutts, 2002).

This study focuses on a Midwestern county, so similar relationships would be expected in regions with a similar socioeconomic makeup, but the applicability of these results to other geographic regions might be limited.

### 1.6 Dissertation organization

This study consists of six chapters. Chapter 1 provides an introduction and overview of the study including the objectives of the research, its relevance to the field of urban and regional planning and its contribution to this field. It also introduces the method and discusses the study’s assumptions and limitations. Chapter 2 includes a review of relevant literature related to multilevel modeling, self-reported health and the built environment. Chapter 3 describes the data sources, variable selection and sample size. Chapter 4 describes the methods used for the analysis. Chapter 5 describes and discusses the results of the analysis. The final chapter discusses the study’s conclusions and potential directions for future research.
2. Literature Review

This chapter will review the relevant literature for this study. A look at how this study sits at the intersection of the planning and public health fields is first. This is followed by a discussion of the literature linking neighborhoods, housing and health outcomes. Next is a review of the literature relevant to the outcome variable, self-reported health. This chapter ends with a description of literature using multilevel models for health outcomes.

2.1 The intersection of the fields of planning and public health

This study sits squarely at the intersection between the fields of planning and public health. Yet the current collaboration between these two fields is not at its best. The fields of urban and regional planning, sometimes called just urban planning or planning, and public health grew out of the extremely poor living conditions and high mortality rates of the 19th century Victorian city. In the beginning, the two fields shared a very clear goal – to improve the health and life expectancy of residents. Both recognized the need to remediate the poor living and working conditions of the population – including poor housing, sanitation, and ventilation – in order to control the spread of disease. For a succinct review of the evolution of both fields, see: Corburn, Confronting the Challenges in Reconnecting Urban Planning and Public Health in the American Journal of Public Health (2004). For a more complete review see: Corburn Reconnecting with our Roots: American Urban Planning and Public Health in the Twenty-first Century in the Urban Affairs Review (2007). Since the 19th century, the leading causes of mortality have shifted from communicable diseases to chronic diseases, largely because of the successful efforts of early public health and planning professionals. From these successful early efforts, the field of public health moved to a focus on a medical model of health improvement, meaning a focus on the prevention of chronic disease, largely abandoning the
focus on improving the built environment. The field of planning moved toward the adoption of zoning regulations to separate residential areas from industrial. Planning also shifted its focus to managing public spaces, zoning regulations, and studies of transportation patterns. According to Corburn, “By the World War II Era, distinct silos emerged in both fields—transportation, housing, economic development, etc., in urban planning and diseases (e.g., cancer, heart disease, diabetes) and individual risk factors (e.g., smoking, drinking, diet) in public health” (2005, p. 112). Remnants of these silos are still present in the way that both fields continue to work.

These same silos can be seen in the work of the planning and health departments in the study area, Hamilton County, although there are small signs of change. The county’s four health departments focus on very diverse issues, from healthcare provision at clinics and schools to emergency preparedness and disease tracking. There is only one staff person at one of these agencies assigned to environmental health, and that person’s attention is split between many tasks, from environmental pollutants (toxic waste, lead, water quality, etc.) implementation of Health Impact Assessments (HIAs; see end of Section 6.3 for a more complete description), and changes to the built environment. While the county’s planning departments are almost exclusively focused on the form and shape of the built environment, health improvement has not risen to the level of high priority. However, it should be noted that the City of Cincinnati’s award-winning comprehensive plan includes a section on health and health improvement, and for the first time leaders in the health community were intimately involved in the process of drafting the plan’s health goals and objectives.

Nationally, with increasing rates of obesity, both fields have devoted significant research to food environments and obesity prevention, specifically the effect of the built environment (sprawl, increasing drive time, unwalkable communities) on the health and well-being of communities. In recent years both fields have started to revisit their thinking to look at the interrelationships between the environment and health outcomes, specifically what is being called
the social determinants of health. This shift widens the net of acknowledged influences on health from only the individual and his or her genetic makeup to a person’s physical environment and socioeconomic circumstances. Certainly genetics are important, but research suggests that genetics account for only a part of person’s health and that the environment in which he or she lives has a large influence. In *The Weight of the Nation*, a film on obesity in America, obesity experts state that genes account for about 60% of health outcomes, and the environment in which one lives makes up the rest. Specifically, the obesity experts discuss the fact that for obesity it is not about nature vs. nurture, it is both – “they are inexorably intertwined” (2012, part 4). Similarly McGinnis, Williams-Russo, and Knickman state that genetic predispositions account for just 30% of early death, compared to 40% for behavioral patterns, 15% for social circumstances, 10% shortfalls in medical care, and 5% for environmental exposures (2002). In the past, the research interests of the fields of regional planning and public health diverged significantly, but with the shared interest in the built environment and its influence on health and health outcomes, the fields are starting to converge again in small ways around very specific issues.

Both fields recognize the importance of the environment people live and work in for good health, and both recognize that there are large disparities based on the social determinants of health for different groups. Given these relatively recent shifts, one might expect great collaboration between the fields of public health and planning. Certainly there is an increased interest in issues that cross the boundaries of the two fields with exponential growth in health and built environment presentations at public health and planning professional conferences. In 2003 *the American Journal of Public Health, the American Journal of Health Promotion, and The Journal of Urban Health* published special issues on the built environment and health.

---

1 *The Weight of the Nation* is a film designed to get communities moving on the obesity epidemic. The film was presented by HBO and the Institute of Medicine (IOM), along with the Centers for Disease Control and Prevention (CDC), National Institutes of Health (NIH), the Michael & Susan Dell Foundation, and Kaiser Permanente.
Relationships Between Neighborhoods, Housing, Health

However, the underlying conversations in each field are different. Public Health researchers and practitioners, in large part, have adopted methodologies from regional planning and integrated them into their work (e.g., GIS and Health Impact Assessments [HIAs]). Planners have also adopted some health monitoring strategies from public health, but there continues to be a call in planning for more integrated interdisciplinary work (Corburn, 2004, 2005, 2007; Northridge, Sclar, & Biswas, 2003). In one of the mostly widely quoted texts on the built environment and health, Botchwey and Trowbridge describe the current situation as:

*The integrated application of public health and planning perspectives will be essential to realize the goal of healthy and sustainable places – buildings, neighborhoods, communities, cities, and regions. Unfortunately, in current practice, planners and designers, who shape the built environment, and public health professionals, who protect the public’s health, rarely interact. Most public health professionals have little experience working with zoning boards, city councils and others who make decisions about the built environment. Few planners know how to analyze the health implications of design, land-use, and transportation planning decisions in a comprehensive manner.* (2011, p. 322)

And

*…that health implications of planning are frequently considered narrowly by planning students, and opportunities to apply measurement and evaluative tools available from public health specialties such as epidemiology and surveillance are missed…most public health students are not trained to consider geographic and social contexts of data related to disease processes.* (2011, p. 326)

Beyond integration among research agendas and professional training from planners and public health professionals, many think that there should be more formal relationships across industry silos (e.g., planning positions in health departments or health-focused positions in planning departments). This institutionalization of roles has not happened in most communities across the country and has left the field of planning calling for change, but with much less research in the built environment and health as compared to the field of public health. Much of the work studying the direct relationship between health and the built environment has happened in the public health field of environmental justice, where researchers have identified
some uses of land that contribute to poorer health and other environmental assets like parks that could improve health (Corburn, 2004, 2007).

In their review of the last 10 years of work in health and the built environment, Jackson, Danneberg, and Frumkin, leaders in the study of the connection between the built environment and health, described a growing field of research, regulatory programs, and consensus statements that acknowledge the role of the built environment in health outcomes, although these statements tend to focus on obesity as the health outcome (2013, p. 1543). They conclude that, “a vibrant, robust movement has arisen,” but there is still more to be done. They identify a need for more research specifically a diversity of geographic scales, research that goes beyond obesity and physical activity, and studies that, “simultaneously investigate human health and well-being and economic and environmental outcomes” (2013, p. 1543). Later in their discussion of the gaps in research, they specifically identify the need for studies on the most vulnerable and people living in substandard housing.
2.2 Relationships between neighborhoods, housing, and health outcomes

The research literature has moved beyond the question of: Do the socioeconomic characteristics of a community affect one’s health? That topic was top of many research agendas in the mid to late 1990s (Robert, 1990). Instead, focus has shifted to how to tease out the complexities and interconnectedness of the individual-community relationship.

Research on the built environment and health varies widely, but much of the recent work is focused on how the built environment can influence physical activity and obesity with much less attention on how neighborhoods influence health in general. Research supports the idea that some features of the built environment are associated with increased physical activity, but making a causal link between the built environment and improved health is still a stretch based on current research. There are cross-sectional studies that demonstrate that certain features of the built environment are associated with changes in physical activity (Lathey, Guhathakurta, & Aggarwal, 2009; Handy, Cao, & Mokhtarian, 2007; King, Toobert, Ahn, Rensicow, Coday, Riebe, Garber, Hurtz, Morton, & Sallis, 2006; Librett, Yore, & Schmid, 2006 [for older people]; Li, Harmer, Cardinal, Bosworth, Acock, Johnson-Shelton, & Moore, 2008). Wallace’s work is regularly cited as substantial and influential on the relationship between deteriorating inner cities and the spread of HIV and tuberculosis (1990, 1997). Several studies have used abandoned buildings as an indicator of neighborhood deterioration (Furr-Holden, Lee, Milam, Johnson, Lee, & Ialongo, 2010; Cohen, Mason, Bedimo, Scribner, Basolo, & Farley, 2003). Cohen et al. created a “broken window” index that measured the deterioration of physical conditions in neighborhoods (2000).

This study aims to separate out the individual versus the community-level characteristics and their unique contributions to self-reported health outcomes, with a particular focus on housing variables. This type of model has been used multiple times in previous research. Pickett and Pearl did a review of similar multilevel studies and found that most studies showed
some association between neighborhood socioeconomic factors and health outcomes; however, these associations were smaller than the associations between individual socioeconomic factors and health (2001). Similarly, Riva and Garnett found that area or neighborhood effects were greater on low-income individuals and women (2007).

Self-reported health has been widely studied, and there have been a number of multilevel models using self-reported health as an outcome variable, similar to this study. A very thorough and widely-cited article by Riva, Gauvin, and Barnett identified 67 multilevel models that looked at adult health between 1998 and 2005 (2007). Thirty-nine of these studies used some form of self-reported health as the outcome variable, and in thirty-seven of those studies significant associations were found between area socioeconomic measures and self-reported health; with areas of higher poverty connected to poorer self-reported health (Riva, Gauvin, & Barnett, 2007).

It is clear from the discussion above that neighborhood conditions can be linked to health outcomes. However, most research studies focus on the socioeconomic or external built environment of people in order to study the relationship between neighborhoods and health. Most Americans (90%) live in urban areas, and on average they spend 90% of their time indoors (Environmental Protection Agency, n.d.). This suggests that when studying the relationship between the built environment and health outcomes it is important to include the internal built environment. While the need to measure internal environments has been recognized (Hancock, 2002; Galea et al., 2005) finding objective measures of internal housing quality has proven difficult.

Very few studies were found that included objective internal residential characteristics, Table 2.1 provides a brief summary of relevant articles. As shown in Table 2.1, Galea et al. included self-reported internal housing conditions in New York City and found connections between non-functioning kitchen facilities, heating breakdowns in winter, and large areas of
peeling plaster or paint and likelihood of depression. They also identified internal housing conditions as a rare addition to built environment multilevel models, “most of the published peer reviewed literature about the relationship between the built environment and health has focused on characteristics of the external built environment” (2005, p. 825). Similarly, Stafford et al. and Drukker and van Os were able to create or use individual survey data that specifically described the condition of housing. Another study by Beck et al., in a study on childhood asthma admission rates, created a set of neighborhood-level home environment markers using economic market indicators (price and demand), available from the census. These neighborhood-level home environment markers were median home value, percentage of renters, percentage of homes vacant, and population density (Beck et al., 2013). Similar census data will be used for this study.
### Table 2.1: Models for measuring internal housing quality

<table>
<thead>
<tr>
<th>Study</th>
<th>Variable description</th>
<th>Data used</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beck et al. (2013)</td>
<td>Neighborhood-level home environment markers</td>
<td>All Neighborhood level</td>
<td>Census data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Median home value</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. % renter-occupied</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Vacant homes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Population density (persons/sq. mile)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Non-functioning kitchen facilities</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Heat breakdowns in winter</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Large areas of peeling plaster or paint</td>
<td></td>
</tr>
<tr>
<td>Stafford et al. (2001)</td>
<td>Housing quality</td>
<td>Individual level</td>
<td>Census data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Had problems with housing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighborhood level</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Constructed Townsend deprivation index (access to a car, housing tenure, unemployment, overcrowding)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drukker and van Os (2003)</td>
<td>Objective housing conditions</td>
<td>1. Person-bedroom index</td>
<td>Census data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Residential type (single room, apartment, townhouse, etc.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Housing perceptions</td>
<td>1. Perception of housing conditions</td>
<td>Community questionnaire</td>
</tr>
</tbody>
</table>
Relationships Between Neighborhoods, Housing, Health

The makeup of a community’s residents (renters versus owners) has historically been a way to describe the general quality of the housing environment and social investment in the neighborhood, with many communities having a strong desire for increased home ownership. In their often-cited literature review on the costs and benefits of home ownership, Dietz and Haurin state:

*Through their investment in the home – and therefore in the local neighborhood – homeowners appear to be overall more involved in their communities. These efforts by homeowners generate benefits for their communities in addition to the benefits for their families. These spillover benefits suggest that the neighborhood homeownership rate itself may produce positive social consequences for communities. Initial empirical research is consistent with this assertion.* (2003, p. 2)

Homeowners have a different investment in their communities than renters because of their large financial investment in and responsibility for their home and engagement in the community. Haurin, Dietz, and Weinberg discuss the fact that homeowners should collectively engage in behaviors that increase their property values, but acknowledge that the neighborhood effects of high homeownership rates is still an “under studied topic,” but that the “preponderance of thought suggests that there is a positive effect” (2002, p. 143). What the literature is clear about are the positive benefits of a high homeownership rate in most communities. What is implicit in this assumption is there are community-level economic and social benefits tied to a higher rate of homeownership, and these same community-level economic and social benefits are not tied to a higher rate of renters. What is not always stated explicitly is that the opposite of homeownership is renting, although there is a body of literature that looks at renting.

Specifically, research has shown that the homes of owners are kept in better shape than the homes of renters (Rohe & Stewart, 1996). This suggests that the percentage of renters in a community could be a measure of neighborhood level housing environment quality in many communities (with more renters signally less invested in the quality of the neighborhood’s housing environment). There is also literature to suggest that the percentage of renters in a community is a sign of poorer housing conditions and much higher rates of poverty (Mallach,
Relationships Between Neighborhoods, Housing, Health

2010). Pendall, Theodos, and Hildner show explicitly the link between a higher percent of renters in a community and poor housing and neighborhood conditions (2014). Malega found that neighborhoods with a higher percentage of rental homes were strongly connected with increasing rates of poverty (2003). Some suggest that home ownership itself gives more access to opportunity than renting (Rohe, Van Zandt, & McCarthy, 2002). However, some authors suggest that homeownership does not benefit all groups in the same ways and that special attention should be paid to the potential burden of home ownership for low-income and minority populations (Herbert & Belsky, 2008). Not all scholars think that homeownership is the answer for neighborhood improvement, some suggest that the assumptions about homeownership are overly optimistic and simplistic (Rohe & Lindblad, 2013; Haurin, Dietz, & Weinberg, 2002). Rohe and Lindblad point out that with the recent housing market bubble, some communities no longer view homeownership as always better than renting (2013). However, Rohe and Lindblad note that despite the recent housing bubble, home ownership can be a powerful community engagement tool, if done well, “Policies that foster positive homeownership experiences and minimize negative experiences therefore have potential to impact not only individual households, but also social capital within communities” (2013, p. 37).

For the purposes of this study, one of the housing environment indicators will be the percentage of renters in a given community. There is a connection between the two community-level variables used for this study (percent renters and percent mobility in a community. Rohe and Lindblad, in their recent review of the literature, identify that many studies have found that homeowners participate in their neighborhood and community organizations at much higher rates; some of this difference is believed to be connected with residential mobility/stability (2013). The literature suggests that owners move less frequently than renters do (Dietz & Haurin, 2003; Rohe, Van Zandt, & McCarthy, 2002). However, Herbert and Belsky describe the relationship between mobility and home ownership this way:
The fact that owners move less than renters do does not mean that the evidence is clear that homeownership causes greater residential stability. In fact, individuals are more likely to buy a home when they know they are less likely to want to move in the near future. In this case, lower expected mobility leads to homeownership, not the other way around. Still, homeownership would be expected to lower mobility in several ways. First, higher transaction costs of moving make owners less inclined to move as their household circumstances change. Second, owners also have greater ability to tailor homes to meet their needs and tastes so many have less need to move to adjust their housing consumption. (2008, p.14)

The link between mobility, housing quality, and health is complex, but has been widely studied. There is a body of literature that focuses on the link between poor neighborhoods with high mobility and a loss of social support structures for neighborhood residents linked to poorer health outcomes (Smith & Mallinson, 1996; Smith, Alexander, & Easterlow, 1997; Howden-Chapman, 2004; Acevedo-Garcia, Osypuk, Werbel, Meara, Cutler, & Berkman, 2004). Coulton explains that, "residential mobility is one of the primary factors driving neighborhood change and can have an important effect on social conditions and quality of life in an area" (2014, p. 260). However, the link between mobility, housing quality, and health outcomes varies depending on neighborhood composition, macroeconomic cycles, and the availability of affordable housing (Howden-Chapman, 2004). In some cases, increased mobility is a sign of upward mobility (e.g., improvements in housing), and sometimes it is not (Winstanley, Thorns, & Perkins, 2002; Coulton, 2014). Acevedo-Garcia et al. conducted a large literature review to test if housing mobility, that is moving poor individuals out of poor neighborhoods, improves the health of those that move (2004). They found that the link may be there, but that the empirical evidence is sparse. Mobility of homeowners will be tested in this study as a measure of the socioeconomic environment for a community.

Previous research has identified a correlation between housing quality and health outcomes, although claims of causality are still rare. Mason and Brown discuss the dangers of using cost-benefit models to determine the value of housing interventions to improve health (2010). Leventhal and Brooks-Gunn found a link between neighborhood income and mental health outcomes, specifically parents and male children who moved to neighborhoods with less
poverty reported much less mental distress (2003). Coley, Leventhal, Lynch, and Kull found that living in poor quality homes takes an emotional toll on children’s well-being due to parental stress (2013). Hamoudi and Dowd found a direct link between the value of single-family homes and physical health for older homeowners, specifically, “our results indicate that increases in housing wealth were associated with better health outcomes for homeowners in late middle age and older” (2013, p. 1039).

2.3 Self-reported health

Most current connections between the built environment and health rely on singular non-multidimensional measures of health (e.g., obesity, asthma prevalence, etc.). There is little use of general health measurement or inclusion of health factors that go beyond physical health condition (e.g., mental and emotional health). However, health researchers have been measuring general health ratings for decades and have very simple, yet sophisticated ways to measure the state of individual health. Self-reported health, or the answer to the question: “In general how would you rate your health, excellent, very good, good, fair, poor?” has been very widely studied. Self-reported health has a very robust and consistent research history of reliably capturing morbidity and mortality across the world and among many different demographic groups (Brunner, 2006; Dowd & Zajacova, 2007; John & Montgomery, 2012; McFadden et al., 2009; Regidor et al., 2010; Singh-Manoux et al., 2007; Eriksson, Undén, & Elofsson, 2001; and Undén & Elofsson, 2006; among many others). It has been show to predict mortality (Idler & Benyamini, 1997; Singh-Manoux et al., 2007; and Undén & Elofsson 2006) and disability (Kaplan et al., 1993). Specifically, Idler and Benyamini state, “Twenty-seven studies in U.S. and international journals show impressively consistent findings. Global self-rated health is an independent predictor of mortality in nearly all of the studies, despite the inclusion of numerous specific health status indicators and other relevant covariates…” (1997, p. 21). Research
suggests that respondents consider mental, physical, and emotional health (Undén & Elofsson 2006). The question is simple to understand and easy to implement.

Some studies have found a difference in self-reported health status between men and women, although study findings are mixed. Undén and Elofsson looked at whether Swedish males and females rated their health differently (2006). In their study, they found that for men, physical, educational, and cultural activities more correlated with self-reported health, while women found more important sleep and doctor visits for determining their self-reported health. They also found a moderately strong (and similar) correlation between age and self-reported health. In Brunner’s review of self-reported health studies, he notes that although small differences have been found in self-reported health between men and women, these differences are minor compared to the study-wide variance in self-reported health (2006).

Singh-Manoux et al. found that self-reported health seems to be related to socio-demographic variables, objective measures of health, and health behaviors in both men and women (2007). And while self-reported health predicted mortality equally well in women and men in their study it explained more of the connection between self-reported health in men than in women (Singh-Manoux et al., 2007). St. John and Montgomery considered whether self-reported health predicts mortality in older adults with and without depression, and they found that self-reported health is predictive of mortality with or without depressive symptoms (2012).

The relationship between self-reported health and mortality by socioeconomic status has been shown to vary by country. McFadden et al. looked at whether the relationship between self-reported health and mortality is different depending on social class (2009). They speculated that adults of differing social classes might have different criteria for rating their health status. In their study of 22,457 adults in the U.K., they found no interaction between social class and self-reported health for men or women. However, in a study done by Dowd and Zajacova in the U.S. in 2007, there was variance in the predictive power of self-reported health on mortality based on differing levels of income and education. There was a strong correlation
Relationships Between Neighborhoods, Housing, Health

with lower self-reported health and mortality for adults with more education and higher income as compared to adults with less education and lower income. They concluded that self-reported health might not be directly comparable between people of different socioeconomic status because they might evaluate their health differently (Dowd & Zajacova, 2007). In another study in France, Singh-Manoux et al. found that the relationship between self-reported health and mortality got weaker as socioeconomic status increased (2007). In yet another study of older adults in Spain, Regidor et al. found that the relationship between self-reported health and mortality was the strongest for older adults with the high levels of education (2010).

There is little debate about the ability of self-reported health to predict mortality and mobility, but there is some debate about what exactly self-reported health measures. It has been widely accepted as a global indicator of health, but there is still debate about what are the critical factors that contribute to self-reported health. While many suggest that self-reported health is influenced by physical, mental, emotional, and social health or that study the influence between variables within the study and self-reported health. There are a limited number of large studies that explicitly attempt to determine the major drivers of respondents’ self-reported health rating. All four of the large studies found physical health or physical functioning was among the strongest drivers of self-reported health (Mavaddat, Kinmonth, Sanderson, Surtees, Bingham, & Khaw, 2011; Singh-Manoux, Martikainen, Ferrie, Zins, Marmot, & Goldberg, 2006; Bailis, Segall, Mahon, Chipperfield, & Dunn, 2001; Jylhä, Guralnik, Ferrucci, Jokela, & Heikkinen, 1998). This is not to suggest that other aspects are not influential, but that physical health (in some cases this included chronic disease and/or physical disability) was the strongest driver of self-reported health rating. Mavaddat, Kinmonth, Sanderson, Surtees, Bingham, and Khaw, note in their study that all dimensions of the SF-36 were associated with self-reported health, confirming that self-reported health is a global health measure that includes measures of physical, mental and social health (2011). However, they found that the strength of the association between self-reported health and physical functioning was more than twice that of mental health and social
functioning, after controlling for age, sex, social class, and chronic disease. In a large analysis of 28,000 participants in the Whitehall (British) and Gazel (French) studies, physical mobility and mental health were the most critical factors in self-reported health (Singh-Manoux, Martikainen, Ferrie, Zins, Marmot, & Goldberg, 2006). Singh-Manoux et al. also concluded that age, early life factors, family history, socio-demographic variables, psychosocial factors, and health behaviors only contributed in a minor way to self-reported health (2006). Similarly, in a National Population Health Survey in Canada, self-reported health rating was closely related to physical factors and experience of symptoms (Bailis, Segall, Mahon, Chipperfield, & Dunn, 2001). In their study of Italian and Finish older adults Jylhä, Guralnik, Ferrucci, Jokela, and Heikkinen found that self-reported health is a useful summary of physical health, but that comparisons of self-reported health across cultures should be made with caution (1998). Given the close relationship between self-reported health and physical health/physical functioning, it should not be surprising that the built environment could play a role in self-reported health rating. How conducive the built environment is to daily activities plays a large role in one’s ability to live, work, and play. If one’s housing or neighborhood limit one’s ability to conduct normal daily activities, quality of life would be expected to suffer.

The strength and consistency of the relationship between self-reported health and mortality has led to a rich, long and extensive literature on self-reported health. There is, however, some debate about how respondents determine their self-reported health. Jylhä suggests that self-reported health follows a conceptual model, whereby respondents consider three major things: 1) What should be included in health, and what are the relevant components of their individual health; 2) They rate their general health, taking into consideration of number of factors (e.g., age, health compared to peers, previous health history, health expectations); 3) They use their cultural norms to determine which on the scale is the normal or ordinary score and evaluate how their situation is in relationship to “normal” (Jylhä, 2009). This thought process then leads to a final selection of a self-reported health rating.
In their response to Jylhä, Huisman and Deeg, while generally supportive of Jylhä’s conceptual framework, suggests that self-reported health is not as well thought out at the time of the question response, but rather, “a snapshot of a moving scene” (2009, p. 653). In a separate article, Bailis, Segall and Chipperfield suggest that self-reported health is, “not only a spontaneous assessment of one’s health status and related practices …. [but also] may be regulated by efforts to achieve one’s relatively important health-related goals” (2003, p. 203).

Jylhä states that her conceptual model suggests that an individual’s self-reported health includes all of the relevant information that the respondent believes describes his or her health. Jylhä herself suggests that this perspective would include, “non-health individual and social characteristics such as social class, education, standard of living, social networks, social capital, and the quality of neighborhoods” (2009, p. 309). These and other “non-health” characteristics help shape individual self-reported health (Jylhä, 2009; Krause, 1996; Manderbacha, Lundberg, & Martikainen, 1999). “People have considerable freedom to decide what information to base their evaluations on” (Jylhä, 2009, p. 309).

For many, health has a wide definition, which includes not only individual health characteristics but also the built environment and socioeconomic environmental characteristics. (This is in line with the definition of the social determinants of health in Section 1.2.) In a locally focused, but not yet published study looking at self-reported health within the primary care setting, over half (53%; N=503) of the adult respondents identified things in their neighborhood that are linked to their self-reported health and what could be done to improve quality of life (Elder, Jacobson, & Chubinski, 2014). These results suggest that people consider, at least partially, their built environment when determining their self-reported health. The section that follows includes multiple examples of previous self-reported health studies that found a link between self-reported health and neighborhood characteristics.


2.4 Health measurement in a multilevel model

This study will use a method called multilevel modeling. Multilevel models have been used extensively to model health and its relationship with the built environment. In their literature review of multilevel models (1998-2005), Riva et al. indicate that most of the studies included age, gender, socioeconomic status, and marital status. In addition, some also included health behaviors, medical conditions, social network, and years of residency in the area (2006). While most studies focused on the general population, there were some that focused their analysis on men, older adults, and certain racial or ethnic groups (Riva et al., 2006). After basic demographic and health characteristics “…economic conditions are the most frequently examined structural factors thought to be relevant for health status over and above individual characteristics” (Wen, Browning, & Cagney, 2003, pg. 843). This look at economics often includes a focus on poverty (Wen, Browning, & Cagney, 2003; Patel et al., 2003; and Galea et al., 2005).

Multilevel modeling is commonly used in health studies; however, each study’s choice of what variables to include to measure health is different. Patel et al.’s outcome variable was self-reported health, but included in their control variables were both chronic conditions and health behaviors (2003). They found that after adjusting for individual characteristics, neighborhood context mattered. Specifically, their respondents (older Mexican Americans) were more likely to rate their health poorly if they lived in neighborhoods that were poor, less densely populated by other Hispanics, or located within 50 miles of the U.S.-Mexican border.

Prince et al. were interested in the relationship between neighborhoods and leisure time physical activity, and as a result included obesity and physical activity in their definition of health (2012). They also found a relationship between health and neighborhood conditions. Specifically, that park areas, crime rates, and food stores have an influence on male and female leisure-time physical activity and overweight/obesity status. Galea et al. focused on mental
Relationships Between Neighborhoods, Housing, Health

health, specifically depression (2005). They also found that, after adjusting for individual characteristics, there was a relationship between poor quality built environment and greater likelihood to report depression in New York City. They specifically identified both the fields of public health and urban planning as the targets of their research, “Future prospective work designed to access potential mechanisms underlying these associations may guide public health and urban planning efforts aim at improving population mental health” (Galea et al., 2005, p. 822).

Other studies included variables that fall classically within the planning literature. Rundle et al. were interested in the relationship between obesity and urban built environment, and it was the only study found that included land use mix (2005). The same study included measures of access to public transit, population, and intersection density. They found that land use mix, density of bus stops, subway stops, and population density were related to obesity, after controlling for individual-level variables. Intersection density, however, was not found to be significant.

Browning and Cagney, well known for their work on multilevel models and social capital, created a social cohesion and informal social control variable in their 2002 and 2005 work in Chicago. Along similar lines, they defined residential instability as the percentage of owner-occupied homes combined with the percentage of new residents to the community (Browning & Cagney, 2002; Cagney et al., 2005). While both studies used self-reported health as their outcome variables, their conclusions were slightly different. The 2002 study, Neighborhood Structural Disadvantage, Collective Efficacy, and Self-Rated Physical Health in an Urban Setting, they found that when they controlled for individual health and socioeconomic conditions there was no significant relationship between self-reported physical health and neighborhood socioeconomic disadvantage. However, they did find that individuals residing in neighborhoods with higher levels of collective efficacy reported better overall health. Their 2005 study focused
on adults over age 55, and they found that collective efficacy was not associated with health; instead neighborhood affluence contributed to self-reported health.

Residential or housing environment has been included in other studies (Cummins et al. 2005, Galea et al. 2005 for depression). Cummins et al. found that some neighborhood conditions in Scotland and England were associated with fair to poor self-reported health: poor quality residential environment, left-wing political environment, low political engagement, high unemployment, lower access to private transport, and lower transportation wealth. Other neighborhood conditions were not significantly associated, including: public recreation facilities, crime, health service provision, access to food stores, and access to banks.

Multilevel models have been used extensively to model health and its relationship with the built environment because of the nested nature of the data these models provide the quantitative structure needed to help understand the interconnections between respondents and their environments. As shown in Figure 1.2, this link is heavily influenced by individual behavior and individual health condition. The compositional effect, that is the variability between neighborhoods rather than within neighborhoods, is measured by the interclass correlation coefficient (ICC, see Section 5.1.1. for more details). In Figure 1.2, the ICC would be 10%; that 10% of the variation in self-reported health would be because of variability between neighborhoods instead of within neighborhoods (90%, the variability driven by individual characteristics). This estimation (90%/10%) is based on the literature for multilevel models which states that ICCs rarely exceed 20% (Snijders and Bosker, 1999). Among multilevel models with an individual health outcome and a neighborhood group variable, the ICCs have varied. Many published articles do not report the ICC. Among those that do, there are a few studies that report ICCs in the 20% range. Li et al. found that 28% of variation on walking was due to between neighborhood differences (2005) and Pampalon, Hamel, De Koninick and Disant found that 20% of perception of problems and social cohesion varies by neighborhood after accounting for individual characteristics (2007). Most studies report ICCs below 5%.
Brown and Cagney found that 5.4% of physical activity can be attributed to neighborhood level variation (2002). Lebel, Pampalon, Hamel, and Thériault reported ICCs of 1.23% for the variation in overweight and obesity by regional characteristics for men and 2.61% for women (2009). Prince et al. found ICCs from 1% - 8% for individual differences in physical activity and overweight/obesity status (split by gender with higher ICCs for men) and neighborhood environment in Canada. Doyle reported an ICC of 1% for the amount of variability in BMI accounted for by variation at the county level. Research has shown that ignoring the non-independence with even an ICC as low as 5% may lead to inflated Type I error (Julian 2001; Barcikowski 1981). Additional examples of research studies with small ICCs include: Myers et. al 2012; Boyle and Willms 1999; Hart, Ecob, and Smith 1997; and Ecob 1996. The ICC for this study is 5.5% and is discussed in Section 5.1.1.
3. Data and Measures

This chapter will cover the data and measures used in this study. A review of the data sources for Level-1 and Level-2 data is first followed by a discussion about how the geographic boundaries were determined. This is followed by a description of the measure for each of the variables considered for the study, starting with the outcome variable: self-reported health status.

3.1 Description of Individual (Level-1) Health Data: The Greater Cincinnati Community Health Status Survey

The Greater Cincinnati Community Health Status Survey (CHSS) has been conducted in the Greater Cincinnati region every three to five years since 1999. The CHSS is conducted by the Institute for Policy Research at the University of Cincinnati and sponsored by Interact for Health (formerly The Health Foundation of Greater Cincinnati). The University of Cincinnati’s Institutional Review Board (protocol number: IRB #10-07-23-01E) approved this study. Standard random-digit-dialing telephone survey methodology (RDD) strategies were used to contact potential respondents. The response rate for the 2010 CHSS was 21%. The CHSS is designed to collect health status data on the 22 counties that surround Cincinnati; including parts of Indiana, Ohio, and Kentucky (see Figure 3.1).
For the 2010 CHSS, a total of 2,246 adults (over age 18) were interviewed by telephone between August 14 and September 27, 2010. Ninety percent of these interviews were done over landlines, ten percent were done with cell phone only users. Post stratification weights are generated that correct for non-response and sampling design. For analyses that require a representative population of the region, weights are provided. The sample was weighted, based on census data for the same populations, for sex, race, age, educational attainment, and geography of residence. After weighting, the sample represents the non-institutionalized population of the region.

The survey is a comprehensive health status survey, including over 100 questions on a wide range of health issues, health behaviors, and health opinions of local adults. The CHSS has been used in research (Chubinski and Carrozza 2012; Ludke, Obermiller, & Horner 2012; Ludke, Obermiller, Rademacher, & Turner 2012) and community report cards (The State of the Community Report, Indicators of Healthy Communities, among others). The CHSS is a valuable local source for self-reported health data; the variables of interest for this study are self-reported health status, age, gender, poverty status, employment status, chronic conditions, health behaviors, and social support of each individual adult respondent.

For the purposes of this study only the unweighted sample for Hamilton County respondents (N=1356) will be used. The decision was made to use the unweighted data because the goal of this study is not to produce representative estimates of population
characteristics. Because the study is using unweighted data, there would be no expectation that the study sample would be representative of the county’s population (as defined in census data).

Not all 1,356 respondents could be included for the purposes of this study (see Table 3.1). Only respondents who provided their residential address or city neighborhood could be attached to a specific census tract and used for this study (N=1295). In the original dataset, respondents who gave their mailing addresses were geocoded, and they were assigned a census tract and county of residence (N=1248). Respondents living in the City of Cincinnati were also asked what neighborhood they lived in, so respondents who did not provide a residential address, but did respond to the question about neighborhood of residence were added to the dataset (N=47). If respondents did not give a mailing address or neighborhood they were assigned a county of residence based on phone exchange, but were not used in the dataset for this study, because they could not be connected with a specific neighborhood within the county (N=61). The statistical software used for this study does not allow for any missing data points. Of the 1,295 respondents, 369 were missing one or more variables used in the final multilevel model; most of these were missing income data. This meant that the final model was calculated using 926 adult responses (see Table 3.1).
Table 3.1. Response counts for study sample

<table>
<thead>
<tr>
<th>Total Number of respondents</th>
<th>Unweighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamilton County:</td>
<td>1,356</td>
</tr>
<tr>
<td>Residential address</td>
<td>1,248</td>
</tr>
<tr>
<td>City neighborhood</td>
<td>47</td>
</tr>
<tr>
<td>No sub-county geography</td>
<td>(61)</td>
</tr>
<tr>
<td>Study sample sub-total</td>
<td>1,295</td>
</tr>
<tr>
<td>Run-time software deletions</td>
<td>(369)</td>
</tr>
<tr>
<td>Final study sample</td>
<td>926</td>
</tr>
</tbody>
</table>

3.2 Description of community data

Neighborhood conditions were modeled using data from the 2010 census and the 2006-2010 American Community Survey. These community-level data, all collected before or during the year 2010, correspond to the same timeframe as the individual data (2010 Greater Cincinnati Community Health Status Survey, described above). The Census Bureau provides the data at the census tract level, these data were aggregated based on the neighborhood groupings (described in the section that follows) to provide the community-level indicators of interest. This specific source for each community-level measure is shown in Table 3.2.
At the community level, this study looks to measure the additive explanatory power of community housing environments after controlling for individual and community-level drivers of health. These community-level drivers of health are called the social determinants of health. Based on the literature described in Chapter 2, four commonly-used measures of community-level socioeconomic conditions were chosen for this study: percent white, percent unemployed, percent in poverty, and percentage of home owners who moved into the neighborhood in the last 5 years (since 2005 = mobility).

Table 3.3: Community-level measures of socioeconomic conditions and housing environment

<table>
<thead>
<tr>
<th>Socioeconomic measures</th>
<th>Housing environment markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>% white</td>
<td>median home value</td>
</tr>
<tr>
<td>% unemployed</td>
<td>% renter occupied homes</td>
</tr>
<tr>
<td>% in poverty</td>
<td>% vacant houses</td>
</tr>
<tr>
<td>% mobility</td>
<td></td>
</tr>
</tbody>
</table>
Similar (Beck et al., 2013) neighborhood-level housing environment markers were identified. These markers were chosen based on a review of the literature described in Chapter 2 and discussions with housing experts. The three neighborhood-level housing environment markers are: percentage of homes vacant, percentage of homes occupied by renters, and median home value.

### 3.3 Determining geographic boundaries

Understanding the neighborhood-based drivers of health requires a decision about what constitutes each respondent’s neighborhood. Researchers have used national and state-based data, but this study aims to analyze data at a finer geographic level: the community level, or groupings of locally identified neighborhoods. In previous research, both ZIP codes and groupings of census tracts have been used to study the influence of environment on self-reported health. While ZIP codes were designed for efficient mail delivery, the communities, or parts of communities contained within each ZIP code are often different with respect to socioeconomic characteristics. Census tracts are smaller and typically more similar with respect to socioeconomic conditions; however, attention has not always been paid to grouping census tracts in a way that is not only homogeneous with respect to socioeconomic factors, but also creates boundaries that are meaningful to the community.

The size of the CHSS sample was not large enough to allow for analysis by individual census tracts; however, the sample is large enough to conduct a sub-county, community-based analysis.

Existing demographic and geographic tools described in Table 3.6 were used to combine the county’s census tracts into 29 geographies based on community-defined boundaries, socioeconomic homogeneity, and proximity. The 29 geographies range in population size from 17,355 to 43,336. The geographic areas are shown in Figure 3.2, and detailed descriptions of each are available in Table A.1 in Appendix A.
The census tracts in Hamilton County were analyzed for race, income, home ownership and geographic location and collapsed into 29 geographic units using a number of existing geographic and demographic tools.

### Table 3.4: Major steps in geographic boundary determination

<table>
<thead>
<tr>
<th>Geographic Aggregation</th>
<th>Source Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identify contiguous tracts</td>
<td>2010 U.S. Census Bureau Hamilton County map with census tracts labeled</td>
</tr>
<tr>
<td>2. Identify broad sections of the county with similar</td>
<td>Social Areas of Cincinnati (Maloney &amp; Auffrey, 2013)</td>
</tr>
<tr>
<td>socioeconomic status</td>
<td></td>
</tr>
<tr>
<td>3. Compare tracts on key census data (race, income, home</td>
<td>U.S. Census Data</td>
</tr>
<tr>
<td>ownership)</td>
<td></td>
</tr>
<tr>
<td>4. Aggregate tracts based on areas with local meaning</td>
<td>Hamilton County Planning Department Population by Census Tract table</td>
</tr>
<tr>
<td>(e.g., communities and neighborhoods that go together)</td>
<td></td>
</tr>
</tbody>
</table>

First, contiguous tracts were identified by consulting the census tract maps from the 2010 census (U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau 2010). This contiguity, paired with the census-tract based socioeconomic analysis available for Hamilton County in *The Social Areas of Cincinnati* (see Figure 2 in Maloney & Auffrey 2013), provided the broad sketches for the more precise geographic determination that happened later in the process.

In order to aggregate census tracts in a way that created geographies that were both meaningful for analysis and encompassed areas that had local meaning, the Hamilton County Planning Department’s 1900-2000 Population by Census Tract data table was consulted. The table identified census tracts by neighborhood (e.g., Norwood, Over-the-Rhine, etc.) and changes to census tract boundaries within the county between 2000 and 2010, including a description of whether those changes were major or minor (Hamilton County Regional Planning...
Relationships Between Neighborhoods, Housing, Health

Commission, 2010). There were only eight tract boundary changes that were classified as major, and those changes were all absorbed within the neighborhood groupings defined by this study. This is important because CHSS respondents were connected to census tracts using 2000 census tract boundaries. As a result, each respondent had to be reassigned a census tract based on 2010 boundaries. The fact that all major boundary changes were taken into consideration means that any respondent of the CHSS that lived within one of the tracts with major boundary changes would be correctly classified to the same community regardless of the major boundary change from the 2000 to 2010 census. The same 1990-2010 population by census tract data table was used to identify the Cincinnati neighborhood/township or census designated place. These definitions of neighborhood, socioeconomic data by census tract from the 2010 census and the 2010 Census Bureau map by census tract were used to combine the county’s census tracts into 29 geographies based on community-defined boundaries, socioeconomic homogeneity, and proximity. The geographic areas are shown below in Figure 3.2 with detailed descriptions of each area in Appendix A, Table A.1.
It is important to note that at the center of the study area is the City of Cincinnati. The City is more densely populated than the outer ring suburbs, and there is significant heterogeneity between proximate communities. In particular, in some of the geographic areas created for this study within the urban core, there was wide variation on measures of socioeconomic status (Point 2, Table 3.4) and key census data (Point 3, Table 3.4). For example, Area 20 includes the neighborhoods of Mt. Auburn, CUF, Heights, Corryville, and Clifton. While these neighborhoods are proximate, they have widely varying socioeconomic profiles (from SES II - SES IV in *The Social Areas of Cincinnati* [Maloney & Auffrey, 2013]). The decision was made to focus on contiguous neighborhoods because the author believed the results of this study could be better applied to existing work in the community; both planning and public health practitioners tend to have a neighborhood or community focus and not a focus on homogeneous communities with no attention to proximity. Further, adults living in heterogeneous communities are likely affected by the heterogeneity of their community.
A variety of different ways to combine census tracts based on the data and patterns within the data was considered. The method used in this study was deemed an acceptable method for setting geographic boundaries in a way that created meaningful geographic areas based on community-understood boundaries, homogeneity within geographies, and heterogeneity between geographies. However, using the same guidelines, different permutations of geographic boundaries could be created.

3.4 Measures

Details on the measurement of each of the variables follow, beginning with the outcome variable and individual level measures. Neighborhood-level measures will be discussed in the section that follows.

3.4.1 Outcome variable: Self-reported health status

The outcome variable of interest is self-reported health, specifically the answer to the question, “In general how would you rate your health, excellent, very good, good, fair, poor?” As described above, this variable has a very robust and consistent research history of reliably capturing morbidity and mortality across the world and among many different specific demographic groups; non-response across all studies is consistently low, and for this study all respondents answered this question (Lundberg & Manderbacka, 1996; Dowd & Zajacova, 2007; among others). It is important to note that for the purposes of this study lower numbers on the self-reported health scale represent a better health rating (1= excellent … 5 = fair).

Table 3.5 shows the descriptive statistics for all of the non-categorical variables in the CHSS data set. The first line of the table shows the descriptive statistics for self-reported health, including the mean (2.69), standard deviation (1.104), and skewness (negative). Skewness measures the symmetry of the distribution. When the mean equals the median, the
Relationships Between Neighborhoods, Housing, Health

distribution is not skewed; when the mean is greater than the median, there is a positive skew or right-tailed distribution; and when the mean is less than the median, there is a negative skew or left-tailed distribution. Self-reported health is slightly skewed to the left, and with a distribution that is more flat than a normal distribution. The remainder of the variables in the table will be discussed below.

**Table 3.5: Descriptive statistics for continuous variables in CHSS data**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Std Dev</th>
<th>Mean</th>
<th>Median</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Health Rating</td>
<td>926</td>
<td>1</td>
<td>5</td>
<td>1.104</td>
<td>2.69</td>
<td>3.00</td>
<td>negative</td>
</tr>
<tr>
<td>Current Age</td>
<td>926</td>
<td>18</td>
<td>91</td>
<td>17.156</td>
<td>53.17</td>
<td>54.00</td>
<td>negative</td>
</tr>
<tr>
<td>Chronic Conditions</td>
<td>926</td>
<td>0</td>
<td>10</td>
<td>1.6982</td>
<td>1.8654</td>
<td>2.00</td>
<td>negative</td>
</tr>
</tbody>
</table>

There is discussion in the literature about whether to model self-reported health as an ordinal or continuous variable. Similar studies that have built a multilevel model with self-reported health as the outcome variable have modeled it as an ordinal variable (Wen, Browning, & Cagney, 2003; Browning & Cagney, 2002; and Patel et al., 2003). However, there are a number of researchers who argue that self-reported health has an underlying continuous structure, “the border separating bad from good health is vague and implies continuity” (Manderbacka 1998), and “methods with self-rated health…suggest that self-rated health forms a continuum” (Manor 2000). The author agrees with the concept that the variable has an underlying continuous structure and intends to model it as a continuous outcome variable.

### 3.4.2 Individual (Level-1) predictors

The multilevel model created for this study includes a collection of individual responses from the CHSS on age, gender, race, poverty status, employment status, chronic conditions,
health behaviors, and social capital of each individual adult respondent. Each of these data items will be described in the section that follows.

**Age:** Respondents were asked, “What is your current age?” For this study, age is a continuous variable from 18-95 years old. Any respondent 96 years old or older was categorized as 95 years old. Table 3.5 shows that the mean for age in this sample is 53.2 years. It is slightly negatively skewed, meaning there is a long tail of older respondents and it has a flatter than normal distribution.

**Race:** Respondents were asked, “Which one of the following would you say best represents your race... White, Black or African-American, Asian, native Hawaiian or other Pacific Islander, American Indian or Alaska Native, or some other race?” For the purposes of this study, this variable will be operationalized as a dichotomous variable White or African American/Other (0= African American/Other, 1 = White). The study sample is 39.8% African American and Other, and 59.5% White (this does not total 100% because N=11, or 1.1% of respondents did not provide race). This distribution is different from the actual makeup of the county. According to the U.S. Census, Hamilton County, Ohio, is 69.5% White alone and 25.9% Black alone. The over-representation of African Americans is due to the fact that the survey was specifically designed to oversample the African American population so that analysts would have a powerful enough sample to comment in a meaningful way on African American health in the region. The Other category (N=26) has been combined with African Americans in this sample because of the similarity in response patterns (for a frequency table for this variable please see Appendix B).

**Gender:** Respondents were asked, “For scientific purposes, I need you to please verbally indicate your sex?” Gender is a dichotomous variable male or female (0= male, 1= female). The dataset has more female (63.6%) than male respondents (36.4%), and this varies from the demographic makeup of the county according to the American Community Survey (48.8% male, 51.2% female) (for a frequency table for this variable, please see Appendix B).


**Poverty Level:** Federal Poverty Level (FPL) is a categorical variable that is calculated for each respondent based on their reported income in the last calendar year and the number of people living in their household (0 = below 100% FPL, 1= between 100% and 200% FPL, and 2 = above 200% FPL). This variable had the largest number of missing responses, either because respondents did not answer the question about their income and/or their household size. Each respondent is classified as below 100% FPL, between 100% and 200% FPL, or above 200% FPL. 20.9% of respondents are below 100% FPL, 22.8% are between 100% and 200% FPL, 56.3% are above 200% FPL (for a frequency table for this variable, please see Appendix B).

**Chronic Conditions:** Respondents of the CHSS were asked if a doctor or health care provider ever told them they had any of the following conditions: asthma, cancer, chronic lung disease, diabetes, heart trouble or angina, high blood pressure or hypertension, high cholesterol or triglycerides, stroke, severe allergies, or depression. Each response to the chronic condition question added one point to the respondent’s chronic condition score. Therefore, chronic health condition is a continuous variable from 0-10. For a histogram of the distribution of the chronic conditions variable please see Appendix B. Appendix B shows that this variable is skewed to the lower end of the scale, with most people reporting between one and three chronic conditions. The mean of this variable is 1.86, meaning on average each respondent reported almost two chronic conditions. This variable has a standard deviation of 1.698. Because of the distribution of this variable it was recalculated as a categorical variable (0=0, 1=1, 2=2, and 3= any response over 2) for a histogram of the original and grouped version of chronic conditions please see Appendix B.

**Employment Status:** Respondents to the CHSS were asked about their employment status with the following question:

---

2 100% FPL in 2009 was $14,570 for a family of two and $22,050 for a family of four.
Last week . . . were you working full-time, part-time, going to school, keeping house, or what?" working full-time; working part-time; with a job but not at work because of temporary illness, vacation, strike; unemployed, laid off, looking for work; disabled, too ill to work; retired; in school, keeping house.

For the purposes of this study, employment will be defined as working full-time, part-time or going to school. Retired, keeping house, unemployed, and disabled will all be considered “unemployed." So the variable employed is dichotomous (0 = unemployed, 1 = employed). Forty-eight percent of the respondents were coded as working full-time, 52% were not.

Social Capital or Social Support: Respondents were asked three questions about the social capital in their community:

Now I am going to read to you some statements about your community. Please tell me if you agree or disagree with each statement. First . . . [INSERT QUESTION] . . . do you agree or disagree?”
People can depend on each other in my community.
Living in my community gives me a secure feeling.
People in my community know they can get help from the community if they are in trouble.

The histograms in Appendix B show the distribution of responses for each of the social support questions on the survey. In each case, the variables are ordinal with values from 1-6, with lower values indicating stronger agreement with the statement and therefore more social ties to their community. Each of the three variables has a similar bimodal distribution with more people agreeing strongly or disagreeing with the statements, but not many responses in the middle of the scale. These three variables were all transformed into dichotomous variables 0 if the respondent disagreed with the question (indicating low or no social support) and 1 if they agreed with the statement (indicating the perceived presence of social support in their neighborhood). The recoded dichotomous variables were then combined into a social support scale. The social support scale could have values from zero to three. A respondent would have a zero value if he or she disagreed with all of the social support questions and conversely a three if he or she agreed with all of the questions. The scale was tested using a factor analysis with a varimax rotation, and it returned one factor, indicating that these questions measure one
Relationships Between Neighborhoods, Housing, Health

phenomena. The reliability of the scale was tested using the Kuder-Richardson 20 (KR-20), a modified Cronbach’s Alpha, test of reliability for dichotomous variables. The scale tested reliable with \( \alpha = 0.722 \). For a more extensive discussion of this social support scale please see Dooley (2009).

**Health Behaviors**: Respondents were also asked about their health behaviors, specifically, whether they meet physical activity guidelines, fruit and vegetable consumption guidelines, if they smoke, and how much alcohol they drink.

*Smoking*: Respondents were asked two questions: first, “Have you smoked at least 100 cigarettes in your entire life,” and second, “Do you now smoke cigarettes every day, some days, or not at all?” If they answered yes to the first question and every day or some days to the second question they were considered current smokers for the purposes of this study. Each respondent was coded as 1 = non-smoker, 0 = smoker. Twenty-seven percent of respondents were classified as current smokers.

*Binge Drinking*: Respondents were asked a series of questions related to alcohol consumption. For purposes of this study, only the binge drinking questions were used. Binge drinking is consuming more than 4 drinks on any one occasion in the last 30 days (for women; 5 drinks for men). Respondents were asked, “Considering all types of alcoholic beverages, how many times during the past 30 days did you have 4 (for women; or 5 for men) or more drinks on an occasion?” Each respondent was coded as 1 = non-binge drinker, 0 = binge drinker. Sixteen percent of the respondents were classified as binge drinkers, 84% were not.

*Physical Activity*: Respondents were asked a series of questions on how much moderate and vigorous activity they participate in during their leisure time. Respondents met the physical activity guidelines if they participate in moderate physical activity at least 30 minutes, 5 days per week or at least 20 minutes, 3 days per week of vigorous activity. To establish whether respondents met these guidelines respondents were asked:
1. Now, thinking about the moderate activities you do in a usual week “when you are not working”… do you do moderate activities for at least 10 minutes at a time, such as brisk walking, bicycling, vacuuming, gardening, or anything else that causes some increase in breathing or heart rate?”

2. “How many days per week do you do these moderate activities for at least 10 minutes?”

3. “On days when you do moderate activities for at least 10 minutes at a time, how much total time per day do you spend doing these activities?”

4. “Now, thinking about the vigorous activities you do in a usual week “when you are not working”…, do you do vigorous activities for at least 10 minutes at a time, such as running, aerobics, heavy yard work, or anything else that causes large increases in breathing or heart?”

5. “How many days per week do you do these vigorous activities for at least 10 minutes?”

6. “On days when you do vigorous activities for at least 10 minutes at a time, how much total time per day do you spend doing these activities?”

Fifty-seven percent of respondents did not meet the physical activity guidelines, 43% did meet the physical activity guidelines. Each respondent was coded as 1 = met physical activity guidelines, 0 = did not meet the physical activity guidelines.

Fruit and Vegetable Consumption: Respondents meet the fruit and vegetable consumption standard if they reported eating at least two servings of fruits and three servings of vegetables per day. Respondents were asked, “A serving of vegetables is a half cup of any vegetable (not including potatoes) or 1 cup of salad greens. In the past week, how many servings of vegetables did you eat, on average, daily?” and “A serving of fruit is defined as a half a cup of sliced fruit or one medium piece of fruit. In the past week, how many servings of fruit did you eat, on average, daily?” Each respondent was coded as 1 = if they met the fruit and vegetable consumption requirements, 0 = if they did not meet the fruit and vegetable guidelines. Seventy-five percent of respondents did not meet the fruit and vegetable consumption guidelines, 25% did.
3.4.3 Neighborhood (Level-2) predictors

The multilevel model for this study includes a collection of census data aggregated by census tract to provide neighborhood-level socioeconomic and housing measures for each of the 29 communities in the study. Each of the variables will be defined below:

First from the 2010 census:

1. Race  
   i. Percent White = the percentage of the population that identify themselves as White alone.

2. Housing  
   i. Percent vacant houses = the percentage of housing units that are vacant.  
   ii. Percent renter occupied houses = the percentage of occupied housing units that are occupied by renters.

From the 2006-2010 American Community Survey (ACS):

3. Housing  
   i. Percent mobility = the percentage of housing units where the homeowner moved into the unit in 2005 or Later. This means any homeowner that has moved into the neighborhood in the last five years.  
   ii. Median home value = median home value was calculated for each neighborhood. This calculation will be discussed below.

4. Income  
   i. Percent unemployed = the percentage of the total adult population (over age 16) which report being unemployed.  
   ii. Poverty status = the percentage of the adult population who report income below 200% of the federal poverty level.

A look at the descriptive statistics for the community-level data (see table below) shows that there are observations for each of the 29 communities for each of the variables and variation across each community. The only variables that are not very skewed are the percentage of renter occupied homes and the percentage of residents living between 100% and 200% FPL. The following variables have a relatively flat distribution with more observations in the extreme values than would be expected with a normal distribution (% White, % renter occupied, % unemployed, % 100%-200% FPL, % over 200% FPL).
Table 3.6: Descriptive statistics for neighborhood predictors

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min (%)</th>
<th>Max (%)</th>
<th>Std Dev (%)</th>
<th>Mean (%)</th>
<th>Median (%)</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>% White</td>
<td>29</td>
<td>14.3</td>
<td>74.6</td>
<td>17.95</td>
<td>54.2</td>
<td>60.2</td>
<td>Negative</td>
</tr>
<tr>
<td>% vacant homes</td>
<td>29</td>
<td>5.0</td>
<td>28.9</td>
<td>6.3</td>
<td>11.4</td>
<td>9.3</td>
<td>Positive</td>
</tr>
<tr>
<td>% renter occupied</td>
<td>29</td>
<td>9.9</td>
<td>64.5</td>
<td>15.1</td>
<td>35.1</td>
<td>36.8</td>
<td>Negative</td>
</tr>
<tr>
<td>% mobility</td>
<td>29</td>
<td>19.1</td>
<td>50.4</td>
<td>6.9</td>
<td>31.6</td>
<td>30.4</td>
<td>Positive</td>
</tr>
<tr>
<td>% unemployed</td>
<td>29</td>
<td>4.0</td>
<td>16.0</td>
<td>3.8</td>
<td>8.2</td>
<td>6.0</td>
<td>Positive</td>
</tr>
<tr>
<td>% under 100% FPL</td>
<td>29</td>
<td>3.1</td>
<td>38.8</td>
<td>10.4</td>
<td>13.9</td>
<td>10.4</td>
<td>Positive</td>
</tr>
<tr>
<td>% 100-200% FPL</td>
<td>29</td>
<td>6.3</td>
<td>26.4</td>
<td>5.4</td>
<td>15.3</td>
<td>14.8</td>
<td>Positive</td>
</tr>
<tr>
<td>% over 200% FPL</td>
<td>29</td>
<td>41.2</td>
<td>90.6</td>
<td>15.1</td>
<td>70.8</td>
<td>75.3</td>
<td>Negative</td>
</tr>
<tr>
<td>Median Home Value</td>
<td>29</td>
<td>$88,468</td>
<td>$312,613</td>
<td>$53,388</td>
<td>$156,511</td>
<td>$144,784</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Median home value was used as a community variable. The ACS provides census tract level values Blalock's formula for calculating the median value from grouped data was used (1979, p. 64-65). Home value Interval data on each census tract (e.g., number of homes in that tract with a value between $10,000 and $15,000) was combined and a group median was calculated based on Blalock's formula:

\[ Md = l + \frac{N - F_i}{f} \]

Where

- \( Md \) = group median
- \( N \) = number
- \( F \) = cumulative number of observations corresponding to the lower limit of the range that contains the median value
- \( f \) = number of cases in the interval containing the median
- \( l \) = the lower limit of the interval containing the median value
- \( i \) = the number of observations in the interval containing the median value

for each geographic area in this study.
The measures for each grouping of census tracts were combined with weighted averaging in order to calculate a percentage for each geographic area. For example, count data from the census at the tract level was combined for the variable percent renter and divided by the total housing units in that geographic area in order to calculate the percent renters for each of the study’s geographic areas (a similar calculation was made for percent vacant). Similarly, an aggregate count of new homeowners to the area was created and divided by the total housing units to calculate the variable percent mobility. The variable percent White was calculated by aggregating the counts of the population who identify themselves as White alone divided by the aggregate count of the total population for the same geography. Poverty calculations were made by aggregating the total count of adults under 100% FPL in each of the combined tracts divided by the aggregate of the total number of adults with poverty measures (a similar calculation was made for under 200% FPL).

This chapter covered the data and measures used in this study including: a review of the data sources and a discussion about how the geographic boundaries were determined. The next chapter will discuss in more detail the methods used for this study.
4 Methods

This chapter will cover the methods used in this study. An introduction to multilevel modeling is first followed by a description of a general multilevel model (Level-1, Level-2, and the mixed model). Next is an explanation of variable centering and the implications of centering on a multilevel model. A discussion about how the parameters are estimated follows. This chapter ends with the analytical plan for the study.

4.1 Introduction to multilevel modeling

Traditional regression approaches (OLS) to measuring the effect of the environment on individuals can only handle one level of study – either the individual or the neighborhood. Multilevel methods are designed to deal with data that are hierarchical or nested in nature. Nested data would include individuals nested in schools, work groups within departments, departments within organizations, students within classes or in the case of this study, adults within neighborhoods. Figure 4.1 shows the nested nature of this study’s dataset – individuals within communities.
As described in the data and measures chapter, each level of measurement has a data construct, so for this study there is a dataset that describes individual adult community members and a dataset that describes the community. Multilevel models are also called hierarchical linear models, empirical Bayes’ models, generalized linear mixed models, random cluster-specific approach models, mixed and random effects models, covariance component models, and random coefficient models. For the purposes of this study, they will be called multilevel models.

Self-reported health has been widely studied, and there have been a good number of multilevel models using self-reported health as an outcome variable, similar to this study. A very thorough and widely-cited review of these models by Riva, Gauvin, and Barnett identified 67 multilevel models that looked at adult health between 1998 and 2005 (2007). Thirty-nine of these studies used some form of self-reported health as the outcome variable. “In all but two studies significant associations were found between at least one measure of area socioeconomic status (SES) and self-reported health. More specifically, less favorable area socioeconomic conditions were associated with poorer self-reported health” (Riva, Gauvin, & Barnett, 2007, p. 857). In addition, Riva, Gavin, and Barnett found that community effects on
Relationships Between Neighborhoods, Housing, Health

health were different depending on demographics, “effects of area deprivation on poor health, unhealthy behaviors, and risk of mortality were often greater among low SES individuals and women” (2007, p. 856). Cohen et al. also postulated that the neighborhood effects would be different for people of differing affluence:

*The physical aspects of a neighborhood create opportunities for people to interact and to informally monitor one another’s behavior. Neighborhoods are where people exercise and purchase their foodstuffs and other consumer products (including illegal substances). Local neighborhood resources are likely to be more important for persons of lower income, because more affluent people have greater mobility, allowing them to travel farther to obtain healthful products as well as social support* (2003, p. 467).

Because of the widespread use of self-reported health in previous research, there is a deep body of research to consult for identifying variables that should be included in the model. The multilevel model will include the variables listed in Table 4.1.

<table>
<thead>
<tr>
<th><strong>Table 4.1: Variables in multilevel model</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level-1: 926 individuals</strong></td>
</tr>
<tr>
<td>Outcome variable: self-reported health status</td>
</tr>
<tr>
<td>Predictors:</td>
</tr>
<tr>
<td>CHSS = Age, gender, race, poverty status, employment status, chronic conditions, health behaviors, and social support</td>
</tr>
<tr>
<td><strong>Level-2: 29 communities</strong></td>
</tr>
<tr>
<td>Predictors:</td>
</tr>
<tr>
<td>Housing Conditions</td>
</tr>
<tr>
<td>Census = Median home value, % vacant, % renter occupied</td>
</tr>
<tr>
<td>Neighborhood Conditions</td>
</tr>
<tr>
<td>Census = racial makeup, poverty rate, mobility (% moved in 2005 or later), employment status</td>
</tr>
</tbody>
</table>

Multilevel modeling has many strengths; first and foremost, this modeling strategy recognizes the fact that proximity causes similarities. For example, students in the same classroom may be more similar or react more similarly because they share the same classroom experience. In traditional regression models, if researchers want to include variables from
Relationships Between Neighborhoods, Housing, Health

multiple levels (e.g. individual and community) they are forced to aggregate the data to the level of analysis, either community data averages attached to individual records or individual data summarized at the community level. This aggregation ignores variation at one level when the researcher chooses to analyze at a different level. When the data are clustered in some way, they are not truly independent of each other, and if this is ignored, standard error measures will be misestimated. This means that the t ratios and p values will be wrong (likely too small), and may lead one to conclude that there is a higher level of precision than is actually the case.

Experts have suggested that as little as 1-2% of similarity can disrupt standard errors (Sayer, 2013). And since the influence of individual differences is always greater than the neighborhood effects, if researchers do not separate out the individual effects variables from the neighborhood effects variables, the neighborhood effects are unlikely to be found (Sayer, 2013).

Multilevel models allow for variation within clusters, in this case between adults living in the same community, and between communities. Multilevel modeling allows for unbalanced data across Level-2, so in this case each community has a different number of responses (from N=16 to N=295; this wide variation is due to the oversampling of several small communities in the original survey design), and that is allowed within this type of model. Multilevel models also allow for the use of any combination of categorical and continuous predictors. Modeling the nested data of this study in a more standard OLS model would result in a violation of the assumption of independence, meaning the data are not independent.

Multilevel modeling is not the only way to address some of these problems. The assumption that all coefficients are the same in all neighborhoods or communities can be avoided by using an analysis of variance (ANOVA) or an analysis of covariance (ANCOVA), but these techniques require additional dummy variables for each neighborhood. For large samples this would mean hundreds of additional regression terms and lead to regression equations that are neither efficient nor parsimonious (Duncan, 1998). In the case of this study, each of the 29
Relationships Between Neighborhoods, Housing, Health

neighborhoods would require a dummy variable (N = 28) for any Level-2 data points. Multilevel modeling is able to deal simultaneously with micro data, in this case individual response data, and macro data, in this case neighborhood-level data. Because multilevel models capture variation at both the individual level and the community level, “the technique provides a way of showing how, and for which types of people, contexts [neighborhood] effects matter” (Duncan, 1998, p. 102). Research that encompasses both planning and health is increasingly recognizing the role that physical environment plays in influencing health. This perspective (neighborhood-level characteristics’ influence on individual health outcomes) is fundamentally multilevel.

Multilevel modeling does have some weaknesses. One is that the dataset needs to be fairly robust. The rule of thumb is 30 groups with at least 30 respondents in each group. There are also some computational decisions that must be carefully made when using multilevel modeling, how variables are modeled at Level-2 and centering (both discussed later in this chapter).

4.2 Description of a general multilevel model³

This study will use a two-level multilevel model. As the name suggests the model includes two sub-models, a Level-1 model (individuals) and a Level-2 model (neighborhoods). This type of model is called a higher-order multilevel model and is the most common type of multilevel model (Tabachnick & Fidell, 2013). There are multiple ways to model the data when building a multilevel model. This section will describe a basic two-level model with one predictor at Level-1 (\(X_{ij}\)) and one predictor at Level-2 (\(W_j\)).

Relationships Between Neighborhoods, Housing, Health

The self-reported health of the $i^{th}$ person living in the $j^{th}$ neighborhood is $Y_{ij}$. Assuming that self-reported health follows a normal distribution, the Level-1 model can be defined as:

**Level-1 model (individuals)**

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \quad [4.1]$$

Where

- $\beta_{0j}$ is the intercept for the outcome variable ($Y_{ij}$) for neighborhood $j$;
- $\beta_{1j}$ is the slope for the relationship in neighborhood $j$ between the outcome variable ($Y_{ij}$) and the Level-1 predictor ($X_{ij}$);
- $X_{ij}$ is a Level-1 predictor for adult $i$ in neighborhood $j$; and
- $r_{ij}$ is the Level-1 random effect or error term.

We assume that the error term is normally distributed with a mean of zero and variance of $\sigma^2$: $r_{ij} \sim N(0, \sigma^2)$.

**Level-2 model (neighborhoods)**

Each of the two Level-1 coefficients, ($\beta_{0j}, \beta_{ij}$) that are included in the Level-1 model become an outcome variable in the Level-2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \quad [4.2]$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \quad [4.3]$$

Where

- $\gamma_{00}$ is the overall intercept, or the grand mean, of $Y_{ij}$;
- $\gamma_{01}$ is the slope between a Level-2 predictor ($W_j$) and $Y_{ij}$;
- $\gamma_{10}$ is the slope between a Level-1 predictor ($X_{ij}$) and $Y_{ij}$;
- $W_j$ is a Level-2 predictor;
- $u_{0j}$ is the unique effect of neighborhood $j$ on the intercept, or the random error for the deviation of the intercept for a specific group from the overall intercept; and
\( u_{ij} \) is the random error or error term for the slope.

We assume \( u_{0j} \) and \( u_{1j} \) are random variables with zero means, variances (\( \tau_{00} \) and \( \tau_{11} \)) and covariance (\( \tau_{01}, \tau_{10} \)).

Substituting equations 4.2 and 4.3 into 4.1 gives us the mixed or combined model (4.4).

**Mixed or Combined model**

\[
Y_{ij} = \gamma_0 + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}X_{ij}W_j + u_{0j} + u_{1j}X_{ij} + r_{ij} \quad [4.4]
\]

Where

\( \gamma_{10}X_{ij} \) is a Level-2 predictor (\( \gamma_{10} \)) multiplied by a Level-1 predictor (\( X_{ij} \))

\( \gamma_{01}W_j \) is a Level-2 predictor (\( \gamma_{01} \)) multiplied by a Level-2 predictor (\( W_j \))

\( \gamma_{11}X_{ij}W_j \) is a Level-2 predictor (\( \gamma_{11} \)) multiplied by the combination of Level-2 and Level-1 predictors (\( X_{ij}W_j \)); this is a cross-level interaction term

\( u_{0j} + u_{1j}X_{ij} + r_{ij} \) are the random error or error term components.

Variance components

\( \tau_{00} \) is the variance among the random intercepts

\( \tau_{11} \) is the variance among the slopes

\( \tau_{10}, \tau_{01} \) are the covariance between slopes and intercepts

Also represented as:

\[
Var\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = T \quad [4.5]
\]

A look at the error structure of Equation 4.4 should help the reader understand the difference between a multilevel model and an OLS model. In OLS, in order to have efficient estimation and accurate hypothesis testing, random errors have to be independent, normally distributed with a constant variance. The random error part of Equation 4.4 has a much more complicated form (\( u_{0j} + u_{1j}X_{ij} + r_{ij} \)). The errors have unequal variances because the variance term (\( u_{0j} + u_{1j}X_{ij} + r_{ij} \)) includes parts (\( u_{0j} + u_{1j} \)) that vary across neighborhoods (Level-2) and...
Relationships Between Neighborhoods, Housing, Health

parts \((X_{ij} + r_{ij})\) that vary across individuals (Level-1). This is why standard regression is not appropriate, but iterative maximum likelihood procedures (described later) can be used (Bryk & Raudenbush, 2002). Bryk and Raudenbush also note that if \(u_{0j} + u_{1j}\) were zero for each neighborhood \((j)\), Equation 4.4 would be equal to an OLS regression model (2002).

It is important to note that each Level-1 coefficient can be modeled in several different ways at Level-2. That is, Equations 4.2 and 4.3 can take on the following forms:

1. A fixed model:
   \[
   \beta_{qj} = \gamma_{q0} \tag{4.6}
   \]
   The fixed model constrains the Level-1 variable as nonrandom or fixed across every Level-2 neighborhood.

2. A randomly varying Level-1 coefficient model:
   \[
   \beta_{qj} = \gamma_{q0} + u_{qj} \tag{4.7}
   \]
   In this case the parameter is modeled as a function of an average value, \((\gamma_{q0})\), plus the random error term \((u_{qj})\) associated with each Level-2 group.

3. A model with a Level-1 coefficient with both non-random and random sources of variation:
   \[
   \beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj} \tag{4.8}
   \]
   Where the parameter, \((\beta_{qj})\), is modeled as a function of an average value \((\gamma_{q0})\), plus an additional predictor \((W_{sj})\) and an error term \((u_{qj})\). In some cases, the edition of the predictor leaves a negligible amount of residual variation. In this case it would make sense to model the parameter as nonrandomly varying as show in Equation 4.9.

4. A model with non-randomly varying Level-1 coefficient:
   \[
   \beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} \tag{4.9}
   \]
The Level-2 equation set will generally include multiple equations, and the type of Level-1 coefficient modeling could be different for each Level-2 equation. The actual dimensions of the covariance matrix for the study depends on the number of Level-1 coefficients that vary randomly. It is also important to note that a different set of Level-2 predictors can be used for each equation in the Level-2 model. These decisions depend on the model specifications and theoretical development of the model.

### 4.3 Centering

Bryk and Raudenbush describe the importance of centering, “In all quantitative research, it is essential that the variables under study have precise meaning so that statistical results can be related to the theoretical concerns that motivate the research. In the case of (hierarchical linear models) multilevel models, the intercept and slopes in the Level-1 model become outcome variables at Level-2. It is vital that the meaning of these outcome variables be clearly understood” (1992, p. 25). In particular, if the zero, or intercept, value of a variable is not meaningful, then the researcher may want to transform that variable so its intercept or zero is meaningful. Centering is the process of transforming an X variable by subtracting a meaningful constant, usually the group or grand mean (Luke, 2004). The most common form of centering is grand mean centering. For example, centering X based on its grand mean, or adjusted for Level-2 means, would look like:

\[ X'_{ij} = (X_{ij} - \bar{X}_{..}) \]  

Where \( \bar{X}_{..} \) is the grand mean.

and now \( X'_{ij} \) is interpreted as a deviation from the mean. For example, in this study if age is grand mean centered, the centered X variable would be interpreted as the value for a respondent of average age. Centering does not affect the underlying relationships between the
variables; it can simply make interpreting the results more in line with real-world relationships. For example, without centering, the variable age could take on a “0” value in this study. Only adult data were included in this study, and the age value in the actual data set could only take on values between 18 and 95. Centering the age variable makes interpreting the age variable more in line with the possible responses within the dataset.

Without centering, the intercept, $Y_{ij}$, would be interpreted as the expected value of the outcome variable when all of the predictors, ($\beta$) are zero. However, as described in the previous paragraph, a zero value is not always a meaningful value for some predictors (e.g. age). When the variables are centered, the equation can be easier to understand, because the intercept represents the expected value when each predictor is equal to its mean (Paccagnella, 2006). The only value that changes when centering is the intercept ($Y_{ij}$).

### 4.4 Parameter estimation

There are three different types of parameters estimated in the multilevel model. The empirical Bayes (EB) estimates of randomly varying Level-1 coefficients; generalized least squares (GLS) estimates of the Level-2 coefficients; and maximum likelihood estimates (ML) of the variance and covariance components. Each of these parameter estimation methods takes into consideration the number of observations for each neighborhood when calculating the parameters. Neighborhoods with a larger number of observations provide more information and as a result more precision.

#### 4.4.1 Level-1 coefficients, $\beta_{qj}$

The Level-1 coefficient estimates are empirical Bayes estimates of randomly varying Level-1 coefficients, $\beta_{qj}$. It is important to note that the use of the Bayesian approach means that a very different assumption than OLS is being made about the underlying structure of the
probability distribution. In classical statistics (OLS) the underlying distribution is based on repeated samples of the data, however, the real distribution is unknown. Bayesians, in contrast, produce a posterior density function, based on the researcher’s prior knowledge of the variable in question and the degree of confidence they have related to various potential outcomes (Kennedy, 2004). The density function relates to the predicted value of the sample data and is not a sampling distribution (as is the case in the OLS model). For this study the empirical Bayes approach is being used; this means that the posterior density function is being constructed from the data themselves (Bryk & Raudenbush, 2002). One of the advantages of Bayesian estimators is that their variance is generally smaller because of the incorporation of additional information; the posterior density function being the additional information (Kennedy, 2004). In fact, Kennedy points out that, “an important feature of the Bayesian approach is the prior information is incorporated in an explicit fashion” (2004, p. 215).

The Level-1 empirical Bayes (EB) coefficient estimates for each unit \( j \) are based-on composite estimates from the data from neighborhood \( j \) and estimates based on data from all neighborhoods (in the study, see Equation 4.11). In comparison, in an OLS model the Level-1 intercept, \( \beta_{0j} \), is calculated for each neighborhood individually. The less reliable the \( Y \) (reliability is discussed in Equations 4.11-4.11a) in a particular neighborhood, the more it can be helpful to use the grand mean (the mean of all neighborhoods) as our estimate. This modeling method is “borrowing strength” from the complete study data set to improve the Level-1 (EB) estimates for each neighborhood (Sayer, 2012; Raudenbush et al., 2000; and Verbitsky, 2007). Empirical Bayes estimates are also called “shrunken estimates” of the Level-1 coefficients (Raudenbush et al., 2000; Verbitsky, 2007; Luke, 2004). This is calculated using the following formula:

\[
\hat{\beta}_{0j}^{EB} = \lambda_j \hat{\beta}_{0j}^{OLS} + (1 - \lambda_j) \hat{\gamma}_{00} \quad [4.11]
\]

Where
\( \hat{\beta}_{0j}^{EB} \) is the empirical Bayesian estimate of \( \beta_{0j} \); 
\( \hat{\beta}_{0j}^{OLS} \) is the OLS estimate of \( \beta_{0j} \); 
\( \hat{\gamma}_{00} \) is the grand mean of \( Y \); and
Relationships Between Neighborhoods, Housing, Health

\[ \lambda_j \] is the reliability or weight of Y in group j. The formula for reliability is:

\[ \lambda_j = \frac{\sigma_{u0}^2}{\left(\sigma_{u0}^2 + \frac{\sigma_r^2}{n_j}\right)} \]  \[4.11a\]

Where

\[ \sigma_{u0}^2 \] is the Level-2 variance or error term;
\[ \sigma_r^2 \] is the Level-1 variance or error term; and
\[ n_j \] is the number of Level-1 respondents per neighborhood.

This shows that the reliability for a particular neighborhood is directly connected to the number of Level-1 respondents (\( n_j \)) in that neighborhood. Returning to Equation 4.11, if the reliability of \( \hat{\beta}_{0j} \) is high (\( \lambda_j \) close to 1), the Empirical Bayesian estimate of \( \hat{\beta}_{0j} \) will be very close to the OLS estimate for that particular neighborhood. The same EB estimation techniques are used to calculate \( \hat{\beta}_{1j} \). If the reliability is low, then the Bayesian estimate will be close to the value for the grand mean of Y (\( \hat{\beta}_{00} \) in the Equation 4.11). The Bayesian estimate will always be between the neighborhood mean and the grand mean; however, as reliability gets worse, the estimate will “shrink” toward the grand mean, hence the name “shrinkage estimates” (Luke, 2004).

4.4.2 Level-2 coefficients, \( \gamma_{qs} \)

Substituting the Level-2 equations for each \( \beta_{qj} \) into the corresponding Level-1 terms, this produces a single linear equation with a complex error structure (consisting of the Level-1 error, \( r_{ij} \), and the Level-2 error, \( u_{qi} \)). As a result of this complex error, estimation of the \( \gamma_{qs} \) requires consideration of the differing reliability or number of observations for each neighborhood. This is done by using a generalized least squares (GLS) estimates for the \( \gamma^s \).

The following equations explain the Level-2 coefficient estimation process. Beginning with the simplest form of the model with no additional variables at Level-1 or Level-2: the null model or the one-way ANOVA (see Equations 4.1-4.3 for specific details on each coefficient):
Relationships Between Neighborhoods, Housing, Health

\[ Y_j = \beta_{0j} + \bar{r}_j \]  \hspace{1cm} [4.12]

Where

\( Y_j \) is the mean for neighborhood \( j \), or the group mean; and
\( \bar{r}_j \) is the mean of error terms for neighborhood \( j \) and is defined as
\[ \bar{r}_j = \frac{\sum r_{ij}}{n_j} \]

The Level-2 equation also does not include any additional explanatory variables (see equation 4.7). Substituting equation a simple equation for \( \beta_{0j} \) ( \( \beta_{0j} = y_{00} + u_{0j} \) ) into Equation 4.12 gives us the combined model:

\[ Y_j = y_{00} + u_{0j} + \bar{r}_j \]  \hspace{1cm} [4.13]

We are assuming that \( u_{0j} \) and \( r_{ij} \) are normally distributed, and the variance of \( Y_j \) is

\[ \text{Var}(Y_j) = \tau_{00} + V_j = \Delta_j \]  \hspace{1cm} [4.14]

Where:

\( V_j \) is the error variance
\( \Delta_j \) is the parameter variance and the error variance combined (see 4.16)
\[ V_j = \sigma^2/n_j \]  \hspace{1cm} [4.15]

If the dataset were balanced, meaning that every neighborhood had the same sample size then \( \Delta_j \) would be equal in each group and the unique, minimum-variance, unbiased estimator of the Level-2 coefficients (the \( \gamma_{qs} \)) would be the OLS estimator. However, given the unbalanced nature of the study data the \( Y_j \)s will have different variances:

\[ \Delta_j = \tau_{00} + V_j \]  \hspace{1cm} [4.16]

\( = \) parameter variance + error variance

It is important to note that while the parameter variance (\( \tau_{00} \)) is equal across neighborhoods (Level-2 units) the error variance (\( V_j \)) is different depending on the number of observations within each neighborhood. As a result the unique, minimum-variance, unbiased estimator of the Level-2 coefficients (the \( \gamma_{qs} \)) would be the weighted least squares estimator
where each neighborhood’s data is weighted in proportion to its precision $\Delta_j^{-1}$. This value shown in equation 4.17 is called the maximum likelihood (ML) estimator.

$$\hat{r}_{00} = \Sigma \Delta_j^{-1} \bar{Y}_j / \Sigma \Delta_j^{-1}. \quad [4.17]$$

This same logic can be used to explain the estimation of the Level-2 coefficients in more complex models using matrices.

Level-1:
$$Y_j = X_j \beta_j + r_j \quad [4.18]$$

Level-2:
$$\beta_j = W_j \gamma + u_j \quad [4.18a]$$

And the combined model:
$$Y = X_i W_j \gamma + X_i u_j + r_j \quad [4.18b]$$

Where:

- $W_j$ is a $(Q+1)$ by $F$ matrix of predictors, $\gamma$ is an $F$ by 1 vector of Level-2 coefficients (fixed effects), $r_j$ is a $F$ by 1 vector of Level-1 errors or random effects, $u_j$ is a $(Q+1)$ by 1 vector of Level-2 errors (or random effects). We assume that the error term $u_j$, is multivariate normal with a mean of 0 and a variance $= \tau_{qq}$. For any combination of random effects, the variance and covariance components are combined into a $(Q+1)$ by $(Q+1)$ dispersion matrix, $T$.

The dispersion of $\beta_j$ is:

$$\text{Var}(\beta_j) = \text{Var}(u_j + e_j) = T + \text{Var}(u_j, v_j) = \Delta_j \quad [4.19]$$

As above (4.16),

$$T = \text{parameter dispersion} + \text{error dispersion}$$

Because the data are not balanced, $\Delta_j$ will differ across groups, so the best estimator of $\gamma$ is the generalized least squares estimator (GLS; Bryk, & Raudenbush, 2002) given by:

$$\hat{\gamma} = (\sum W_j' \Delta_j^{-1} W_j)^{-1} \sum W_j' \Delta_j^{-1} \hat{\beta}_j \quad [4.20]$$
The GLS estimator weights each neighborhood by $\Delta_j^{-1}$, that neighborhood's precision matrix, which is also the inverse of the variance-covariance matrix (Bryk & Raudenbush, 2002).

It is important to note that with OLS the solution is minimizing the sum of squared residuals. However, with GLS the solution is minimizing a weighted sum of squared residuals. Kennedy describes GLS as, “Observations that are expected to have large residuals because of the variances of their associated disturbances are known to be large are given a smaller weight. Observations whose residuals are expected to be large because other residuals are large are also given smaller weights” (2004, p. 117).

4.4.3 Maximum likelihood estimates of variance and covariance components, $\sigma^2$ (Level-1) and $T$ (Level-2)

“When designs are unbalanced (as is typically the case), iterative numerical procedures must be used to obtain efficient estimates, usually via maximum likelihood” (Bryk & Raudenbush, 1992, p. 44). This study has an unbalanced data set, meaning the number of respondents in each neighborhood is different as are the distributions of the predictors in each neighborhood. Traditional methods for variance and covariance estimation will not yield efficient estimates given the unbalanced nature of the data.

Maximum likelihood estimates of a set of parameters is simply the set of parameters that gives the highest probability of obtaining the observed data (Kennedy, 2004). HLM 7, the software utilized for this study, “…uses an expectation maximization (EM) algorithm with Fisher scoring used for every fifth iteration” (Garson, 2013, p. 28). Raudenbush and Bryk point out that the estimates for Level-1 and Level-2 coefficients require knowing the variance components, while estimation of the variance components requires knowing both levels of coefficients – this interrelationship suggests the need for an iterative process (2002). The EM assumes that $Y_{ij}$ is known and $u_j$ is not. If, however, $u_j$ was known estimation of the variance components would
be straightforward. The equations below illustrate the EM process for a null model or an ANOVA.

\[
\hat{\tau} = \frac{\sum_{j=1}^{J} u_j^2}{J} \quad [4.21]
\]

\[
\sigma^2 = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - u_j)}{\sum_{j=1}^{J} n_j} \quad [4.22]
\]

However, \( u_j \) is not known, so it must be estimated based on the initial estimates of parameters.

For the purposes of this exercise, the superscript \( (0) \) indicates an estimate. Each \( u_j \) is assumed to be normally distributed with mean \( u_j^* \) and variance \( V_j^* \).

\[
u_j^* = \lambda_j^{(0)} (\bar{Y}_j - \gamma^{(0)}) \quad [4.23]
\]

\[
V_j^* = \tau^{(0)} (1 - \gamma_j^{(0)}) \quad [4.24]
\]

The E step (“expectation” of the EM algorithm) creates estimates based on the value of \( Y \) and the current parameter estimates (equations 4.25, 4.26, and 4.27).

\[
E[\sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - u_j)] | Y, \gamma^{(0)}, \tau^{(0)}, \sigma^2^{(0)}] = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - u_j^*) \quad [4.25]
\]

\[
E[\sum_{j=1}^{J} u_j^2 | Y, \gamma^{(0)}, \tau^{(0)}, \sigma^2^{(0)}] = \sum_{j=1}^{J} (u_j^* + V_j^*) \quad [4.26]
\]

\[
E \left[ \sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - \hat{\gamma} - u_j)^2 | Y, \gamma^{(0)}, \tau^{(0)}, \sigma^2^{(0)} \right] = \sum_{j=1}^{J} \sum_{i=1}^{n_j} ((Y_{ij} - \gamma^0 - u_j^*)^2 + V_j^*) \quad [4.27]
\]

The M step (“maximization” of the EM algorithm) substitutes these estimates into the Equations 4.23 and 4.24. Verbitsky describes this process well, “the algorithm iterates between the E-step and M-step until the difference in observed-data log likelihood between two consecutive iterations falls below some specified tolerance level” (2007, p. 13). Using maximum likelihood multilevel modeling chooses the estimates of \( \gamma, \tau, \) and \( \sigma^2 \) such that the likelihood of observing the actual data \( Y \) is maximized (Bryk & Raudenbush, 2002).
4.5 Analytical plan

This methodological summary only covers the types of models used in this study (see Table 4.2). This study includes three models – Model 1: the unconditional model; Model 2: the one-way ANCOVA with Random Effects; and Model 3: an extension of the random-effects ANCOVA with Level-2 covariates.

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Parameters at Level-1 besides intercept ($\beta_{1j}$)</th>
<th>Parameters at Level-2 ($W_j$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-way ANOVA with Random Effects</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Null model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully unconditional model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>One-way ANCOVA with Random Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Extension of the random-effects ANCOVA with Level-2 covariates</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For this study, three models will be tested, which represent three hypotheses. Model 1 is called the unconditional model, because there no parameters at Level-1 or 2. The unconditional model is a typical first model when using multilevel modeling techniques. It will measure the proportion of variance in the outcome variable (self-reported health) that is accounted for by the grouping variable (neighborhood). This means that self-reported health and its variation across neighborhoods will be estimated using an intercept-only model, and no other predictors. If there is no variation in self-reported health by neighborhood, then there is no need for a multilevel model; a simple one-level regression model would suffice.
Model 2 includes just the Level-1 predictors. This Level-1 model with seven variables is designed to be an inclusive model (Bingenheimer & Raudenbush, 2004). Chapter 2 presented the research to suggest that each of these variables have been strongly linked to health outcomes. The variables in the model include chronic conditions, race, non-smoking, physical activity, social support, employment, and poverty status. Age, gender, fruit and vegetable consumption and binge drinking were dropped from the model; this is discussed in greater detail in the next chapter. This model will test the amount of variance in self-reported health that is explained with just the study’s Level-1 predictors.

For each model, goodness of fit will be tested using a deviance score. The deviance score is obtained by multiplying the natural log of the likelihood by minus two (-2LL). The deviance score is a measure of goodness of fit between the model and the data (Luke, 2004). A smaller deviance score indicates a better fitting model. The deviance score does not have meaning on its own, but can only be used to compare one model that is nested within another (meaning all of the variables from one model appear in another). In the case of this study, all of the model variables in Model 1 (the unconditional model) appear in Model 2, so the deviance score can be used to evaluate goodness of fit. This will be discussed in the next chapter.

Model 3 is the final model and includes all of the Level-1 predictors (exactly like Model 2) and the best fitting neighborhood/housing characteristics. This model tests the amount of variance that is explained when we add all of the Level-1 predictors as well as all of the Level-2 characteristics. At Level-2, the only neighborhood variables that were significant were percent mobility and percent renter. The variables racial makeup, under 100% FPL, unemployment rate, median home value and percent vacant were dropped from the model. This will be discussed in greater detail in the following chapter.
5. Findings and Discussion

This chapter will cover the findings from this study, along with some discussion of the implications of the results. A methodological summary is first followed by the final multilevel model and the hypotheses tested. Next is a discussion of the model assumptions and error checking. The chapter ends with the interpretation and discussion of the results from the model.

5.1 Methodological summary

This section will describe in detail each phase of multilevel model building and the modeling decisions that were made to best fit the data.

5.1.1 Model 1: Unconditional model

As described in the previous chapter the first model is called the unconditional model because it measures the proportion of the variance in the outcome variable (self-reported health) that is accounted for by the grouping variable (neighborhood). The model takes the following form:

Model 1: Unconditional Model

\[ H_0: \text{There is no variation in self-reported health by neighborhood} = \text{Interclass correlation coefficient (ICC; explained below)} = 0, \ u_{0j} \text{ is not significant (this means that the intercept (}\beta_{0j}\text{), self-reported health, is not significantly affected by neighborhood association).} \]

Level-1:

\[ SRH_{ij} = \beta_{0j} + r_{ij} \quad [5.1] \]

Where

\( SRH_{ij} \) is the self-reported health status for the \( i^{th} \) adult in the \( j^{th} \) neighborhood;
\( \beta_{0j} \) is the mean self-reported health in the \( j^{th} \) neighborhood; and
\( r_{ij} \) is the Level-1 random effect.

Level-2:

\[
\beta_{0j} = \gamma_{00} + u_{0j} \tag{5.2}
\]

Where

\( \gamma_{00} \) is the grand mean;

\( u_{0j} \) is the Level-2 random effect for neighborhood \( j \).

The unconditional model is used to calculate the interclass correlation coefficient (ICC). The ICC is used to justify multilevel modeling and test for compositional effect; that is the variability between neighborhoods rather than within neighborhoods. This is also called a one-way ANOVA model. This model demonstrates how much of the outcome variability is at Level-1 and 2; the \( \sigma^2 \) represents the variability between individuals within their neighborhood and the \( \tau_{00} \) represents the between-neighborhood variability (Bryk & Raudenbush, 1992). The ICC is given by the formula:

\[
\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} \tag{5.3}
\]

For this study, the estimated Level-1 variance is 1.14 and it is 0.066 for Level-2. Therefore the ICC = \( (0.066/(0.066+1.14)) = 0.055 \); this suggests that neighborhood composition accounts for 5.5% of the variability in self-reported health among adults in this study. This ICC is in line with previous studies (see discussion at the end of Section 2.4). While not large, this ICC still suggests that multilevel modeling is necessary. Research has shown that ignoring the non-independence with even an ICC as low as 5% may lead to inflated Type I error (Julian, 2001; Barcikowski, 1981). Examples of research studies with small ICCs include: Myers et al...
The Level-2 random effect ($u_{0j}$) is 0.066 and is significant (p-value <0.01). This means that self-reported health is significantly affected by neighborhood and that the null hypothesis is rejected, that neighborhood does not affect their residents’ self-reported health. In addition, the estimated grand mean across neighborhoods, $\gamma_{00}$, was 2.56 and was significant. This means that on average people across all neighborhoods rate their health between very good (2) and good (3). This model also calculates the initial deviance statistic (-2LL) and other coefficients used as a comparison to more complex models. The deviance score for this model (discussed in the previous chapter) is 3877.44 and a more complex model with more predictors would be expected to reduce significantly this deviance value indicating an overall improvement of the model fit.

5.1.2 Model 2: Level-1 model/One-way ANCOVA with random effects

In this section and the one that follows, the model refinement steps will be described in detail. Both Model 2 and 3 required several iterations to arrive at the best fitting final model.

Model 2 includes just the Level-1 predictors. This Level-1 model provides the explanatory variables for individual characteristics. The initial version of this model included 11 variables, all of which were allowed to vary across neighborhoods. This model takes the following form:

Model 2: Level-1 parameters only

$H_0$: individual characteristics do not increase the fit of the model. There is no significant difference between the model 1 and model 2 ($\chi^2$ test of model deviances).
Level-1:

\[ SRH_{ij} = \beta_{0j} + \beta_{1j}Race_{ij} + \beta_{2j}Employed_{ij} + \beta_{3j} (\text{Poverty}_{ij} - \overline{\text{Poverty}}_{..}) + \beta_{4j}NonSmoking_{ij} + \beta_{5j}PhysicalActivity_{ij} + \beta_{6j} (\text{Chronic Conditions}_{ij} - \overline{\text{Chronic Conditions}}_{..}) + \beta_{7j} (\text{Social Capital}_{ij} - \overline{\text{Social Capital}}_{..}) + \beta_{8j} \text{Fruit and Vegetable Consumption}_{ij} + \beta_{9j} \text{Binge Drinking}_{ij} + \beta_{10j} (\text{Age}_{ij} - \overline{\text{Age}}_{..}) + \beta_{11j} \text{Gender}_{ij} + r_{ij} \]

Where

- \( \beta_{0j} \) = mean self-reported health in the \( j \)th neighborhood after controlling for the other variables in the model
- \( \beta_{qj} (q = 0, 1, ..., Q) \) are Level-1 coefficients

\( (\text{Poverty}_{ij} - \overline{\text{Poverty}}_{..}) \) = Poverty status grand mean centered; note all of the non-binary ordinal variables are grand mean centered. Centering was discussed in greater detail in the previous chapter.

Level-2:

\[ \beta_{0j} = \gamma_{00} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + u_{1j} \]
\[ \beta_{2j} = \gamma_{20} + u_{2j} \]
\[ \beta_{3j} = \gamma_{30} + u_{3j} \]
\[ \beta_{4j} = \gamma_{40} + u_{4j} \]
\[ \beta_{5j} = \gamma_{50} + u_{5j} \]
\[ \beta_{6j} = \gamma_{60} + u_{6j} \]
\[ \beta_{7j} = \gamma_{70} + u_{7j} \]
\[ \beta_{8j} = \gamma_{80} + u_{8j} \]
\[ \beta_{9j} = \gamma_{90} + u_{9j} \]
\[ \beta_{10j} = \gamma_{100} + u_{100j} \]
\[ \beta_{11j} = \gamma_{110} + u_{11j} \]
Where
\[ \gamma_{qs} (q = 0, 1, \ldots, S_q) \] are Level-2 coefficients;
\( u_{ij} \) are Level-2 random effects.

Running this model yielded a collection of individual parameters (\( \beta s \)) that were not statistically significant (gender, binge drinking, fruit and vegetable consumption, and age); because these four parameters were not statistically significant, the corresponding variables were all dropped from the model.

The model was re-run without the four variables with insignificant parameters described above (gender, binge drinking, fruit and vegetable consumption, and age). Results from this run of the model showed that most of the variance components (\( u_{ij} \)) were not significant. This suggests that most of the variables did not vary significantly at Level-2, or across neighborhoods. As a result, the variance term at Level-2 was removed for each of the following variables: employment, poverty, race, and chronic disease. Model 2 was run for a third time, and again the results suggested little variation across neighborhoods for Level-1 characteristics, so the remaining variance components (the \( u_{ij} \)'s for non-smoking and physical activity) were fixed at Level-2. This was the final Level-1 model: seven variables all (\( u_{ij} \)'s) fixed (or non-variable across neighborhoods) at Level-2. The final equations for Model 2 takes the following form:

**Model 2: Level-1 parameters only**

**Level-1:**

\[
SRH_{ij} = \beta_{0j} + \beta_{1j} \text{Race}_{ij} + \beta_{2j} \text{Employed}_{ij} + \beta_{3j} (\text{Poverty}_{ij} - \overline{\text{Poverty}}) + \beta_{4j} (\text{Non-Smoking}_{ij} - \overline{\text{Non-Smoking}}) + \beta_{5j} (\text{Physical Activity}_{ij} - \overline{\text{Physical Activity}}) + \beta_{6j} (\text{Chronic Conditions}_{ij} - \overline{\text{Chronic Conditions}}) + \beta_{7j} (\text{Social Capital}_{ij} - \overline{\text{Social Capital}}) + r_{ij}
\]
Relationships Between Neighborhoods, Housing, Health

Level-2:
\[ \beta_{0j} = \gamma_{00} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} \]
\[ \beta_{2j} = \gamma_{20} \]
\[ \beta_{3j} = \gamma_{30} \]
\[ \beta_{4j} = \gamma_{40} \]
\[ \beta_{5j} = \gamma_{50} \]
\[ \beta_{6j} = \gamma_{60} \]
\[ \beta_{7j} = \gamma_{70} \]

Where
\[ u_{ij} \] are Level-2 random effects. It is important to note that all of the Level-1 parameters, with the exception of the intercept, \( \beta_{0j} \), were modeled as fixed effects (no error term) at Level-2.

The deviance score for this model was 2313.92, a significant improvement over the fit value for the unconditional model (3877.44, \( p<0.01 \)). The final version of Model 2 also fit the data better than the original Model 2 (with all 11 variables all varying at Level-2).

5.1.3 Model 3: Final model/Extension of the random-effects ANCOVA with Level-2 covariates

After identifying the Level-1 model that best fit the data (Model 2), Model 3 will add the Level-2 (neighborhood) variables of interest. The first set of Level-2 variables included for this model, as suggested by the literature, are shown below. The Level-1 model remains unchanged (model 2 above). The Level-2 intercept equation (for \( \beta_{0j} \)) now includes all of the Level-2 variables (both neighborhood socioeconomic conditions and housing characteristics). This model is only testing differences in the intercepts across neighborhoods, so the remainder of the Level-2 equations remain fixed (without error terms).
Model 3: Final model

\( H_0: \) Neighborhood conditions do not add explanatory power to the model \( u_{0j} = 0; \) There is no significant difference between Model 2 and Model 3 (\( \chi^2 \) test of model deviances).

Level-1:

\[
SRH_{ij} = \beta_{0j} + \beta_{1j} Race_{ij} + \beta_{2j} Employed_{ij} + \beta_{3j} (Poverty_{ij} - Poverty_{j})
+ \beta_{4j} Non-Smoking_{ij} + \beta_{5j} Physical Activity_{ij}
+ \beta_{6j} (Chronic Conditions_{ij} - Chronic Conditions_{j})
+ \beta_{7j} (Social Capital_{ij} - Social Capital_{j}) + r_{ij}
\]

Level-2:

\[
\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Median Home Value}_{ij} - \text{Median Home Value}_{j})
+ \gamma_{02}(\text{Percent Vacant}_{ij} - \text{Percent Vacant}_{j}) + \gamma_{03}(\text{Percent Renter}_{ij}
- \text{Percent Renter}_{j}) + \gamma_{04}(\text{White}_{ij} - \text{White}_{j}) + \gamma_{05}(\text{Poverty Rate}_{ij}
+ \text{Poverty Rate}_{j}) + \gamma_{06}(\text{Percent Mobility}_{ij} - \text{Percent Mobility}_{j})
+ \gamma_{07}(\text{Unemployment Rate}_{ij} - \text{Unemployment Rate}_{j}) + u_{0j}
\]

\[
\beta_{1j} = \gamma_{10} \\
\beta_{2j} = \gamma_{20} \\
\beta_{3j} = \gamma_{30} \\
\beta_{4j} = \gamma_{40} \\
\beta_{5j} = \gamma_{50} \\
\beta_{6j} = \gamma_{60} \\
\beta_{7j} = \gamma_{70}
\]

Initial results show that a collection of neighborhood variables (Level-2 variables) was not statistically significant. Because removal of variables changes the significance of other variables in the model, variables were removed in a step-wise fashion. First, the variables for the percent White, and percent over 200% FPL were removed because they were not significant. Results from the next iteration of the model showed unemployment rate and median home value remained not statistically significant, so they were also removed from the model. The model ran again and the variable: percent vacant remained insignificant, so this variable was removed from the model. At this point, the Level-2 model included the variables: percent under 100% FPL, percent renter, and percent mobility. They were all non-significant because
they are all correlated and in essence representing a similar phenomenon – poverty. The variables percent under 100% FPL and percent renter were strongly collinear. This suggested that there needed to be a selection between the variables under 100% FPL or percent renter in the model. Both variables had performed similarly in the final model, so the choice was made to remove the variable under 100% FPL and keep the variables: percent renter and mobility in the model. The variables: percent renter and percent of the population under 100% of FPL perform similarly in this model, but the variable percent renter has more clear implications for policy change. This final model is shown below:

Level-1:

\[ SRH_{ij} = \beta_{0j} + \beta_{1j}Race_{ij} + \beta_{2j}Employed_{ij} + \beta_{3j}(Poverty_{ij} - \overline{Poverty}) + \beta_{4j}Non-Smoking_{ij} + \beta_{5j}Physical\ Activity_{ij} + \beta_{6j}(Chronic\ Conditions_{ij} - \overline{Chronic\ Conditions}) + \beta_{7j}(Social\ Capital_{ij} - \overline{Social\ Capital}) + r_{ij} \]

Level-2:

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}(Percent\ Mobility_{j} - \overline{Percent\ Mobility}) + \gamma_{02}(Percent\ Renters_{j} - \overline{Percent\ Renters}) + u_{0j} \]

\[ \beta_{1j} = \gamma_{10} \]
\[ \beta_{2j} = \gamma_{20} \]
\[ \beta_{3j} = \gamma_{30} \]
\[ \beta_{4j} = \gamma_{40} \]
\[ \beta_{5j} = \gamma_{50} \]
\[ \beta_{6j} = \gamma_{60} \]
\[ \beta_{7j} = \gamma_{70} \]

The deviance score for this model is 2310.6, a significant improvement over the fit value for the unconditional model (Model 1; 3877.44, p<0.01). While the deviance score for this model is lower than for Model 2 (2313.92), suggesting a better fit, the difference is not
significant. The implications of this slight, but not statistically significant, improvement will be discussed later in this chapter.

5.2 Model assumptions

This section will discuss the assumption checks that were conducted for this model.

Each of these models was calculated using HLM 7 Scientific Software International (2011) set to full maximum likelihood. Full maximum likelihood was used, because the models are all nested within each other and to allow for comparability across models (i.e., use of the deviance statistic).

Following Tabachnick and Fidell’s checklist for multilevel modeling (2013, p. 852), the following issues will be reviewed in this section:

1. Adequacy of sample sizes and missing data
2. Normality of distributions at Level-1 and Level-2
3. Absence of multicollinearity
4. Independence of errors

5.2.1 Adequacy of sample sizes and missing data

First, the adequacy of sample sizes and missing data will be evaluated. This study includes 29 neighborhoods; the per neighborhood sample sizes for this study varied from 10 to 217. The model has seven Level-1 predictors and two Level-2 predictors. Tababachnick and Fidell (2013), and Sayer (2010) recommend 10 observations per variable; that means for this study each neighborhood should have had 70-100 individuals. Only two neighborhoods reach that standard. Tababachnick and Fidell suggest that model convergence could be an issue without a dataset this large (2013). This study did not have convergence issues.
Relationships Between Neighborhoods, Housing, Health

The final dataset used for this study did not include any missing data, because respondents with missing data values were dropped by the software (see discussion in the data and measures chapter for more details).

5.2.2 Normality of distributions at Level-1 and Level-2

Next, the normality of distributions at Level-1 and Level-2 will be evaluated. In multilevel modeling, checks of normality are necessary variable by variable and with the whole model working together. All of the Level-1 variables are categorical, with 3-4 categories, or dichotomous. Many of these variables are not normally distributed, however a review of the univariate variables does not indicated a wide spread on any variables except poverty status, social support and non-smoking. While the distributions are not perfectly normal, transformations of the variables did not improve fit, therefore the choice was made to model the untransformed versions of the variables for greater interpretability. And while the distributions are not normal, Tabachnick and Fidell suggest that if at least 10% of the individuals are in the least frequent response category, for categorical variables, the variable is usable (2013). At Level-2 both variables are continuous and normally distributed (see Appendix A for details). A check of multivariate normality or the distribution of the whole model was attempted. The standard test for multivariate normality is a Q-Q plot of Chi-square distribution values by Mahalanobis distance (the distance between the estimate and its expected value). If this plot is close to a 45-degree line the model is considered multivariate normal. A check for multivariate normality was attempted, however, because of the nature of the data points (all categorical and most dichotomous; see Table 5.2) this test is not appropriate (Sayer, 2014). The resulting fit value was exactly the same for many neighborhoods because of the categorical nature of the data (see Figure 5.1). In Figure 5.1, seven of 29 neighborhoods are plotted on the graph; the
remaining 23 neighborhoods were plotted at the x,y intercept (0,0). Because all of the Level-1 variables are categorical, the best error checking method is to review the univariate variables and look for a wide distribution on the data (Sayer, 2014). As described at the beginning of this paragraph, the distributions of the data are within an acceptable range for the model. The reason this test does not work for the data is because the model is trying to fit a regression line to each neighborhood using variables from each level. At Level-1 none of the variables are modeled as varying across neighborhoods, therefore it is all based on Level-2 variation. In essence, the Level-1 variables are just holding constant the variables of interest for individuals.

Violations of normality do not affect the coefficient estimates but could influence the estimation of standard errors. That is why this model uses the robust error estimates provided.
in HLM, which take into consideration the non-normality of the variables. Using robust standard errors helps to correct for non-normality. When comparing the non-robust and robust errors, one sees that the coefficients don’t change, but the standard errors do – this is the HLM software trying to model the categorical Level-1 model variables; because of the categorical variables, it is more difficult for the software to fit the model. Comparing the standard errors to the robust errors is like the difference between a mean and a median in correcting for outliers (Sayer 2014).

5.2.3 Absence of multicollinearity

Next, multicollinearity will be evaluated. Multicollinearity is of particular concern when cross-level interaction terms are used in multilevel models (Tabachnick & Fidell, 2013). While several cross-level terms were explored in early versions of this study, no cross-level terms are in the final model. One of the major risks of multicollinearity is an inability of the model to converge. This was not a problem for any of the models covered in this study. An additional risk of multicollinearity is that the marginal interpretation of the parameters ($\beta$s) can be questionable. While there is no evidence of multicollinearity, one solution for multicollinearity is to center variables (Tabachnick & Fidell, 2013). All variables that were appropriate for centering were centered in this model, for ease of interpretation.

5.2.4 Independence of errors

The independence of errors between Level-1 (individual health characteristics) and Level-2 (neighborhood-level characteristics) in multilevel modeling is tested by reviewing the interclass correlation coefficient (ICC). The ICC for this study is 5.5%, a level that suggests a relationship
5.3 Results

This section will discuss the results of the multilevel models from this study. Table 5.1 shows the results from each of the three models. First, for Model 1, the null model, our null hypothesis was that there was no variation in self-reported health across neighborhoods. This null hypothesis is rejected based on an ICC (5.5%) that suggests there is variation by neighborhood and that multilevel modeling is necessary. In addition the $u_{0j}$ is significant (p-value <0.01) suggesting that self-reported health is significantly affected by neighborhood of residence. In addition, the estimated grand mean across neighborhoods, $\gamma_{00}$, was 2.57 and was significant; this means that on average people across all neighborhoods rate their health between very good (2) and good (3). The deviance score for this model is 3877.44.

A look at the Model 2 coefficients and standard errors in Table 5.1 begins to tell a story of the interrelatedness of individual characteristics and self-reported health in the study sample. Model 2 is the model that includes only the individual characteristics of the adult respondents. The null hypothesis for this model was that individual characteristics do not increase the fit of the model, meaning that Model 1 was a better fit. The test of this hypothesis includes a look at the significance of each variables’ coefficients (for Model 2) and a comparison of the Model 1
and Model 2 deviance statistics. This null hypothesis is rejected, because each of the parameters are statistically significant (p<0.01) and the deviance statistic for Model 2 (2313.9) is significantly smaller than the deviance statistic for Model 1 (3877.44).

Table 5.1: Results with Robust Standard Errors *(italicized variables grand mean centered in the model)*

<table>
<thead>
<tr>
<th></th>
<th>Model 1: null model</th>
<th>Model 2: Level-1 model</th>
<th>Model 3: final model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level -1</strong></td>
<td>Coef.</td>
<td>S.E.</td>
<td>Coef.</td>
</tr>
<tr>
<td>Chronic Conditions</td>
<td>0.391**</td>
<td>0.019</td>
<td>0.388**</td>
</tr>
<tr>
<td>Race</td>
<td>-0.266**</td>
<td>0.029</td>
<td>-0.194**</td>
</tr>
<tr>
<td>Non-smoking</td>
<td>-0.268**</td>
<td>0.048</td>
<td>-0.268**</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>-0.327**</td>
<td>0.057</td>
<td>-0.331**</td>
</tr>
<tr>
<td>Social Support</td>
<td>-0.110**</td>
<td>0.030</td>
<td>-0.106**</td>
</tr>
<tr>
<td>Employment</td>
<td>-0.262**</td>
<td>0.044</td>
<td>-0.253**</td>
</tr>
<tr>
<td>Poverty</td>
<td>-0.192**</td>
<td>0.030</td>
<td>-0.186**</td>
</tr>
<tr>
<td><strong>Level – 2</strong></td>
<td>-</td>
<td>-</td>
<td>Coef</td>
</tr>
<tr>
<td>% Renter</td>
<td></td>
<td></td>
<td>0.007**</td>
</tr>
<tr>
<td>% Mobility</td>
<td></td>
<td></td>
<td>-0.011*</td>
</tr>
<tr>
<td>Intercept, (\beta_0)</td>
<td>2.576**</td>
<td>0.060</td>
<td>3.330**</td>
</tr>
<tr>
<td>Intercept Variance Component, (\tau_{00})</td>
<td>0.06568</td>
<td>0.00003</td>
<td>0.00003</td>
</tr>
<tr>
<td>ICC</td>
<td>5.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance Statistic (# of parameters)</td>
<td>3877.44</td>
<td>2313.9**+ (10)</td>
<td>2310.6**+ (12)</td>
</tr>
<tr>
<td>Proportion of reduction of variance</td>
<td>99.95%</td>
<td></td>
<td>99.95%</td>
</tr>
</tbody>
</table>

** = significant at 99% CI, * = significant at 95% CI
+ = as compared with model 1/null model

The data in Table 5.1 show the coefficients and standard errors for Model 3. Model 3 includes the neighborhood variables as well as individual characteristics. The null hypothesis for Model 3 is that neighborhood conditions do not add explanatory power to the model. The test of this hypothesis includes reviewing the significance of each variable’s coefficient, as well
as the remaining error to be explained in the model at Level-2 and a comparison of the deviance statistics. Although both of the neighborhood variables are statistically significant and the Model 2 deviance statistic is larger than the Model 3 deviance statistic, the difference is not statistically significant ($\chi^2$). Therefore, the null hypothesis is not rejected. This will be discussed further later in this chapter.

A review of the proportion of reduction of variance, that is the amount of additional error variance explained by a more complicated model, suggests that Model 2 does a much better job of modeling the data than Model 1, but that Model 2 and Model 3 perform similarly.

The formula for the proportion of reduction in unexplained variance is as follows:

$$\text{pseudo } R^2 = \frac{\tau_{00}(\text{unconditional}) - \tau_{00}(\text{conditional})}{\tau_{00}(\text{unconditional})}$$  \[5.4\]

For this study the proportion of reduction in unexplained variance is $(0.06568-0.00003)/0.06568 = .99954 = 99.95\%$. This suggests that Model 2 explains 99.95\% more of the variance than Model 1. However, the addition of Level-2 variables does not improve the pseudo $R^2$, it remains 99.95\%, (comparing Model 1 to Model 3). This suggests that model 3 is not significantly better at modeling the data in the study than Model 2.

### 5.4 Interpretation and discussion of results

This section explains how to interpret the results of Model 1, the null model, followed by a description of how to interpret the model coefficients based on variable type. This is followed by an explanation of how to interpret the results of Models 2 and 3. Next is a discussion of the magnitude of influence of each variable. The section ends with a comparison of what changed between Models 2 and 3.
In interpreting the data from these models, it is important to remember that the outcome variable is self-reported health, with lower responses on a five point scale indicating better health (1 = excellent health, 5 = poor health).

### 5.4.1 Interpretation of Model 1 results

First a review of Model 1, the unconditional or null model. This is a typical first model for a multilevel modeling study because it measures the proportion of variance in the outcome variable (self-reported health) that is accounted for by the grouping variable (neighborhood). The model is an intercept-only model; there are no additional predictors. This model is also used to check the assumption of independence by calculating the interclass correlation coefficient (ICC; see Equation 5.3). The ICC calculated for this study is 5.5%, suggesting that neighborhood composition accounts for 5.5% of the variability in self-reported health among adults in this study. This ICC is in line with previous studies (see discussion at the end of Section 2.4). While not large, this ICC suggests that multilevel modeling is necessary. In addition, the Level-2 random effect ($u_{0j}$) is 0.066 and is significant (p-value <0.01). This means that self-reported health is significantly affected by neighborhood and that the null hypothesis is rejected, that neighborhood does not affect their residents’ self-reported health. In addition, the estimated grand mean across neighborhoods, $y_{00}$, was 2.56 and was significant. This means that on average people across all neighborhoods rate their health between very good (2) and good (3). The goodness of fit or deviance score (discussed in the previous chapter) for this model is 3877.44.
5.4.2 Interpretation of model coefficients based on variable type

Models 2 and 3 are more complicated to interpret than Model 1. In order to understand the results from Models 2 and 3, it is critical to understand how to interpret the coefficients for different variable types and for variables that were centered and not centered. Understanding how to interpret different variable types is necessary before being able to interpret the variable’s influence on the model. Variable type is also critical for checking the model assumptions (Section 5.2). Three Level-1 variables are categorical (scales with 3-4 possible categories) poverty status, chronic conditions and social support. Four Level-1 variables are dichotomous (race, employment status, smoking, and physical activity) and both Level-2 variables are continuous. These differences in variable types can be seen in Table 5.2.

Table 5.2: Variable Type *(italicized variables grand mean centered in the model)*

<table>
<thead>
<tr>
<th></th>
<th>Categorical</th>
<th>Dichotomous</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level-1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic Conditions</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Non-smoking</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Physical Activity</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Social Support</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Level-2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Renter</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>% Mobility</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

The Level-1 variables poverty, social support, chronic disease and Level-2 variables percent renter and percent mobility were all grand mean centered *(italicized in Tables 5.1 and 5.2, see Equation 4.10)* the remaining Level-1 variables in the model are dichotomous;
dichotomous variables are not typically centered. Both Level-2 variables were centered. Centering transforms the variable by subtracting a meaningful constant, in this case the grand mean, or the mean of the whole dataset. After centering, the variable is interpreted as the expected value when each predictor is equal to its mean (Paccagnella, 2006). The only value that changes when centering is the intercept ($Y_{ij}$). For a complete discussion of centering, please see Section 4.3.

Next, the influence of each variable on self-reported health is reviewed (see Table 5.1). First a discussion about how to interpret the categorical variables chronic conditions, social support, and poverty (all Level-1 variables). The chronic conditions scale ranges from 0-3, with respondents receiving a 0 if they do not report that a physician or other healthcare provider has ever told them they had one of the ten conditions (asthma, cancer, chronic lung disease, diabetes, heart trouble or angina, high blood pressure or hypertension, high cholesterol or triglycerides, stroke, severe allergies, or depression) asked about on the survey; respondents receive a 3 on the scale if they report more than 2 chronic conditions (please see Section 3.4.2 for more details). The variable chronic conditions has a coefficient value of 0.388; this can be interpreted as an additional increase of one category on the chronic condition scale is related to a 0.388 increase in self-reported health score (meaning a poorer health rating), when all other variables in the model are held constant.

This same interpretation method works for social support and poverty at Level-1. The social support scale ranges from 0-3 with a higher value indicating a higher level of perceived social support or social capital. A one-point increase on the social support scale is connected to a -0.106 change in self-reported health rating (or an improved self-reported health rating), when all other variables in the model are held constant. The poverty scale ranges from 0-2 based on respondents’ reported income in the last year and household size. The scale values represent below 100% FPL (0), between 100% FPL and 200% FPL (1) and above 200% FPL (2). A one-point increase on the poverty scale (or increased income) is connected to a -0.186 change in
self-reported health or a better health rating, when all other variables in the model are held constant.

Dichotomous model variables are interpreted differently. The dichotomous variables in the model include race, non-smoking, physical activity, and employment (all Level-1 variables). When interpreting the meaning of a dichotomous variable, the coefficient in the model represents the value for the difference between the two categories. For example for Race (0= African American/Other, 1= White) the difference in self-reported health for those in the African American/Other category results is 0.194 point increase in self-reported health (worse health), when all other variables in the model are held constant. This suggests that the expected self-reported health difference between respondents in the study who are African American/Other and Whites is 0.194, with being White connected to better self-reported health, when all other variables in the model are held constant. Similarly, the expected self-reported health effect of being a current smoker versus being a non-smoker is a -0.268. This suggests that the expected self-reported health difference between respondents in the study who are current smokers and those who are current non-smokers is 0.268, with non-smoking being connected to better self-reported health, when all other variables in the model are held constant. The difference between being physically active or not is expected to be a difference of -0.331 on the self-reported health scale, with respondents who are physically active connected to better self-reported health, when all other variables in the model are held constant. Finally the difference between being employed or not is expected to be a difference of -0.253 on the self-reported health scale with respondents who are employed connected to better self-reported health, when all other variables in the model are held constant. All of the Level-1 variables are significant at 99% confidence interval for Model 2.
5.4.3 Interpretation of Model 2 results

Next a general review of Model 2 (individual characteristics only). The results for Model 2 can be found in the middle columns of Table 5.1. First, all of the coefficient signs are in the direction expected. In this model, having more chronic conditions is linked to poorer self-reported health while being a non-smoker, physically active, having more social support, being employed and not in poverty are linked to better self-reported health. Having a race of African American/Other is linked to poorer health outcomes. The intercept for Model 2 is 3.330; this can be interpreted as the predicted self-reported health for an individual with an average score on the chronic conditions, social support, and poverty scales who is an African American/Other, non-physically active, unemployed, and a smoker. This intercept represents a self-reported health score of between good (3) and fair (4).

At Level-1, all of the signs of the coefficients are as would be expected, with higher values for the chronic conditions scale and being African American/Other, a smoker, not physically active, and unemployed connected to poorer reports of self-reported health. While higher values on the poverty scale (increased income) and social support scale (increased social support) are connected with better reports of self-reported health. At Level-1 all of the coefficients are significant at p<0.01 level. A look at the size of the coefficients indicates that chronic conditions (0.388), physical activity (-0.331), and not smoking (-0.268) have the highest coefficients in absolute terms, and therefore the greatest potential impact on self-reported health.

The model for this study had a relatively large number of Level-1 variables. Because individual characteristics are so critical in self-reported health, they are important to include when modeling the relationship between self-reported health and neighborhood characteristics. Further, failing to include individual characteristics when seeking to understand the relationship between the neighborhood characteristics and self-reported health outcomes would be leaving
out what researchers believe is approximately 90% of the influence on self-reported health (Figure 1.2).

Gender, binge drinking, fruit and vegetable consumption, and age were dropped from the model because of non-significance. The most surprising of these variable omissions is age, because the literature demonstrates very clearly a consistent relationship between increased age and poorer self-reported health. Based on the literature review, age was expected to be an influential individual-level variable, because many strong contributors to better health are strongly influenced by age (e.g. physical activity, drinking, diet, chronic diseases). When the model is run with age as the only individual Level-1 variable, age is significant. Age is collinear with several other variables in the model, and when those variables are included (chronic disease and physical activity) the influence of age on self-reported health is captured by other variables. In this dataset, age was highly collinear with other variables in the model, and as a result those variables were sufficient to model the variance presented by age. Age modeled on its own is significant and predictive of self-reported health.

The literature shows mixed results for gender on self-reported health (see Section 2.3), and as a result it is not surprising that it was not significant. There is limited literature on vegetable consumption and binge drinking in a multilevel model, but they were tested in this study since poor diet and excessive alcohol consumption are part of the leading causes of early mortality. In retrospect fruit and vegetable consumption (2 servings of fruit and 3 servings of vegetables on an average day) measured with the questions in this study is probably a poor measure of quality of diet (75% of the study sample did not meet the guidelines). Binge drinking (4 drinks on one occasion in the last 30 days for women; 5 drinks for men) was reported by a small part of the study sample (16%) and was heavily skewed by age (more young people report binge drinking). While binge drinking can be unhealthy and dangerous, heavy drinking (defined as having more than an average of one drink per day for a woman and two drinks per
day for a man) or long-term heavy use of alcohol is likely a better measure of excessive alcohol use, but that variable was not available at the time of the study.

The Level-1 variables that remained in the model where chosen to represent the socioeconomic status of the respondents (race, employment, poverty), their health condition (chronic conditions), health behaviors (smoking and physical activity) and the amount of social support they perceive in their lives. All of these variables are well established in the literature and in previous health-focused multilevel models as important individual characteristics.

### 5.4.4 Interpretation of Model 3 results

Next a review of Model 3 results. Model 3 includes the addition of two community-level variables (percent renter and percent mobility). A review of the coefficient values in Table 5.1 shows that all of the coefficient signs are in the direction expected based on the literature review, with the exception of mobility. This exception will be discussed in the following section. Both Level-2 variables are continuous percentages (see Table 5.2), and both variables were centered. Continuous variables are interpreted as percentage point increases. For example, for the variable, percent renters, for each percentage point increase in the percentage of renters in a community (e.g., going from 50% to 51% renters) there would be a 0.007 point increase in self-reported health (worse self-reported health). Similarly, with percent mobility, for each percentage point increase in the percentage of homeowners who have moved to the neighborhood in the last five years, there would be a -0.011 decrease in self-reported health (or better self-reported health).

In Model 3 all of the Level-1 variables remain significant at the 99% confidence interval. In addition, the Level-2 variable percent renter is significant at the 99% confidence interval and percent mobility is significant at the 95% confidence interval.
The intercept for Model 3 is 3.268; this can be interpreted as the predicted self-reported health for an individual with an average score on the chronic conditions, social support, and poverty scales who is an African American/Other, non-physically active, unemployed, smoker in a community with an average percentage of renters and mobility. This score is between a self-reported health rating of good (3) and fair (4).

At Level-2 the socio-economic variables racial makeup, poverty rate and employment rate and the housing variables median home value and percent of vacant homes were dropped from the model. The most surprising of these variables was the poverty rate; as discussed in Chapter 2, there is extensive literature linking high poverty neighborhoods to poor self-reported health (Wen, Browning, & Cagney, 2003; Patel et al., 2003; and Galea et al., 2005). However, as discussed in Section 5.1.3, when building Model 3, poverty rate (under 100% FPL) was strongly collinear with percent renter. Both variables performed similarly in the model, but percent renter was kept because the author believed it had more clear implications for policy change. The literature would suggest that the remaining variables on this list could have been significant, but they were not tested as a group in previous studies reviewed by the author.

The Level-2 variables that remained in the model where chosen to represent the socioeconomic community-level measures (mobility) and to describe the community’s housing environment (see Section 2.2 and 2.3 for complete discussion of these variables). While the author hoped that a larger number of community variables would be statistically significant in the model, the two Level-2 variables (percent renter and percent mobility) are well established in the literature and in previous health-focused multilevel models as important community-level characteristics. The literature suggests that the percentage of renters in a community is a sign of poorer housing conditions and much higher rates of poverty (Mallach, 2010). Pendall, Theodos, and Hildner (2014) show explicitly the link between a higher percent of renters in a community and poor housing and neighborhood conditions. Malega (2003) found that neighborhoods with a higher percentage of rental homes were strongly connected with
increasing rates of poverty. Within the dataset used for this study, the percent of renters within the community was highly correlated with the percentage of community residents reporting incomes under 100% of FPL. This suggests that within the data used for this study there is a strong and direct link between the percentage of renters in a community and the percentage of poverty.

The link between mobility, housing quality, and health is complex, but has been widely studied. There is a body of literature that focuses on the link between poor neighborhoods with high mobility and a loss of social support structures for neighborhood residents linked to poorer health outcomes (Smith & Mallinson, 1996; Smith, Alexander, & Easterlow, 1997; Howden-Chapman, 2004; Acevedo-Garcia, Osypuk, Werbel, Meara, Cutler, & Berkman, 2004). Coulton explains that, “residential mobility is one of the primary factors driving neighborhood change and can have an important effect on social conditions and quality of life in an area” (2014, p. 260).

At Level-2, percent renter has the expected sign, with a higher percentage of renters connected with worse self-reported health. However, mobility has the opposite sign to what was expected, with a greater percentage of new homeowners to the community in the last five years connected to better self-reported health. Both Level-2 variables are continuous percentages, and both variables were centered. So their coefficients can be interpreted as percentage point increases. For example, for the percent of renters in a neighborhood, for each percentage point increase in the percentage of renters in a community (e.g., going from 50% to 51% renters) there would be a 0.007 point increase in self-reported health. Similarly with percent mobility, for each percentage point increase in the percentage of residents who have moved to the neighborhood in the last five years, there would be a -0.011 decrease in self-reported health.

The tests of model fit suggest that both Models 2 and 3 are significantly better at fitting the data than Model 1, but that Model 3 is not significantly better than Model 2. However, the coefficients for both Level-2 variables, added in Model 3, are significant, suggesting that percent
renter and percent mobility are community characteristics that are linked to the self-reported health of a community’s residents. This section will discuss how model 3 fits the data.

Figure 5.2 shows a plot of two neighborhoods from the study: a high mobility and low mobility neighborhood, as the percent renters increases (horizontal axes; please note that percent renters is centered hence the negative values on the horizontal axes) self-reported health increases (this means that the respondents are reporting worse self-reported health) at the same rates (same slope of lines) across high and low mobility neighborhoods, however the intercept/ outcome is different in the two neighborhoods.

![Figure 5.2: Comparing a low and high mobility neighborhood](image)

This suggests that the lower mobility neighborhoods have a higher (worse) self-reported health rating (approx. 0.11 higher rating on SRH) than high mobility neighborhoods. As stated earlier, the previous literature on the link between mobility and self-reported health suggested that the expected relationship would be a link between increased mobility and poorer health outcomes. This study, instead, found the opposite, that increased mobility was linked to better
self-reported health. This could potentially be explained by comparing the definition of mobility used in previous studies with that used in this study. For this study mobility was defined (using census data) as the percent of homeowners who moved into the neighborhood in the last five years. It is important to note that mobility in the case of this study is the mobility of homeowners, not the mobility of all residents. In Browning and Cagney’s multilevel model (2002), where more mobility was linked to poorer self-reported health, mobility was defined as years of residency of each survey participant in the neighborhood and taken from a large community survey done in Chicago. Similarly in Cagney’s multilevel model (2002), mobility was measured using a residential stability scale constructed from census data on housing tenure and housing occupied by owners, and in contrast more mobility was linked to poorer self-reported health. The definition of mobility, for this study, only measures mobility of homeowners. The literature reviewed in Chapter 2 showed that there is a strong link between more homeowners and more neighborhood stability and better health outcomes. This will be discussed further in the next chapter.

5.4.5 Comparing the influence of each model variable

Each model variable is calculated on a different scale, so they are not directly comparable. In order to compare the influence of each variable, the percent change in self-reported health for a one point change in each variable was calculated (holding all other variables constant, see Table 5.3). Table 5.3 is organized by model level (1 and 2) and in descending order of magnitude from chronic conditions (12%) to social support (3%) at Level-1 and percent mobility (0.34%) to percent renter (0.21%) at Level-2.
Table 5.3 shows the magnitude of difference from the Level-1 variables (all ranging from 3%-12% changes) to the Level-2 variables, which are all less than 1%.

Table 5.3: Percent change in self-reported health for a one-point change in each variable

<table>
<thead>
<tr>
<th></th>
<th>Percent change in self-reported health</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level -1</strong></td>
<td></td>
</tr>
<tr>
<td>Chronic Conditions</td>
<td>12%</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>10%</td>
</tr>
<tr>
<td>Non-smoking</td>
<td>8%</td>
</tr>
<tr>
<td>Employment</td>
<td>8%</td>
</tr>
<tr>
<td>Poverty</td>
<td>6%</td>
</tr>
<tr>
<td>Race</td>
<td>6%</td>
</tr>
<tr>
<td>Social Support</td>
<td>3%</td>
</tr>
<tr>
<td><strong>Level – 2</strong></td>
<td></td>
</tr>
<tr>
<td>% Mobility</td>
<td>0.34%</td>
</tr>
<tr>
<td>% Renter</td>
<td>0.21%</td>
</tr>
</tbody>
</table>

The effect size of both of the Level-2 (neighborhood) variables are very small, in particular when one compares the magnitude of the coefficients to the Level-1 variables (0.388 for chronic conditions compared to 0.007 for percent renters; or the percent change in self-reported health from table 5.3: 12% for chronic conditions and 0.34% for percent mobility). In this model, both of the Level-2 variables are close to zero, yet still statistically significant to the model. This signals that in this study renting and moving matter at the neighborhood level, but that their influence on self-reported health is very small.
5.4.6 Comparing Models 2 and 3

In comparing Models 2 and 3, the addition of the Level-2 variables shifts some coefficients at Level-1 in Model 3. While the coefficient for non-smoking is unchanged and the coefficients for chronic conditions, physical activity, social support, and poverty all shift slightly (less than 0.005) from Model 2 to Model 3, the coefficient for employment goes down by 0.009 from Model 2 to Model 3. The addition of Level-2 variables has the most influence on race, the race coefficient changes from -0.266 to -0.194 (a change of 0.072). This suggests that of all of the individual variables, race is the most influenced by environmental characteristics. Or rather, that when environmental characteristics are added, the link between race and self-reported health is less strong. This will be discussed further in the following chapter.
6. Conclusions and Future Research

This dissertation investigated whether modeling housing environment quality proved to have additive explanatory power on self-reported health after controlling for other powerful individual drivers of health. The research presented here has implications for those interested in improving the self-reported health of a community. This chapter will review the connection between this study and existing literature, discuss the study’s contribution to the literature, its shortcomings, and potential for future research.

6.1 Conclusions

This section will discuss the overall conclusions from the model. A two-level multilevel model assessed the effects of individual and neighborhood housing and socioeconomic conditions on self-reported health. Level-1 of the model was adult respondents to the 2010 Greater Cincinnati Community Health Status Survey, who could be geocoded to a specific neighborhood in Hamilton County, OH (N=926). Unweighted data were used and only respondents who answered all of the questions included in this study remained in the study sample. Level-2 of the model was geographic areas created by the author for this study. The areas are ecologically meaningful geographic areas based on community-understood boundaries, homogeneity within geographies and proximity. There are 29 communities at Level-2. Multilevel modeling was implemented using HLM software, Version 7.

Multilevel modeling allows for data that is collected at different levels. For example, people living together in the same neighborhood could potentially violate the assumptions of independence in more standard OLS models. Multilevel modeling takes into consideration these dependencies by calculating intergroup variance, differences in average group response (intercept, see Figure.5.2 for example) and differences in group variance (differing slope).
original Model 2 included Level-1 variance across neighborhoods (inclusion of $u_{0j}$ at Level-2) but the variance was not significant so all Level-1 variables were modeled as fixed at Level-2; this means that the model did not include differences in slopes between different neighborhoods, only intercepts.

Model assumptions were reviewed and checked. The model has seven Level-1 predictors and two Level-2 predictors, which means that only two of 29 neighborhoods meet the standard of 10 observations per variable. However the model did not have convergence problems as a result, probably because the variables were all modeled as fixed at Level-2. There were no missing data at the point of model assumption checking, however 369 respondents were dropped from the model by the software because they were missing responses to one of the Level-1 variables (largely missing responses to one of the two poverty questions). Non-normality was present in most categorical and dichotomous Level-1 variables, but transformations of the variables did not improve the fit, therefore untransformed variables were used. Level-2 distributions were acceptable. Robust standard errors were used to help correct for non-normality. The interclass correlation coefficient of 5.5% suggests that there is a relationship between individuals and their communities and, as a result, a need for multilevel modeling.

Initially when building Model 1, 11 variables were included, all of which were allowed to vary across neighborhoods. Gender, binge drinking, fruit and vegetable consumption, and age were not significant, so the corresponding variables were dropped from the model. In step-wise fashion, the non-significant variance terms were removed until reaching the final Level-1 model with all seven variables fixed at Level-2. This model was a significantly better fit (2313.92, $p<0.01$) to the data, as compared to Model 1. Model 3 includes the addition of the neighborhood-level socioeconomic and housing environment measures. Seven neighborhood-level measures were tested. It was expected that many of the Level-2 variables included would prove to be influential on self-reported health, but the only significant factors in this study were
percent renter and percent mobility. In step-wise fashion, median home value, percent vacant, percent White, poverty rate and unemployment rate were dropped for non-significance. Model 3 was a significant improvement over the fit of Model 1 (deviance score = 2310.6, p<0.01) and was a better fit than Model 2, but not significantly better. Both of the neighborhood variables are statistically significant, suggesting that percent renter and percent mobility are neighborhood characteristics that influence the self-reported health of adult neighborhood residents. Table 5.1 summarizes the three models evaluated. Having more chronic disease and a higher percentage of renters in a neighborhood are connected with worse self-reported health, while being White, not smoking, being physically active, having more social support, being employed, with more income and a neighborhood with more mobility are connected with better self-reported health.

Model 3 (including Level-2 variables) is not significantly better fit to the data than Model 2 (with only Level-1 variables), however, the individual significance of the parameter estimates ($\beta$) suggests that percent mobility and percent renter may be important environmental characteristics to consider when trying to model self-reported health.

The focus of this study is on whether modeling environmental factors, such as housing quality, has additive explanatory power on self-reported health after controlling for other powerful drivers of health (poverty, unemployment, chronic conditions, health behaviors, etc.). The results of this study are in line with much of the literature on what contributes to improved self-reported health. The Level-1 factors (individual health characteristics) contribute more to the variability in self-reported health (95% of the influence) compared to 5% for environmental (Section 2.4 discusses ICCs from similar studies). This study found that chronic health conditions, race, smoking, physical activity, community support, employment status, and poverty status contribute to changes in self-reported health. At Level-2, percent mobility and percent renter proved to be significant variables. Percent renter is one of the model’s housing variables, suggesting that housing conditions in a neighborhood influence self-reported health. Percent
Relationships Between Neighborhoods, Housing, Health

...mobility was also significant and was one of the model’s socioeconomic environment variables suggesting that socioeconomic conditions in the neighborhood influence self-reported health; percent mobility, however, did not influence self-reported health in the way anticipated.

6.2 Relationship of results to theory

This section will review the connection between this study’s results and the theory presented in earlier chapters. The results of this study suggest that housing quality is influential on self-reported health, even after controlling for socioeconomic conditions.

Referring back to the public health pyramid presented in Chapter 1 (and again below) shifting individual behavior (e.g. getting people to stop smoking or to be more physically active) is very difficult and can be a costly public health intervention. This compares to environmental interventions, which have less influence on each individual, but are relatively easy to implement and can have an effect on many people over a long period of time. Specific housing and socioeconomic interventions will be discussed later in this chapter.
Figure 6.1: The Health Impact Pyramid

Working at the top layers of the pyramid is difficult and costly to sustain, while working at the bottom two layers: “changing the context to make individuals’ default decisions healthy” and “socioeconomic factors” are relatively easy to implement and can have an effect on many people over a long period of time. This study and many others support work at this level of the pyramid. According to the National Center for Healthy Housing & American Public Health Association,

_Housing is one of the best known and documented determinants of health. The affordability, location, and quality of housing have all been independently linked to health. Poor quality housing and blighted neighborhoods diminish property values, increase crime, and erode the cohesiveness and political power of communities. Despite the critical role of housing in public health, attention to U.S. housing conditions remains incommensurate with its importance to our wellbeing._ (2014, p. 1)

This study’s results are similar to others that looked at the contributing factors to self-reported health. Consistent with the literature, this study found that as individual poverty increases (or income decreases) self-reported health gets worse (McFadden et al., 2009; Dowd
& Zajacova, 2007; Wilkinson & Marmot, 2003; Wen, Browning, & Cagney, 2003; Patel et al., 2003; and Galea et al., 2005), that being unemployed is detrimental to self-reported health (Wilkinson & Marmot, 2003) as are chronic conditions (Patel et al., 2003). However, healthy behaviors (getting exercise and not smoking) and social capital help to improve self-reported health (Singh-Manoux et al., 2007; Patel et al., 2003; Prince et al., 2012; Browning & Cagney, 2002; Cagney et al., 2005). Unlike previous research, a link between gender and self-reported health was not found (Singh-Manoux et al., 2007; and Undén & Elofsson, 2006). As described above, the race variable changed the most with the addition of neighborhood level data. These results are in line with previous research (Cagney, Browning, & Wen, 2005; for older African Americans in Chicago) that suggests that neighborhood conditions could be a mediating factor in the influence of race on self-reported health.

At Level-2, the connection between an increased percentage of renters in a community and poorer health outcomes is in line with previous research (Beck et al., 2013), but the connection between higher mobility and better self-reported health outcomes does not match with previous research with a more encompassing definition of mobility (Browning & Cagney, 2002). However, studies with definitions of mobility more similar to this one have found a similar relationship between increased mobility and better self-reported health (Cagney et al., 2005). A housing expert in Hamilton County suggested that these results might reflect a still recovering housing market (e.g. an unusual amount of mobility across the county); claiming that 2010 (the year the data for this study were collected) was one of the worst years for the housing market in Hamilton County (K. Schwab, personal communications, September 5, 2014). The Working in Neighborhoods (WIN) report for Hamilton County substantiates an unstable housing environment in the county, "...around 67% of municipalities in Hamilton County and 48% of neighborhoods in the City of Cincinnati experienced an increase in completed foreclosures in 2010... an increasing number of these foreclosures are taking place outside the low- and moderate-income neighborhoods" (2010, pg. 2). Beyond local variations in the housing market,
mobility can vary by neighborhood and over time with larger economic trends. The Census Bureau found that mobility for the United States was the lowest (35.4%) in the history of the Current Population Survey (CPS) for the period of this study (2005-2010; Ihrke & Faber, 2012). This suggests that this study was conducted during a year of recovery from the great recession, so the neighborhood-level findings are reflective of the conditions at the time, and might not be reflective of conditions during more stable or typical economic periods.

This dissertation rests partially on the premise that more renters in a community is a sign of a poorer community housing environment and that this poorer housing environment (or built environment) is linked to poorer health outcomes. This link between percent of renters, housing conditions and poor health is inextricably linked with poverty, both individual and community-level poverty. That is, individuals living in poverty are more likely to be renters and are also more likely to have poorer self-reported health. Communities of high poverty are more likely to have a higher percentage of renters and poorer health outcomes. The literature is very clear on the link between poorer health outcomes and higher rates of poverty, both individually and living in communities of high poverty (Wilkinson & Marmot, 2003; Wen, Browning & Cagney, 2003; Patel et al., 2003; and Galea et al., 2005). There is also literature to suggest that the percentage of renters in a community is a sign of poorer housing conditions and much higher rates of poverty (Mallach, 2010). Pendall, Theodos, and Hildner (2014) show explicitly the link between a higher percent of renters in a community and poor housing and neighborhood conditions. Malega (2003) found that neighborhoods with a higher percentage of rental homes were strongly connected with increasing rates of poverty. Within the dataset used for this study, the percent of renters within the community was highly correlated with the percent of community residents reporting incomes under 100% of FPL. This suggests that within the data used for

---

4 It is important to note that this report includes data for both renters and owners and that renters were much more mobile than owners.
this study there is a strong and direct link between the percentage of renters in a community and the percentage of poverty.

The link between poor housing environments in a community and poor self-reported health would not be a surprise to anyone working in the field. The conceptual link this study is trying to make is that a higher percent of renters is linked to poorer neighborhood housing conditions, which is linked to poorer self-reported health. This entire chain of relationships is heavily influenced by poverty. A high percentage of renters (in many, but not all communities) can be linked to higher poverty neighborhoods and poorer built environment conditions. Adults living in poor housing conditions and in higher poverty generally report poorer self-reported health scores. Causality is difficult to decipher, but the relationship between the community-level measures and individual self-reported health is consistent. Rates of individual and neighborhood poverty cannot, however, be disconnected from the discussion, because poverty is such a strong and consistent driver of health outcomes at both the neighborhood and individual level.

The research suggests that an increasing percentage of renters is bad for the quality of a neighborhood’s built environment. Housing makeup is quite complex and community-specific, and while percent renters might be an important marker, it cannot be the only critical measure. Researchers have suggested, and the results of this study support, the fact that the percentage of renters in a community is a proxy for persistent poverty. Housing environments can help or exacerbate rising neighborhood poverty (Pendall, Theodos, & Hildner, 2014) and the variable percent renter is one way to monitor shifting built environment characteristics. Monitoring community housing environment changes is important because of the link to health outcomes. There is a deep body of research in the literature about the role of affordable housing in poverty alleviation and about the best strategies for implementing affordable housing policies in communities of varying makeups. This study did not include a measure for affordable housing so cannot comment on the role affordable housing could play in self-reported health, although
the author acknowledges that in future research this could be an important variable to consider.
Pendall, Theodos, and Hildner (2014) state that the built environment can exacerbate or protect against rising neighborhood poverty, and they recommend that policy to reduce concentrated poverty should consider changes to the housing stock.

At the introduction of this study, the relationships of interest were presented in Figure 1.2. After conducting the study, there are a number of modifications to be made to Figure 1.2. First, although in line with similar studies (see Section 2.4 for discussion of ICCs), the percentage of influence of Level-2 characteristics was lower than anticipated at 5.5%. Secondly, the model including Level-2 characteristics was not significantly better than the Level-1 model so the arrows showing the connection between the environmental characteristics have been reduced to dotted lines to show a relationship, but one that is less strong than the arrows from the individual characteristics to self-reported health.
Further, this chart makes each of the four categories of characteristics seem distinct and unrelated and while that separation is necessary (and elegant) for a mathematical model like this one a next step in research could be to explore the inter-relationships between individual characteristics and the environment one lives in.

These results are based on data collected for Hamilton County, Ohio, in 2010. The results of this study might be generalizable to similar mid-sized Midwestern counties, with similar socioeconomic and housing conditions. However, it is important to note that Hamilton, County Ohio, has a high percentage of renter-occupied housing (41%); the percentage is even higher in the City of Cincinnati (situated at the center of Hamilton County, 61%). This compares to 32% in the State of Ohio or 35% in across the nation (U.S. Census Bureau, 2010 Census). In the model, the variable, percent renter, was collinear with the variable percent in poverty, suggesting that the percentage of renters in a community is a proxy for poverty. While Hamilton County’s poverty rates (18%) are similar to the nation’s (15%) and the state’s (16%), the City of Cincinnati has much higher poverty rates (30%) (U.S. Census Bureau, 2009-2013 American Community Survey 5-Year Estimates). Poverty is a strong driver of health outcomes; therefore,
the results from this study would not be expected to be generalizable to communities with a significantly larger or smaller housing rental community or with a significantly larger or smaller percentage of residents in poverty.

6.3 Policy implications

This section will review the potential policy implications of the results of this study, specifically actions that policymakers could take. There are three specific recommendations in this section:

1. Increased measurement of the health benefits of investments in socioeconomic and housing improvements;
2. High quality evaluations of pilot projects linking housing quality improvement to health;
3. Formally reconnect the fields of public health and planning.

These recommendations will be discussed further in the pages that follow.

Recommendation 1: Increased measurement of the health benefits of investments in socioeconomic and housing improvements. There are significant financial investments in the study area (Hamilton County, Ohio) in improving socioeconomic and housing conditions. These neighborhood development projects could also improve health, but there is often little to no measurement of the health benefits of infrastructure, housing and poverty reduction programs. In their recent book chapter, Evaluating the Social Determinants of Health in Community Development Projects, Fleming, Karaza, and Wysen suggest that the goal most community development projects include is improving one of the social determinants of health, yet measurement and evaluation of progress toward these health improvement goals is almost not existent. (2014). Policymakers should consider measuring the health benefits of their investments in socioeconomic and housing improvements.
Recommendation 2: High-quality evaluations of pilot projects linking housing quality improvement and health. There are numerous recent examples of organizations focusing on housing, and the social determinants, in hopes of influencing people’s health through healthier home environments:

1. UnitedHealth Group, the nation’s largest health insurance provider, is investing $150 million to build low-income housing in twelve different states. The insurer recognizes the connection between the social determinants of health and health outcomes, stating that “…those without stable homes are sick more often, have more undiagnosed illnesses and are more likely to wind up seeking expensive care in emergency departments” (Baxter 2014, n.p.).

2. The Robert Wood Johnson Foundation (RWJF), one of the nation’s largest philanthropic health organization has shifted their organization’s focus to a Culture of Health and describe this shift, “…our health is influenced greatly by education, housing, income, and numerous other factors outside of the health care we receive.” (Robert Wood Johnson Foundation, n.d.). As leaders in the field of health philanthropy and health policy, their shift signals their belief that the socioeconomic and housing environments are a critical piece to health improvement.

3. Several Federal Government Departments have acknowledged the link between housing quality and health, with healthy homes programs (The United States Department of Agriculture [USDA], U.S. Department of Housing and Urban Development [HUD], Centers for Disease Control and Prevention [CDC]).

4. The Urban Land Institute (ULI), a nonprofit research and education membership organization for real estate development professionals, has focused much of its organizational energy on a building healthy places initiative – noting the direct relationship between real estate development and health. In a recent publication, “Housing in America: integrating housing, health, and resilience in a changing
Relationships Between Neighborhoods, Housing, Health

“environment,” they note, “…a safe, resilient, and healthy home and community constitute a holistic system. The level of physical and psychological health and resilience can either support or undermine the entire system” (2014, p. 6).

5. Local Initiatives Support Corporation (LISC), a national community development corporation with a presence in Hamilton County, Ohio, has an increasing recognition that where people live affects their health. Through their affiliates, they are collecting evidence to show that affordable housing can be the first step to improved health outcomes (K. Schwab, personal communications, September 5, 2014).

6. The National Center for Healthy Housing (NCHH) and the American Public Health Association (APHA) have co-authored the National Healthy Housing Standard. The guidebook is designed to help reconnect the housing and public health sectors and use evidence-based strategies for improving housing conditions.

Some of these programs target individual health improvement. For example, many of these initiatives focus on creating internal housing environments that are healthier (e.g. removing lead, reducing asthma triggers). These types of programs are targeted at improved housing environments for individuals with a longer-term goal of increasing the total housing stock of healthy homes. Given the results for this study and others that show that individual health condition is a strong driver of self-reported health, these programs could improve the health of the individuals who live in the homes treated.

Other programs listed above target a more population-level improvement, that is, wholesale improvement in the neighborhood. While the results from this study and others suggest that neighborhood conditions are important, but a weaker driver of health outcomes, these wider environmental changes are likely to affect many more people than the one-house-at-a-time programs.

Specifically, the MacArthur Foundation recognized the need for more research in the field of housing (in their case, affordable housing). They describe their research network as,
“aiming to understand how investments in affordable housing have a host of other social and economic outcomes beyond merely providing shelter, especially in children and families” (n.d., n.p). As exemplified by the list above, many non-traditional organizations are entering the neighborhood development field (e.g., UnitedHealth Group) and several national neighborhood developers have added a health lens to their work (e.g., ULI, LISC). High quality evaluations of the impact of current or new projects linking housing quality improvements with health outcomes would be very informative to the field. In their book chapter, *Evaluating the Social Determinants of Health in Community Development Projects*, Fleming, Karasz, and Wysen suggest that, “public health and community development professionals can advance healthy communities together if health considerations are incorporated into project planning, implementation, and evaluation” (2014, p. 360). They emphasize the importance of incorporating health measures and evaluation from the early stages of the project.

**Recommendation 3:** Formally reconnecting the fields of public health and planning.

Chapter 2 of this study discussed in detail the shared origins and interests of the fields of planning and public health. This study sits squarely at the intersection of the two fields looking at how the socioeconomic and housing environments influence self-reported health. Yet there is evidence to suggest there is much work to be done to formally link the two fields (Corburn, 2004, 2005, 2007; Northridge, Sclar, & Biswas, 2003). Many scholars believe that coordination across the silos of public health and planning could yield a higher return on investment for both fields on projects focusing on the bottom two layers of the health impact pyramid (at the “Changing the context to make individuals’ default decisions healthy” and “socioeconomic factors,” Figure 1.1). There are a number of specific ways that the two fields could work together. First is through the integration of planning and public health goals in a region or city’s comprehensive plan. The American Planning Association (APA) has launched a project to study the integration of public health goals and measurement into a region’s comprehensive
Relationships Between Neighborhoods, Housing, Health

plan. The results are mixed, but APA suggests that this is an area where collaborative work should continue:

…the public health world’s awareness of the importance of the built environment and planning for public health has significantly evolved. By comparison, the planning profession has taken somewhat longer to turn their attention to public health concerns and is doing so in some areas more quickly than in others. The use of data to inform and prioritize planning policies, as well as to assist in the setting of benchmarks for success, is one major area in which public health could provide significant input to the comprehensive planning process. While this evaluation shows that much work still needs to be done to bring the two fields of planning and public health back together, many of the plans evaluated here show promise for the future and signify the current and future creation of more livable, sustainable communities for everyone. (Ricklin, et al., 2012, p.21).

In addition, the group has provided a model for integrating health into a comprehensive plan. The two fields could work more closely on Health Impact Assessments (HIAs).

According to the Health Impact Project, HIAs are systematic way to bring together data, public health expertise and often details on the built environment to identify the, often unidentified, health effects of proposed new laws, regulations, projects, and, often, urban design or transportation projects (Wernham, 2014). These assessments bring together the opinions and expertise of many actors, which allows for a robust and rich discussion about the effects of new projects or policy on health. If done early enough in a process, HIAs can improve the health benefits of the final project. This is an area where planners and public health professionals have already started working together, but additional collaboration would build relationships that could benefit other community projects in the future. In particular, a local project on the health effects of a high percentage of rental properties in a community would be an excellent follow-on project to this study.
6.4 Research contributions

The focus of this study is on whether modeling housing quality proves to have additive explanatory power on self-reported health after controlling for other powerful drivers of health. This study informs the discussion on the area effects on health with a focus on housing as well as the relative influence of individual and neighborhood health characteristics on self-reported health. The study is among the first multilevel models done in the field of planning (Doyle et al., 2006; Van Dyck et al., 2010; and Sharma, 2008). The ecologically meaningful neighborhood clusters are fairly unique within multilevel modeling (Prince et al., 2012). The relatively low-level geography is a strength of the study, along with the fact that the geographic boundaries were selected to coincide with groupings of neighborhoods that are meaningful to the community. With smaller geographic areas than most similar studies, this study provides results at a level that is more meaningful for community-level action. And this is the only known model of this type within a mid-sized city in a Midwestern county.

Some existing studies use data sources at each level of the model that are from multiple years, some with data collection dates that range over a period of ten years or more. The preponderance of data for this study all comes from the same year (2010).

6.5 Research limitations and future research

This study has a number of limitations. The sample size for this project, while large (N = 926) varied significantly across neighborhoods (from N = 16 to N = 295). More observations per neighborhood would have provided increased power to measure the influence of variables of interest, in particular the Level-2 neighborhood variables. If the model had a larger number of observations in each neighborhood, the model would have had increased power and may have been able to identify Level-1 variable variation at Level-2.
Relationships Between Neighborhoods, Housing, Health

This study uses cross-sectional data, so the results reflect the effect of individual health characteristics and environmental factors at a point in time. Longitudinal data would allow the study of the effect of long-term exposure to various environmental characteristics.

Given that Hamilton County saw an unusual number of vacancies in 2010 in all neighborhoods (WIN, 2010), the measurement of mobility in this model might not reflect neighborhood mobility during a more stable housing period. In addition, mobility did not influence self-reported health in the direction expected, but this study’s definition of mobility (homeowner-focused) is slightly different from previous studies.

The results of this study suggest that: individual and neighborhood-level characteristics influence self-reported health. It is important to note that while the following characteristics were identified in this study as individual characteristics: chronic conditions, race, smoking, physical activity, social support, employment status, and poverty status – many of these characteristics are not entirely under the control of the individual person – but can be heavily influenced by environmental factors. It was not possible in this study to determine the direction of causality on the individual health characteristics.

The method used for setting geographic boundaries created geographic areas based on community-understood boundaries, homogeneity within geographies, and heterogeneity between geographies; however, using the same guidelines, different permutations of geographic boundaries could be created. In addition, the boundaries are arbitrary with respect to the area of environmental influence on health. Furthermore, the causal direction of neighborhood effects on health is unclear. Is the direction of causality from the neighborhood to the individual or does a certain type of person self-select into a poor neighborhood?

There are a number of directions that this study could spur in future research.

Given the discussion above about the definition of mobility in this study, testing multiple definitions of mobility for their influence on self-reported health would be an interesting future direction for this research. Coulton (2014) suggests that in addition to looking at residential
Relationships Between Neighborhoods, Housing, Health

mobility, researchers should consider residential instability, housing unit turnover, and neighborhood change as measures of mobility. In his work, focused on young adults, How Neighborhoods Affect Health, Well-Being and Young People’s Futures, Galster acknowledges that, “to advance our understanding of neighborhood effects, research must turn to more nuanced methods, including longer-term horizons and a broader palette of neighborhood measures” (2014, p.2). Galster suggests the need for more studies that incorporate more mixed-methods. In particular an area for future research would be understanding from residents of neighborhoods in Hamilton County how they see the impact of increasing or decreasing percentages of renters and mobility. Doing this type of study qualitatively would provide rich results to compare if these influences are the same in different neighborhoods.

The use of a multilevel model takes into account the nesting or clustering of the data (that is people within neighborhoods), by including random effects or an additional error structure (this is discussed in Chapter 4). However multilevel modeling methods assume that these random effects are independent. Given the spillover effect between neighborhoods, it is unlikely that the random effects are independent. There have been relatively few studies that consider spatial aspects of the data within a multilevel model (Verbitsky-Savitz & Raudenbush, 2009; and Verbitsky, 2007); in future research it would be informative to consider nested data with spatial considerations.

The data from this study represent adults in one Midwestern county; it would be useful to extend the research to other Midwestern counties to see if the relationship between Level-1 and Level-2 and self-reported health are consistent across communities.

The results from this study suggest that health policy changes that target changes in the built environment in order to influence health improvements should expect only small improvements in individual self-reported health at a point in time. However, if effective, the small point-in-time improvements could cumulatively lead to improvements in self-reported health. However, the full impact of built environment changes on self-reported health could only
be measured conclusively with longitudinal data; looking at influence of individual health characteristics and community characteristics cumulatively and over time. With longitudinal data, one could also test what happens if someone moves from a community with a less healthy built environment to a community with a healthier built environment – does their self-reported health improve?

While self-reported health’s link to actual health outcomes has been well studied, health is a multi-faceted issue. Does the impact of one’s environment on self-reported health vary at different points in one’s life? Are there points when the influence of the built environment is more critical (e.g., with advancing age or disability, when you do not have a car)? These questions all represent research questions that would build on this work. The community of researchers using self-reported health is large and the connection between self-reported health and actual health is strong, but strategies for improving self-reported health remain elusive. The number of studies that look at interventions to improve self-reported health over time are limited (Duberg, Hagberg, Sunvisson, & Möller, 2013; Ichida, Hirai, Kondo, Kawachi, Takeda, & Endo, 2013; Pisinger, Ladelund, Glümer, Toft, Aadahl, & Jørgensen, 2009). Most studies explore point-in-time factors. However, in the longitudinal studies available, self-reported health remains stable for most participants. For researchers interested in working across the boundaries of the fields of planning and public health with a focus on improving self-reported health, there is still much to study and much to learn. But the power of moving someone or a whole neighborhood to a healthier self-reported health is worth the effort.

This study includes a call to reconnect the fields of planning and public health, but what is practical and actionable given the current reality in Hamilton County? In most places, Hamilton County among them, the fields have separate university departments and separate governmental offices that are not in the practice of working closely together. In addition, Hamilton County is known for its extreme decentralization of local government, with a very high number of local governments; there are multiple planning departments within the
municipalities/cities/local governments within the county. There are also four different health departments within the county. Not only are these academic and governmental bodies separate but each local planning department and health department also has a slightly different focus or set of regulations to enforce. For example, the City Health Department provides healthcare services through their network of five Federally Qualified Health Center sites, coordinates most of the disaster preparedness for the county, and does a lot of the lead remediation. In contrast the County Health Department provides no direct healthcare, but is responsible for some health promotion activities as well as tracking epidemiological data. Planning departments are more similar in their responsibilities, but each municipality has different areas of focus (growth, sustainability, jobs, zoning, etc.). Beyond being very siloed, most public health and planning departments are minimally staffed. This staffing does not allow for an expansion of services or the addition of complex collaborations with other departments, unless a significant amount of additional funding was made available. While health is important, it rarely rises to the top of the list in planning department priorities. Similarly while many public health departments realize the importance of the built environment, environmental health is rarely a dedicated focus of this county’s health departments. When it is, as in the case of the City Health Department, the focus of the one staff person is split between environmental pollutants (toxic waste, lead, water quality, etc.), implementation of HIAs (see end of Section 6.3 for a discussion of HIAs), and changes to the built environment. So while the results of this study suggest that improvements in the built environment, specifically housing quality and socioeconomic conditions of a community, could lead to improvements in self-reported health, the author has concerns about the practicality or action ability of these results.

Upon reflection there are many barriers to making this work a reality. The first is the historic lack of coordination and collaboration across planning and health departments, and while there is interest in shaping the built environment, there seems to be no real desire to change these historic patterns. The second is a lack of funding for initiatives focused on re-
shaping the built environment. While there are national funding streams available for communities whose thinking is this sophisticated, the funding streams are not yet large enough to motivate communities to do the heavy lifting on infrastructure change to make this kind of new movement happen. The funding available is not sizeable enough for the kind of massive cultural change required in governmental infrastructure and community coordination to reach a cultural tipping point where coordination across planning and public health is the normal protocol for built environment projects and not unusual. Third, it is unclear how much citizens are demanding these changes. Residents want to live in high-quality housing in a safe and prosperous community, but it is doubtful that the general public understands that these changes could come from a more coordinated effort of the planning and health departments. Without a highly motivated and demanding set of constituents, it would take a very motivated and politically sophisticated leader to make this kind of systematic change happen. Further, it might not be enough to simply have better coordination between the public health and planning departments; community-level improvements will require engaged residents. While some communities have very engaged and active community councils, others do not. Fourth, it would be helpful to have a coordinated national discussion on what has proven successful in various types of communities. It would be particularly helpful to have suggestions of built environment changes that can be implemented at a lower cost. Adding high quality housing, or upgrading current housing stock can be very expensive, as is adding sidewalks or changing streetscapes. There must be lower-cost changes that signal to the community an intention to improve the built environment and socioeconomic climate that can be done quickly and effectively to start a cultural shift in a community; to increase community-member interest and engagement in the process and in community improvement; and to build momentum for larger community investment.

So in answer to the original question - this study includes a call to reconnect the fields of planning and public health, but is that practical and actionable given the current reality in
Relationships Between Neighborhoods, Housing, Health

Hamilton County? The answer is that there significant barriers to making this happen locally, however, a committed leader or team of community members could work to show pockets of change in order to argue for a more systematic shift in the local standards of practice. If you believe, as this author does, that the built environment affects health and that talented and well-meaning public health and planning-focused practitioners, working together, could improve the built environment, then it is worth outlining the barriers and seeking ways to overcome them with the hope of improved health for all members of the community.
References


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


Relationships Between Neighborhoods, Housing, Health


U.S. Census Bureau; American Community Survey, 2009-2013 American Community Survey 5-Year Estimates, Table DP03; generated by Jennifer Chubinski; using American FactFinder: http://factfinder2.census.gov; (11 March 2015).

U.S. Census Bureau; Census 2010, Table DP-1; generated by Jennifer Chubinski; using American FactFinder: http://factfinder2.census.gov; (11 March 2015).

Relationships Between Neighborhoods, Housing, Health

http://www2.census.gov/geo/maps/dc10map/tract/st39_oh/c39061_hamilton/DC10CT_C39061_000.pdf.


Relationships Between Neighborhoods, Housing, Health


## Appendix A – Detailed list of geographic boundaries

### TABLE A.1. Combinations of census tracts based on community-defined neighborhoods.

<table>
<thead>
<tr>
<th>Approx. Total Pop</th>
<th>Name</th>
<th>Tracts</th>
</tr>
</thead>
<tbody>
<tr>
<td>36,022</td>
<td>Lincoln Heights, Arlington Heights, Lockland, Reading, Dillonvale, Ross moyne, Deer Park, Silverton, Madisonville</td>
<td>227, 228, 274, 232.01, 232.10, 232.22, 236, 237.01, 237.02, 238, 242</td>
</tr>
<tr>
<td>20,525</td>
<td>Springdale, Glendale, Sharonville, Evendale</td>
<td>223.01, 224, 230.01, 230.02, 231</td>
</tr>
<tr>
<td>28,123</td>
<td>Pleasant Hills, Springfield Twp, Finneytown, Wyoming, Springdale</td>
<td>221.01, 221.02, 222, 226.01, 226.02, 225</td>
</tr>
<tr>
<td>42,671</td>
<td>Forest Park, Springdale, Pleasant Run Farm, New Burlington, Mount Healthy Heights, Greenhills</td>
<td>219, 216.02, 216.03, 216.04, 217.01, 217.02, 218.01, 218.02</td>
</tr>
<tr>
<td>28,687</td>
<td>White Oak, Finneytown, New Burlington, Northbrook, Skyline Acres, Mt. Healthy, North College Hill</td>
<td>205.05, 207.01, 207.41, 207.42, 207.05, 207.61, 207.62, 207.07</td>
</tr>
<tr>
<td>37,464</td>
<td>Northgate, Groesbeck, White Oak, Northbrook</td>
<td>205.01, 205.02, 205.04, 206.01, 206.02</td>
</tr>
<tr>
<td>23,914</td>
<td>Taylor Creek, Dry Ridge, Dunlap, Pleasant Run, Monfort, Heights, Taylor Creek, Bridgetown</td>
<td>208.11, 208.12, 208.02, 209.01, 209.02, 210.01</td>
</tr>
<tr>
<td>27,473</td>
<td>Mt. Airy, Montfort Heights, White Oak, Bridgetown, Monfort Heights</td>
<td>204.01, 204.03, 204.04, 211.01, 211.02</td>
</tr>
<tr>
<td>27,232</td>
<td>Addyston, Cleves, North Bend, Grandview, Hoven, Mack, Miami Heights, Dent, Covedale</td>
<td>102.01, 212.02, 210.03, 107, 210.02, 212.01</td>
</tr>
<tr>
<td>23,778</td>
<td>Westwood, Bridgetown, Covedale, West Price Hill</td>
<td>104, 105, 106, 213.02, 213.04, 213.03, 214.01, 214.21, 214.22</td>
</tr>
<tr>
<td>32,664</td>
<td>Riverside, Saylor Park, Delhi Hills, Delshire, West Price Hills</td>
<td>263, 92, 93, 94, 95, 96, 97, 98, 99.01, 99.02, 103, 91</td>
</tr>
<tr>
<td>33,590</td>
<td>Queensgate, Lower/East/West Price Hill, Westwood, Sedamsville, Riverside</td>
<td>88, 100.03, 100.04, 100.05, 100.02, 101, 102.02, 100.01</td>
</tr>
<tr>
<td>26,188</td>
<td>Westwood, East Westwood</td>
<td>28, 74, 75, 77, 78, 79, 85.01, 85.02, 86.01, 272, 89, 87</td>
</tr>
<tr>
<td>22,090</td>
<td>Camp Washington, Northside, Mt. Airy, Fay Apartments, South Cummingsville, Millevale, North Fairmount, English Woods, South Fairmount</td>
<td>213.02, 213.04, 213.03, 214.01, 214.21, 214.22, 263, 92, 93, 94, 95, 96, 97, 98, 99.01, 99.02, 103, 91</td>
</tr>
<tr>
<td>Approx. Total Pop</td>
<td>Name</td>
<td>Tracts</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>21,895</td>
<td>Spring Grove Village, College Hill, Mt. Airy, White Oak, Finneytown</td>
<td>73, 81, 82.01, 82.02, 83, 84, 111</td>
</tr>
<tr>
<td>19,305</td>
<td>Hartwell, Cartage, Winton Hills, Elmwood Place, St. Bernard</td>
<td>60, 61, 80, 257, 258</td>
</tr>
<tr>
<td>21,425</td>
<td>Pleasant Ridge, Kennedy Heights, Golf Manor, Amberley Village</td>
<td>57.01, 57.02, 58, 59, 233, 234</td>
</tr>
<tr>
<td>32,182</td>
<td>Avondale, Bond Hill, North Avondale, Roselawn, Paddock Hills</td>
<td>270, 63, 64, 65, 66, 68, 69, 110, 62.02, 271, 67, 34, 62.01</td>
</tr>
<tr>
<td>33,524</td>
<td>Mt. Auburn, CUF, Heights, Corryville, Clifton</td>
<td>18, 22, 23, 25, 26, 27, 29, 30, 32, 33, 70, 71, 72</td>
</tr>
<tr>
<td>18,441</td>
<td>Westend, CBD Riverfront, Over The Rhine, Pendleton</td>
<td>2, 264, 265, 7, 9, 10, 11, 269, 16, 17, 3.01, 2.02, 4, 6, 8</td>
</tr>
<tr>
<td>19,195</td>
<td>Mt. Adams, Walnut Hills, E. Walnut Hills, Evanston,</td>
<td>268, 19, 20, 267, 36, 37, 38, 39, 40, 41, 42, 12, 13, 21, 35</td>
</tr>
<tr>
<td>40,036</td>
<td>Mt. Lookout, Columbia Tusculum, Hyde Park, Oakley, Madisonville</td>
<td>47.01, 48, 49, 50, 51, 52, 53.01, 53.02, 54, 55, 56, 108, 53</td>
</tr>
<tr>
<td>17,355</td>
<td>East End, Linwood, California, Turpin Hills CDP, Mt Washington, Sherwood CDP, Salem Heights</td>
<td>47.02, 266, 45, 46.04, 46.05, 46.02, 46.03, 46.01, 44, 43</td>
</tr>
<tr>
<td>30,181</td>
<td>Blue Ash, Sharonville, Kenwood, Montgomery, Rossmoyne</td>
<td>235.01, 235.21, 235.22, 239.01, 240.01, 240.02</td>
</tr>
<tr>
<td>31,557</td>
<td>Brecon, Highpoint, Sixteen Mile Stand, Remington, Loveland Park, Indian Hill</td>
<td>223.02, 243.01, 243.21, 243.22, 243.03</td>
</tr>
<tr>
<td>22,099</td>
<td>Madeira, Concord Hills, Plainville, Fairfax, Mariemont, Milford</td>
<td>241, 244, 245, 259, 247, 248, 273</td>
</tr>
<tr>
<td>43,336</td>
<td>Anderson Twp, Dry Run, Turpin Hills, Mt Washington, Sherwood, Fruit Hill, Salem Heights, Cherry Grove, Forestville, Coldstream</td>
<td>249.01, 249.02, 250.01, 250.02, 251.01, 251.02, 251.03, 251.04</td>
</tr>
<tr>
<td>19,205</td>
<td>Norwood</td>
<td>252, 253, 254.01, 254.02, 255, 256</td>
</tr>
<tr>
<td>22,217</td>
<td>Harrison, Blue Jay Miamitown, New Baltimore, New Haven, Elizabeth, Hooven</td>
<td>260.01, 260.02, 261.01, 261.02, 262</td>
</tr>
</tbody>
</table>
A quick look at several variables of interest by geography is revealing. Most researchers who understand the socioeconomic structure of the county would not be surprised by Figure A.1, a map of this study’s geographies by percentage of population under 200% of the Federal Poverty Level. Darker polygons on the map indicate higher concentrations of people under 200% FPL, with a strong concentration of poverty in the urban core and much lower rates in the outer-ring eastern suburbs.

**Figure A.1 Percentage of community residents under 200% FPL**

Source: 2006-2010 American Community Survey

The next map shows the percentage of homes that are vacant by community. Again, the darker polygons indicate higher percentages of vacant homes, and again, the highest rates of vacant homes are located in the urban core and first ring suburbs. In contrast to the previous map, however, the lowest rates of vacant homes lie in the western outer-ring suburbs.
Figure A.2. Percentage of vacant homes by community

A third map shows what researchers call the mobility rate. This is the percentage of people who have moved into the community since 2005 (in the five years before the CHSS). These are the newcomers to the community, and significant turnover in residents might signal that a community is undesirable. This map is a little harder to interpret with widespread variation across the communities.

Source: 2010 Census
Figure A.3. Percentage of residents who have moved in since 2005

Source: 2006-2010 American Community Survey
Appendix B – An analysis of histograms for select variables

The histogram below shows the distribution of responses for self-reported health rating. This variable can have a value of 1, 2, 3, 4, or 5. It is slightly skewed to the lower end of the scale, which are higher health rankings. So respondents tend to rate their health more positively (1 = excellent…. 5 = poor).

The histogram below shows the distribution of responses current age (mean = 53.7). It is slightly negatively skewed, meaning older respondents and has a flatter than normal distribution.
The table below shows the frequency distribution of the race variable within the CHSS data.

<table>
<thead>
<tr>
<th>Race</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American &amp; Other</td>
<td>380</td>
<td>39.3</td>
<td>39.8</td>
<td>39.8</td>
</tr>
<tr>
<td>White</td>
<td>575</td>
<td>59.5</td>
<td>60.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>955</td>
<td>98.9</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>System</td>
<td>11</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>966</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table below shows the frequency distribution of the sex variable within the CHSS data.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>352</td>
<td>36.4</td>
<td>36.4</td>
<td>36.4</td>
</tr>
<tr>
<td>Female</td>
<td>614</td>
<td>63.6</td>
<td>63.6</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>966</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>
Relationships Between Neighborhoods, Housing, Health

The table below shows the frequency distribution of the poverty status variable within the CHSS.

<table>
<thead>
<tr>
<th>Poverty Status</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% FPL and below</td>
<td>202</td>
<td>20.9</td>
<td>20.9</td>
<td>20.9</td>
</tr>
<tr>
<td>Between 100% and 200% FPL</td>
<td>220</td>
<td>22.8</td>
<td>22.8</td>
<td>43.7</td>
</tr>
<tr>
<td>2.00</td>
<td>544</td>
<td>56.3</td>
<td>56.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>966</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

The histogram below shows the distribution of responses for the chronic condition variable. It is heavily skewed to the lower end of the scale, with most people reporting between one and three chronic conditions. The mean of this variable is 1.86, meaning on average each respondent reported almost two chronic conditions. This variable has a standard deviation of 1.698. It is negatively skewed with a long right tail and has a relatively high peak.

Because of the distribution of the chronic conditions variable, I created a grouped categorical chronic conditions variable, shown in the histogram below and described in the data and methods chapter.
The histograms below shows the distribution of responses for each of the social capital questions on the survey. In each case the variables are ordinal with values from 1-6 with lower values indicating stronger agreement with the statement and therefore more social ties to their community. Each of the three variables has a similar bimodal distribution with more people agreeing strongly or disagreeing with the statements, but not many responses in the middle of the scale. In this case, it does not make much sense to report the descriptive statistics, because these variables are not normally distributed.
The social capital questions were transformed into a single categorical variable. The specifics of this transformation are described in the data and measures chapter; however, the histogram of the new variable is presented below.
The histograms below help to show the distributions of a selected group of Level-2, census variables.
Relationships Between Neighborhoods, Housing, Health

Employment Status

Poverty

![Histogram of Employment Status]

![Histogram of Poverty]

147
Relationships Between Neighborhoods, Housing, Health