Understanding the Role of Residential Self-Selection in the Relationship between the Built Environment and Physical Activity in Urban Residential Areas

By
Unnur Gudnadottir

A dissertation submitted in partial fulfillment of
The requirements for the degree of

Doctor of Philosophy
(Population Health Sciences)

at the
UNIVERSITY OF WISCONSIN-MADISON
2018

Date of final oral examination: 5/21/2018

The dissertation is approved by the following members of the Final Oral Committee:
Kristen Malecki, Assistant Professor, Population Health Sciences
Kirsten Beyer, Assistant Professor, Division of Epidemiology, Institute for Health and Equity, Medical College of Wisconsin
Lisa Cadmus-Bertram, Assistant Professor, Kinesiology
Mari Palta, Professor, Population Health Sciences
Paul Peppard, Associate Professor, Population Health Sciences
Kristin Thorleifsdottir, Assistant Professor, Planning and Landscape Architecture

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ACKNOWLEDGEMENTS

I would like to express my profound appreciation to those who contributed to the completion of this dissertation. I would like to thank my advisor, Dr. Kristen Malecki, for believing in me and guiding me through the process of writing this dissertation. I am very grateful for the continued support, valuable feedback, and encouragement I have received from her the last five years. Further, I would like to express my gratitude to the other members of my Ph.D. committee, for providing their valuable time and expertise during the course of this work. Sincere thank you to Dr. Mari Palta, Dr. Paul Peppard, Dr. Kirsten Beyer, Dr. Kristin Thorleifsdottir, and Dr. Lisa Cadmus-Bertram.

I also want to thank my family for their continued support and understanding throughout this process. Without my husband's encouragement and the gracious help from my parents, completing this dissertation would not have been possible.

Lastly, I want to recognize all the hard working people at the Survey of the Health of Wisconsin (SHOW) for the data used in this dissertation and for funding my education. Funding for SHOW in 2014-2016 came from the Wisconsin Division of Public Health, the Wisconsin Partnership Program PERC Award (233 PRJ 25DJ), the National Institutes of Health's Clinical and Translational Science Award (5UL 1RR025011). This dissertation work falls under the SHOW IRB protocol (2013-0251) currently approved through 11/9/2018. The manuscript proposal for this work was approved by SHOW on 4/19/2016.

ABSTRACT

Background: Research has shown an association between numerous features in the built environment and physical activity. However, most of these studies ignore the fact that individuals do not randomly sort into residential areas. This introduces the problem of residential self-selection. This dissertation aims to gain further understanding of residential self-selection and its role in the relationship between the built environment and physical activity. Understanding these associations is necessary in order to develop effective policies and allocate public funding to addressing low levels of physical activity and related chronic diseases.

Methods: Data come from the 2016 cohort of the Survey of the Health of Wisconsin (SHOW). Confirmatory factorial analysis (CFA), latent profile analysis (LPA), linear regression, and structural equation modeling (SEM) were used to achieve the aims of this dissertation.

Results: Two profiles emerged with LPA, based on individual reasons for moving to current residence: “active lifestyle” profile and “non-active lifestyle” profile. Individuals in the active lifestyle profile took on average more steps per minute and reported sitting less than those in the non-active lifestyle profile. Further, individuals who reported reasons related to an active lifestyle as being important when deciding to move to their current residence lived in more walkable environments than those who did not report these reasons as being important. When walkability was objectively measured, the association between reasons and walkability of current residential area was stronger among high-income block groups than low-income block groups. Finally, no association was found between walkability and average steps per minute (both objectively measured) when adjusted for residential self-selection and sociodemographic factors in SEM.

Conclusions: These results suggest that residential self-selection based on a predisposition to an active lifestyle exists, but its extent may vary based on economic status of the residential area. This may lead to an inflated association between the built environment and physical activity, even when adjusted for common sociodemographic factors. Different types of policies might have to be implemented based on community economic status.

INTRODUCTION

Over the past several decades there has been growing interest in the role of the built environment in shaping individual health behaviors and outcomes (Diez Roux & Mair, 2010; Jackson, Dannenberg, & Frumkin, 2013). Physical activity has been a major component of this research. In general, review papers on the built environment and physical activity agree that there is sufficient evidence to recommend creating physical activity friendly residential environments in order to increase levels of physical activity (Kärmeniemi, Lankila, Ikäheimo, Koivumaa-Honkanen, & Korpelainen, 2018; Saelens & Handy, 2008). However, the vast majority of studies examining this relationship are based on cross-sectional data and do not account for the non-random selection of individuals into neighborhoods (i.e. residential self-selection). Thus, many of these studies may be reporting biased estimates of the relationship between the built environment and physical activity, because they do not take self-selection into account.

The increased interest in studying the built environment in relation to physical activity stems in part from the fact that levels of physical activity have been declining in the past few decades, and this decline has been partly attributed to the built environment (Brownson, et al., 2005). In the last 60-70 years, substantial changes have taken place in the built environment of the United States (U.S.) causing people to become increasingly dependent on cars as their primary form of transportation (e.g. Brownson, et al., 2005; Muller, 1995). After World War II, people started increasingly moving from areas in central cities to the suburbs, with around 15% of the total U.S. population living in suburbs in 1940 compared to about 50% in 2000 (U.S. Bureau of the Census, 2000). Suburbs commonly have lower population density than residential areas in central cities and consist largely of single-family homes (Ann, 2012). Additionally,

suburbs often have zoning ordinances that separate residential land use from other types of land use such as commercial development (Schilling & Linton, 2005). These characteristics of suburbs make it difficult for suburban residents to run their daily errands on foot or by bike and make them more dependent on their personal automobiles.

Around the time people were migrating from the city to the suburbs, the National Interstate and Defense Highway Act was authorized. This act led to fundamental urban planning and community design changes in the U.S., as cities and towns were increasingly designed and organized around the automobile (Muller, 1995). Large highway infrastructures were designed to rapidly transport individuals across the U.S. by car. Many smaller highways were transformed into interstates, and cities were divided by interstates or had looping highway systems built around them. From about this time, car ownership and daily vehicle miles travelled per person started to increase (NHTS Highlights Report, 2001). In 2000, over 90% of U.S. households owned/leased a car, with 57% of households having two or more cars (U.S. Bureau of the Census, 2000.). Further, data from the 2001 National Household Transportation Survey showed that in 2000 a typical person traveled, on average, 35 miles a day in a personal vehicle, compared to 5 miles a day in 1950 (NHTS Highlights Report, 2001).

However, recent surveys show that growing numbers of Americans prefer to live in walkable environments where they can run their daily errands by foot or bike (American Planning Association, 2014; Urban Land Institute, 2015). A survey of 1040 U.S. adults age 21-65 with at least two years of college education, showed that 56% of participants aged 21-34 and 46% of participants aged 50-65 preferred living in a walkable neighborhood. Less than ten percent of all participants reported seeing themselves living in a car-dependent neighborhood in

the future, despite 40% of them currently living in a car-dependent neighborhood (American Planning Association, 2014). Thus, it seems that Americans are increasingly favoring dense and walkable residential areas as opposed to auto-oriented suburbs.

BACKGROUND

Chronic Diseases, Physical Activity, and Sedentary Time

In the U.S., chronic diseases[1] are among the leading causes of death (Murray et al., 2013). In 2012, almost half of the U.S. adult population were living with one or more chronic diseases (Ward, Schiller, & Goodman, 2014) and it is estimated that 86% of all health care spending in 2010 was for people living with a chronic disease (Gerteis et al., 2014).

Regular physical activity has been shown to reduce the risk of a number of chronic diseases such as hypertension, coronary heart disease, type 2 diabetes and some types of cancer (U.S. Department of Health and Human Services, 1996; Warburton, Nicol, & Bredin, 2006a). Physical activity is defined as any bodily movement that is produced by skeletal muscles that result in spending energy (Caspersen, Powell, & Christenson, 1985). Physical activity can include daily activities such as household chores, walking and biking during leisure time or when travelling between places, and running for exercise.

Mounting evidence indicates that sedentary time is a distinct risk factor, independent of physical activity, for poor health outcomes and some chronic diseases (F. B. Hu, Li, Colditz, Willett, & Manson, 2003; Thorp, Owen, Neuhaus, & Dunstan, 2011). Sedentary behavior is any

[1] Defined as a condition that lasts a year or more, require an ongoing medical care and results in functional limitation, for an example: heart disease, cancer, stroke, diabetes, and obesity (Goodman et al., 2013)

waking behavior characterized by sitting, lying or reclining and very low levels of energy expenditure (Tremblay et al., 2017). Common forms of sedentary behavior are sitting and watching TV, sitting and using a computer, and sitting while commuting by car.

Despite a widespread knowledge of the health benefits of physical activity and a growing awareness of the adverse effects of a sedentary lifestyle, inadequate physical activity and high volumes of sedentary time are common. Self-reported physical activity data from the 2013 Behavioral Risk Factor Surveillance System (BRFSS) indicate that about half of U.S. adults meet the aerobic physical activity recommendations provided in the 2008 Physical Activity Guidelines for Americans[2] (Centers for Disease Control and Prevention (CDC), 2008). However, accelerometer-derived data from the National Health and Nutritional Survey (NHANES) indicate that only about five percent of U.S. adults meet the aerobic physical activity recommendations (Troiano, Berrigan, Dodd, MâSse, et al., 2008). Additionally, the accelerometer-derived data from NHANES show that U.S. adults spend the majority of their time in sedentary behaviors (Matthews et al., 2008).

The Built Environment and Walkability

Individually focused interventions aimed at increasing physical activity levels have shown limited success in creating a population-level shift in physical activity (Kahn et al., 2002; Orleans, 2000). Therefore, researchers and public health representatives have increasingly focused on the built environment as a possible target to boost population physical activity levels.

[2] The 2008 Physical Activity Guidelines recommend doing aerobic physical activity and muscle-strengthening activities. Aerobic physical activities: 150 minutes a week of moderate-intensity physical activity (e.g. brisk walking) or 75 minutes a week of vigorous-intensity activity (e.g. running). Muscle-strengthening activities: (e.g. yoga, sit-ups or push-ups, using free weights or weight machines) 2 or more days a week (CDC, 2008).

If the built environment influences physical activity levels, there may be potential for far-reaching health benefits for the population as a whole.

The built environment consists of the structured and planned aspects of our environment, including all buildings, infrastructure, and spaces that are made or modified by people (Barton, 2009). Handy (2005) conceptualized the built environment as consisting of three components: land use patterns, transportation infrastructure, and design features. Together these three components generate need and provide opportunities for travel and physical activity. *Land use patterns* refers to the spatial distribution of human activities often defined by density (e.g. number of people living within a specific area) or land use mix (i.e. variety of land use such as residential, retail, industrial). *Transportation infrastructure* refers to the physical infrastructure and service that provide connection between activities such as streets, sidewalks, and rail systems. The transportation system component also refers to the quality of the infrastructure and service, such as safety and speed. *Design* refers to the aesthetics of the environment, such as the design of buildings and streetscapes and overlays both land use patterns and transportation systems components.

Walkability of a built environment can be defined as the extent to which characteristics of the built environment may be conducive to various forms of physical activity, such as walking/biking for transport, recreation or to exercise (Carr, Dunsiger, & Marcus, 2010a; Duncan, Aldstadt, Whalen, Melly, & Gortmaker, 2011; Leslie et al., 2007a). A walkable environment is commonly characterized by high street connectivity, high population density and mixed land use with close proximity to services (Leslie et al., 2007a; Saelens, Sallis, Black, & Chen, 2003a). A variety of measures have been used to capture the walkability of an

environment in research on walkability and physical activity, with each method having its strengths and limitations with respect to measurement bias, feasibility and content validity.

Measures of Walkability

Perceived measures of the environment obtained by self-administered questionnaires or by phone interview are frequently used. The most commonly used questionnaire is the Neighborhood Environment Walkability Scales (NEWS) developed by Saelens and colleagues (Saelens, Sallis, Black, & Chen, 2003b) and adapted by others (Bailey et al., 2014). Geographic information system (GIS) based measures of the built environment are also frequently used. The GIS-based measures are usually derived from existing data sources that have some spatial reference and provide information about population density, land use mix, access to amenities etc. Finally, direct observation of the environment using neighborhood-based audit tools have been used (Malecki et al., 2014), but less frequently. As expected, given all of the different available measurement approaches, conceptualization of walkability varies greatly among studies and the different disciplines interested in the subject. Consequently, it has been challenging to generalize and validate results from studies on walkability and physical activity (Hall & Ram, 2018; McCormack & Shiell, 2011). As an attempt to improve comparability of studies, researchers have increasingly been using the GIS based Walk Score® to estimate walkability.

Walk Score® is a publicly available measure based on distance to amenities, taking intersection density and block length into account (“Walk Score Methodology,” 2016). Walk Score® ranges from 0-100, with higher scores indicating greater walkability. Amenities within 0.25 miles of origin receive a high score, and amenities located further away from the origin

receive lower scores, with amenities located 1.5 miles or farther from the origin receiving a 0 score. The Walk Score® methodology then includes penalties for low intersection density and long blocks, two features in the built environment that have been found to work as barriers to walking (Berrigan, Pickle, & Dill, 2010). The distance scoring system is based on travel surveys and average walking speeds. It is estimated that pedestrians walk approximately 3 miles per hour, thus it takes them 5 minutes to walk 0.25 miles, and 20 minutes to walk 1 mile (Turner, Shunk, & Hottenstein, 1998; "Walk Score Methodology," 2016). A commonly accepted rule-of-thumb in the walking literature is that individuals are not willing to walk to amenities reachable in 30 minutes or more of walking, i.e. located 1.5 miles or further from origin (Talen & Koschinsky, 2013). In 2010, the algorithm for Walk Score® was revised and is now based on network distance (i.e. following street lines) instead of Euclidean distance (i.e. straight-line/crow-flies distances). Further, the revised version of the Walk Score® to certain extent takes into account the pedestrian friendliness of the walk route (by means of intersection density and the length of blocks), and it gives amenities that are highly correlated with walking greater weight when calculating the score ("Street Smart Walk Score," 2011).

Walk Score® has been found to correlate with a number of GIS-based measures commonly used to operationalize walkability, such as access to walkable amenities, intersection density, population density, and access to public transit (Carr, Dunsiger, & Marcus, 2010b; Carr et al., 2010a; Duncan et al., 2011). However, Walk Score® has been criticized for not taking into account the safety and aesthetics of the environment, both features that have been found to correlate with physical activity (Foster & Giles-Corti, 2008; Handy, 2005; McCormack et al., 2004). Most likely because of these limitations of Walk Score, it has been found that Walk Score

overestimates true walkability in low-income neighborhoods (Koschinsky, Talen, Alfonzo, & Lee, 2017). Commonly, low-income neighborhoods are associated with higher crime rates, and environmental aesthetics (e.g. trees along sidewalks, clean streets etc.) are often lacking/impaired (Neckerman et al., 2009).

Studies on Walkability and Physical Activity

Studies have shown associations between physical activity and numerous perceived and objectively measured attributes of the built environment, walkability being one of them (Humpel, Owen, & Leslie, 2002; McCormack et al., 2004; Owen, Humpel, Leslie, Bauman, & Sallis, 2004). As an example, Saelens et al. (2003) found that adults living in a high-walkable neighborhood (characterized by high residential density, mixed land use, and good street connectivity) in San Diego, California had 70 more minutes of accelerometer-assessed physical activity over a week than adults living in a low-walkable neighborhood in the same city. Further, a Belgian study showed that living in a high-walkable neighborhood (defined by a number of GIS-based measures) was associated with more accelerometer-based minutes of moderate-to-vigorous physical activity, and more self-reported recreational walking and walking/biking for transport (Van Dyck et al., 2010).

A recent systematic review of studies using GIS-based measures to operationalize walkability and pedometer/accelerometer-assessed daily steps to estimate physical activity in adults supported the notion that walkability is positively associated with daily number of steps (Hajna, Ross, Brazeau, et al., 2015). The systematic review was based on six studies that took place in either Europe or Asia. The authors of the review conducted a study in Canada and found

no association between walkability and accelerometer-assessed daily steps (Hajna, Ross, Joseph, Harper, & Dasgupta, 2015). Hajana and colleagues concluded that their contradicting results might stem from the fact that their study took place in a country with different environment/culture than the studies in the review. For an example, Canadians rely heavily on their cars for transport while individuals in Europe and Asia (where the previous studies took place) use other types of transport more frequently. However, Hajana and colleagues observed a positive relationship between walkability and utilitarian walking and found that individuals who reported doing some utilitarian walking did accumulate more steps.

Studies on walkability estimated by Walk Score® and physical activity (measured in various ways) show mixed results. The relationship between Walk Score® and utilitarian walking/biking appears to be quite consistent. However, the evidence for an association between Walk Score® and other types of physical activity (e.g. walking during leisure time and total walking) are less consistent (Hall & Ram, 2018). Tuckel & Milczarski (2015) examined the association between Walk Score® and numerous self-reported measures of physical activity in a nationally representative sample of 1224 American adults. They found that Walk Score® was positively associated with utilitarian walking, but not recreational walking or total walking. A study using data from the Multi-Ethnic Study of Atherosclerosis (MESA) of 4638 adults living in six cities in the U.S., found that individuals in residential areas with a low Walk Score® were less likely to report walking for transport or during their leisure time than individuals in residential areas with a high Walk Score® (Hirsch, Moore, Evenson, Rodriguez, & Roux, 2013). Further, Brown et al. (2013) found a positive association between Walk Score® and utilitarian

walking, utilitarian walking time, and meeting the physical activity recommendations through utilitarian walking among 391 recent healthy Cuban immigrants living in Florida.

To my knowledge, only two studies have examined the relationship between Walk Score® and accelerometer-assessed steps. Hajna and colleagues used data from a representative-sample of 2949 Canadian adults and found no association between Walk Score® and number of accelerometer-assessed steps (Hajna, Ross, Joseph, et al., 2015). However, Duncan et al. (2016) examined the association between Walk Score® and accelerometer-assessed steps at the trip level (as opposed to the individual level) among 227 adults living in Paris, France. Duncan and colleagues tracked participants with GPS receivers and estimated Walk Score® for each origin and destination of every trip (6969 trips) and found that a higher Walk Score® of the origin or the destination of each trip was associated with a greater number of cumulative steps per trip.

The Primary Limitation of the Literature: Residential Self-Selection

In general, review papers on the built environment and physical activity conclude that there is sufficient evidence to recommend the implementation of policies that create more walkable neighborhoods and provide increased access to places for physical activity (Kärmeniemi et al., 2018; Saelens & Handy, 2008). However, most of these studies are observational and based on cross-sectional data, making it challenging/impossible to draw inferences regarding causal effects of the built environment on physical activity. Additionally, many of these studies leave out the role of individual choices in physical activity behavior.

Up until 1980, theories focusing on interpersonal factors were dominant in the area of physical activity research (Simons-Morton, McLeroy, & Wendel, 2011). Psychological

behavioral theories such as the Theory of Planned Behavior (TPB) were popular. The TPB states that individual behavior is driven by interpersonal factors, such as attitudes and norms that lead to behavioral intention and finally the behavior itself (Fishbein & Ajzen, 1975). In TPB and other similar behavioral theories, the environment is viewed as an interaction term rather than a driver of the behavior. That is, the environment is considered to serve as a facilitator/barrier to the behavior being performed while the behavior itself is considered to be driven by interpersonal factors.

In recent years, some researchers interested in the influence of the built environment on physical activity have revived consideration of individual choices. They have pointed out that the positive association commonly seen between the built environment and physical activity, might stem at least in part from residential self-selection (McCormack & Shiell, 2011; Oakes, 2014). The term "residential self-selection" refers to the notion that people who live in a residential area that supports physical activity decided to live there because they knew the environment would support their physically active lifestyle. If that is the case, it is not the physical activity friendly environment that is causing the residents to be physically active. Rather, it is the residents' predisposition to physical activity that is generating the positive association between the built environment and physical activity. Consequently, observational studies that do not account for residential self-selection might be wrongly implying or over-estimating a causal effect of the built environment on physical activity.

How to Account for Residential Self-Selection?

Studies from the fields of demography and housing show that individuals do not randomly sort into residential areas (Bishop, 2009; Clark, 2009). Originally, researchers mainly focused on socio-demographic variables such as age, education, and income when trying to explain and predict residential location. More recent studies have shown that lifestyle preferences (meaning preferences towards particular way of living) are also important determinants of individuals' decisions on where to live (Thomas, 2011; Walker & Li, 2006). Additionally, individuals tend to select into residential areas based socioeconomic resources and discrimination (Sampson & Sharkey, 2008).

Considering the fact that individuals do not sort into residential areas randomly (Bishop, 2009; Clark, 2009; Walker & Li, 2006), it is essential to account for residential self-selection when attempting to establish a causal relationship between the built environment and physical activity. The optimal way to control for residential self-selection would be with a randomized-controlled study, where participants are randomly assigned to live in areas that differ only in terms of opportunities to be physically active. However, due to logistical reasons and costs, no randomized-controlled study of this kind has been done.

Another study design that would offer more robust evidence of causality is a quasi-experiment. A quasi-experiment would measure physical activity levels for the same individuals before and after changes in the built environment and compare these levels to physical activity levels of individuals living in another area where no changes were made (i.e. control group). Few studies of this kind have been published in peer-reviewed journals and to my knowledge, none of them have assessed the individuals' motivation or willingness to change their behavior (become more physically active) before the changes in the environment were made. Therefore, it is still

possible with this design that any changes in the intervention area were due to residents being motivated to be more physically active.

Due to limited availability of experimental or quasi-experimental data, researchers have taken various innovative approaches to accounting for residential self-selection with observational data (McCormack & Shiell, 2011). Some have used a longitudinal design (Giles-Corti et al., 2013) while others have used cross-sectional designs and tried to account for residential self-selection with instrumental variables (Greenwald & Boarnet, 2001) or propensity scoring (Cao, Handy, & Mokhtarian, 2006).

Previous Attempts to Account for Residential Self-Selection within Observational Data

A few longitudinal studies have been conducted where individuals' physical activity levels were compared before and after the individuals moved to a different neighborhood. Giles-Corti et al. (2013) surveyed participants before they moved to a new neighborhood and approximately twelve months after they moved into the new neighborhood. Giles-Corti and her team found that self-reported transport-related and recreational walking increased among participants who moved to a neighborhood with access to more destinations (e.g. parks, sport field, post office, supermarket etc.) than their old neighborhood. Another similar study that considered the association between built environment and physical activity with cross-sectional and longitudinal data found a negative association between urban sprawl and physical activity in the cross-sectional data, but this association was non-significant in the longitudinal data (Lee, Ewing, & Sesso, 2009). Despite many benefits of a longitudinal design, the design itself does not account for residential self-selection. Individuals who do move within a longitudinal study may be

moving to a neighborhood that they know will support their already physically active lifestyle or they are moving because they are motivated to adapt a physically active lifestyle.

Other methods that researchers have been using to control for residential self-selection with observational data are instrumental variables and propensity scores. An instrumental variable is a variable that 1) has causal effects on the exposure, 2) only affects the outcome through the exposure, and 3) is not associated with any confounders of the association between the exposure and the outcome (Szklo & Nieto, 2004). While instrumental variables can be powerful in controlling for endogeneity due to unobserved characteristics, their effectiveness depends on the validity of the variable (i.e. that the variable meets all of the three previously mentioned criteria). Finding a valid instrumental variable can be extremely challenging. Greenwald & Boarnet (2001) performed instrumental variable analysis to address the relationship between urban setting (exposure) and travel behavior (outcome). The instrumental variables they used were aggregated-level variables that were associated with urban setting but not travel behavior such as per capita income, percent of population with college education, percent of Black and Hispanic persons, and percent of housing units classified as rural. The instrumental variable analysis indicated that urban setting positively influenced travel behavior. Arguably, some of the instrumental variables may not meet all of the three necessary criteria. For an example, per capita income and percent of population with college education are likely to be associated with residential self-selection that is a confounder of the association between urban setting and travel behavior, thus failing the third instrumental variable criterion.

Propensity score methods (in this context) model the probability of living in an environment that is supportive of physical activity given observed individual characteristics, and

the probabilities (propensity scores) are then used as covariates in the regression analysis of built environment and physical activity. Cao (2010) used propensity scores when examining the association between neighborhood type and recreational walking. A conditional probability that an individual resided in a traditional vs. suburban neighborhood was estimated based on characteristics of the participant (including residential preferences) and included as a covariate in the analysis. Cao concluded that neighborhood type (traditional/suburban neighborhood) influenced recreational walking over residential self-selection. However, the challenge with propensity score methods is that they are based on observed characteristics and can therefore only control for residential self-selection to the extent that variables included in the model used to create the propensity scores capture determinants of selection into an environment that supports physical activity (Joffe & Rosenbaum, 1999; Oakes & Church, 2007). Therefore, including propensity scores in the analysis does not do much more than traditional methods of adjusting for socio-demographic factors.

None of the methods mentioned above have been fully successful in controlling for the possibility of residential self-selection in the relationship between the built environment and physical activity. Another possible approach is to ask individuals directly about their reasons for moving to their current neighborhood and include that information in statistical analyses on the relationship between the built environment and physical activity.

Direct Questioning: Ask about Residential Preferences

The first studies that used the approach of direct questioning came from the field of transportation, most of them examining the association between the built environment and mode

of transport. In these studies, participants were asked about their lifestyle preferences (relating to work, family, money and time use etc.), travel attitudes (e.g. attitudes toward having your own car, ridesharing and public transportation), and preferred neighborhood characteristics (important to have stores within walking distance from their home, to live in a safe or attractive neighborhood etc.) (Cao et al., 2006; Schwanen & Mokhtarian, 2005a).

Handy, Cao, & Mokhtarian (2006) used data from residents of eight neighborhoods in Northern California to examine the relationship between residential location and walking behavior. They observed a relationship between characteristics of the residential location and walking behavior even after adjusting for residential preferences. Further, they found that participants who expressed a preference for a neighborhood with opportunities to be physically active and having stores within walking distance walked more frequently to the store than those who did not express these preferences. However, preference for a neighborhood with high accessibility or having opportunities to be physically active were not found to be associated with environmental characteristics of the residential location. Bagley & Mokhtarian (2002) used data from residents of five diverse neighborhoods in the San Francisco Bay Area and found an association between environmental characteristics of the neighborhood and travel behavior. However, they found that preferences for certain environmental characteristics/mode of travel (e.g. pro-high density, pro-driving, pro-transit) and a certain type of lifestyle (e.g. adventurer, homebody, outdoor enthusiast) had greater impact on travel behavior than environmental characteristics of the neighborhood. In general, studies from the field of transportation indicate that residential self-selection based on preferences for a certain type of lifestyle/environmental characteristics/mode of travel does occur and explains part of the association between the built

environment and travel, although an independent effect of the built environment on travel behavior does exist (Cao, Mokhtarian, & Handy, 2009).

However, even though someone prefers a certain type of neighborhood/lifestyle/mode of travel, it does not necessarily mean that the environment that they live in reflect these preferences. The process of choosing a place to live is commonly a household decision (not an individual one), based on trade offs (the best house vs. best location etc.) within budget constraints (Clark & Dieleman, 1996). A study by Schwanen & Mokhtarian (2004) found that about 25 percent of the participants lived in a neighborhood that did not match their land-use preferences; for example, some lived in a low-density suburban area but stated a preference for living in a high-density urban area. Therefore, questions about residential preferences are not a good measure of residential self-selection. If the questions used to account for residential self-selection, when examining the association between the built environment and physical activity, are a poor representative of residential self-selection, it will result in imperfect adjustment.

Direct Questioning: Ask about Reasons for Moving to Current Residence

More recently, researchers started asking people about their reasons for moving to their current neighborhood to be able to account for residential self-selection more carefully. Frank, Saelens, Powell, & Chapman (2007) were among the first to ask participants about their reasons for moving to their current neighborhood when examining the association between neighborhood walkability and travel behavior. Frank and collages found that if walkability was an important reason for moving to the current neighborhood, the number of walking trips was higher irrespective of the walkability of the neighborhood.

A number of additional studies in the public health literature have utilized questions about reasons for residential location to account for residential self-selection when examining the association between the built environment and physical activity (Baar, Romppel, Igel, Brähler, & Grande, 2015; McCormack et al., 2012; Owen et al., 2007; Saelens et al., 2012; Sallis, Saelens, et al., 2009; Van Dyck, Cardon, Deforche, Owen, & De Bourdeaudhuij, 2011). The majority of these studies have taken place outside the U.S. For an example, Owen et al. (2007) examined the association between GIS-based walkability, reasons for residential location, and self-reported walking among 2650 adults living in an Australian city. Individuals who reported choosing their neighborhood because of its walkability (e.g. nearby shop and services, ease of walking, and closeness to public transportation) were more likely to report walking for transport and recreation. However, individuals who lived in highly walkable neighborhoods tended to walk more for transport (not recreation) than individuals living in low walkable neighborhoods, irrespective of reasons for residential location. Similarly, Van Dyck et al. (2011) used data from 412 adults living in Belgium and found that walkability was associated with physical activity independent of importance of walkability characteristics. They found living in a high-walkable neighborhood was associated with more accelerometer-accessed moderate-to-vigorous physical activity and self-reported minutes of active travel (no difference in self-reported recreational walking and moderate-to-vigorous leisure-time physical activity), independent of importance of walkability characteristics.

Taken together, if the studies found a relationship between the built environment and physical activity in the first place, the relationship did hold, but showed attenuation, in most cases after adjustment for reasons for moving to neighborhood responses. However, these studies

controlled for residential self-selection using a rather limited set of questions that cover only a few reasons that an active person might think about when contemplating whether a residential location under consideration is supportive of their physically active lifestyle. In addition, the wording for some of the questions was not specific enough. For an example, many questions include the wording/phrase “close to” or “near to” some type of amenities such as job or shops etc. The term “close/near to” can mean very different things to different people; it can mean it takes 10-15 minutes to drive to these amenities or it can mean it is possible to walk or bike there. These issues can result in an imperfect adjustment of residential self-selection.

Research Gap

A number of knowledge gaps regarding residential self-selection and its role in the relationship between the built environment and physical activity exist in the public health literature. First, further studies are needed to understand the possible pathways of residential self-selection and its role in the association between the built environment and physical activity. Researchers have pointed out that the literature could benefit from using more complicated models that support modeling many relationships at the same time to try to understand possible pathways (Ball, Timperio, & Crawford, 2006; Bauman, Sallis, Dzewaltowski, & Owen, 2002; Ding & Gebel, 2012; McCormack & Shiell, 2011). Residential self-selection has been identified as a potential confounder in the relationship between the built environment and physical activity (Frank et al., 2007; Handy et al., 2006; McCormack & Shiell, 2011). To be a confounder, residential self-selection needs to be associated with the built environment and it needs to influence physical activity (Szklo & Nieto, 2004). A few studies in the public health literature have examined the

association between residential self-selection and physical activity (Owen et al., 2007; Saelens et al., 2012; Van Dyck et al., 2011). However, to my knowledge, only one study has examined the relationship between residential self-selection and the built environment (Baar et al., 2015).

Second, studies need to include more specific questions about reasons related to a physically active lifestyle to be able to adjust for the possible confounding of residential self-selection. If the questions are not specific enough they may not be able to adjust fully for residential self-selection, and therefore leave room for residual confounding[3], meaning that a spurious/inflated relationship between the built environment and physical activity may still be observed.

Third, more studies that take place in the U.S. are needed, since most previous studies incorporating reasons for location choice have taken place in other countries. Context can play an important role in studies on the built environment and physical activity. A study by Sallis and colleagues found that variables often used to define the built environment (i.e. transit stops, shops, recreational facilities) varied greatly between 11 countries examined. Also, living conditions, such as the cost of owning a car and the convenience of using public transport, vary greatly among countries (Kenworthy & Laube, 1999).

Finally, studies on the implication of residential self-selection in different socioeconomic environments are scarce. Socioeconomic status has consistently been identified as an important predictor of health in the population (Feinstein, 1993). Further, individuals tend to sort into residential areas based on socioeconomic resources and discrimination, particularly in the U.S. where segregation by socioeconomic status is significant (Reardon, Townsend, & Fox, 2017;

---

[3] "Residual confounding occurs when the construct validity of the variable used for adjustment is not ideal: that is, the variable is an imperfect marker of the true variable one wishes to adjust for" (Szklo & Nieto, 2004, p. 174)

Sampson & Sharkey, 2008; Watson, 2009). It is likely that the extent of residential self-selection varies by socioeconomic status of the residential area since higher income groups may be able to satisfy more personal criteria when selecting a residential location (Levine & Frank, 2007). In addition, environmental aesthetics (e.g. trees along sidewalks, clean streets etc.) that have been found to be associated with increased physical activity often vary greatly between high- and low-income neighborhoods (Neckerman et al., 2009).

SPECIFIC AIMS

This dissertation seeks to address previously stated research gaps and consequently gain further understanding of residential self-selection and its role in the relationship between the built environment and physical activity. Specific aims of this dissertation are:

1) Investigate whether numerous questions about reasons for moving to current residence produce distinct residential self-selection profiles and test whether physical activity levels and sedentary time differ between the profiles.

2) Determine whether individuals who report reasons related to a physically active lifestyle being important when moving to their current residential area reside in areas with high walkability. Further, test whether the relationship between reasons and walkability differs between high- and low-income block groups.

3) Examine the association between walkability and physical activity and to what extent reasons related to a physically active lifestyle (i.e. residential self-selection) may influence that relationship.

SIGNIFICANCE AND INNOVATION

The ultimate goal of studies on built environment and physical activity is to identify features in the built environment that do lead to greater physical activity and ultimately better health. The results from these studies are then expected to be used to guide interventions and policies. We live in a world with limited resources (e.g., dollars, time) and thus it is very important that we use the available resources wisely and focus on proven interventions to improve the health of the population (Kindig, 2007). Therefore, it is important to know whether the association between the built environment and physical activity holds when accounting for residential self-selection. Additionally, we aim to understand the extent of the residential self-selection and if it differs by different environmental socioeconomic characteristics is crucial. If the relationship between the built environment and physical activity does not hold when properly adjusted for residential self-selection, focusing our resources on public health campaigns that aim at targeting individual's attitudes and preferences might be preferable to focusing on changing the environment. However, if the relationship does hold, understanding how the relationship varies by neighborhood context can help in designing targeted interventions for those who are most at risk for lack of physical activity and sedentary lifestyle.

The proposed study seeks to provide further understanding of the pathways of residential self-selection based on individual predisposition to physical activity and how it may influence

the association between the built environment and physical activity. The study will examine if various reasons for moving to current residence produce distinct residential self-selection profiles and if these profiles differ on physical activity (aim 1). This will help us understand whether there are truly groups of people who emphasize reasons related to a physically active lifestyle when choosing a residential location and if they are more physically active than the individuals who do not belong to this profile. Further, the study will investigate the relationship between residential self-selection and walkability to determine if people who report reasons related to a physically active lifestyle as important in determining their current residence do, in fact, live in a more walkable environment. Additionally, it will be tested whether the relationship between reasons and walkability differs by high- and low-income neighborhoods (aim 2). Finally, the proposed study will examine two possible pathways of residential self-selection and the relationship between walkability and physical activity simultaneously with a method uncommonly employed in public health, called structural equation modeling (aim 3).

To achieve aims 2 and 3, the proposed study will use more detailed questions about reasons for moving to the current residence related a physically active lifestyle than employed in previous studies. This will allow for more careful adjustment of residential self-selection and reduce the risk of residual confounding. Furthermore, since lack of consistency in measurements of the built environment is a problem (Hall & Ram, 2018; McCormack & Shiell, 2011), an increasingly popular and publicly available measure of the built environment (Walk Score®) will be used to facilitate replication. Finally, accelerometer-assessed measures of physical activity will be used in lieu of self-reported measures, which have been used in many of the previous

studies in the literature and have been shown to be less accurate (Prince et al., 2008; Troiano, Berrigan, Dodd, MâSse, et al., 2008).

The proposed study will add to the literature by examining important pathways of residential self-selection in more detail than done in previous studies, by using data from the U.S., and by examining the implications of residential self-selection in different economic environments (i.e. high- vs. low-income block groups). Understanding the potential confounding role of residential self-selection in the relationship between walkability and physical activity can help in refining and moving the existing literature forward towards more effective and informed public health interventions.

THEORY AND CONCEPTUAL FRAMEWORK

The basic principle underlying studies on the built environment and physical activity is that environmental factors can influence individual physical activity levels. This is a relatively recent idea, as before the 1980s physical activity was mainly believed to be under the control of the individual and represent individual choices (Simons-Morton et al., 2011). In 1986, Bandura introduced the Social Cognitive Theory (SCT, see model in Appendix 1), which represented a new way of thinking about human behavior. According to the central concept of SCT, reciprocal determinism, behavior is influenced by personal factors and the environment, the personal factors and the environment influence each other, and are also influenced by the behavior. The interactions among these three factors do not occur simultaneously and the strength of the influences between each pair of factors is not necessarily the same (Bandura, 1986).

The SCT set the stage for ecological theories, which are widely used within public health today. Ecological theories place individual behaviors within a broad context without removing the importance of interpersonal factors. They acknowledge that individual choices (and consequently behaviors) may be influenced by the environment in which they take place. For an example, The Social Ecological Model (SEM, see model in Appendix 1) for Health Promotion assumes that health behavior is influenced by interpersonal factors that are influenced by the physical and social environments that the individual lives in and that these environments are under the influence of policies (Stokols, 1996; Stokols, Grzywacz, McMahan, & Phillips, 2003).

Building on the SCT and SEM for Health Promotion, the conceptual model for the proposed study is presented in Figure 1. As opposed to SCT where all relationships are viewed as bidirectional, the relationships between residential self-selection and physical activity, and residential self-selection and the built environment, are shown as unidirectional. The rationale behind the unidirectional relationships between residential self-selection and physical activity, and residential self-selection and the built environment, is the temporal sequence, viz., residential self-selection refers to reasons people moved to their current residence. Conversely, built environment and physical activity are obtained after the people moved to their current residence. The relationship between built environment and physical activity is also shown as unidirectional since being more physically active is not expected to change the built environment. That thought is in line with the SEM for Health Promotion where individuals are believed being capable of influencing changes in their environment but not directly with the health behavior of interest (i.e. physical activity) (Simons-Morton et al., 2011).

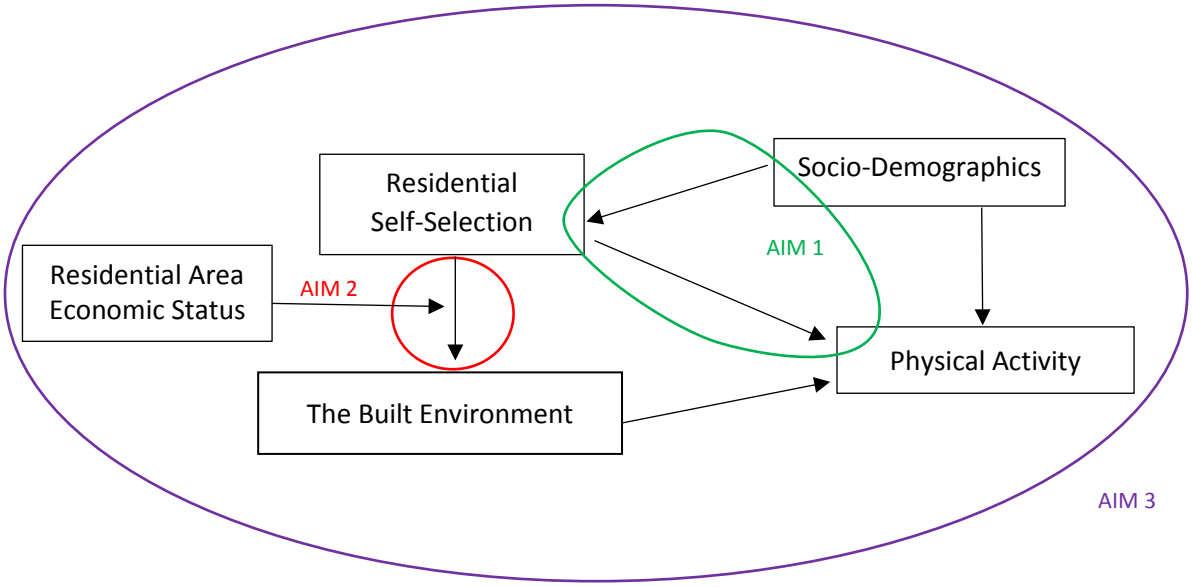


Figure 1. Conceptual model for the proposed study.

METHODS

Data Source

All aims of the proposed study will use data from the Survey of the Health of Wisconsin (SHOW). SHOW is an ongoing repeated annual cross-sectional population-based health examination survey of civilian, non-institutionalized residents of Wisconsin (WI) of all ages. Details on SHOW protocols have been published previously (Nieto et al., 2010). For the first six years of the survey (2008-2013) a two-stage, probability-based cluster sampling approach was used to randomly select households and each year consisted of a representative sample of the state of WI. In 2014, the sample strategy was shifted to a three-stage, probability-based cluster sampling approach where a three-year sample is representative of the state of WI. The three stages involve 1) counties, 2) 2010 Census block groups, and 3) households.

Participants are recruited via an in-home visit (following a mailed introductory letter to the household) from a centrally trained and certified field staff interviewer. The survey consists of household interview, self-administered questionnaire (SAQ), physical exam, and accelerometry and biosample collection. All study protocols are approved by the University of Wisconsin Health Sciences Institutional Review Board and all participants provide written informed consent. Randomly selected audio tapes from interviews are reviewed by SHOW staff on an ongoing basis to monitor interviewer adherence to the protocols.

The sample for the proposed study is limited to adults (≥18 years old) who participated in SHOW in 2016 and were living in an urban area according to the 2010 U.S. Census Bureau's urban-rural classification system. The reason for limiting the sample to one year is that in 2016 three additional questions related to the topic of interest (residential self-selection) were added to SHOW. Further restricting the sample to only urban participants was done since very limited variability was found in the main measure used for walkability among rural participants (see appendix 2).

In 2016, 478 urban adult residents were recruited into SHOW. Due to missing data, the number of participants in the analyses range from 258 to 383. Most missing data stem from participants not participating in or not having a valid wear time in the accelerometry section (n=177) or not completing the SAQ (n=97). A comparison of characteristics of individuals with valid accelerometry and self-administered questionnaire data (n=269) with those of individuals who did have this data missing (n=209) are presented in appendix 3. In 2016, SHOW collected data from four counties in WI: Milwaukee, Brown, Eau Claire, and Waushara; no participants from Waushara are included in the analyses since the whole area is classified as rural.

Key Measures

Residential-self-selection is measured with a self-administered questionnaire that asks about participants' reasons for moving to their current residence: "*Thinking back when you moved to your current residence, at that time, how important were each of the following factors in your decision to move to your current residence?*" Participants are presented with 19 different items/neighborhood characteristics and indicate their answer on a 4-point scale ranging from "*Not at all important*" to "*Very important*". The questionnaire can be seen in Appendix 4. A residential self-selection scale ranging from 1-4 is created in aim 2 and 3. A higher score on the scale indicates greater importance of living in an environment that supports a physically active lifestyle.

Sixteen items in the questionnaire originate from previous work by Dr. Kirsten Beyer, Dr. Paul Rathouz, and colleagues in 2013. They developed these questions when working on a grant proposal and brainstorming factors influencing people's decisions to move to specific neighborhoods. The questions represented a broad spectrum of reasons identified in the literature and by the team, without a specific focus on physical activity. In anticipation of the proposed study, three new items developed by Unnur Gudnadottir were added to the questionnaire in 2016. The three items were created in an effort to be able to identify more carefully participants that moved to their current residence because they knew they could be physically active there.

Walkability of the residential environment is estimated with one objective and one subjective measure. Studies have found moderate to low levels of agreement between subjective and objective measures of the built environment (Ball et al., 2008; Gebel et al., 2011; McCormack et

al., 2008; Lackey et al., 2009; Kirtland et al., 2003), and discordance between measurements is even greater in urban areas (Bailey et al., 2014). It is unclear to what extent subjective and objective measures of the built environment assess similar constructs or whether they assess different constructs. Tuckel & Milczarski (2015) found that a subjective measure of walkability was associated with walking for transport, recreational walking, and total walking but an objective measure of walkability was only associated with walking for transport. Some researchers have suggested that both types of measurement should be included in future studies (Bailey et al., 2014; Mcginn et al., 2007). The Pearson correlation between the objective (Walk Score®) and subjective (self-reported walking distance to a number of facilities) measures used in this dissertation is -0.62. The negative association stems from that higher scores on the objective measure indicate greater walkability but higher scores on the subjective measure indicate less walkability.

Objective measure of walkability - the objective measure used is Walk Score® (www.walkscore.com). Walk Score® is a publicly available measure based on network distance (i.e. following street lines) to amenities, taking intersection density and block length into account ("Walk Score Methodology," 2016). Walk Score® ranges from 0 to100, with a higher score indicating greater walkability. For every Walk Score®, hundreds of walking routes are generated using proprietary algorithms and various data streams (e.g. Google, Education.com, Open Street Map, the U.S. Census, Localeze, and places added by the SSWS users) to find nearby amenities.

Amenities are divided into five categories: educational (e.g., schools), food (e.g., restaurants, coffee shops), retail (e.g., grocery stores, pharmacies, convenience, and bookstores),

recreational (e.g., parks and gyms), and entertainment (e.g., movie theaters). Amenities get points based on the how far away from the origin they are located; amenities within .25 miles (i.e. 5 minute walking distance) are given maximum points, while amenities further away than .25 miles get fewer points and amenities located 1.5 miles (i.e. 30 minute walk) or further away from the origin get no points. Amenities that have been found to be highly correlated with walking (that is, places that people tend to walk to) are given more weight. The Walk Score® also takes into account the number of amenities in the same category, for an example having more than one grocery store in a walking distance will not increase the SSWS but having more than one restaurant will. Finally, the Walk Score® methodology includes penalties for low intersection density and long blocks.

In 2010, the algorithm for Walk Score® was revised but a systematic review of studies on Walk Score® and physical activity concluded that both versions yielded the same pattern of results (Hall & Ram, 2018). The previous version of Walk Score® has been shown to correlate positively with several objective GIS-based measures: street connectivity (intersection density, street density, and average block length), population density, and access to public transit (Carr et al., 2010b). It has also been shown to correlate highly (r = 0.56-0.71) with access to walkable amenities (Carr et al., 2010a).

The benefits of using Walk Score® over other available objective measures of the built environment, such as “homegrown” GIS-based measures are that Walk Score® is a standard and publicly available measure. Getting access to GIS based measures can be difficult and these measures are often expensive. Furthermore, researchers using GIS based measures tend to use their “own” measure causing difficulties when comparing results across different studies.

Subjective measure of walkability – the subjective measure used for walkability is derived from a self-administered 18 item questionnaire that asks participants to indicate how many minutes it would take them to walk from their home to the nearest specified types of facilities. The questionnaire can be seen in Appendix 5. Participants indicate their answer on a six-point scale ranging from "*0-5 minutes*", "*6-10 minutes*", "*11-20 minutes*", "*21-30 minutes*", "*More than 30 minutes*" to "*None within walking distance*". The questionnaire includes facilities such as supermarket, trail for walking or biking, post office, school, and restaurant. A self-reported walkability scale is created from the 18 questions. Scores on the scale range from 1-6, with higher scores indicating less walkability. An average score of 2.5 or less is considered having high walkability since neighborhoods are commonly considered walkable when services that individuals need to be able to access on a daily bases are within 10 minutes of walking (Talen & Koschinsky, 2013).

Physical activity and sedentary time are estimated with two different methods, accelerometers and via self-report. Despite having several limitations (see review in Pedišić & Bauman, 2014) accelerometer-assessed physical activity measures have been deemed to provide more accurate and precise estimates of true physical activity than self-reported measures (Prince et al., 2008). However, it was decided to include both types of measures in aim 1 since it was of interest to see if the different types of measures would provide different patterns of results. A Pearson correlation between the self-reported and accelerometer-assessed measures of physical activity and sedentary time used in this study is 0.39 and 0.34, respectively.

Measures of total physical activity levels are used as opposed to commonly used measures, such as minutes spent in moderate and vigorous physical activity. This ensures that all types of physical activity, even low intensity activity are captured. It has been argued that total physical activity is the most salient correlate of health benefits and is therefore what matters from a public health perspective (as opposed to whether individuals walk or drive to work, for example) (Committee on Physical Activity, Health, Transportation, and Land Use, 2005; Hajna, Ross, Brazeau, et al., 2015).

Accelerometer-assessed measures - Actigraph WGT3X-BT devices (Actigraph Corporation, Pensacola FL) were used to obtain an objective measure of physical activity and sedentary time. Participants were instructed to wear the device on the right hip for seven continuous days. The accelerometers are programmed to collect data in one-second intervals but are aggregated to 60-second intervals for wear time validation, scoring, and analyses. Accelerometer data are processed using ActiLife version 6.13. The adult Freedson cut points are used for wear time validation. A participant who wears the devices for at least 10 hours for 3 out of 7 days is considered having a valid wear time (Freedson, Melanson, & Sirard, 1998). Only participants with valid wear time are included in the analyses.

Physical activity is operationalized as the average number of steps a participant takes every minute while wearing the actigraph device. Step count is an accepted and one of the most commonly used measures of physical activity (Bassett, Toth, LaMunion, & Crouter, 2017). Sedentary time is operationalized as the average hours participant spends in sedentary intensity a day. The Actigraph device has been shown to provide an accurate step count (Chow, Thom,

Wewege, Ward, & Parmenter, 2017) and to provide useful estimates of sedentary time in populations (Healy et al., 2011).

Self-reported measures - Self-reported physical activity is estimated with an interviewer-administered version of the Global Physical Activity Questionnaire (GPAQ; Armstrong & Bull, 2006). The GPAQ contains questions about physical activity within three domains: 1) household/domestic, 2) sport/fitness/leisure, and 3) transportation. Each domain includes questions about frequency and duration of activity. Additionally, the first two domains include questions about intensity of activity (moderate or vigorous). Participants are asked to think about a typical week while answering the questions and only report activity that lasted for at least 10 minutes continuously. Total physical activity is calculated by adding the amount of reported activity within each domain resulting in a variable displaying hours of physical activity within a week across all domains. Sedentary time is assessed with one question: "How much time do you usually spend sitting on a typical day?" and operationalized as total hours spent sitting on a typical day. The GPAQ is considered an appropriate instrument to assess self-reported physical activity and sedentary time in populations (Bull, Maslin, & Armstrong, 2009).

Statistical Analyses

Aim 1: Analyses for aim 1 are conducted in Mplus 7.4 and SAS 9.4. Mplus is used to conduct a latent profile analysis (LPA) on the 19 residential self-selection questions. LPA derives a categorical variable that represents groups of individuals who share a similar residential self-selection profile. LPA identifies mutually exclusive groups that maximize between-group

variance and minimize within-group variance based on model fit criteria (Collins & Lanza, 2010). Then four regression analyses with four different outcomes, i.e. accelerometer-assessed and self-reported total physical activity and sedentary time are run in Mplus. The new categorical variable derived from LPA is the main independent variable of interest while a number of sociodemographic characteristics are included as covariates. Finally, SAS is used to compare sociodemographic characteristics of the distinct residential self-selection profiles using the Mann-Whitney U test (for continuous variables) and chi-square test (for categorical variables).

To find the optimal number of profiles, LPA with 1-5 profiles are conducted and each solution evaluated by the following fit indices: Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMRT; Lo, Mendell, & Rubin, 2001), Akaike information criteria (AIC; Akaike, 1974), sample size-adjusted Bayesian information criteria (sBIC; Schwarz, 1978), and entropy. A significant (i.e. p-value < 0.05) LMRT test indicates that a model with more profiles is a better fit than a model with one fewer profile (e.g., a 3 profile model is better than a 2 profile model). Smaller values of AIC and sBIC indicate better model fit and higher values of entropy (range: 0-1) indicate greater precision of assigning latent profile membership. In addition to these indices, conceptual interpretations of the resultant profiles are considered when determining the number of profiles.

When running the LPA, group specific weights that reflect the measurement error of the categorical variable derived from the LPA are created. These weights are then included in the regression analyses to account for the measurement error of the categorical variable, which indicates the different residential self-selection profiles. More details about creation of the weights and their use in the regression can be found in (Asparouhov & Muthén, 2014). The aim

of the regression analyses is to test if the residential self-selection profiles differ in levels of physical activity and sedentary time. Maximum likelihood estimates with robust standard errors (MLR estimator) are used to account for the non-normality of the indicators and the outcomes in the regression. Adjustment for non-independence between participants is made by specifying Census 2010 block groups as clusters (i.e. higher level sampling units).

All analyses are restricted to individuals with complete data. Depending on the outcome, the sample size ranges from 259-361. Boxplots in SAS are used to identify outliers; values above 1.5 times the interquartile range are considered to be outliers (Aguinis, Gottfredson, & Joo, 2013).

Aim 2: All statistical analyses for aim 2 are conducted in SAS 9.4. To compare characteristics of individuals living in walkable versus non-walkable areas, unadjusted bivariate analysis are conducted using Chi-square analysis (for categorical variables) and independent t-test (for continuous variables).

Two confirmatory factor analyses (CFA) are conducted using PROC CALIS. One CFA is conducted to test if the six items from the reasons for moving to current residence questionnaire identified as being relevant to a physically active lifestyle load on one factor. Another CFA is conducted to test if the eighteen questions from the self-reported walkability questionnaire load on one factor. For both of the CFAs, a polychoric correlation matrix with unweighted least-squares (ULS) factor extraction is used since the variables are on an ordinal scale (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Han, Neilands, & Dolcini, 2011). Model fit is assessed using four fit indices: 1) Standard root-mean-square residual (SRMR); 2) Goodness of

fit index (GFI); 3) Adjusted GFI (AGFI); and, 4) Normed fit index (NFI). For SRMR a value less than 0.08 is considered a good fit, and a value of 0.95 or greater for GFI, AGFI, and NFI suggests good model fit (Hu & Bentler, 1999). Items have to have a minimum factor loading of 0.3 to be considered being a part of the factor, but higher factor loadings are preferred (Kline, 2015).

To evaluate the association between walkability, residential self-selection, and neighborhood economic status, linear regression (PROC SURVEYREG with DOMAIN statement) models are used, with the residential self-selection scale as the outcome and Walk Score and self-reported walkability scale as independent variables (in separate models). An interaction term between each of the walkability variables and average block group income is used to test if the association between walkability and residential self-selection is moderated by census block group economic status. An analysis stratified by average block group income is performed for significant interaction terms.

To account for SHOW's multistage sampling design, survey procedures are used in SAS (PROC SURVEYREG) and cluster and stratum variables are specified. Further, a subpopulation (urban residents who participated in 2016) is specified with a DOMAIN statement to indicate that the data in the analyses are a subset of a population-based sample. All analyses were restricted to individuals with complete data (N=388).

Aim 3: Analyses for aim 3 are conducted using Mplus 7.4 and SAS 9.4. SAS is used to calculate descriptive statistics for the study sample and Pearson's bivariate correlations between the main variables of interest. Mplus is used to run a confirmatory factor analysis (step 1) on six variables

from the residential self-selection questionnaire and a structural equation model (step 2) to examine the association between residential self-selection, walkability, and physical activity simultaneously. This two-step process has been recommended when constructing and testing structural equation models (Anderson & Gerbing, 1988).

While running the models in Mplus, the WLSMV estimator is used. The WLSMV estimator provides weighted least square (WLS) parameter estimates using a diagonal weight matrix with robust standard errors and mean- and variance-adjusted chi-square statistics. It is recommended to use the robust WLS method when analyzing categorical or ordered data that are not normally distributed (T. A. Brown, 2014), as is the case in the current study. In step 1, variables have to have a minimum factor loading of 0.3 to be considered indicative of the latent variable, but higher factor loadings are preferred (Kline, 2015). The fit of the models is evaluated using various fit indices that are considered appropriate to use with WLSMV. The fit indices and cutoff values used in this study are 1) Chi-square test (χ^2) with a *p*-value greater than .01; 2) Root Mean Square Error of Approximation (RMSEA) less than .06; 3) Comparative Fit Index (CFI) equal or greater than .96; 4) Tucker-Lewis Index (TLI) equal or greater than .95; and 5) Weighted Root Mean Square Residual (WRMR) equal to or less than 1. The proposed cutoff values come from a study by Yu (Yu, 2002). The goal is to keep the models as parsimonious as possible; residuals and modification indices are used to decide on alternative models. The fit of alternative models are compared with the chi-square difference test (DIFFTEST command) (Kline, 2015).

To account for SHOW's multistage sampling method, both cluster and stratum variables are included in the analyses. Further, a SUBPOPULATION statement is included to indicate the

sample of interest, i.e. individuals who participated in SHOW in 2016 and live in an urban area. All analyses are restricted to individuals with complete data. Boxplots in SAS are used to indicate outliers, values above 1.5 times the interquartile range are considered to be outliers (Aguinis et al., 2013).

AIM 1: Residential Self-Selection Profiles, Physical Activity and Sedentary Time: A Latent Profile Analysis

Abstract

Studies on the built environment and physical activity often fail to account for potential issues of residential self-selection, thus leading to biased estimates of the association between the built environment and physical activity. Few studies have examined predictors of residential self-selection and its association with physical activity/sedentary time. Latent profile analysis was utilized to explore whether 19 self-reported reasons for moving to current residence produced distinct residential self-selection profiles. Sociodemographic characteristics, physical activity and sedentary time were compared between profiles. Data from the 2016 cohort of the Survey of the Health of Wisconsin were used, restricted to adults living in urban areas (n=383). Two residential self-selection profiles emerged and were labeled: non-active lifestyle and active lifestyle. The profiles were similar on sociodemographic characteristics except that a higher portion of individuals in the non-active lifestyle profile owned/leased a car. The active lifestyle profile had higher physical activity levels (accelerometer-assessed estimates of average steps per minute) and lower sedentary time (self-reported) than the non-active lifestyle profile. These results provide some evidence of residential self-selection that may lead to fallacious association between the built environment and physical activity even when adjusted for common sociodemographic factors.

Introduction

Physical activity is any bodily movement that is produced by skeletal muscles and results in energy expenditure. It includes daily activities such as household chores, walking and biking during leisure time or when travelling between places, and running for exercise (Caspersen et al., 1985). Physical activity has been shown to have health benefits and reduce risk for many chronic diseases including coronary heart disease, hypertension, some cancers, and diabetes (U.S. Department of Health and Human Services, 1996; Warburton, Nicol, & Bredin, 2006b). In the United States (U.S.) chronic diseases are among the leading causes of death (Murray et al., 2013) and it is estimated that individuals with chronic diseases account for 86% of all health care spending (Gerteis et al., 2014).

In contrast with physical activity, sedentary behavior is any behavior characterized by sitting, lying or reclining and low levels of energy expenditure (Tremblay et al., 2017). Common forms of sedentary behavior are sitting and watching TV, sitting and using a computer, and sitting while commuting by car. Mounting evidence indicates that sedentary time is a distinct risk factor, independent of physical activity, for poor health outcomes and some chronic diseases (Hu et al., 2003; Thorp et al., 2011). Despite the health benefits of physical activity being widely known and growing awareness of the adverse effects of a sedentary lifestyle, U.S. adults spend the majority of their waking time in sedentary behaviors and less than one-half are getting the recommended level of physical activity (Centers for Disease Control and Prevention (CDC), 2008; Matthews et al., 2008; Troiano, Berrigan, Dodd, Mâsse, et al., 2008).

The lack of physical activity and the high amount of sedentary time have partly been blamed on the environment that people live in (Brownson, Boehemer, & Luke, 2005). Studies

have shown a positive association between numerous features of the built environment and physical activity (Humpel et al., 2002; McCormack et al., 2004), leading some researchers to conclude that higher levels of physical activity in an area are a result of the built environment in that particular area (Hirsch et al., 2014; Saelens et al., 2003a). Other researchers have pointed out that the observed association between the built environment and physical activity might stem from residential self-selection. Residential self-selection being that individuals who are predisposed to a physically active lifestyle choose to live in a built environment that is more conducive to such behavior, leading to a non-causal association between the built environment and physical activity (Oakes, 2014).

A number of cross-sectional studies examining the association between the built environment and physical activity have attempted to account for residential self-selection. These studies have commonly used questions about participant environmental preferences (e.g. preferences for having stores close by or opportunities for physical activity). However, even though someone prefers a certain type of environment it doesn't mean that the environment that they live in necessarily reflects these preferences. The process of choosing a place to live is commonly a household decision (not an individual one), based on trade offs (the best house vs. best location etc.) within budget constraints (Clark & Dieleman, 1996). A study by Schwanen & Mokhtarian (2004) found that about 25 percent of the participants lived in a neighborhood that did not match their environmental preferences; for example, some lived in a low-density suburban area but stated a preference for living in a high-density urban area. Therefore, questions about residential preferences are not a good measure of residential self-selection and researcher

have started using questions about reasons for moving to current residence to operationalize residential self-selection.

Previous studies, which have mostly taken place outside the U.S., have shown an association between reasons for moving to current residence and physical activity. For example, a German study on college students living in the city of Leipzig showed that reasons for moving to current residence were related to self-reported physical activity levels (Baar et al., 2015). Further, Christiansen et al. (2014) found that Danish adults living in Aarhus who reported reasons related to destination accessibility being important when they decided to move to their current residence, reported more frequently walking for transport than those who did not report these reasons being important. One of the few U.S. studies on this topic found that individuals reporting reasons related to physical activity being important when moving to their current residence had higher levels of self-reported and accelerometer-assessed physical activity (Saelens et al., 2012).

Studies that have specifically examined the importance of different reasons for moving to current residence indicate that the importance of reasons related to a physically active lifestyle may differ between countries. A study on Belgian adults showed that walkability characteristics of a neighborhood (e.g. closeness to shops, work/school, amount and quality of sidewalks, traffic safety) were considered similarly important as house price and the desire to live in a quiet neighborhood. The walkability characteristics were considered more important than living close to family and friends or living near the neighborhood people grew up in (Van Dyck et al., 2011). Similar pattern was found among Danish adults, “getting around easily by bicycle”, “getting around easily on foot” and “short distances to green areas” were considered being the most

important reasons after “safe neighborhood” (Christiansen et al., 2014). While in a U.S. study participants tended to report importance of reasons related to an environment supportive of physical activity relatively low (e.g “Ease of walking” and “Near outdoor recreation”). Things such as “Low crime” and “Affordability” weighed more heavily in the decision process (Frank et al., 2007). Findings suggest context and culture may play a role in determining residential location and subsequently residential self-selection. Further studies, especially in the U.S. since most previous studies have taken place in other countries, are needed to gain further insight into possible pathways and importance of residential self-selection.

The overall goals of the current study were two-fold. First, explore weather nineteen reasons reported for moving to current residence produce distinct residential self-selection profiles using latent profile analysis (LPA). Second, test whether physical activity and sedentary time differs by residential self-selection profiles. We used data from a well-characterized population study to test our hypothesis. We hypothesized that we would identify distinct residential self-selection profiles with varying values for importance of reasons related to a physically active lifestyle when moving to current residence and that these profiles would differ by sociodemographic characteristics. Furthermore, we hypothesized that individuals belonging to the profile/s indicating high importance of reasons related to a physically active lifestyle when moving to current residence would have higher levels of physical activity and lower levels of sedentary time compared to the other profile/s.

Using LPA allows us to capture the complexity of residential location choice as opposed to investigating one reason for moving to current neighborhood at a time. To our knowledge, this is the first study using LPA modeling for residential self-selection profiles, but this technique has

been used previously to produce neighborhood environment profiles (Adams et al., 2011; McDonald et al., 2012; Norman et al., 2010; Todd et al., 2016).

Methods

Participants

Data for this study come from the 2016 cohort of the Survey of the Health of Wisconsin (SHOW). Details on SHOW protocols have been published previously (Nieto et al., 2010). In brief, SHOW is a cross-sectional population-based health examination survey of civilian, non-institutionalized residents of Wisconsin of all ages. SHOW consists of the following sections: in-home interview accompanied by physical measurements, self-administered questionnaire, accelerometry, and collection of biological samples. Participants are recruited from a random sample of households using a three-stage stratified cluster sampling approach. The three stages involve 1) counties, 2) Census 2010 block groups, and 3) households. While SHOW collects data on urban and rural populations, the focus of this study is on participants 18 years old and older living in an urban area (N=478). Due to participant drop off in the self-administered questionnaire and the accelerometry section, the number of participants in the analyses range from 259-383. All SHOW study protocols were approved by the University of Wisconsin Health Sciences Institutional Review Board and all participants provided written informed consent as part of the survey.

Measures

Exposure Variable. The main exposure of interest was residential self-selection profile. The residential self-selection profile variable was generated with LPA of residential self-selection questions (see Table 2). The questions were self-administered and asked participants about their reasons for moving to the current neighborhood: "*Thinking back to when you moved to your current residence, at that time, how important were each of the following factors in your decision to move to your current residence?*" Participants were presented with 19 different items/neighborhood characteristics and indicated their answer on a 4-point scale ranging from "*Not at all important*" to "*Very important*". Examples of items are "*Parks and open spaces nearby*", "*Affordable housing*" and "*Ease of walking or biking to services (such as work, school, grocery stores, restaurants, etc.)*".

Outcome Variables. The outcomes of interest were physical activity and sedentary time, both outcomes were accelerometer-assessed and self-reported. Accelerometer-assessed physical activity and sedentary time were measured with Actigraph WGT3X-BT device (Actigraph Corporation, Pensacola FL). Participants were instructed to wear the device on the right hip for seven continuous days. Physical activity was operationalized as the average number of steps a participant took every minute while wearing the actigraph device. Sedentary time was operationalized as the average hours participant spent in sedentary intensity a day. Accelerometer data were processed using ActiLife version 6.13. Participants who wore the devices for at least 10 hours for 3 out of 7 days were considered to have a valid wear time (Freedson et al., 1998). Only participants with valid wear time were included in this analysis.

The Actigraph device has been shown to provide an accurate step count (Chow et al., 2017) and to provide useful estimates of sedentary time in populations (Healy et al., 2011).

Self-reported physical activity and sedentary time were estimated with an interviewer-administered version of the Global Physical Activity Questionnaire (GPAQ; Armstrong & Bull, 2006). The GPAQ contains questions about physical activity within three domains: 1) household/domestic, 2) sport/fitness/leisure, and 3) transportation. Each domain includes questions about frequency and duration of activity; additionally, the first two domains include questions about intensity of activity (moderate or vigorous). Participants were asked to think about a typical week while answering the questions and only report activity that lasted for at least 10 minutes continuously. Total physical activity was calculated by adding the amount of reported activity within each domain resulting in a variable displaying hours of physical activity within a week across all three domains. Sedentary time was assessed with one question "How much time do you usually spend sitting on a typical day?" and operationalized as total hours spent sitting on a typical day. The GPAQ is considered an appropriate instrument to assess self-reported physical activity and sedentary time in populations (Bull et al., 2009).

Demographic Characteristics and Other Covariates. The following sociodemographic variables were included in the study: gender (female, male), age (years), race/ethnicity (non-white, white), marital status (not married, married), education (high school or less, some college, college graduate), car ownership (no, yes=owns or lease a car), kids living in household (no, yes), month of survey administration (1-12), poverty income ratio (ratio of midpoint of household income range to annual poverty level, the higher the ratio the financially better off is the individual).

Poverty levels were calculated using 2016 poverty guidelines from the U.S. Department of Health and Human Services (U.S. Department of Health and Human Services., 2017). Participant address was used to categorize participant into urban or rural residents based on the 2010 U.S. Census Bureau's urban-rural classification system.

Statistical Analyses

Mplus 7.4 software was used to conduct a two-step latent profile analysis (LPA) with a distal outcome using the BCH method (Asparouhov & Muthén, 2014a). LPA derives a categorical latent variable that represents groups of individuals who share a similar residential self-selection profile. LPA identifies mutually exclusive groups that maximize between-group variance and minimize within-group variance based on model fit criteria (Collins & Lanza, 2010). Maximum likelihood estimates with robust standard errors (MLR estimator) were used to account for the non-normality of the indicators and the outcomes in the regression. Adjustment for non-independence between participants was made by specifying Census 2010 block groups as clusters (i.e. higher level sampling units).

In the first step, a latent profile model was estimated with the 19 residential self-selection questions. Simultaneously, the group specific BCH weights, which reflect the measurement error of the categorical variable, were created. We ran LPA with 1-5 profiles. To determine the optimal number of profiles each model was evaluated using Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMRT; Lo, Mendell, & Rubin, 2001), Akaike information criteria (AIC; Akaike, 1974), sample size-adjusted Bayesian information criteria (sBIC; Schwarz, 1978), and entropy. A significant (i.e. p-value < 0.05) LMRT test indicates that a model with more profiles

is a better fit than a model with one fewer profile (e.g., a 3 profile model is better than a 2 profile model). Smaller values of AIC and sBIC indicate better model fit and higher values of entropy (range: 0-1) indicate greater precision of assigning profile membership. In addition to these indices, conceptual interpretation of the resultant profiles was considered when determining the number of profiles.

In the second step, four linear regression models were estimated using the BCH weights as training data. The aim of the regression analyses was to test if the residential self-selection profiles identified in the first step did differ in levels of physical activity and sedentary time. Covariates included in the regression models were age, gender, race, education, poverty income ratio, marital status, kids living in household, car ownership, and month of participation in survey.

The sociodemographic characteristics of the distinct residential self-selection profiles were compared using Mann-Whitney U test (for continuous variables) and chi-square test (for categorical variables) in SAS 9.4 using the most likely profile membership. Six values in the outcome variable average steps/minute were identified as outliers (above 1.5 times the interquartile range); these values were set to the highest values, i.e. the 1.5 times the interquartile range (Aguinis et al., 2013). All analyses were restricted to individuals with complete data.

Results

Estimation of the Number of Profiles

All the model fit indices (see Table 1) indicated that a 2-profile solution fit the data better than a 1-profile solution. The 2- profile solution had significant (p<.001) LMRT, lower values of AIC

(19192.89 vs. 20572.33) and sBIC (19237.86 vs. 20601.79) than the 1- profile solution. The 3-profile solution had a slightly lower AIC and sBIC than the 2-profile solution, however, LMRT was insignificant (p>.05) and the 2-profile solution had a slightly higher entropy (0.92 vs. 0.89). Based on the insignificant LMRT, higher entropy and the conceptual interpretation of the profiles the 2- profile solution was deemed superior.

The average profile probabilities for most likely profile membership were 0.98 and 0.97 for class 1 and 2, respectively. These probabilities indicated good assignment accuracy.

Table 1. Model fit statistics for latent profile analysis with different number of classes. Urban residents in SHOW 2016 (n=383).

Solution	LMRT (*p*)	AIC	sBIC	Entropy
1 profile		20572.33	20601.79	
2 profiles	1407.61 (<0.001)	19192.89	19237.86	0.92
3 profiles	364.75 (0.340)	18865.08	18925.54	0.89
4 profiles	241.36 (0.624)	18661.69	18737.66	0.88
5 profiles	188.65 (0.575)	18511.45	18602.93	0.89

Note. LMR = Lo-Mendell-Rubin Adjusted Likelihood RatioTest, AIC = Akaike Information Criterion, sBIC = sample size adjusted Bayesian Information Criterion

Residential Self-Selection Profiles

Based on the most likely profile membership, 57.4% (n=220) of the study population was assigned to profile 1 and 42.6% (n=163) to profile 2. Individuals in profile 2 had on average higher item response means on all of the 19 residential self-selection questions than individuals in profile 1. The greatest difference in the item response means was on questions 12-19 (see Table 2). Questions number 12, 14, 15, 17, 18 and 19 can be considered being relevant to physically active lifestyle and therefore profile 2 was labeled as “active lifestyle” and profile 1 as “non-active lifestyle.”

Overall and within the profiles, the reasons for moving to current neighborhood that had the highest response means were “Safe, low-crime area”, “Affordable housing”, and “Attractive, well-kept homes.” In general, the item response means were rather low for the six questions relevant to a physically active lifestyle (12, 14, 15, 17, 18 and 19).

Table 2. Item response means and standard errors for the residential self-selection questions for the total sample and for each profile. Urban residents in SHOW 2016 (n=383).

		Total sample n=383 Mean (SE)	**Profile 1** n=220 Mean (SE)	**Profile 2** n=163 Mean (SE)	**Mean difference**
1	Close to work/job	2.41 (1.10)	2.24 (0.08)	2.65 (0.09)	0.41
2	Lots of trees and other greenery	2.50 (1.02)	2.21 (0.08)	2.91 (0.10)	0.71
3	Close to friends/family	2.23 (1.05)	1.99 (0.07)	2.57 (0.10)	0.58
4	High quality schools	2.44 (1.25)	2.23 (0.09)	2.74 (0.14)	0.51
5	Affordable housing	3.18 (0.95)	3.00 (0.07)	3.43 (0.09)	0.43
6	Friendly neighbors	2.80 (0.97)	2.49 (0.06)	3.21 (0.10)	0.72
7	Safe, low-crime area	3.37 (0.82)	3.14 (0.06)	3.68 (0.06)	0.54
8	Attractive, well-kept homes	3.07 (0.88)	2.79 (0.05)	3.46 (0.08)	0.68
9	Amenities such as community center or church	2.16 (1.01)	1.85 (0.06)	2.59 (0.11)	0.74
10	Privacy and quiet	2.91 (0.89)	2.68 (0.05)	3.22 (0.10)	0.54
11	High racial or ethnic diversity	1.66 (0.94)	1.36 (0.08)	2.06 (0.12)	0.70
12	Parks and open spaces nearby	2.54 (0.99)	2.04 (0.07)	3.24 (0.06)	1.20
13	Lots of things to do	2.38 (0.93)	1.90 (0.08)	3.04 (0.09)	1.13
14	Outdoor recreational opportunities	2.25 (1.03)	1.64 (0.06)	3.09 (0.08)	1.45
15	Easy access to public transit	1.89 (1.10)	1.49 (0.09)	2.45 (0.14)	0.96
16	Easy access to health care or other services	2.25 (1.05)	1.76 (0.06)	2.94 (0.12)	1.18
17	Ease of walking or biking to services	2.14 (1.07)	1.52 (0.06)	2.99 (0.11)	1.47
18	Opportunities for outdoor activities during free time	2.32 (0.98)	1.73 (0.06)	3.15 (0.07)	1.42
19	Easy access to a gym or other workout facility	1.75 (0.96)	1.28 (0.04)	2.39 (0.11)	1.11

Sample and Profile Characteristics

The average age of the participants was approximately 50 years; the majority of the participants were white (83.81%), married (54.97%), owned/leased a car (81.72%), did not have a child living in the household (65.54%) and were highly educated (52.36% college graduates).

The two self-selection profiles did not differ significantly on sociodemographic characteristics (see Table 3), except that more individuals in profile 1 (non-active lifestyle profile) owned or leased a car compared to profile 2 (85.5% vs. 76.7%). When the profiles were compared on physical activity and sedentary measures, individuals in profile 1 (non-active lifestyle profile) reported on average 49 more minutes of sitting per day than individuals in profile 2 (U=28486.00, p=0.022).

Table 3. Characteristics of the sample and the two residential self-selection profiles (significance test comparing the profiles). Urban residents in SHOW 2016 (n=383).

	Total Sample n=383	**Profile 1** n=220	**Profile 2** n=163	***p*-value**[a]
Age in years (mean, SD)	50.26 (17.59)	48.82 (17.22)	52.19 (17.94)	0.061
Poverty income ratio (mean, SD)	3.28 (2.47)	3.31 (2.52)	3.25 (2.40)	0.870
Child living in household %	34.46	36.82	31.29	0.260
Male gender %	42.04	42.27	41.72	0.913
White race %	83.81	85.91	80.98	0.196
Married %	54.97	53.64	56.79	0.540
Own/lease a car %	81.72	85.45	76.69	0.028
Education %				0.233
High school or less	26.70	27.27	25.93	
Some college	20.94	23.64	17.28	
College graduate	52.36	49.06	56.79	
Accelerometer-assessed measures (mean, SD)				
Average steps/minute	7.51 (3.43)	7.23 (3.29)	7.90 (3.59)	0.139
Average sedentary hours/day	8.85 (1.59)	8.92 (1.54)	8.75 (1.64)	0.348
Self-reported measures (mean, SD)				
Total activity hours/week	14.81 (20.32)	15.22 (21.97)	14.24 (17.92)	0.364
Total sitting time hours/day	6.17 (3.34)	6.52 (3.51)	5.71 (3.06)	0.022

[a] *p*-values are from two-sided z approximation Mann-Whitney U test for continuous variables and chi-square test for categorical variables, comparing profile 1 and 2.

Adjusted Comparisons of Outcomes by Profile

The two residential self-selection profiles differed in physical activity when physical activity was accelerometer-assessed but not when self-reported (see Table 4). Individuals in profile 2 (active lifestyle profile) took an average one more steps per minute than individuals in profile 1 (χ^2= 5.531, p=0.019). The residential self-selection profiles differed in sedentary behavior when sedentary behavior was self-reported but not when accelerometer-assessed. Participants in profile 1 (non-active lifestyle profile) reported sitting on average 45 minutes more on a typical day than participants in profile 2 (χ^2= 7.052, p=0.008).

Table 4. Mean values of physical activity and sedentary time. Urban residents in SHOW 2016 (n=259-361[a]).

	Profile 1		Profile 2			
	Mean	**SE**	**Mean**	**SE**	***β***	***p*-value**[b]
Accelerometer-assessed:						
Average steps/minute	9.16	1.32	10.20	1.45	1.04	0.019
Average sedentary hours/day	8.07	0.73	7.83	0.71	-0.24	0.275
Self-reported:						
Total active hours/week	13.14	0.03	14.09	0.02	0.95	0.692
Sitting time hours/day	6.03	0.89	5.28	0.88	-0.75	0.008

Note. Regression models were adjusted for age, gender, race, education, poverty income ratio, marital status, kids in household, car ownership, and month of participation.

β = regression coefficient; SE= standard error.

[a]Sample size differs based on the outcome being used.

[b]*p*-values are from a Wald test testing if the profile specific intercept are equal.

Discussion

The results support the notion that there is a group of individuals who places greater emphasis on living in an environment that is conducive to physical activity. We found two distinct profiles in our sample; individuals in the profile labeled “active lifestyle” tended to report reasons related to access to amenities, opportunities for outdoor activities, and active commute being more important than individuals in the profile labeled “non-active lifestyle.” This is in accordance with our hypothesis that stated that we would identify distinct residential self-selection profiles with varying values for importance of reasons related to a physically active lifestyle when moving to current residence.

Despite the individuals in the active profile report having placed greater importance on reasons related to active lifestyle than individuals in the non-active profile, the same reasons for

moving within the active and the non-active profile had the highest response means (i.e. “Safe, low-crime area”, “Affordable housing” and “Attractive, well-kept homes”). The reasons that can be considered being relevant to an active lifestyle (e.g. “Ease of walking or biking to services” and “Parks and open space nearby”) were on average not considered of highest importance. This can indicate that in general environmental features that support an active lifestyle are not considered important or that there is a lack of residential areas that support physical activity in this part of the U.S. The low relative importance of reasons related to an active lifestyle fits with the one previous study conducted in the U.S. (Frank et al., 2007) but not with previous studies from Europe (Christiansen et al., 2014; Van Dyck et al., 2011).

The profiles did not differ on most of the sociodemographic variables except car ownership, where greater portion of individuals in the non-active lifestyle profile owned/leased a car. The lack of differences between the profiles is surprising since commonly certain sociodemographic characteristics have been associated with specific types of neighborhoods, such as people of certain racial/ethnic background are more likely to live in inner city neighborhoods (National Research Council, 1990). Even though the profiles in our study do not directly represent certain types of neighborhoods, we would expect the individuals within each profile to live in a neighborhood with similar characteristics. However, the sociodemographic similarities between the groups are in line with the recent notion in the housing literature that sociodemographic variables are not adequate to predict residential location (Heijs, Deursen, Leussink, & Smeets, 2011; Walker & Li, 2006).

Adjusted comparison of the profiles showed that the active lifestyle profile had higher levels of physical activity and lower values of sedentary time than the non-active lifestyle

profile. However, a statistical difference was only found between the profiles when physical activity was accelerometer-assessed and when sedentary time was self-reported. Although results comparing of activity levels by profiles are not conclusive, the results show a consistent trend indicating that the active lifestyle profile is more physically active and less sedentary than the non-active lifestyle profile. This is in line with previous studies showing that individuals who emphasize reasons related to a physically active lifestyle being important when moving to their current residence tend to have higher levels of physical activity (Baar et al., 2015; Christiansen et al., 2014; Saelens et al., 2012). However, this is the first study to examine the relationship between reasons for residential location and sedentary time.

There are several limitations of the study that are worth mentioning. First, although SHOW is a population-based study, our analysis only included part of the whole SHOW sample and therefore we were not able to provide population based estimates nor generalize our results to the whole population of WI. Second, all participants were asked retrospectively about the importance of reasons for moving to their current neighborhood. There is a possibility that some of the participants have never moved and the reliability of their answers is questionable. Additionally, it is possible that the retrospective reporting of the importance of the reasons for moving are biased by the participant experiences of living in their current residential environment. Finally, the distribution of the objectively measured and self-reported measures of physical activity were not completely normally distributed, even after the self-reported variable was log transformed. However, after careful analysis of the distributions they were deemed close enough to normal distribution, nevertheless, the exact values of the regression coefficient should be interpreted with caution.

The key strength of our study is the novel approach to use LPA to categorize individuals based on their reasons for moving to their current residence and then examine the association between residential self-selection profiles and physical activity. A few studies have used a similar approach to study the effect of the built environment on physical activity with the goal of capturing the complexity of the built environment better than examining one characteristic of the environment at a time (Adams et al., 2011; McDonald et al., 2012; Nelson, Gordon-Larsen, Song, & Popkin, 2006; Norman et al., 2010). Another strength of our study is the use of accelerometer-assessed and self-reported measures of physical activity and sedentary time. Even though the use of accelerometers has been considered the gold standard for measuring physical activity and sedentary time in population studies for some time, the majority of the literature on the association between the built environment and physical activity is based on self-reported measures of physical activity (Durand, Andalib, Dunton, Wolch, & Pentz, 2011; McCormack et al., 2004). Finally, the study uses questions that ask about specific reasons for moving to current neighborhood, not general preferences for a neighborhood like many previous studies on residential self-selection have done (Bagley & Mokhtarian, 2002b; Kitamura, Mokhtarian, & Laidet, 1997; Schwanen & Mokhtarian, 2005b).

Together the results of our study suggest that there is a group of individuals who places greater importance on living in an environment that supports a physically active lifestyle. The group has similar sociodemographic characteristics as the other group who does not emphasize reasons related to a physically active lifestyle to the same extent. Finally, the group that emphasizes reasons related to a physically active lifestyle tends to be more active than the other group. These results provide some evidence that residential self-selection based on predisposition

to an active lifestyle is a real phenomenon. Further, the results indicate that adjusting for sociodemographic variables when exploring the association between the built environment and physical activity might not be enough to control for self-selection into residential areas. It is important to figure out if there is a causal relationship between the built environment and physical activity or not. We live in a world with limited resources (e.g., dollars, time) and therefore it is very important that we use the available resources wisely and focus on proven interventions to improve the health of the population (Kindig, 2007). If the commonly reported relationship between the built environment and physical activity is spurious, focusing our resources on public health campaigns that aim at targeting individual's attitudes and preferences might be preferable.

AIM 2: Disentangling the Relationship between Residential Self-Selection Related to a Physically Active Lifestyle, Walkability and Block Group Income

Abstract

Purpose: Evidence on how residential self-selection influences associations between the built environment and physical activity is lacking. We examined the association between residential self-selection related to a physically active lifestyle (RSS-PA), residential walkability, and potential interactions by block group economic status.

Methods: Data from adult urban residents in the 2016 cohort of Survey of the Health of Wisconsin (SHOW) were used. A RSS-PA scale (range: 1-4) was created from six self-reported reasons related to a physically active lifestyle for moving to current residence. Objective (Walk Score®, range: 0-100) and self-reported measures (range: 1-6) of walkability and a 5-year estimate for average income per block group were used. Multivariate linear regression was performed to explore the association between residential self-selection, walkability, and block group income.

Results: We observed a relationship between the RSS-PA scale and both self-reported (β=-0.256, p<0.0001) and objective measures (β=0.008, p=0.005) of walkability. The association between RSS and walkability was moderated by block group income when walkability was objectively measured, but not when self-reported. A significant association was found between objectively measured walkability and RSS in high-income block groups (β=0.012, p<0.0001) but not in low-income block groups (β<-0.001, p=0.988).

Conclusions: This study indicates that self-selection into residential environments based on favoring an environment that supports an active lifestyle exists, but that the magnitude of the self-selection differs depending on the economic status of the residential area.

Introduction

Walkability of a built environment can be defined as the extent to which characteristics of the built environment may be conducive to individuals for various forms of physical activity, such as walking/biking to and from places, for recreation or to exercise (Carr et al., 2010a; Duncan et al., 2011; Leslie et al., 2007b). Current studies have documented an association between walkability and physical activity, leading many to believe that walkability of neighborhoods has a causal impact on the physical activity levels of the individuals living in the neighborhood (Hirsch et al., 2014, 2013; Saelens et al., 2003b). However, these studies face a number of methodological challenges, including limited adjustment for residential self-selection (Oakes, 2004). Residential self-selection related to a physically active lifestyle (RSS-PA) is the notion that people who live in a walkable area decided to live there because they knew the environment would support their physically active lifestyle. If that is the case, it is not the walkable environment that is causing the residents to be physically active. Rather, it is the residents' predisposition to physical activity that is generating the positive association between walkability and physical activity.

A limited amount of research has examined this hypothesized mechanism of self-selection into areas based on predisposition to a physically active lifestyle. These studies show mixed results depending on how residential self-selection is measured. Handy, Cao, &

Mokhtarian (2006) asked participants living in eight neighborhoods in Northern California about their preferences for neighborhood characteristics. Handy and colleagues found no association between preferences for having physical activity options in the neighborhood or accessibility and the walkability of the residential location. Further, Schwanen & Mokhtarian (2004) found a significant mismatch between individual environmental preferences and the environment of their current neighborhood. They found that one of every four participants lived in a neighborhood that did not reflect their environmental preferences. However, Baar, Romppel, Igel, Brähler, & Grande (2015) asked German college students living in the city of Leipzig about the reasons they moved to their current neighborhood and found that utilitarian reasons (closeness to facilities, affordability etc.) were associated with self-reported walkability of their neighborhood. Further studies are needed on this possible pathway of residential self-selection. Especially studies that ask about reasons for moving to current residence since it appears that dissonance between environmental preferences and residential location is common (Schwanen & Mokhtarian, 2004). Asking about reasons for moving to current neighborhood as opposed to environmental preferences may be important when attempting to account for residential self-selection in the relationship between the built environment and physical activity, to minimize residual confounding.

If self-selection into residential environments based on predisposition to an active lifestyle exists, it may vary based on the economic status of the neighborhood. In the U.S., there is a great deal of segregation by economic status (Watson, 2009) and higher income groups may be able to satisfy more personal criteria when selecting a residential location (Levine & Frank, 2007). In addition, environmental aesthetics (e.g. trees along sidewalks, clean streets etc.) that

have been found to be associated with increased physical activity often vary greatly between high- and low-income neighborhoods (Neckerman et al., 2009). Studies on the implication of residential self-selection in different economic environments are lacking.

The main goal of this study was to address current gaps in the literature and gain further understanding of the role that residential self-selection plays in studies on residential built environment and physical activity. First, characteristics of individuals living in areas with different walkability were examined. Next, the association between a residential self-selection scale (RSS-PA; operationalized as the importance of six reasons related to a physical active lifestyle for moving to current residence), subjective and objective measures of walkability, and block group economic status were examined. It was hypothesized that individuals living in high walkable areas would report greater importance of reasons specifically related to a physically active lifestyle (i.e. score higher on the RSS-PA scale) than individuals living in low walkable areas. Further, we hypothesized that the relationship between the RSS-PA scale and walkability would be moderated by block group economic status.

Material and Methods

Study Sample

Data for this study came from the 2016 cohort of the Survey of the Health of Wisconsin (SHOW) Details on SHOW protocols have been published previously (Nieto et al., 2010) and are available at www.show.wisc.edu. In brief, SHOW is a population-based health examination survey of civilian, non-institutionalized residents of Wisconsin of all ages and consists of a household interview, self-administered questionnaire (SAQ), physical exam, and accelerometry

and biosample collection. Participants are recruited from a random sample of households using a three-stage stratified cluster sampling approach. The three stages involve 1) counties, 2) Census 2010 block groups, and 3) households. While SHOW collects data on urban and rural populations, the focus of this study is on SHOW participants 18 years old and older living in an urban area (N=478). Due to missing data mainly on residential self-selection items collected by self-administered questionnaire and income, 357 participants were included in the final analyses. All SHOW study protocols were approved by the University of Wisconsin Health Sciences Institutional Review Board and all participants provided written informed consent as part of the survey.

Measures

Residential self-selection related to a physically active lifestyle (RSS-PA) was operationalized using six questions about reasons for moving to current residence. The questions were self-administered and asked about participant reasons for moving to their current residence: "*Thinking back when you moved to your current residence, at that time, how important were each of the following factors in your decision to move to your current residence?*" Participants were presented with 19 different items/neighborhood characteristics and indicated their answer on a 4-point scale ranging from "*Not at all important*" to "*Very important.*" Six items reflecting factors previously found in the literature to be associated with residential built environment and physical activity were included in the study. The six items were "*Outdoor recreational opportunities*", "*Parks and open spaces nearby*", "*Easy access to public transit*", "*Ease of walking or biking to services (such as work, school, grocery stores, restaurants, etc.)*", "*Easy*

access to a gym or other work-out facility (such as yoga studio, running track, pool, etc.)" and "*Opportunities for outdoor activities during free-time (such as walking, rollerblading, biking, playing outdoor games, etc.)*". A residential self-selection scale ranging from 1-4 was created from the six questions. A higher score on the scale indicates greater importance of living in an environment that supports a physically active lifestyle.

Walkability. One objective and one subjective measure were used to assess the walkability of the residential environment.

The *objective measure* that was used is the Walk Score®. Walk Score was designed by the software development company Front Seat Management and measures the walkability of the built environment using a number of data sources such as Open Street Map, Google and U.S. Census ("Walk Score Methodology," 2016). Walk Score is a score between 0-100 (higher score indicates greater walkability) and is calculated from distance to amenities accessible from the participant's home based on street network/connectivity. For every Walk Score, hundreds of walking routes are generated using proprietary algorithms and various data streams to find nearby amenities. Amenities get points based on the distance from participant home address. Amenities that are highly correlated with walking get more weight and Walk Score also takes into account the number of amenities in the same category. Furthermore, Walk Score takes into account two physical features of the street network that have been found to be associated with walking, i.e. intersection density and block length (Berrigan et al., 2010). Front Seat Management has categorized Walk Score into five categories: 0-24 very car dependent (almost all errands require a car), 25-49 car-dependent (most errands require a car), 50-69 somewhat walkable (some errands can be accomplished on foot), 70-89 very walkable (most errands can be

accomplished on foot), 90-100 Walker's paradise (daily errands do not require a car). When Walk Score is used as a categorical variable in this study the first two categories will be combined into one car dependent category (0-49) and the last two into one walkable category (70-100).

Walk Score, has been shown to be associated with other objective measures of walkability, such as access to walkable amenities and intersection density (Carr et al., 2010a; Duncan et al., 2011). In August 2016, Walk Score was obtained for each participant through an application-programming interface offered by Front Seat Management using geographic coordinates for the participant's home address. To mask identification of SHOW participants when obtaining the Walk Score, ten additional random addresses were fed into the application program for every SHOW participant address.

The *self-reported measure* walkability was derived from a self-administered 18 item questionnaire that asks participants to indicate how many minutes it would take them to walk from their home to the nearest of a particular facility. Participants indicate their answer on a six-point scale ranging from "*0-5 minutes*", "*6-10 minutes*", "*11-20 minutes*", "*21-30 minutes*", "*More than 30 minutes*" and "*None within walking distance*". The questionnaire includes facilities such as supermarket, trail for walking or biking, post office, school, and restaurant. A self-reported walkability scale was created by averaging each participant score on the 18 questions. Scores on the scale range from 1-6, with higher score indicating less walkability. An average score of 2.5 or less was considered having high walkability since neighborhoods are commonly considered walkable when services that individuals need to be able to access on a daily bases are within 10 minutes of walking (Talen & Koschinsky, 2013).

Sociodemographic variables. The following self-reported sociodemographic variables were included in the study: gender (male, female), age (years), race/ethnicity (white, non-white), marital status (married, not married), education (high school or less, some college, college graduate), own/lease a car (yes, no), kids living in household (yes, no), duration lived at address (0-1 year, 1-3 years, 3-10 years, >10 years), poverty income ratio (PIR= ratio of midpoint of household income range compared to federal annual poverty levels in the United States). Higher PIR is reflective of greater household income. The poverty guidelines come from the Department of Health and Human Services (U.S. Department of Health and Human Services., 2017).

Geocoded addresses were used to categorize participants into urban or rural areas based on the 2010 U.S. Census Bureau's urban-rural classification system and to obtain information about the neighborhood economic status. To represent the neighborhood economic status, a 5-year estimate from the 2016 American Community Survey for average yearly per capita income per block group was used. Individuals were categorized into "high-income" and "low-income" block groups based on the ACS data. Individuals living in a block group with average yearly block group income equal to or less than $25,951 were placed in the "low-income" category. Individuals living in a block group with average yearly block group income greater than $25,951 were placed in the "high-income" category. The cut-off dollar amount used was the mean average yearly block group income for the study sample, that dollar-amount was similar to the average yearly block group income ($25,881) of the state where most of the participants lived in.

Statistical Analyses

All statistical analyses were conducted in SAS (Version 9.4). Unadjusted bivariate comparisons of individuals living in walkable areas versus non-walkable areas were conducted using Chi-square tests (for categorical variables) and independent t-test (for continuous variables).

Two confirmatory factor analyses (CFA) were conducted using PROC CALIS to justify creation of RSS-PA scale and self-reported walkability scale. One CFA was conducted to test if the six items from the residential self-selection questionnaire identified as being relevant to a physically active lifestyle loaded on one factor. Another CFA was conducted to test if the eighteen questions from the self-reported walkability questionnaire loaded on one factor. For both of the CFAs, a polychoric correlation matrix with unweighted least-squares (ULS) factor extraction was used since the variables were on an ordinal scale (Forero et al., 2009; Han et al., 2011). Model fit was assessed using multiple fit indices that are considered appropriate to use with USL extraction ("The CALIS Procedure," 2011) including 1) Standard root-mean-square residual (SRMR); 2) Goodness of fit index (GFI); 3) Adjusted GFI (AGFI); and, 4) Normed fit index (NFI). For SRMR a value less than 0.08 indicates a good model fit, and a value of 0.95 or greater for GFI, AGFI and NFI suggest good model fit (Hu & Bentler, 1999). Items had to have a minimum factor loading of 0.3 to be considered as part of the factor, but higher factor loadings were preferred (Kline, 2015).

To evaluate the association between the walkability and residential self-selection, and neighborhood economic status, we used linear regression (PROC SURVEYREG with DOMAIN statement) models with residential self-selection as the outcome and Walk Score and self-reported walkability scale as independent variables (in separate models). An interaction term between each of the walkability variables and the binominal high/low block group income

variable was used to test if the association between walkability and residential self-selection was moderated by census block group economic status. A stratified analysis by average block group income was performed for significant interaction terms. To account for SHOW's multistage sampling method, CLUSTER and STRATA statements were specified. Covariates that were considered for inclusion in the regression models were: age, gender, race, poverty income ratio, education, car ownership, kids living in household, marital status, duration lived at address. All analyses were restricted to individuals with complete data.

Results

Confirmatory Factor Analyses (CFA)

CFA of the six self-selection variables loading on one factor indicated a good fit. Factor loadings ranged from 0.563 to 0.906 (mean=0.759). Fit indices were within the suggested range: GFI=0.988, AGFI=0.972, NFI=0.981 (>0.95) and SRMR= 0.067 (<0.08). The Cronbach Alpha for the RSS-PA scale was 0.85.

CFA of the eighteen self-reported walkability questions loading on one factor also indicated a reasonable fit. Factor loadings ranged from 0.390 to 0.819 (mean=0.655). Fit indices were within the suggested range: GFI=0.973, AGFI=0.966, NFI=0.966 (>0.95) and SRMR= 0.078 (<0.08). The Cronbach Alpha for the self-reported walkability scale was 0.91.

Sample Characteristics and Characteristics by Walkability

Sample characteristics are displayed in Table 1. Overall, individuals in this sample were white (84.31%), well educated (73.88% some college or college graduate), owned or leased a car

(82.35%), did not have children living in the household (65.83%), and had lived at their current address for at least one year (84.72%).

When comparing individuals living in a walkable environment according to Walk Score (38.14%) to individuals living in car dependent environment (see Table 1), individuals living in walkable environment were worse off financially, the block group they lived in had on average lower income, were more likely to be non-white, not married, and not own/lease a car. Further, individuals living in a walkable environment, according to Walk Score, scored higher on the RSS-PA scale and all six of the individual variables in the scale except "*Outdoor recreational opportunities*".

Comparing individuals living in a high walkable area according to self-report (18.57%) to individuals living in a low walkable area (see Table 1), individuals living in a high walkable areas were younger, and more likely to be male. Individuals living in a high walkable areas also scored higher on the RSS-PA scale and all six variables in the scale except "*Easy access to public transit*".

Table 1. Sample characteristics by Walk Score® and self-reported walkability. Urban residents in SHOW 2016 (N=388).

	Total sample	**Walk Score®**			**Self-reported walkability**		
		Car dependent[a] N=240	**Walkable** N=148	*p*-value[c]	**Low**[b] N=285	**High** N=65	*p*-value[c]
Age in years (mean, SD)	50.1 (17.54)	51.34 (18.12)	48.31 (16.44)	0.099	51.86 (17.53)	43.09 (14.26)	<.0001
Poverty income ratio (mean, SD)	3.33 (2.46)	3.62 (2.56)	2.89 (2.23)	0.005	3.36 (2.46)	3.55 (2.52)	0.571
Average BG income in $1000 (mean, SD)	25.95 (10.75)	28.00 (8.19)	22.63 (13.31)	<0.001	25.69 (10.03)	27.08 (12.97)	0.420
Living in high-income BG %	47.94	59.58	29.05	<.0001	46.67	50.77	0.550
Children live in household %	34.79	33.64	34.46	0.914	35.44	38.46	0.647
Male gender %	42.01	43.18	41.89	0.970	38.95	55.38	0.015
White race/ethnicity %	83.76	94.55	66.89	<.0001	83.86	87.69	0.440
Married %	54.78	62.27	42.18	<.0001	56.34	50.77	0.415
Own/lease a car %	81.44	89.09	69.59	<.0001	84.56	80.00	0.369
Education %				0.952			0.596
High school or less	26.61	26.36	27.21		54.65	27.69	
Some college	20.67	20.00	21.09		22.54	16.92	
College graduate	52.71	53.64	51.70		52.82	55.38	
Duration lived at				0.123			0.132
0-1 years	15.28	12.8	20.27		13.73	16.92	
1-3 years	20.98	20.59	21.62		20.42	26.15	
3-10 years	23.06	23.11	22.97		22.18	29.23	
>10 years	40.67	44.12	35.14		43.66	27.69	

(Continued)

Table 1 (continued)

	Total sample	Walk Score®			Self-reported walkability		
		Car dependent[a] N=240	Walkable N=148	*p*-value[c]	Low[b] N=285	High N=65	*p*-value[c]
RSS-PA scale (mean, SD)	2.15 (0.77)	1.99 (0.69)	2.41 (0.84)	<.0001	2.04 (0.71)	2.45 (0.83)	<.0001
Items from RSS-PA scale: (mean, SD):							
Parks and open space nearby	2.54 (0.99)	2.45 (0.93)	2.69 (1.07)	0.021	2.47 (0.97)	2.80 (0.95)	0.015
Outdoor recreational opportunities	2.25 (1.03)	2.21 (0.98)	2.32 (1.12)	0.307	2.13 (0.98)	2.61 (1.09)	0.001
Easy access to public transit	1.89 (1.10)	1.58 (0.92)	2.41 (1.17)	<.0001	1.81 (1.04)	1.98 (1.12)	0.231
Ease of walking or biking to services	2.14 (1.07)	1.88 (0.93)	2.57 (1.15)	<.0001	1.98 (1.01)	2.68 (1.13)	<.0001
Opportunities for outdoor activities…	2.32 (0.98)	2.21 (0.91)	2.51 (1.07)	0.006	2.23 (0.93)	2.63 (1.08)	0.003
Easy access to a gym or other work-out…	1.75 (0.96)	1.62 (0.86)	1.96 (1.09)	0.002	1.64 (0.89)	2.06 (1.05)	0.001

RSS-PA= Residential self-selection related to a physically active lifestyle; BG= Census block group.
[a]Car dependent= Walk Score< 50; Walkable= Walk Score > 50.
[b]High walkability= self-reported walkability scale ≤ 2.5; Low walkability= self-reported walkability scale > 2.5.
[c]*p*-value is based on t-test for continuous variables and chi-square test for categorical variables.

Regression Analyses – Residential Self-Selection and Walkability

Table 2 shows the results for the linear regression with Walk Score as the main independent variable of interest and the RSS-PA scale as the outcome. In the adjusted model, individuals living in an environment with higher Walk Score scored on average higher on the RSS-PA scale than individuals who live an environment with lower Walk Score (β=0.008, p=0.0008), independent of sociodemographic factors. Additionally, individuals who owned/leased a car scored on average lower on the RSS-PA scale than individuals who did not own/lease a car (β=-0.392, p=0.0003).

Further, the association between Walk Score and the RSS-PA scale was found to be moderated by block group income (β= -0.011, p=0.033). A stratified analysis by block group income showed a positive significant association between Walk Score and the RSS-PA scale in high-income block groups (β= 0.012, p<0.0001) but no significant association was found in the low-income stratum (β= -0.00004, p=0.988).

Table 2. Linear regression models of the association between Walk Score® and the RSS-PA scale. Urban residents in SHOW 2016 (N=356).

	Unadjusted β (SE)	**Adjusted** β (SE)	**Adjusted w/interaction** β (SE)
Walk Score®	0.009 (0.002)***	0.008 (0.002)**	0.012 (0.002)***
Age (years)		0.005 (0.003)	0.007 (0.003)*
Poverty income ratio		0.011 (0.018)	0.006 (0.016)
Marital status			
Not married		-0.125 (0.102)	-0.112 (0.113)
Married		*Ref*	*Ref*
Race/ethnicity			
White		-0.038 (0.128)	-0.182 (0.156)
Non-white		*Ref*	*Ref*
Own/lease a car			
Yes		-0.389 (0.090)**	-0.478
No		*Ref*	*Ref*
Length of residency		-0.095 (0.061)	-0.091 (0.060)
Walk Score*BG income			
Low-income			-0.011 (0.005)*
High-income			*Ref*

Note. RSS-PA= Residential self-selection related to a physically active lifestyle, SE=standard error, BG= block group.
*p<0.05, **p<0.001, ***p<0.0001

Table 3 shows the results for the linear regression with self-reported walkability as the main independent variable of interest and the RSS-PA scale as the outcome. On average, individuals reporting their residential area not being walkable scored lower on the RSS-PA scale (β=-0.257, p<0.0001), independent of sociodemographic factors. Further, older individuals tended to score higher on the RSS-PA scale (β=0.008, p=0.0349), individuals who had lived longer at their current residence tended to score lower on the RSS-PA scale (β=-0.107, p=0.0402) as well as individuals owning/leasing a car (β=-0.398, p=0.0009). The association

between self-reported walkability and the RSS-PA scale was not moderated by block group income (β=0.084, p=0.393).

Table 3. Linear regression models of the association between self-reported walkability scale and the RSS-PA scale. Urban residents in SHOW 2016 (N=323).

	Unadjusted β (SE)	**Adjusted** β (SE)	**Adjusted w/interaction** β (SE)
Walkability scale	-0.216 (0.041)**	-0.256 (0.040)***	-0.296 (0.061)**
Age (years)		0.008 (0.004)*	0.009 (0.004)*
Poverty income ratio		0.0002 (0.019)	-0.005 (0.017)
Marital status			
Not married		-0.051 (0.092)	-0.023 (0.101)
Married		*Ref*	*Ref*
Race/ethnicity			
White		-0.132 (0.093)	-0.180 (0.105)
Non-white		*Ref*	*Ref*
Own/lease a car			
Yes		-0.398 (0.095)**	-0.425 (0.092)**
No		*Ref*	*Ref*
Length of residency		-0.107 (0.048)*	-0.101 (0.046)*
Walk Score*BG income			
Low-income			0.084 (0.096)
High-income			*Ref*

Note. RSS-PA= Residential self-selection related to a physically active lifestyle, SE=standard error, BG= block group.
*p<0.05, **p<0.001, ***p<0.0001

Discussion

To our knowledge, this is the first study in the United States to examine the relationship between reasons (specifically related to physical active lifestyle) for moving to current residence and characteristics of the respondent's current residential area. Further, this is among the first studies to examine whether the proposed relationship between reasons for moving to current residence and walkability differ by block group economic status. Using data from adult urban residents in

SHOW, we observed that individuals who lived in a walkable environment (objectively measured and self-reported) reported reasons related to a physically active lifestyle more important (i.e. scored higher on the RSS-PA scale) than individuals who lived in a less walkable environment when adjusted for a number of sociodemographic factors. When walkability was objectively measured (Walk Score), not when self-reported, the association between walkability and residential self-selection did differ depending on the income of the block group. No association was found between reasons for moving to current residence and Walk Score among individuals living in low-income block groups while strong association was observed among individuals living in high-income block groups.

On average, individuals living in a walkable environment according to Walk Score as opposed to car dependent were worse off financially, were more likely to live in a low-income block groups, and were more likely to be non-white. According to these results, lack of walkability is not something that is associated with low-income block groups or areas where racial minorities live. However, these results need to be interpreted with caution since Walk Score has found to be positively associated with crime (Carr, Dunsinger & Marcus, 2010b) and high crime rates are commonly found in low-income neighborhoods with high percent of racial minorities (Neckerman et al., 2009). If individuals are not safe while being outside in their neighborhood, it is unlikely that they will walk to the available amenities in their neighborhood even though the amenities are within walking distance.

The contradicting results found between the self-reported and objective measures of walkability when testing for interaction between high/low block group income and walkability may not be surprising. Previous studies have commonly found low agreement between

subjective and objective measures of environmental features used to operationalize walkability (Bailey et al., 2014; Ball et al., 2008; Gebel, Bauman, & Owen, 2009; Lackey & Kaczynski, 2009; McCormack, Cerin, Leslie, Toit, & Owen, 2008). These studies have indicated that self-reported measures of walkability capture the pleasantness of walking to a greater extent than the Walk Score and other objective measures (Tuckel & Milczarski, 2015). It is possible that the observed interaction between Walk Score and block group income is highlighting the limitation of the Walk Score as a measure of walkability. Additionally, Walk Score has been criticized for failing to capture the aesthetics of the environment, i.e. if it is aesthetically pleasing and enjoyable to walk in, if the streets are clean, if trees line the sidewalks etc. Safety and aesthetics are more commonly lacking in low-income neighborhoods than in high-income neighborhoods (Neckerman et al., 2009). Therefore, the non-significant association found between Walk Score and residential self-selection among low-income block groups might stem from that low-income block groups with high Walk Score are not truly walkable. Individuals living in these block groups do not perceive their environment as walkable and thus did not report moving to the area because of it being supportive of a physically active lifestyle. In support of this, Koschinsky et al (2017) found that Walk Score overestimated the walkability of low-income neighborhoods with high access to amenities.

Apart from walkability, the RSS-PA scale was found to be associated with age, owning/leasing a car, and time lived at current residence. In our study, older participants tended to score slightly higher on the RSS-PA scale than younger participants. In a Belgian study by Van Dyck, Cardon, Deforche, Owen, & De Bourdeaudhuij (2011) the same pattern was observed, i.e. older participants reported reasons related to a walkable environment being more

important than younger participants. In that Belgian study, women and individuals with less education were found to place greater importance on walkability-related reasons for moving to current neighborhood. In our study, however, we did not find any association between gender or education and the RSS-PA scale. Gender and education were not included in the final models shown in results. In our study, individuals who owned/leased a car scored lower on the RSS-PA scale as well as individuals who had lived at their current residence for a longer time. No previous studies have examined the relationship of car ownership and length of residency with residential self-selection related to a physically active lifestyle and therefore these associations cannot be compared to other results.

There are a number of limitations that need to be kept in mind when interpreting the results of this study; the study utilized cross-sectional data and therefore individuals were asked about their reasons for moving to their current residence after they had lived there. Living in the area might have colored or changed their thinking about why they decided to move there. Furthermore, some individuals might never have moved, they might have lived in the same residential area for their whole life. If that is the case, the individuals may have made up their answers based on the residential area they live in. Both of these issues might lead to inflation in the estimated association between residential self-selection and Walk Score. Finally, previously mentioned flaws with the Walk Score, i.e. not taking into account the safety of the environment. On the other hand, the study has number of noteworthy strengths such as the use of objective measures of the environment, specific questions about reasons for moving to current residence as opposed to general environmental preferences for neighborhood, and use of a random sample of individuals of varies ages, possible making the results more generalizable.

Conclusions

This study provides further understanding of the mechanism of residential self-selection into environments that support a physically active lifestyle. The results support the notion that the walkability of the environment that individuals live in reflect their preference for an environment that supports an active lifestyle (i.e. residential self-selection exists). However, the magnitude of the residential self-selection differs depending on the economic status of the block group. Therefore, it is important to keep in mind the whole process of self-selection into residential areas when exploring the association between the built environment and physical activity. Not everyone has the luxury of being able to choose where to live. Because of residential segregation, individuals who live in poor neighborhoods are often forced to stay in poor neighborhoods and do not have the privilege to decide a residential location based on their preferences for being physically active. If residential self-selection exists in previous studies on the association between walkability and physical activity, the studies might be overestimating the relationship between walkability and physical activity. Especially if the studies took place in wealthy residential areas.

AIM 3: Residential Self-Selection Related to a Physically Active Lifestyle, Walkability and Physical Activity: A Structural Equation Model

Abstract

Introduction: Evidence on how residential self-selection influences the association between the built environment and physical activity is lacking. The present study explored the association between walkability and physical activity while taking residential self-selection related to a physically active lifestyle (RSS-PA), block group economic status, and safety from crime into account.

Methods: Data came from 478 urban residents in the Survey of the Health of Wisconsin (SHOW 2016). A RSS-PA scale (range: 1-4) was created from five questions about reasons for moving to the current residence. Higher score indicates greater importance of reasons related to a physically active lifestyle. Walk Score® (range: 0-100) was used to estimate walkability and accelerometer-assessed number of steps was used to estimate physical activity. Participants were categorized into high- and low-income block groups based on average yearly block group income. The relationships of interest were tested in a structural equation model.

Results: A positive association was found between the RSS-PA scale and number of steps (β=.25, p<.0001) and Walk Score and the RSS-PA scale but only among individuals living in high-income block groups (β=.50, p<.0001). No association was found between Walk Score and number of steps (β=.06, p=.481). An indirect effect of safety from crime on number of steps via the RSS-PA scale was observed (β= .04, p=0.003) but no direct effect.

Conclusions: Findings suggest that residential self-selection based on predisposition to an active lifestyle may explain associations between walkability and physical activity, but the extent of it may vary by economic status.

Introduction

Chronic diseases are a major cause of morbidity and mortality nationwide and it is estimated that costs related to caring for people living with a chronic disease account for 86% of all health care spending (Gerteis et al., 2014). Regular physical activity has been shown to play an important role in decreasing the risk of a number of chronic diseases (Committee, 2008). Despite that, only about 5% of adults living in the United States (U.S.) are getting the recommended daily amount of physical activity (Troiano, Berrigan, Dodd, MâSse, et al., 2008). In recent years, there has been growing interest in improving the walkability of the environment as a means of increasing physical activity (Jackson et al., 2013; U.S. Department of Health and Human Services, 2015).

Walkability is a complex construct, which has been defined broadly as the ability of the built environment to support a physically active lifestyle (Carr et al., 2010b). The conceptualization of walkable environment varies greatly between studies and therefore it has been challenging to generalize and validate results from studies on walkability and physical activity (Hall & Ram, 2018; McCormack & Shiell, 2011). As one attempt to improve comparability of studies, researchers have increasingly been using Walk Score® to estimate walkability. Walk Score is a publicly available measure based on distance to amenities and street connectivity, taking block length and intersection density into account (“Walk Score Methodology,” 2016). Walk Score has been found to correlate with a number of environmental

characteristics commonly used to operationalize walkability, such as access to amenities, intersection density, population density, and access to public transit (Carr et al., 2010b, 2010a; Duncan et al., 2011). However, a major limitation of Walk Score is that is has been found to be positively related to crime rates (Carr et al., 2010b). Lack of safety from crime has been identified as a barrier to physical activity (Foster & Giles-Corti, 2008; Handy et al., 2006; McCormack et al., 2004) and therefore it is important to include additional measures of safety when using Walk Score as a measure of walkability.

A number of studies have shown a positive association between Walk Score and physical activity (S. C. Brown et al., 2013; Hirsch et al., 2013; Manaugh & El-Geneidy, 2011; Tuckel & Milczarski, 2015). However, research on the association between walkability and physical activity often do not take the possibility of confounding by residential self-selection into account, and to our knowledge, no study examining the association between Walk Score and physical activity has accounted for residential self-selection. Residential self-selection being that people who are motivated to be physically active are more likely to move to walkable neighborhoods and show higher levels of physical activity. This self-selection may lead to a significant correlation being found between walkability and physical activity levels despite there not being a causal relationship (Oakes, 2014).

A number of cross-sectional studies examining the association between the built environment (using other measures than Walk Score) and physical activity have attempted to account for residential self-selection. These studies have commonly used questions about participant environmental preferences (e.g. preferences for having stores close by or opportunities for physical activity) or reasons for moving to current residence as a measure of

residential self-selection. In most of these studies, the association between the built environment and physical activity has stayed even after adjusting for environmental preferences or reasons for moving to current residence (Baar et al., 2015; Christiansen et al., 2014; Handy et al., 2006; McCormack & Shiell, 2011). However, these previous studies may not have been fully successful in adjusting for the potential confounding of residential self-selection since majority of them used questions about environmental preferences or limited set of questions about reasons for moving to current residence. Asking about preferences instead of reasons for moving can cause problems since often individuals do not live in an environment that is reflective of their preferred environmental characteristics (Schwanen & Mokhtarian, 2004). If the questions used for adjustment (i.e. environmental preferences) are an imperfect marker of the confounding variable (i.e. residential self-selection) that you want to adjust for, a residual confounding can occur, i.e. confounding still exist (Szklo & Nieto, 2004). Using limited set of questions that only cover a few reasons that draws an active person into an environment that supports physical activity can cause the same problem. Further studies that use more specific questions when adjusting for residential self-selection in the relationship between the built environment and physical activity are therefore needed.

To be identified as a confounder, a variable needs to be associated with the independent variable of interest (i.e. walkability of the built environment) and influence the dependent variable of interest (i.e. levels of physical activity) (Szklo & Nieto, 2004). Even though residential self-selection has been identified as a potential confounder of the association between the built environment and physical activity (Frank et al., 2007; Handy et al., 2006) few studies have actually examined the relationship between residential self-selection and the built

environment or physical activity. The few studies that have done so show mixed results depending on how residential self-selection is measured. Baar and colleagues asked German college students living in the city of Leipzig about the reasons they moved to their current neighborhood and found that reasons for choosing a neighborhood were related to self-reported characteristics of their neighborhood and self-reported physical activity levels (Baar et al., 2015). Christiansen et al. (2014) found that Danish adults living in Aarhus who reported reasons related to destination accessibility being important when they decided to move to their current neighborhood, reported more frequently walking for transport than those who did not report these reasons being important. Handy and colleagues asked adults living in eight neighborhoods in Northern California about their preferences for neighborhood characteristics instead of reasons for moving to current neighborhood. They found no association between neither preferences for physical activity options or accessibility with environmental characteristics of residential location. However, individuals who expressed preference for accessibility and options for physical activity in the neighborhood walked more frequently to the store (Handy et al., 2006). Additional studies are needed to examine these possible pathways of residential self-selection.

If self-selection into residential environments based on predisposition to an active lifestyle exists, it may vary based on the economic status of the neighborhood. In the U.S., there is a great deal of segregation by economic status (Watson, 2009) and higher income groups may be able to satisfy more personal criteria when selecting a residential location (Levine & Frank, 2007). In addition, environmental aesthetics (e.g. trees along sidewalks, clean streets etc.) that have been found to be associated with increased physical activity often vary greatly between high- and low-income neighborhoods (Neckerman et al., 2009). Sallis and colleagues found

some evidence of residential self-selection varying by neighborhood income among individuals living in 32 neighborhoods in Seattle, WA and Baltimore, MD. In their study, they found that the association between walkability and walking for transport was weaker among adults living in low-income block group than high-income bloc groups (Sallis, Saelens, et al., 2009). Apart from this study, studies on residential self-selection and economic status are scarce.

To sum up, studies have found a positive relationship between the built environment and physical activity even after accounting for residential self-selection (Baar et al., 2015; Christiansen et al., 2014; Handy et al., 2006; McCormack & Shiell, 2011). However, these studies might have used insufficient methods to account for residential self-selection and additional studies are needed with more careful adjustment of residential self-selection. Further studies examining the possible pathways of residential self-selection and how it might differ by economic status are also needed to gain further understanding of residential self-selection and how it may influence the relationship between the built environment and physical activity.

To address these gaps in the literature, the present study tested a conceptual model (shown in Figure 1) that explored the association between walkability and physical activity while taking residential self-selection related to a physically active lifestyle (RSS-PA), block group economic status, and safety from crime into account. Researchers have pointed out, that the literature could benefit from using more complicated conceptual and statistical models that take into account possible moderators and mediators as well as correlates of physical activity to try to understand possible pathways (Ball et al., 2006; Bauman et al., 2002; Ding & Gebel, 2012; McCormack & Shiell, 2011). We hypothesized that: 1) Walkability would be positively associated with physical activity independent of RSS-PA, 2) RSS-PA would be positively related

to physical activity, 3) Walkability would be positively related to RSS-PA but the association would be moderated by block group economic status, 4) Perceived safety from crime would be positively associated with physical activity.

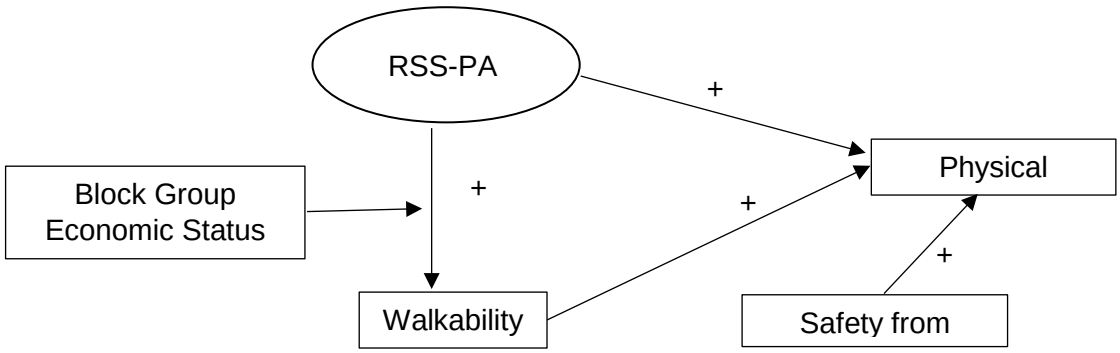


Figure 1. Conceptual model of associations between residential self-selection related to a physically active lifestyle (RSS-PA), walkability, and physical activity.

Methods

Study Sample

Data for this study came from the 2016 cohort of the Survey of the Health of Wisconsin (SHOW). Details on SHOW protocols have been published previously (Nieto et al., 2010) and are available at www.show.wisc.edu. In brief, SHOW is a population-based health examination survey of civilian, non-institutionalized residents of Wisconsin of all ages. SHOW consists of a household interview by trained field staff, self-administered questionnaire, physical exam, objective physical activity assessment using accelerometers and biological sample collection. Participants are recruited from a random sample of households using a three-stage stratified cluster sampling approach. The three stages involve 1) counties, 2) Census 2010 block groups, and 3) households. While SHOW collects data on urban and rural populations, the focus of this

study is on participants 18 years of age and older living in urban areas (N=478). Due to participant drop off in the self-administered questionnaire and accelerometer data collection, the number of participants in the analyses range from 258-383. All SHOW study protocols were approved by the University of Wisconsin Health Sciences Institutional Review Board and all participants provided written informed consent as part of the survey.

Measures

Residential self-selection related to a physically active lifestyle (RSS-PA). As a part of the self-administered questionnaire in SHOW participants are asked about reasons for moving to current residence: "*Thinking back when you moved to your current residence, at that time, how important were each of the following factors in your decision to move to your current residence?*" Participants were presented with 19 different items/neighborhood characteristics and indicated their answer on a 4-point scale ranging from "*Not at all important*" to "*Very important*". Six questions reflecting features previously found in the literature to be associated with neighborhood built environment and physical activity were included in the study: "*Outdoor recreational opportunities*", "*Parks and open spaces nearby*", "*Easy access to public transit*", "*Ease of walking or biking to services (such as work, school, grocery stores, restaurants, etc.)*", "*Easy access to a gym or other work-out facility (such as yoga studio, running track, pool, etc.)*" and "*Opportunities for outdoor activities during free-time (such as walking, rollerblading, biking, playing outdoor games, etc.)*". A scale ranging from 1-4 was created from these questions. A higher score on the scale indicates greater importance of living in an environment that supports a physically active lifestyle.

Walkability of the built environment that each participant lives in was estimated with Walk Score®. Walk Score was designed by the software development company Front Seat Management and measures the walkability of built environment using a number of data sources such as Open Street Map, Google, and U.S. Census ("Walk Score Methodology," 2016). Walk Score is calculated from distance to amenities accessible from the participant's home based on street network/connectivity. For every Walk Score, hundreds of walking routes are generated using proprietary algorithms and various data streams to find nearby amenities. Amenities get points based on the distance from participant home address. Amenities that are highly correlated with walking get more weight when Walk Score is calculated and Walk Score also takes into account the number of amenities in the same category. Furthermore, Walk Score takes into account two physical features of the street network that have been found to be associated with walking, i.e. intersection density and block length (Berrigan et al., 2010). The final measures is a score between 0-100 (higher score indicates greater walkability). Walk Score, has been associated with other objective measures of walkability, such as access to walkable amenities and intersection density (Carr et al., 2010a; Duncan et al., 2011). In August 2016, Walk Score was obtained for each participant through an application-programming interface offered by Front Seat Management using geographic coordinates for participant's home address. To mask identification of SHOW participants when obtaining the Walk Score, ten additional random addresses were fed into the application program for every SHOW participant address.

Physical activity was operationalized as average number of steps a participant took every minute. Step count is an accepted method of assessing physical activity and is among the most commonly used measures of physical activity (Bassett et al., 2017). Number of steps was estimated with Actigraph WGT3X-BT device (Actigraph Corporation, Pensacola FL). Participants were instructed to were the device on the right hip for seven continuous days. Accelerometer data were processed using ActiLife version 6.13. Participants who wore the devices for at least 10 hours for 3 out of 7 days were considered having a valid wear time (Freedson et al., 1998). Only participants with valid wear time were included in this analysis. The Actigraph device has been shown to be provide an accurate step count (Chow et al., 2017).

Sociodemographic variables. The following self-reported sociodemographic variables were included in the analysis: gender, age (years), race/ethnicity (white, non-white), marital status (married, not married), education (high school or less, some college, college graduate), car ownership (yes=owns or lease a car, no), kids living in household (yes, no), poverty income ratio (PIR= ratio of midpoint of household income range compared to federal annual poverty levels in the United States). Higher PIR is reflective of greater household income. The poverty guidelines come from the Department of Health and Human Services (U.S. Department of Health and Human Services., 2017). Safety from crime was estimated with one variable "*How safe from crime is your community for walking or riding a bike?*" (not at all safe/not very safe= 0 and somewhat safe/ very safe=1). Geocoded addresses were used to categorize participants into urban or rural residents based on the 2010 U.S. Census Bureau's urban-rural classification system and to obtain information about neighborhood economic status. To represent the

neighborhood economic status, a 5-year estimate from the 2016 American Community Survey (ACS) for average yearly per capita income per block group was used. Individuals were categorized into “high-income” and “low-income” block groups based on the ACS data. Individuals living in a block group with average yearly block group income equal to or less than $25,951 were placed in the “low-income” category. Individuals living in a block group with average yearly block group income greater than $25,951 were placed in the “high-income” category. The cut-off dollar amount used was the mean average yearly block group income for the study sample, that dollar-amount was similar to the average yearly block group income ($25,881) of the state where most of the participants lived in.

Statistical Analyses

Descriptive statistics for the study sample and Pearson’s bivariate correlations were computed in SAS (version 9.4). To test our conceptual model, we ran a structural equation model (SEM) in Mplus (version 7.4) using WLSMV estimator. The WLSMV method provides weighted least square (WLS) parameter estimates using a diagonal weight matrix with robust standard errors and mean- and variance-adjusted chi-square statistics. It is recommended to use robust WLS method when analyzing categorical or ordered data that are not normally distributed as we were in the current study (Brown, 2014). To account for SHOW’s multistage sampling method both cluster and stratum variables were included in the analyses. Further, a SUBPOPULATION statement was included to indicate the sample of interest, i.e. individuals who participated in SHOW in 2016 and live in an urban area.

We followed a recommended two-step process when testing our conceptual model (Anderson & Gerbing, 1988). First, we ran a confirmatory factor analysis to test the measurement model of our conceptual model. The measurement model consisted of one factor that was based on six indicator variables (reasons for moving to current residence). The loading of the first indicator variable was set to one to set the metric of the latent scale. Indicator variables had to have a minimum factor loading of 0.3 to be considered being part of the scale, but higher factor loadings were preferred (Kline, 2015). Correlation was allowed between all the indicator variables. In the second step, we ran a SEM with the selected measurement model and the observed variables of interest.

The fit of the measurement model and the SEM were evaluated using various fit indices that are considered appropriate to use with WLSMV. The fit indices and cut-off values used in this study were: 1) Chi-square test (χ^2) with a *p*-value greater than .01; 2) Root Mean Square Error of Approximation (RMSEA) less than .06; 3) Comparative Fit Index (CFI) equal or greater than .96; 4) Tucker-Lewis Index (TLI) equal or greater than .95; 5) Weighted Root Mean Square Residual (WRMR) equal to or less than 1. The proposed cut-off values come from a study by Yu (Yu, 2002). The goal was to keep the models as parsimonious as possible, residuals and modification indices were used to decide on alternative models. The fit of alternative models were compared with chi-square difference test (DIFFTEST command) (Kline, 2015).

Demographic variables included in the SEM were carefully selected after reviewing the literature and examining what variables were important to include in this specific sample. The variables that were considered for inclusion as having direct path on to residential self-selection and average steps/minute were: age, gender, race, poverty income ratio, education, car

ownership, children living in the household, and marital status. An interaction term between Walk Score and mean block group income was used to test if the association between Walk Score and RSS-PA was modified by census block group economic status. Six values in the outcome variable were identified as outliers (above 1.5 times the interquartile range), these values were set to the highest values, i.e. the 1.5 times the interquartile range (Aguinis et al., 2013). All analyses were restricted to individuals with complete data.

Results

Descriptive Statistics

The age of the participants in the study ranged from 18 to 90 years (mean=50.26, SD=17.59). Other characteristics of the study sample are presented in table 1. Briefly, the majority of the sample was white (83.81%), married (54.97%), owned/leased a car (81.72%), had some college education (73.06%), did not have children under 18 years of age living in the household (65.4%).

A significant correlation was observed between the RSS-PA scale and Walk Score ($r=0.21$, $p<0.001$) and the RSS-PA scale and average steps/minute ($r=0.16$, $p<0.05$). No correlation was found between Walk Score and average steps/minute ($r=-0.04$, $p>0.05$).

Table 1. Sample demographics and correlation between main constructs of the conceptual model. Urban residents in SHOW 2016 (n=383).

	Mean (SD)/ %	Safety	Steps	WS	RSS-PA
Age in years	50.26 (17.59)				
Poverty income ratio	3.28 (2.47)				
Average block group income in $1000	25.95 (10.75)				
Kids live in household	34.46				
Male gender	42.04				
White race/ethnicity	83.81				
Married	54.97				
Own/lease a car	81.72				
Education					
High school or less	26.70				
Some college	20.70				
College graduate	52.36				
RSS-PA scale	2.20 (0.81)				
RSS-PA items:					
Parks and open space nearby	2.54 (0.99)				
Outdoor recreational opportunities	2.25 (1.03)				
Easy access to public transit	1.89 (1.10)				
Ease of walking or biking to services	2.14 (1.07)				
Opportunities for outdoor activities during free-time	2.32 (0.98)				
Easy access to a gym or other work-out facility	1.75 (0.96)				
Above average block group income	47.94	**.34**	.08	**-.28**	.06
Safety from crime (Safety)	3.37 (0.76)		-.03	**-.41**	.06
Average steps/minute (Steps)	5.75 (3.43)			-.04	**.21**
Walk Score (WS)	38.51 (23.19)				**.16**

Note: RSS-PA= Residential self-selection related to a physically active lifestyle; SD= Standard Deviation
Boldface indicates statistical significance ($p<0.05$)

Measurement Model

A confirmatory factor analysis of the six reasons for moving to current residence items in the measurement model indicated an inadequate fit ($\chi^2= 69.568$, $p<0.0001$; RMSEA=0.133;

CFI=0.982, TLI=0.970; WRMR=1.363) despite all items loading significantly on the latent factor. It was decided to remove the item (*Easy access to public transit*) with the lowest factor loading (0.567) from the analysis and based on the modification indices, allow correlation between the measurement errors of item "*Outdoor recreational opportunities*" and "*Parks and open spaces nearby*". With these modifications, four out of five modification indices indicated a good fit (χ^2= 13.047, *p*=0.011; RMSEA=0.077; CFI=0.997, TLI=0.993; WRMR=0.549), factor loadings were all significant and ranged from 0.628 to 0.964 (mean=0.787). To keep the measurement model as parsimonious as possible, no further modifications were done.

Structural Equation Model

Because of how the methodology of Walk Score works and the design of the SHOW sample, the relationship between the RSS-PA scale and Walk Score had to be modeled with the RSS-PA scale as the outcome instead of Walk Score. Upon initial testing, fit indices for the modified version of the conceptual model with a one-way path from walkability to RSS-PA suggested poor model fit, with two out of the five fit indices indicating inadequate fit (χ^2= 112.078, *p*= 0.0003; RMSEA= 0.044 (95% CI=0.029-0.058); CFI= 0.971, TLI= 0.962; WRMR= 1.207). Modification indices indicated that adding a path from safety from crime to the RSS-PA scale would improve the fit. The fit indices for the modified model (shown in Figure 2) indicated a better fit than the original model (χ^2= 90.348, *p*= 0.0205; RMSEA= 0.033 (95% CI=0.013-0.048); CFI= 0.984, TLI= 0.979; WRMR= 0.995). Additionally, an alternative model without the safety variable was estimated in order to create the most parsimonious model possible, this model also had a good fit (χ^2= 83.851, *p*= 0.0278; RMSEA= 0.032 (95% CI=0.011-0.048); CFI=

0.986, TLI= 0.981; WRMR= 0.979). However, a chi-square difference test between the two models with adequate fit, the model including the path between safety from crime and RSS-PA and the alternative model without the safety from crime variable, showed that including the safety variable improved the fit (χ^2_{diff}= 34.174, p< 0.0001). The final model with standardized coefficients is presented in Figure 2, unstandardized coefficients and coefficients for demographic variables in the model are shown in Table 2.

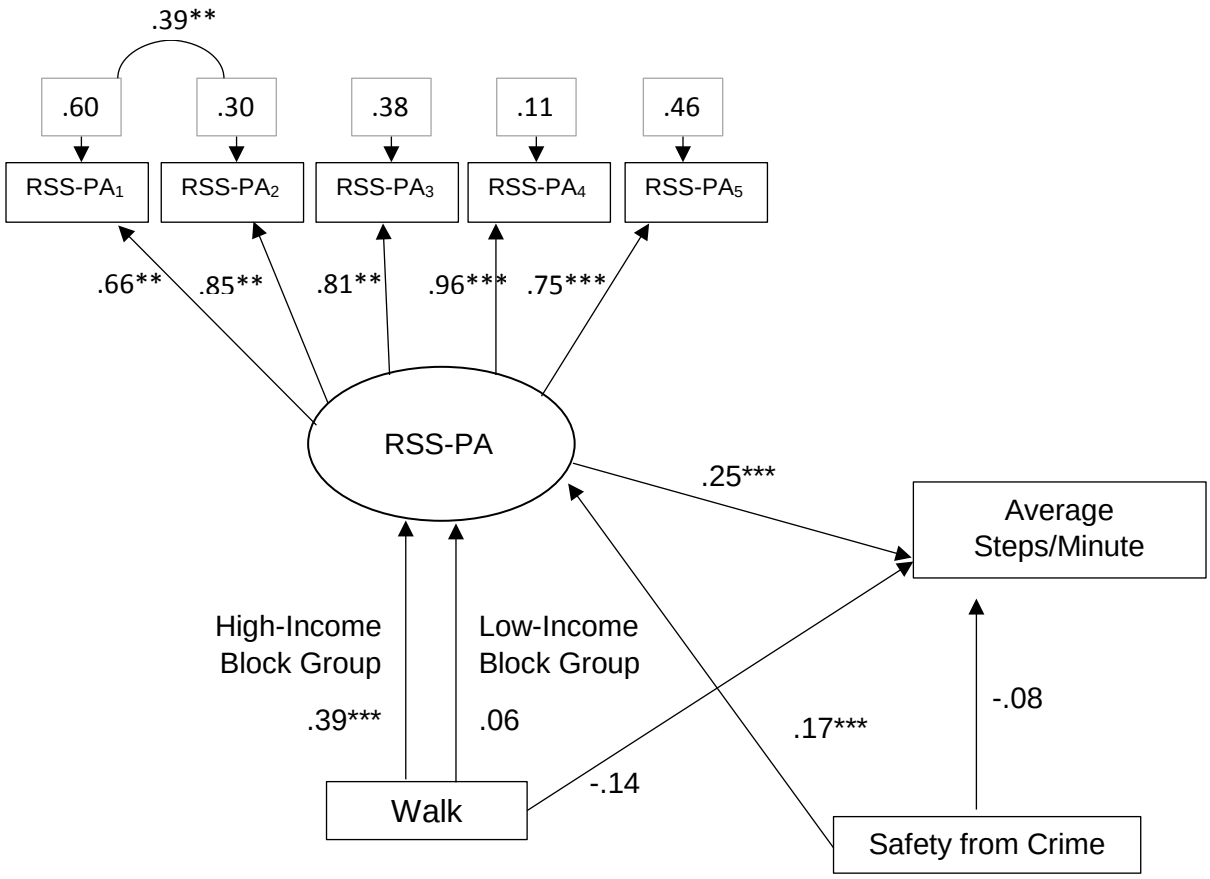


Figure 2. A structural equation model of the residential self-selection related to a physically active lifestyle scale (RSS-PA), Walk Score, and average steps/minute. Fit indices: χ^2= 90.348, p= 0.0205; RMSEA= 0.033 (95% CI=0.013-0.048); CFI= 0.984, TLI= 0.979; WRMR= 0.995.

Note: Standardized coefficients are shown. For reason of clarity, covariance and socio-demographic variable are not shown. *p<0.05, **p<0.001, ***p<0.001.
RRS_1= Parks and open space nearby, RRS_2= Outdoor recreational opportunities,
RRS_4= Ease of walking or biking to services, RRS_5= Opportunities for outdoor activities during free-time,
RRS_6= Easy access to a gym or other work-out facility.

The final model had a significant positive path from the RSS-PA scale to average steps/minute (β=.25, p<0.001). A positive significant path between Walk Score and the RSS-PA scale was also observed among individuals living in high-income block groups (β=.50, p<0.0001), but the same path among individuals living in low-income block groups was non-significant (β=.06, p=0.488). The path between Walk Score and average steps/minute was non-significant (β=-.14, p=0.118). The path between safety from crime to average steps/minute was non-significant (β=-.08, p=0.291), while a significant positive path from safety from crime to the RSS-PA scale was observed (β=.17, p<0.0001).

The demographic variables included in the final model were all significant (see Table 3). Older participants took on average fewer steps/minute (β= -.24, p<0.0001), males took on average more steps than females (β=.12, p=0.031), participants who were better off financially took on average more steps/minute (β=.21, p<0.0001), married participants took on average more steps/minute (β=.15, p=0.001), college graduates took on average fewer steps than participants with no college education (β= -.18, p<0.0001). Participants who owned/leased a car scored significantly lower on the RSS-PA scale (β= -.23, p<0.0001) and older participants tended to score higher on the RSS-PA scale (β=.16, p=0.002). Individuals who had lived longer at their current address scored on average lower on the RSS-PA scale (β= -.15, p=0.010).

A significant negative indirect effect was observed between average block group income and average steps/minute via RSS-PA (β= -.09, p=0.006), a positive indirect effect between safety from crime and average steps/minute via RSS-PA (β= .04, p=0.003). No indirect effect was found between Walk Score and average steps/minute via RSS-PA (β= .02, p=0.510). The

final model accounted for a total of 14% of the variance in the RSS-PA scale and 17% variance in average steps/minute, indicating a medium effect size (Cohen, 1988).

Table 2. Unstandardized and standardized parameters from the final structural equation model[a]. Urban residents in SHOW 2016 (n=363).

Model Parameters	**Unstandardized** Estimate (SE)	**Standardized** Estimate (SE)	*p*-value[b]
Direct effect on average steps			
RSS-PA	1.15 (.31)	.25 (.06)	**<.0001**
Walk Score	-.02 (.01)	-.14 (.09)	.295
Safety from crime	-.36 (.34)	-.08 (.08)	.30
Age	-.04 (.01)	-.24 (.06)	**<.0001**
Gender	.76 (.35)	.12 (.06)	**.031**
Poverty	.26 (.08)	.21 (.06)	**.001**
Married	.95 (.29)	.15 (.04)	**.001**
Some college	-.05 (.38)	-.01 (.05)	.893
College graduate	-1.14 (.27)	-.18 (.04)	**<.0001**
Direct effect on RSS-PA			
Walk Score	.00 (.00)	.06 (.09)	.481
Block group income	-.24 (.08)	-.35 (.10)	**.002**
Block group income*Walk Score	.01 (.00)	.33 (.11)	**.005**
Safety from crime	.16 (.03)	.17 (.03)	**<.0001**
Own/lease a car	-.41 (.09)	-.23 (.05)	**<.0001**
Age	.01 (.00)	.16 (.05)	**.002**
Length of residency	-.09 (.03)	-.15 (.06)	**.008**
Indirect effect on average steps			
Walk Score via RSS-PA	.002 (.003)	.015 (.022)	.508
Safety from crime via RSS-PA	.177 (.061)	.041 (.014)	**.004**
Block group income via RSS-PA	-.271 (.102)	-.086 (.031)	**.008**
Residual variance			
Average steps	8.35 (.56)	.83 (.05)	**<.0001**
RSS scale	0.40 (.04)	.86 (.03)	**<.0001**

Note: SE= Standard Error, RSS-PA= Residential self-selection related to a physically active lifestyle
Boldface indicates statistical significance ($p<0.05$)
[a]Model fit: $\chi^2= 90.348$, $p= 0.0205$; RMSEA= 0.033 (95% CI=0.013-0.048); CFI= 0.984, TLI= 0.979; WRMR= 0.995.
[b]*p*-values for unstandardized coefficients

Discussion

In attempt to address a number of gaps in the literature and gain further understanding of RSS-PA and its role in the relationship between walkability and physical activity, the present study explored the association between walkability and average number of steps while taking RSS-PA, block group income, and safety from crime into account. To our knowledge, this is the first study in the U.S. to examine the association between these variables simultaneously.

As hypothesized, the RSS-PA scale was positively associated with number of steps, i.e. individuals who reported reasons related to a physically active lifestyle being important when moving to their current residence took on average more steps/minute than individuals who did not report these reasons being important. These results are consistent with results from similar studies that have taken place in Denmark (Christiansen et al., 2014), Germany (Baar et al., 2015) and the U.S (Saelens et al., 2012). Further, in support of our hypothesis, walkability was associated with RSS-PA, individuals living in an environment with high Walk Score tended to report reasons related to a physically active lifestyle being important reasons for moving to their current residence. To our knowledge, this is among the first study in the U.S. showing a relationship between reasons for moving to current residence and the characteristics of the residential area. A previous study from Germany showed an association between reasons for moving to current neighborhood and walkability of the neighborhood among college students (Baar et al., 2015). Although, that study used self-reported measures of the environment and did handle the questions about reasons for moving differently than was done in this study.

Furthermore, we found that the association between the RSS-PA scale and walkability was modified by block group economic status, observing only association between Walk Score

and reasons for moving in high-income block groups, not low-income block groups. These result support the idea that high-income groups may be able to satisfy more personal criteria when selecting a place to live (Levine & Frank, 2007). Another interpretation of the modification of block group income might be that it is highlighting the limitations of Walk Score as a measure of walkability. Additionally, to not taking into account safety like mentioned in the introduction, Walk Score has been criticized for failing to capture the pleasantness of walking, e.g. if the environment is aesthetically pleasing (e.g. trees along sidewalks, clean streets etc.). Safety and aesthetics are things that are commonly lacking in low income neighborhoods as opposed to high income neighborhoods (Neckerman et al., 2009). Therefore, the non-significant association found between Walk Score and residential self-selection among low-income block groups might stem from that low-income block groups with high Walk Score are not walkable in reality. Therefore, individuals living in these block groups do not perceive their environment as walkable and thus did not report moving to the area because of it being supportive of a physically active lifestyle. A study on 115 walkable neighborhoods in Washington, DC metro area showed that Walk Score tended to overestimate the walkability of low-income neighborhoods with high access to amenities (Koschinsky et al., 2017).

Contrary to our hypothesis, walkability, as measured by Walk Score, was not found to be associated with number of steps. The Pearson correlation coefficient did not differ significantly from zero and the association in the SEM was non-significant. A recent review paper on Walk Score by Hall & Ram (2018), reported that while previous studies have shown a consistent relationship between Walk Score and walking for transport, association between Walk Score and other types of physical activity/walking (i.e. leisure, total walking) have been less consistent. A

possible explanation for no association between walkability and number of steps is that our outcome measure is not specific enough. Ideally, we would have used a measure of physical activity that took place around the participants' residence, i.e. in the area that the Walk Score refers to. Unfortunately, the dataset that we had access to and most other available datasets do not have separate objective estimates of physical activity by location. Furthermore, walkability is a complex construct that is challenging to operationalize and have some researchers argued that Walk Score is only a proxy for estimating walkability and have recommended using supplementary measures of walkability with Walk Score (Carr et al., 2010a; Duncan et al., 2011). However, lack of comparability between studies on the built environment and physical activity is a big challenge in this field (Brownson, Hoehner, Day, Forsyth, & Sallis, 2009; Ding & Gebel, 2012; Gebel, Bauman, & Petticrew, 2007; McCormack et al., 2004) and if supplementary measures are used with Walk Score the comparability of results from different studies is compromised.

Safety from crime was found to be indirectly (through RSS-PA), but not directly associated with number of steps, indicating that the effect of safety from crime on walking is mediated through RSS-PA. Previous studies on direct effects of safety from crime on walking show mixed results (Foster & Giles-Corti, 2008; Saelens & Handy, 2008). One of the speculations about the reasons for non-significant association between safety from crime and physical activity found in a number of studies is that the effect might be mediated through individual characteristic variables (Foster & Giles-Corti, 2008). Our results support that idea.

No relationship between safety from crime and Walk Score was specified in our SEM model, however, we believe it is worth mentioning that a negative correlation was observed

between Walk Score and perceived safety from crime. The negative correlation indicates that areas that individuals tend to perceive as not being safe for walking or riding a bike are commonly labeled as walkable areas according to Walk Score. A similar pattern was observed in a study by Carr, Dunsiger & Marcus, they found a positive relationship between Walk Score and police department reports of crime. The authors concluded that this association highlighted the limitation of Walk Score and encouraged caution in the use of Walk Score (Carr et al., 2010b).

When creating the RSS-PA scale the item "*Easy access to public transit*" did not fit well in with the rest of the items and was not included in the scale. It is important to keep in mind that this might be specific to the sample of this study. Context can play an important role in studies on the built environment and physical activity, a study by Sallis and colleagues found that variables often used to define the built environment (i.e. transit stops, shops, recreational facilities) varied greatly between 11 countries examined (Sallis, Bowles, et al., 2009). Therefore, if this scale were to be used in a different population, it might be important to include the "*Easy access to public transit*" item in the RSS-PA scale.

There are a number of things that need to be kept in mind when interpreting the results of this study; the study utilized cross-sectional data, so individuals were asked about their reasons for moving to their current residence after they had lived there. Living in the neighborhood might have colored or changed their thinking about why they did decide to move there. Furthermore, some individuals might never have decided to move to their current residence, they might have lived there their whole life. If that is the case, the individuals may have made up their answers based on the environment they live in. Both of these issues might lead to inflation in the estimated association between reasons for moving (i.e. RSS-PA) and Walk Score. Finally,

previously mentioned flaws with the measure used for measuring walkability and physical activity. Walk Score does not take into account all the aspects of a walkable environment and we are measuring average steps taken per minute while wearing the accelerometer steps, not steps that took specifically place in the environment which the participant lives in.

On the other hand, the study has number of noteworthy strengths such as the use of objective measures of the environment and physical activity, using self-reported measures can often complicate the interpretation of study results since self-reported measures tend to be colored by both individual characteristics and behaviors (Lin & Moudon, 2010; Prince et al., 2008). Our study also utilizes specific questions about reasons for moving to current residence as opposed to general environmental preferences for neighborhood. Using specific questions should improve the construct validity of the RSS-PA measure since individuals do not always live in an environment that has their preferred environmental characteristics (Schwanen & Mokhtarian, 2004). Finally, our study includes a random sample of individuals of various ages, possible making the results more generalizable.

Conclusions

Taken together our findings suggest that a residential self-selection based on predisposition to an active lifestyle exist but the extent of it may vary by block group economic status. Unfortunately, we were not able to examine the possible impact of RSS-PA on the association between walkability and physical activity since no association was found between these constructs in our study. It is important to considered safety from crime when estimating how physically active

friendly the environment is even though it might not influence physical activity directly, it might influence physical activity through other variables.

DISCUSSION

Regular physical activity has been shown to be important for overall health, yet, the majority of American adults are not regularly physically active (U.S. Department of Health and Human Services, 1996). Based on studies on the built environment and physical activity, making the built environment more physically active friendly has been identified as a potential means of increasing physical activity levels (Kärmeniemi et al., 2018; Saelens & Handy, 2008). However, most of these studies are based on cross-sectional data and do not account for the possibility of residential self-selection. Thus, many of them show biased estimates of the relationship between the built environment and physical activity. The selection of individuals into residential areas (i.e. residential self-selection) based on their predisposition to a physically active lifestyle has been identified as one of the main challenges of the built environment and physical activity literature (Oakes, 2004). Nevertheless, not many studies have examined the importance of residential self-selection, its magnitude and its implication on the association between the built environment and physical activity. The overall goal of this dissertation was to gain further understanding of residential self-selection and its role in the relationship between the built environment and physical activity.

With using data from a well-characterized random sample of urban residents, evidence of residential self-selection based on predisposition to a physically active lifestyle was observed. A group of individuals who placed greater importance on living in an environment supportive of a physically active lifestyle was identified. These individuals had on average higher levels of physical activity and lower levels of sedentary time. Further, individuals who placed greater importance on living in an environment supportive of a physically active lifestyle tended to live in a walkable area. Additionally, some evidence of residential self-selection varying between

economic status of the residential area were found. If residential self-selection is not accounted for in observational studies on the relationship between the built environment and physical activity the strength of the relationship is likely overestimated. Understanding the associations between residential self-selection, the built environment, and physical activity is important in order to spend the limited resources (e.g., money, time) that exist on effective policies to improve the health of the population. Further, understanding how these relationships vary by neighborhood context can help in designing targeted interventions for those who are most at risk for sedentary lifestyle.

This dissertation examined residential self-selection in relation with the built environment and physical activity with three different approaches. First, it was examined whether nineteen questions about reasons for moving to current residence produced distinct residential self-selection profiles and if physical activity levels and sedentary time differed by the profiles. Second, a residential self-selection scale was created from six reasons identified as being related to a physically active lifestyle. Then that scale was used to examine the association between the residential self-selection based on predisposition to a physically active lifestyle and walkability of participant current residential area. Additionally, the association between the residential self-selection scale and walkability was examined by block group economic status. Finally, a structural equation model was created to examine the association between walkability and physical activity and the previously examined associations between residential self-selection and walkability, and residential self-selection and physical activity, simultaneously. Block group economic status and perceived safety was also included in the model.

Summary of Results

Aim 1 sought to examine weather distinct residential self-selection profiles existed based on reported reasons for moving to current residential location, and whether physical activity levels and sedentary time differed by the profiles. It was hypothesized that there would be a group of people who did emphasize reasons related to features that facilitate an active lifestyle when choosing a residential location, and that these individuals would have higher levels of physical activity than those who did not emphasize these reasons. In the analysis for aim 1, nineteen different questions about reasons for moving to current residence were examined using a unique latent profile analysis (LPA) approach. Using LPA allowed me to capture the complexity of residential location choice as opposed to looking at one reason for moving to current residential area at a time. To my knowledge, this is the first study using LPA modeling for residential self-selection profiles but this technique has been used previously to produce neighborhood environment profiles (Adams et al., 2011; McDonald et al., 2012; Norman et al., 2010; Todd et al., 2016).

Two distinct profiles of individuals with varying emphasis on selecting residential location based on features that facilitate an active lifestyle were identified. The profile with greater emphasis on features that facilitate an active lifestyle had higher levels of accelerometer-assessed physical activity and lower levels of self-reported sitting time. A few studies have investigated the effect of physically active supportive environment being a reason for residential location and its association with physical activity (Baar et al., 2015; Christiansen et al., 2014; Frank et al., 2007). All of these studies have found a relationship between reasons for moving to current residence and physical activity/travel behavior. However, these previous studies were

limited because they used self-reported physical activity data, included only a limited set of questions about reasons for moving to current residence, and the majority of them were conducted outside of the U.S. In aim 1, I addressed these gaps by using more specific questions about reasons for moving to current residence and objective measures of physical activity (accelerometer data). Findings suggest that within a general population of urban residents a distinct residential self-selection profiles with varying emphasis on an environment that facilitates an active lifestyle do exist. The individuals who do pick a residential area based on how physically active friendly the environment is tend to have higher levels of physical activity.

Despite a residential profile with heavy emphasis on environmental features related to a physically active lifestyle was identified, most of the questions related to a physically active lifestyle did not rank high in importance (see appendix 6). For an example the question: "Ease of walking or biking to services" ranked number sixteen out of the nineteen reasons for moving to current residence. This may indicate that in general, environmental features that support an active lifestyle are not considered important among adults living in Wisconsin or that there is a lack of residential areas that are physically active friendly. The low importance of reasons related to a physically active lifestyle corresponds with findings from a previous study conducted in the U.S. (Frank et al., 2007). But, is in contrast with previous studies conducted in Europe where reasons related to a physically active lifestyle have ranked among the most important reasons for residential location (Christiansen et al., 2014; Van Dyck et al., 2011).

The goal of aim 2 was to examine the relationship between reasons for moving to current residence and the walkability of the residential area and test whether the association differed by economic status of the residential area. Six questions about reasons for moving to current

residence identified as being related to a physically active lifestyle were used to create a residential self-selection scale. Higher scores on the scale indicated greater importance of living in an environment that is conducive to physical activity. It was hypothesized that individuals who scored high on the residential self-selection scale would live in a more walkable environment than individuals who scored low on the scale. Further, it was hypothesized that the association between the residential self-selection scale and walkability would differ by the economic status of the residential area.

Studies on reasons for moving to current residence and actual residential environment are scarce but a few studies using self-reported general environmental preferences as a surrogate for residential self-selection and actual environmental characteristics of current residential area do exist. Interestingly, these studies showing that individuals often do not live in an environment that reflect their environmental preferences (Bagley & Mokhtarian, 2002a; Schwanen & Mokhtarian, 2004), suggesting that residential self-selection would not be an important source of bias. However, one study in Germany examined reasons for moving to current residence (as a surrogate for residential self-selection) and current residential environment and showed an association between reasons for moving and self-reported environmental characteristics among college students. Socioeconomic status of the residential area might be important in the context of residential self-selection since not everyone has the opportunity to choose where to live because of their limited socioeconomic resources and residential segregation (Reardon et al., 2017; Sampson & Sharkey, 2008; Watson, 2009). Studies on economic status and residential self-selection are almost non-existent in the public health literature except for one study by

Sallis, Saelens, et al. (2009) that found some evidence of potential differences in residential self-selection between high- and low-income block groups.

The analyses for aim 2 showed a positive association between the residential self-selection scale and walkability, independent of individual sociodemographic characteristics. The association did differ between high- and low-income block groups when walkability was measured with Walk Score but not when self-reported. The observed moderation of block group income when Walk Score was used to measure walkability, supported the hypothesis that individuals in high-income neighborhoods have a greater ability to meet more of their personal criteria when choosing a residential location. However, the absence of moderation by block group income when walkability was self-reported might be indicating that the observed moderation is a result of the limitations of Walk Score as a measure of walkability. Walk Score is limited as an objective measure of walkability because it does not take into account the safety or aesthetics of the environment. Safety from crime and aesthetics of the environment are both features that have been found to be associated with physical activity (Foster & Giles-Corti, 2008; S. Handy, Cao, & Mokhtarian, 2005; G. McCormack et al., 2004) and those features are often lacking in low-income block groups (Neckerman et al., 2009). Therefore, the non-significant association found between Walk Score and residential self-selection among low-income block groups might stem from that low-income block groups with high Walk Score are not walkable in reality. Therefore, individuals living in these block groups do not perceive their environment as walkable and thus did not report moving to the area because of it being supportive of a physically active lifestyle. The self-reported measure of walkability might be capturing the true walkability (accounting for social and economic barriers to physical activity) to a greater extent

than the Walk Score. Thus, no modification of block group income on the relationship between walkability and self-selection are seen when walkability is self-reported. A study by Koschinsky et al. (2017) found that Walk Score overestimated the walkability of low-income neighborhoods with high access to amenities. Results from aim 2 are significant and add to the literature showing a relationship between reasons for moving to current residence and actual walkability of the environment and that this association might differ by economic status of the block group. The results also indicate that confounding by residential self-selection might mainly be an issue in high-income block groups, not low-income block groups.

If residential self-selection mainly exists in high-income residential areas, estimates from previous studies on the association between the built environment and physical activity might not be inflated if the studies took place in lower income areas. Further, considering different types of interventions/policies in areas depending on their economic status might be important. In low-income residential areas where residential self-selection does not exist, removing barriers to physical activity such as improving aesthetics of the environment might be effective in increasing levels of physical activity while that might not be the case in high-income residential areas.

Aim 3 expands upon the preceding aims by examining the association between walkability and physical activity in addition to the previously studies relationships in aim 1 and 2. Aim 3 is among the first studies to examine the relationship between residential self-selection, the built environment, and physical activity simultaneously using a rarely used method in public health, structural equation modeling. One study in Germany had previously examined these associations simultaneously relying on self-reported measures of walkability and physical

activity and found associations between reasons for moving to current residence, walkability, and physical activity (Baar et al., 2015). Additionally, to examine the association between residential self-selection, the built environment, and physical activity, block group economic status and perceived safety from crime were included in the structural equation model in aim 3. Building on findings from aim 2, block group economic status was included as a moderator of the relationship between residential self-selection and walkability. Further, in order to understand how safety from crime may be related to walkability and physical activity, perceived safety from crime was included as a predictor of physical activity. Lack of safety from crime has been identified as an important barrier to physical activity, despite studies showing mixed results (Foster & Giles-Corti, 2008; Handy et al., 2006; McCormack et al., 2004).

The final structural equation model showed a positive association between reasons for moving to current residence and walkability of the residential environment (i.e. Walk Score) among high-income block groups, however, not among low-income block groups. These results are consistent with associations found in aim 2 suggesting that residential self-selection may vary between different economic environments. A positive association was also found between reasons for moving to current residence and physical activity (similar to aim 1), indicating that residential self-selection is indeed a potential confounder in the relationship between walkability and physical activity, but, only among individuals living in high-income block groups. Further, a positive indirect effect between perceived safety from crime and physical activity via reasons for moving to current residence was observed. Indicating that safety is important for physical activity even though it might not be directly related to physical activity and might be one of the reasons for mixed results from previous studies on safety and physical activity (Foster & Giles-

Corti, 2008). No association was found between walkability and physical activity and therefore it was not possible to estimate the total effect of residential self-selection on physical activity. Despite that, the results from aim 3 add to the literature by being the first study in the U.S. examining these relationships simultaneously, using more specific measures of reasons related to a physically active lifestyle, using objective measures of walkability and physical activity, and taking block group economic status and safety into account.

Inconsistencies between Analyses

There are some inconsistencies between the analyses for aim 2 and 3 that are worth mentioning. When creating the residential self-selection scale in aim 2, the factor loadings and the fit indices in the confirmatory factor analysis conducted in the SAS software indicated a good fit when all six variables were set to load on one factor. However, when the same six variables were set to load on one factor in a confirmatory factor analysis in the Mplus software the fit indices indicated a poor fit. This resulted in a six-item residential self-selection scale being used in aim 2 but a five-item scale in aim 3. The discrepancy in the results of the confirmatory factor analyses might stem from the fact that SAS and Mplus provide different fit indices, the fit indices provided in Mplus are somewhat more advanced, i.e. taking into account more aspects of the model such as penalizing for complexity of the model.

Another discrepancy is that in the regression models in aim 2 a number of sociodemographic variables were included (e.g. length of residency, gender) even though they were not significant in the model. However, only the sociodemographic variables that were identified as being important in this sample were included in the structural equation model in aim

3. The reason for including so many sociodemographic variables in the models in aim 2 is to give future studies on this topic some reference because studies on predictors of residential self-selection are limited. However, it was not feasible to include all of these sociodemographic variables in aim 3 because when creating a structural equation model you strive for keeping it as parsimonious as possible and including non-significant variables in the model can result in a poor model fit. Despite these discrepancies, the magnitude of the observed associations between reasons for moving to current residence, walkability (i.e. Walk Score), and block group income are comparable between the aim 2 and 3.

In the conceptual model for the dissertation, there is a one-way arrow from reasons for moving to current residence to walkability so the plan was to use Walk Score as the outcome in aim 2 and 3. However, when working on aim 2 it was discovered that it was impossible to use Walk Score as an outcome in a regression model because of the methodology of Walk Score works and the sample strategy of SHOW. In short, in the third stage of the SHOW sampling strategy, approximately 25 households are randomly picked from each Census block group and most households within the same block group will be assigned exactly the same Walk Score. Having the exact same values for different households/individuals results in incorrect regression estimates. Using Walk Score as the predictor instead of the outcome in aim 2 should be acceptable since the regression is only modeling association.

In aim 3, when using Walk Score as an "outcome" instead of a predictor resulted in a model with residential self-selection in the pathway between Walk Score and physical activity. However, residential self-selection should not be identified as a mediator for two reasons. First, no indirect effects from walkability on physical activity via residential self-selection were found.

Second, even though there is a tradition for making inferences about causal pathways in structural equation models they are showing associations when used with cross-sectional data (Bollen & Pearl, 2013).

Finally, somewhat contradicting methods, LPA and CFA, were used to analyze questions about reasons for moving to current residence in the different aims. The CFA assumes that the underlying latent variable is continues while LPA assumes a discrete latent variable/profiles. Use of these different methods was justified by that in aim 2 and 3 only six out of nineteen questions were analyzed. The six selected questions referred to a similar topic and were expected to represent one continuous latent variable. The nineteen questions analyzed in aim 1 refer to many diverse topics and were not expected to have an underlying continuous latent variable.

Strengths and Limitations

Overall, this dissertation addresses important gaps in understanding the role of residential self-selection in the relationship between the built environment of residential areas and physical activity. The findings may inform community-based and individual-level interventions aimed at promoting physical activity. The study builds on a randomly selected well-characterized cohort of individuals, uses state of the art measures of physical activity and more specific questions of residential self-section than previous studies. Additionally, it uses novel analytic methods to examine the association between residential self-selection, the built environment, and physical activity. Furthermore, this study is conducted from a public health lens using data from a well-established health survey. Majority of research in this area has been largely considered in the area of transportation.

The context of the study is also important. The majority of previous studies on residential self-selection have taken place in Europe where infrastructure for travel, convenience of using public transport and the cost of owning a car greatly differs from the U.S. (Christiansen et al., 2014; Kenworthy & Laube, 1999). The studies that have taken place in the U.S. have mainly looked at large urban metropolitan areas of the west or east coast of the U.S. The data for this study come from urban areas in Wisconsin, both from smaller urban areas and a larger metropolitan area. Wisconsin's urban areas are similar to many other urban areas in the mid-West and central U.S. in that many urban areas that were once thriving have now lost most of their economic base and individuals living in these areas are largely dependent on vehicles for travel.

Despite these strengths, there are a number of things that need to be kept in mind when interpreting the findings of this dissertation. Specific issues regarding each aim have been mentioned in the respective chapters. Here, overarching issues across the analyses are reiterated. First, all the analyses are based on cross-sectional data and, therefore, causal relationships cannot be established. Because the data are cross-sectional, individuals were asked about their reasons for moving to their current residential area after they had lived in the area. Living in the area might have colored or change their thinking about why they did decide to move there. Furthermore, some individuals might never have decided to move to their current residence since they might have lived there their whole life. If that is the case, the individuals may have made up their answers based on the environment of their residential area. If either/both of those issues are influencing the questions about reasons for moving to current residence the association between reasons and walkability is overestimated.

Second, this dissertation focused solely on one aspect of the built environment, its ability to support a physically active lifestyle, i.e. walkability. Two measures of walkability were used, Walk Score and self-reported walkability. Both of these measures mainly focus on access to amenities, i.e. if amenities are within a walking distance. Even though greater access to amenities has been show to correlate with higher physical activity, there are a number of other factors that contribute to creating a walkable environment (Speck, 2013). According to Speck (2013), a walkable environment has to create the opportunity for individuals to take a walk that meets the following four criteria: be useful, safe, comfortable, and interesting. The absent of an association between Walk Score and physical activity in aim 3 may stem from the fact that the measure being used, Walk Score is not reflective of the true walkability of the environment. Further, the contradictory results in aim 2 between analysis with Walk Score and self-reported walkability may stem from the possibility that the self-reported measure is capturing some more aspects of a walkable environment, e.g. the environment being safe and comfortable.

Third, context and demographic characteristics of the study sample need to be kept in mind when interpreting the results. While this can and should be considered a strength, this can also be a limitation for several reasons. First, despite only including SHOW participants living in urban areas, 67% of participants live in a car-dependent environment and 81% own/lease a car. This is not surprising since cities and towns in the U.S. are designed around cars and majority of U.S. households own/lease a car (Muller, 2004; Newman, 1996). Whereas in Denmark and many other European countries, car ownership is much lower and a well-developed cycling infrastructure is in place (Christiansen et al., 2014; Kenworthy & Laube, 1999). Further, the study sample had limited racial diversity, with 84% being white. This limited variability in

walkability, car ownership, and racial diversity prevented stratified analyses by block group income in aim 2 being possible and limits the generalizability of the results.

Fourth, even though accelerometer-based measures of physical activity are considered more accurate than self-reported measures, accelerometer-based measures of physical have their limitations. For an example, accelerometers do not capture some common activities such as biking and carrying loads, also individuals may change their level of activity when wearing the accelerometer because they know they are being monitored. Additionally, it is common that large portion of participants do not want to wear the accelerometer or do not wear it long enough to be considered having a valid wear time (Pedišić & Bauman, 2014). The issue of invalid wear time is a problem in the data being used for this study, where around 31% of participants did not want to participate or did not wear the accelerometer for long enough. This percent is in the higher range of invalid wear time reported from other studies (6-32%) (Pedišić & Bauman, 2014). Participants with invalid wear time were inevitably excluded from the analysis where the outcome was accelerometer-assessed physical activity and may limit the generalizability of the results to only those who adhere to the accelerometer wear instructions.

Finally, additional missing data stem from participants failing to return their self-administered questionnaire, which includes questions about reasons for moving to current residence back to SHOW. Because of the drop-off in the accelerometer section and the self-reported questionnaire between 20-46% (depending on the outcome of interest) of eligible participants are not included in the analysis. It was considered to impute the missing values to be able to include more participants in the analyses. However, the latent profile analysis used in aim 1 cannot handle multiple imputed data and it was not considered appropriate to use single

imputation for this amount of missing data. Using data with single imputation will result in underestimation of standard errors and biased significance tests (Johnson & Young, 2011).

Conclusions

In sum, this study strengthens the evidence of residential self-selection being a confounder in the relationship between the built environment and physical activity, especially in high-income areas. Additionally, it highlights the importance of using reasons for moving to current residence instead of environmental preferences when attempting to adjust for residential self-selection.

Despite no association being found between the objective GIS based measure used in this study to estimate walkability of the built environment and physical activity in this study, it is unlikely that no association exist in reality. It is likely that the lack of association stems from the limitations in the walkability measure used and/or the discrepancy in the reference point of the walkability measure (measure of residential area) and the physical activity measure (overall physical activity). However, previous studies showing a relationship between various features of the built environment and physical activity without accounting for residential self-selection are most likely overestimating the strength of the association.

Making environments more supportive of physical activity may results in increased levels of physical activity, at least among those who are motivated to be physically active, then the environment may work as an enabler. However, it is unlikely that changes in the environment alone will motivate those who are not motivated to be active, more active. Behavior change is a complicated process that takes time (Prochaska, DiClemente, & Norcross, 1992) but can be influenced by social networks and culture (Stokols, 1996). It is possible that non-motivated

individuals who live in an environment that undergoes changes making it more supportive of physical activity may become more active over time. They may become motivated and consequently more active when they see their motivated neighbors start walking/biking to the store and go for a run during their free time etc. Research have shown that frequently seeing others people being active is associated with higher levels of physical activity (King et al., 2000).

Evidence indicates that sedentary individuals are more likely to adopt inexpensive and moderately intense activities that fit into their daily life (e.g. walking/biking to the store) than adopting more vigorous physical activity (e.g. to the gym) (Laitakari, Vuori, & Oja, 1996; Owen & Bauman, 1992; Shephard, 1997). For those individuals, it may be particularly important to try to make neighborhoods more physically active friendly. It has also been shown that lifestyle interventions (i.e. encourage physical activity in daily life) in contrast to structured interventions (i.e. target individuals through structured programs) may result in more long-term maintenance of physical activity (Andersen et al., 1999; A. L. Dunn et al., 1998; Andrea L. Dunn et al., 1999). Thus, trying to find ways to make it easier for individuals to be active during their daily routine is of great interest.

The request for more walkable environments in the U.S. is increasing and the demand is not being met (American Planning Association, 2014; Urban Land Institute, 2015). The current built environment of the U.S. has been shaped by long-standing policies such as zoning regulations, land use ordinances and funding criteria for transportation infrastructure. Policy changes are needed to create new and change existing residential areas to be more supportive of physical activity. Even though the changes alone might not lead to non-motivated individuals to

become physically active, at least it will allow those who are motivated to incorporate more activity into their daily lives.

Future Studies

Looking ahead, the literature on the built environment and physical activity would benefit from more studies focusing on policy-relevant changes in the environment and physical activity levels to make studies more translational. Maybe researchers need to stop focusing on refining measures of the built environment and rather focus on what parts of the environment can be changed with policies. Consequently, studies on those policy-relevant factors and physical activity levels while accounting for residential self-selection are needed. Adjusting for residential self-selection using as specific measures as possible to prevent residual confounding.

Additionally, future studies would benefit from using measures of physical activity with information on where it took place to be able to match the environmental and physical activity measures better. The majority of existing studies that have used objective measures of physical activity do not have information on where the activity took place, i.e. if it was in the residential area, close to work or some other places. Walk Score has been found to be most consistently related to self-reported active travel behavior (i.e. bike/walk to work) as opposed to other types of physical activity. This consistently observed relationship is most likely caused by the fact that active travel commonly starts/ends in the residential area which the Walk Score refers to. No association was found between Walk Score and average number of steps in aim 3, but individuals living in an area with higher Walk Score were more likely to report active travel (data not shown).

In addition, most existing studies have only focused on environmental characteristics of the residential area, thus missing other places that physical activity may take place (e.g. around work place). Duncan et al. (2016) used an innovative approach to examine the association between walkability and walking among adults living in Paris (Duncan et al., 2016). They had participants wear a GPS receiver and an accelerometer for seven days to track their location and physical activity. Walk Score was then obtained for origin and destination of every trip. Finally, they examined the association between Walk Score and walking at the trip level instead of examining it at the individual level like most studies have done. Duncan and colleagues found that trips that either originated or ended in an area with high Walk Score had higher number of steps compared to trip not originating or ending in an area with high Walk Score. In the analysis, adjustment for individual and neighborhood characteristics was made but not for residential self-selection. Further studies, using innovative approaches like this might help to move the literature on the built environment and physical activity forward and strengthen the evidence for implementing policies that aim at making the built environment more physically active friendly.

REFERENCES

Adams, M. A., Sallis, J. F., Kerr, J., Conway, T. L., Saelens, B. E., Frank, L. D., … Cain, K. L. (2011). Neighborhood environment profiles related to physical activity and weight status: A latent profile analysis. *Preventive Medicine*, *52*(5), 326–331. https://doi.org/10.1016/j.ypmed.2011.02.020

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

American Planning Association. (2014). *Investing in Place for Economic Growth and Competitiveness*.

Andersen, R. E., Wadden, T. A., Bartlett, S. J., Zemel, B., Verde, T. J., & Franckowiak, S. C. (1999). Effects of Lifestyle Activity vs Structured Aerobic Exercise in Obese Women: A Randomized Trial. *JAMA*, *281*(4), 335–340. https://doi.org/10-1001/pubs.JAMA-ISSN-0098-7484-281-4-joc80705

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Armstrong, T., & Bull, F. (2006). Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ). *Journal of Public Health*, *14*(2), 66–70. https://doi.org/10.1007/s10389-006-0024-x

Asparouhov, T., & Muthén, B. (2014a). Auxiliary variables in mixture modeling: Three-step approaches using M plus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 329–341.

Asparouhov, T., & Muthén, B. (2014b). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes*, *21*(2).

Baar, J., Romppel, M., Igel, U., Brähler, E., & Grande, G. (2015). The independent relations of both residential self-selection and the environment to physical activity. *International Journal of Environmental Health Research*, *25*(3), 288–298. https://doi.org/10.1080/09603123.2014.938029

Bagley, M. N., & Mokhtarian, P. L. (2002a). The impact of residential neighborhood type on travel behavior: a structural equations modeling approach. *The Annals of Regional Science*, *36*(2), 279–297.

Bagley, M. N., & Mokhtarian, P. L. (2002b). The impact of residential neighborhood type on travel behavior: a structural equations modeling approach. *The Annals of Regional Science*, *36*(2), 279–297.

Bailey, E. J., Malecki, K. C., Engelman, C. D., Walsh, M. C., Bersch, A. J., Martinez-Donate, A. P., … Nieto, F. J. (2014). Predictors of discordance between perceived and objective neighborhood data. *Annals of Epidemiology*, *24*(3), 214–221. https://doi.org/10.1016/j.annepidem.2013.12.007

Ball, K., Jeffery, R. W., Crawford, D. A., Roberts, R. J., Salmon, J., & Timperio, A. F. (2008). Mismatch between perceived and objective measures of physical activity environments. *Preventive Medicine*, *47*(3), 294–298. https://doi.org/10.1016/j.ypmed.2008.05.001

Ball, K., Timperio, A. F., & Crawford, D. A. (2006). Understanding environmental influences on nutrition and physical activity behaviors: where should we look and what should we count? *International Journal of Behavioral Nutrition and Physical Activity*, *3*, 33. https://doi.org/10.1186/1479-5868-3-33

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Prentice-Hall.

Barton, H. (2009). Land use planning and health and well-being. *Land Use Policy*, *26*, *Supplement 1*, S115–S123. https://doi.org/10.1016/j.landusepol.2009.09.008

Bassett, D. R., Toth, L. P., LaMunion, S. R., & Crouter, S. E. (2017). Step Counting: A Review of Measurement Considerations and Health-Related Applications. *Sports Medicine (Auckland, N.Z.)*, *47*(7), 1303–1315. https://doi.org/10.1007/s40279-016-0663-1

Bauman, A. E., Sallis, J. F., Dzewaltowski, D. A., & Owen, N. (2002). Toward a better understanding of the influences on physical activity: The role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American Journal of Preventive Medicine*, *23*(2), 5–14. https://doi.org/10.1016/S0749-3797(02)00469-5

Berrigan, D., Pickle, L. W., & Dill, J. (2010). Associations between street connectivity and active transportation. *International Journal of Health Geographics*, *9*, 20. https://doi.org/10.1186/1476-072X-9-20

Bishop, B. (2009). *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt.

Bollen, K. A., & Pearl, J. (2013). Eight Myths About Causality and Structural Equation Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301–328). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_15

Brown, S. C., Pantin, H., Lombard, J., Toro, M., Huang, S., Plater-Zyberk, E., … Szapocznik, J. (2013). Walk Score®. *American Journal of Preventive Medicine*, *45*(2), 202–206. https://doi.org/10.1016/j.amepre.2013.03.021

Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research, Second Edition*. Guilford Publications.

Brownson, R. C., Boehemer, T. K., & Luke, D. A. (2005). DECLINING RATES OF PHYSICAL ACTIVITY IN THE UNITED STATES: What Are the Contributors? *Annual Review of Public Health*, *26*(1), 421–443. https://doi.org/10.1146/annurev.publhealth.26.021304.144437

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the Built Environment for Physical Activity: State of the Science. *American Journal of Preventive Medicine*, *36*(4, Supplement), S99-S123.e12. https://doi.org/10.1016/j.amepre.2009.01.005

Bull, F. C., Maslin, T. S., & Armstrong, T. (2009). Global Physical Activity Questionnaire (GPAQ): Nine Country Reliability and Validity Study. *Journal of Physical Activity and Health*, *6*(6), 790–804. https://doi.org/10.1123/jpah.6.6.790

Cao, X. (2010). Exploring Causal Effects of Neighborhood Type on Walking Behavior Using Stratification on the Propensity Score. *Environment and Planning A*, *42*(2), 487–504. https://doi.org/10.1068/a4269

Cao, X., Handy, S. L., & Mokhtarian, P. L. (2006). The Influences of the Built Environment and Residential Self-Selection on Pedestrian Behavior: Evidence from Austin, TX. *Transportation*, *33*(1), 1–20. https://doi.org/10.1007/s11116-005-7027-2

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings. *Transport Reviews*, *29*(3), 359–395. https://doi.org/10.1080/01441640802539195

Carr, L. J., Dunsiger, S. I., & Marcus, B. H. (2010a). Validation of Walk Score for estimating access to walkable amenities. *British Journal of Sports Medicine*, bjsports69609. https://doi.org/10.1136/bjsm.2009.069609

Carr, L. J., Dunsiger, S. I., & Marcus, B. H. (2010b). Walk Score™ As a Global Estimate of Neighborhood Walkability. *American Journal of Preventive Medicine*, *39*(5), 460–463. https://doi.org/10.1016/j.amepre.2010.07.007

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Reports*, *100*(2), 126–131.

Centers for Disease Control and Prevention (CDC). (2008). Prevalence of self-reported physically active adults--United States, 2007. *MMWR. Morbidity and Mortality Weekly Report*, *57*(48), 1297–1300.

Chow, J. J., Thom, J. M., Wewege, M. A., Ward, R. E., & Parmenter, B. J. (2017). Accuracy of step count measured by physical activity monitors: The effect of gait speed and anatomical placement site. *Gait & Posture*, *57*, 199–203. https://doi.org/10.1016/j.gaitpost.2017.06.012

Christiansen, L. B., Madsen, T., Schipperijn, J., Ersbøll, A. K., & Troelsen, J. (2014). Variations in active transport behavior among different neighborhoods and across adult lifestages. *Journal of Transport & Health*, *1*(4), 316–325. https://doi.org/10.1016/j.jth.2014.10.002

Clark, W. A. V. (2009). Changing Residential Preferences across Income, Education, and Age Findings from the Multi-City Study of Urban Inequality. *Urban Affairs Review*, *44*(3), 334–355. https://doi.org/10.1177/1078087408321497

Clark, W. A. V., & Dieleman, F. M. (1996). *Households and Housing: Choice and Outcomes in the Housing Market*. Transaction Publishers.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, New Jersey: L*. Erlbaum.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: with applications in the social behavioral, and health sciences*. Hoboken, N.J: Wiley.

Committee on Physical Activity, Health, Transportation, and Land Use. (2005). *Does the built environment influence physical activity?: examining the evidence*. Washington, D.C: Transportation Research Board.

Committee, P. A. G. A. (2008). Physical activity guidelines advisory committee report, 2008. *Washington, DC: US Department of Health and Human Services*, *2008*, A1-H14.

Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health: Neighborhoods and health. *Annals of the New York Academy of Sciences*, *1186*(1), 125–145. https://doi.org/10.1111/j.1749-6632.2009.05333.x

Ding, D., & Gebel, K. (2012). Built environment, physical activity, and obesity: What have we learned from reviewing the literature? *Health & Place*, *18*(1), 100–105. https://doi.org/10.1016/j.healthplace.2011.08.021

Duncan, D. T., Aldstadt, J., Whalen, J., Melly, S. J., & Gortmaker, S. L. (2011). Validation of Walk Score® for Estimating Neighborhood Walkability: An Analysis of Four US Metropolitan Areas. *International Journal of Environmental Research and Public Health*, *8*(11), 4160–4179. https://doi.org/10.3390/ijerph8114160

Duncan, D. T., Méline, J., Kestens, Y., Day, K., Elbel, B., Trasande, L., & Chaix, B. (2016). Walk Score, Transportation Mode Choice, and Walking Among French Adults: A GPS, Accelerometer, and Mobility Survey Study. *International Journal of Environmental Research and Public Health*, *13*(6), 611. https://doi.org/10.3390/ijerph13060611

Dunn, A. L., Garcia, M. E., Marcus, B. H., Kampert, J. B., Kohl, H. W., & Blair, S. N. (1998). Six-month physical activity and fitness changes in Project Active, a randomized trial. *Medicine and Science in Sports and Exercise*, *30*(7), 1076–1083. https://doi.org/10.1097/00005768-199807000-00009

Dunn, Andrea L., Marcus, B. H., Kampert, J. B., Garcia, M. E., Iii, H. W. K., & Blair, S. N. (1999). Comparison of Lifestyle and Structured Interventions to Increase Physical Activity and Cardiorespiratory Fitness: A Randomized Trial. *JAMA*, *281*(4), 327–334. https://doi.org/10-1001/pubs.JAMA-ISSN-0098-7484-281-4-joc80889

Durand, C. P., Andalib, M., Dunton, G. F., Wolch, J., & Pentz, M. A. (2011). A systematic review of built environment factors related to physical activity and obesity risk: implications for smart growth urban planning. *Obesity Reviews*, *12*(5), e173–e182. https://doi.org/10.1111/j.1467-789X.2010.00826.x

Feinstein, J. S. (1993). The Relationship between Socioeconomic Status and Health: A Review of the Literature. *The Milbank Quarterly*, *71*(2), 279–322. https://doi.org/10.2307/3350401

Fishbein, M., & Ajzen, I. (1975). Belief, attitudes, intention, and behavior. *An Introduction to Theory and Research. Massachussets: Addison-Wesley*.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Foster, S., & Giles-Corti, B. (2008). The built environment, neighborhood crime and constrained physical activity: An exploration of inconsistent findings. *Preventive Medicine*, *47*(3), 241–251. https://doi.org/10.1016/j.ypmed.2008.03.017

Frank, L. D., Saelens, B. E., Powell, K. E., & Chapman, J. E. (2007). Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Social Science & Medicine*, *65*(9), 1898–1914. https://doi.org/10.1016/j.socscimed.2007.05.053

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and Science in Sports and Exercise*, *30*(5), 777–781.

Gebel, K., Bauman, A. E., & Petticrew, M. (2007). The Physical Environment and Physical Activity: A Critical Appraisal of Review Articles. *American Journal of Preventive Medicine*, *32*(5), 361-369.e3. https://doi.org/10.1016/j.amepre.2007.01.020

Gebel, K., Bauman, A., & Owen, N. (2009). Correlates of Non-Concordance between Perceived and Objective Measures of Walkability. *Annals of Behavioral Medicine*, *37*(2), 228–238. https://doi.org/10.1007/s12160-009-9098-3

Gerteis, J., Izrael, D., Deitz, D., LeRoy, L., Ricciardi, R., Miller, T., & Basu, J. (2014). *Multiple chronic conditions chartbook*. AHRQ Publications.

Giles-Corti, B., Bull, F., Knuiman, M., McCormack, G., Van Niel, K., Timperio, A., … Boruff, B. (2013). The influence of urban design on neighbourhood walking following residential relocation: longitudinal results from the RESIDE study. *Social Science & Medicine (1982)*, *77*, 20–30. https://doi.org/10.1016/j.socscimed.2012.10.016

Greenwald, M., & Boarnet, M. (2001). Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*, 33–41. https://doi.org/10.3141/1780-05

Hajna, S., Ross, N. A., Brazeau, A.-S., Bélisle, P., Joseph, L., & Dasgupta, K. (2015). Associations between neighbourhood walkability and daily steps in adults: a systematic review and meta-analysis. *BMC Public Health*, *15*, 768. https://doi.org/10.1186/s12889-015-2082-x

Hajna, S., Ross, N. A., Joseph, L., Harper, S., & Dasgupta, K. (2015). Neighbourhood walkability, daily steps and utilitarian walking in Canadian adults. *BMJ Open*, *5*(11), e008964. https://doi.org/10.1136/bmjopen-2015-008964

Hall, C. M., & Ram, Y. (2018). Walk score® and its potential contribution to the study of active transport and walkability: A critical and systematic review. *Transportation Research Part D: Transport and Environment*. https://doi.org/10.1016/j.trd.2017.12.018

Han, L., Neilands, T. B., & Dolcini, M. M. (2011). University of California, San Francisco. Retrieved from http://www.lexjansen.com/wuss/2001/WUSS01044.pdf

Handy, S. (2005). Critical assessment of the literature on the relationships among transportation, land use, and physical activity. *Transportation Research Board and the Institute of Medicine Committee on Physical Activity, Health, Transportation, and Land Use. Resource Paper for TRB Special Report, 282*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.2645&rep=rep1&type=pdf

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, *10*(6), 427–444. https://doi.org/10.1016/j.trd.2005.05.002

Handy, S., Cao, X., & Mokhtarian, P. L. (2006). Self-Selection in the Relationship between the Built Environment and Walking: Empirical Evidence from Northern California. *Journal of the American Planning Association*, *72*(1), 55–74. https://doi.org/10.1080/01944360608976724

Healy, G. N., Clark, B. K., Winkler, E. A. H., Gardiner, P. A., Brown, W. J., & Matthews, C. E. (2011). Measurement of Adults' Sedentary Time in Population-Based Studies. *American Journal of Preventive Medicine*, *41*(2), 216–227. https://doi.org/10.1016/j.amepre.2011.05.005

Heijs, W., Deursen, A. van, Leussink, M., & Smeets, J. (2011). Re-searching the labyrinth of life-styles. *Journal of Housing and the Built Environment*, *26*(4), 411. https://doi.org/10.1007/s10901-011-9237-6

Hirsch, J. A., Moore, K. A., Clarke, P. J., Rodriguez, D. A., Evenson, K. R., Brines, S. J., … Diez Roux, A. V. (2014). Changes in the built environment and changes in the amount of walking over time: longitudinal results from the multi-ethnic study of atherosclerosis. *American Journal of Epidemiology*, *180*(8), 799–809. https://doi.org/10.1093/aje/kwu218

Hirsch, J. A., Moore, K. A., Evenson, K. R., Rodriguez, D. A., & Roux, A. V. D. (2013). Walk Score® and Transit Score® and Walking in the Multi-Ethnic Study of Atherosclerosis. *American Journal of Preventive Medicine*, *45*(2), 158–166. https://doi.org/10.1016/j.amepre.2013.03.018

Hu, F. B., Li, T. Y., Colditz, G. A., Willett, W. C., & Manson, J. E. (2003). Television Watching and Other Sedentary Behaviors in Relation to Risk of Obesity and Type 2 Diabetes Mellitus in Women. *JAMA*, *289*(14), 1785–1791. https://doi.org/10.1001/jama.289.14.1785

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Humpel, N., Owen, N., & Leslie, E. (2002). Environmental factors associated with adults' participation in physical activity: A review. *American Journal of Preventive Medicine*, *22*(3), 188–199. https://doi.org/10.1016/S0749-3797(01)00426-3

Jackson, R. J., Dannenberg, A. L., & Frumkin, H. (2013). Health and the Built Environment: 10 Years After. *American Journal of Public Health*, *103*(9), 1542–1544. https://doi.org/10.2105/AJPH.2013.301482

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited Commentary: Propensity Scores. *American Journal of Epidemiology*, *150*(4), 327–333.

Johnson, D. R., & Young, R. (2011). Toward Best Practices in Analyzing Datasets with Missing Data: Comparisons and Recommendations. *Journal of Marriage and Family*, *73*(5), 926–945. https://doi.org/10.1111/j.1741-3737.2011.00861.x

Kahn, E. B., Ramsey, L. T., Brownson, R. C., Heath, G. W., Howze, E. H., Powell, K. E., … Corso, P. (2002). The effectiveness of interventions to increase physical activity: A systematic review1,2. *American Journal of Preventive Medicine*, *22*(4, Supplement 1), 73–107. https://doi.org/10.1016/S0749-3797(02)00434-8

Kärmeniemi, M., Lankila, T., Ikäheimo, T., Koivumaa-Honkanen, H., & Korpelainen, R. (2018). The Built Environment as a Determinant of Physical Activity: A Systematic Review of Longitudinal Studies and Natural Experiments. *Annals of Behavioral Medicine*, *52*(3), 239–251. https://doi.org/10.1093/abm/kax043

Kenworthy, J. R., & Laube, F. B. (1999). Patterns of automobile dependence in cities: an international overview of key physical and economic dimensions with some implications for urban policy. *Transportation Research Part A: Policy and Practice*, *33*(7–8), 691–723. https://doi.org/10.1016/S0965-8564(99)00006-3

Kindig, D. A. (2007). Understanding Population Health Terminology. *Milbank Quarterly*, *85*(1), 139–161. https://doi.org/10.1111/j.1468-0009.2007.00479.x

King, A. C., Castro, C., Wilcox, S., Eyler, A. A., Sallis, J. F., & Brownson, R. C. (2000). Personal and environmental factors associated with physical inactivity among different racial–ethnic groups of U.S. middle-aged and older-aged women. *Health Psychology*, *19*(4), 354–364. https://doi.org/10.1037/0278-6133.19.4.354

Kitamura, R., Mokhtarian, P. L., & Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation*, *24*(2), 125–158.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Koschinsky, J., Talen, E., Alfonzo, M., & Lee, S. (2017). How walkable is Walker's paradise? *Environment and Planning B: Urban Analytics and City Science*, *44*(2), 343–363. https://doi.org/10.1177/0265813515625641

Lackey, K. J., & Kaczynski, A. T. (2009). Correspondence of perceived vs. objective proximity to parks and their relationship to park-based physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *6*, 53. https://doi.org/10.1186/1479-5868-6-53

Laitakari, J., Vuori, I., & Oja, P. (1996). Is long-term maintenance of health-related physical activity possible? An analysis of concepts and evidence. *Health Education Research*, *11*(4), 463–477. https://doi.org/10.1093/her/11.4.463

Lee, I.-M., Ewing, R., & Sesso, H. D. (2009). The Built Environment and Physical Activity Levels: The Harvard Alumni Health Study. *American Journal of Preventive Medicine*, *37*(4), 293–298. https://doi.org/10.1016/j.amepre.2009.06.007

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007a). Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health & Place*, *13*(1), 111–122. https://doi.org/10.1016/j.healthplace.2005.11.001

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007b). Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health & Place*, *13*(1), 111–122. https://doi.org/10.1016/j.healthplace.2005.11.001

Levine, J., & Frank, L. D. (2007). Transportation and land-use preferences and residents' neighborhood choices: the sufficiency of compact development in the Atlanta region. *Transportation*, *34*(2), 255–274. https://doi.org/10.1007/s11116-006-9104-6

Lin, L., & Moudon, A. V. (2010). Objective versus subjective measures of the built environment, which are most effective in capturing associations with walking? *Health & Place*, *16*(2), 339–348. https://doi.org/10.1016/j.healthplace.2009.11.002

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767–778. https://doi.org/10.1093/biomet/88.3.767

Malecki, K. C., Engelman, C. D., Peppard, P. E., Nieto, F., Grabow, M. L., Bernardinello, M., … Martinez-Donate, A. (2014). The Wisconsin Assessment of the Social and Built Environment (WASABE): a multi-dimensional objective audit instrument for examining neighborhood effects on health. *BMC Public Health*, *14*(1), 1165. https://doi.org/10.1186/1471-2458-14-1165

Manaugh, K., & El-Geneidy, A. (2011). Validating walkability indices: How do different households respond to the walkability of their neighborhood? *Transportation Research Part D: Transport and Environment*, *16*(4), 309–315. https://doi.org/10.1016/j.trd.2011.01.009

Matthews, C. E., Chen, K. Y., Freedson, P. S., Buchowski, M. S., Beech, B. M., Pate, R. R., & Troiano, R. P. (2008). Amount of Time Spent in Sedentary Behaviors in the United States, 2003–2004. *American Journal of Epidemiology*, *167*(7), 875–881. https://doi.org/10.1093/aje/kwm390

McCormack, G. R., Cerin, E., Leslie, E., Toit, L. D., & Owen, N. (2008). Objective Versus Perceived Walking Distances to Destinations: Correspondence and Predictive Validity. *Environment and Behavior*, *40*(3), 401–425. https://doi.org/10.1177/0013916507300560

McCormack, G. R., Friedenreich, C., Sandalack, B. A., Giles-Corti, B., Doyle-Baker, P. K., & Shiell, A. (2012). The relationship between cluster-analysis derived walkability and local recreational and transportation walking among Canadian adults. *Health & Place*, *18*(5), 1079–1087. https://doi.org/10.1016/j.healthplace.2012.04.014

McCormack, G. R., Giles-Corti, B., Lange, A., Smith, T., Martin, K., & Pikora, T. J. (2004). An update of recent evidence of the relationship between objective and self-report measures of the physical environment and physical activity behaviours. *Journal of Science and Medicine in Sport*, *7*(1, Supplement 1), 81–92. https://doi.org/10.1016/S1440-2440(04)80282-2

McCormack, G. R., & Shiell, A. (2011). In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *The International Journal of Behavioral Nutrition and Physical Activity*, *8*, 125. https://doi.org/10.1186/1479-5868-8-125

McDonald, K., Hearst, M., Farbakhsh, K., Patnode, C., Forsyth, A., Sirard, J., & Lytle, L. (2012). Adolescent physical activity and the built environment: A latent class analysis approach. *Health & Place*, *18*(2), 191–198. https://doi.org/10.1016/j.healthplace.2011.09.004

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An Ecological Perspective on Health Promotion Programs. *Health Education & Behavior*, *15*(4), 351–377. https://doi.org/10.1177/109019818801500401

Muller, P. O. (2004). *Transportation and Urban Form - Stages in the Spatial Evolution of the American Metropolis*. Retrieved from https://trid.trb.org/view.aspx?id=756060

Murray, C. J., Abraham, J., Ali, M. K., Alvarado, M., Atkinson, C., Baddour, L. M., … Birbeck, G. (2013). The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *Jama*, *310*(6), 591–606.

National Research Council. (1990). *Inner-City Poverty in the United States*. National Academies Press.

Neckerman, K. M., Lovasi, G. S., Davies, S., Purciel, M., Quinn, J., Feder, E., … Rundle, A. (2009). Disparities in Urban Neighborhood Conditions: Evidence from GIS Measures and Field Observation in New York City. *Journal of Public Health Policy*, *30*, S264–S285.

Nelson, M. C., Gordon-Larsen, P., Song, Y., & Popkin, B. M. (2006). Built and Social Environments: Associations with Adolescent Overweight and Activity. *American Journal of Preventive Medicine*, *31*(2), 109–117. https://doi.org/10.1016/j.amepre.2006.03.026

Newman, P. (1996). Reducing automobile dependence. *Environment and Urbanization*, *8*(1), 67–92.

Nieto, F. J., Peppard, P. E., Engelman, C. D., McElroy, J. A., Galvao, L. W., Friedman, E. M., … Malecki, K. C. (2010). The Survey of the Health of Wisconsin (SHOW), a novel infrastructure for population health research: rationale and methods. *BMC Public Health*, *10*(1), 1.

Norman, G. J., Adams, M. A., Kerr, J., Ryan, S., Frank, L. D., & Roesch, S. C. (2010). A Latent Profile Analysis of Neighborhood Recreation Environments in Relation to Adolescent Physical Activity, Sedentary Time, and Obesity: *Journal of Public Health Management and Practice*, *16*(5), 411–419. https://doi.org/10.1097/PHH.0b013e3181c60e92

Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social Science & Medicine*, *58*(10), 1929–1952. https://doi.org/10.1016/j.socscimed.2003.08.004

Oakes, J. M. (2014). Invited Commentary: Repeated Measures, Selection Bias, and Effect Identification in Neighborhood Effect Studies. *American Journal of Epidemiology*, *180*(8), 785–787. https://doi.org/10.1093/aje/kwu231

Oakes, J. M., & Church, T. R. (2007). Invited Commentary: Advancing Propensity Score Methods in Epidemiology. *American Journal of Epidemiology*, *165*(10), 1119–1121. https://doi.org/10.1093/aje/kwm072

Orleans, C. T. (2000). Promoting the Maintenance of Health Behavior Change: Recommendations for the Next Generation of Research and Practice. *Health Psychology*, *19*(1), 76–83.

Owen, N., & Bauman, A. (1992). The Descriptive Epidemiology of a Sedentary Lifestyle in Adult Australians. *International Journal of Epidemiology*, *21*(2), 305–310. https://doi.org/10.1093/ije/21.2.305

Owen, N., Cerin, E., Leslie, E., duToit, L., Coffee, N., Frank, L. D., … Sallis, J. F. (2007). Neighborhood walkability and the walking behavior of Australian adults. *American Journal of Preventive Medicine*, *33*(5), 387–395. https://doi.org/10.1016/j.amepre.2007.07.025

Owen, N., Humpel, N., Leslie, E., Bauman, A., & Sallis, J. F. (2004). Understanding environmental influences on walking: Review and research agenda. *American Journal of Preventive Medicine*, *27*(1), 67–76. https://doi.org/10.1016/j.amepre.2004.03.006

Pedišić, Ž., & Bauman, A. (2014). Accelerometer-based measures in physical activity surveillance: current practices and issues. *Br J Sports Med*, bjsports-2013-093407. https://doi.org/10.1136/bjsports-2013-093407

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *5*, 56. https://doi.org/10.1186/1479-5868-5-56

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In Search of How People Change. *American Psychologist*, 13.

Reardon, S. F., Townsend, J., & Fox, L. (2017). A Continuous Measure of the Joint Distribution of Race and Income Among Neighborhoods. *RSF*, *3*(2), 34–62. https://doi.org/10.7758/RSF.2017.3.2.02

Saelens, B. E., & Handy, S. L. (2008). Built Environment Correlates of Walking: A Review. *Medicine and Science in Sports and Exercise*, *40*(7 Suppl), S550–S566. https://doi.org/10.1249/MSS.0b013e31817c67a4

Saelens, B. E., Sallis, J. F., Black, J. B., & Chen, D. (2003a). Neighborhood-Based Differences in Physical Activity: An Environment Scale Evaluation. *American Journal of Public Health*, *93*(9), 1552–1558. https://doi.org/10.2105/AJPH.93.9.1552

Saelens, B. E., Sallis, J. F., Black, J. B., & Chen, D. (2003b). Neighborhood-Based Differences in Physical Activity: An Environment Scale Evaluation. *American Journal of Public Health*, *93*(9), 1552–1558. https://doi.org/10.2105/AJPH.93.9.1552

Saelens, B. E., Sallis, J. F., Frank, L. D., Cain, K. L., Conway, T. L., Chapman, J. E., … Kerr, J. (2012). Neighborhood environment and psychosocial correlates of adults' physical activity. *Medicine and Science in Sports and Exercise*, *44*(4), 637–646. https://doi.org/10.1249/MSS.0b013e318237fe18

Sallis, J. F., Bowles, H. R., Bauman, A., Ainsworth, B. E., Bull, F. C., Craig, C. L., … Bergman, P. (2009). Neighborhood Environments and Physical Activity Among Adults in 11 Countries. *American Journal of Preventive Medicine*, *36*(6), 484–490. https://doi.org/10.1016/j.amepre.2009.01.031

Sallis, J. F., Saelens, B. E., Frank, L. D., Conway, T. L., Slymen, D. J., Cain, K. L., … Kerr, J. (2009). Neighborhood built environment and income: Examining multiple health outcomes. *Social Science & Medicine*, *68*(7), 1285–1293. https://doi.org/10.1016/j.socscimed.2009.01.017

Sampson, R. J., & Sharkey, P. (2008). Neighborhood selection and the social reproduction of concentrated racial inequality. *Demography*, *45*(1), 1–29.

Schwanen, T., & Mokhtarian, P. L. (2004). The extent and determinants of dissonance between actual and preferred residential neighborhood type. *Environment and Planning B: Planning and Design*, *31*(5), 759–784. https://doi.org/10.1068/b3039

Schwanen, T., & Mokhtarian, P. L. (2005a). What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography*, *13*(1), 83–99. https://doi.org/10.1016/j.jtrangeo.2004.11.001

Schwanen, T., & Mokhtarian, P. L. (2005b). What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled. *Transportation Research Part D: Transport and Environment*, *10*(2), 127–151. https://doi.org/10.1016/j.trd.2004.11.002

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shephard, R. J. (1997). What is the optimal type of physical activity to enhance health? *British Journal of Sports Medicine, 31*(4), 277–284. https://doi.org/10.1136/bjsm.31.4.277

Simons-Morton, B., McLeroy, K. R., & Wendel, M. L. (2011). *Behavior theory in health promotion practice and research*. Jones & Bartlett Publishers.

Speck, J. (2013). *Walkable City: How Downtown Can Save America, One Step at a Time*. Macmillan.

Stokols, D. (1996). Translating social ecological theory into guidelines for community health promotion. *American Journal of Health Promotion, 10*(4), 282–298.

Stokols, D., Grzywacz, J. G., McMahan, S., & Phillips, K. (2003). Increasing the health promotive capacity of human environments. *American Journal of Health Promotion, 18*(1), 4–13.

Street Smart Walk Score. (2011, January 27). Retrieved March 11, 2016, from http://blog.walkscore.com/2011/01/preview-street-smart-walk-score/

Szklo, M., & Nieto, F. J. (2004). Epidemiology: Beyond the Basics. *Sudbury, Massachussets: Jones and Bartlett Publishers*, p363-377.

Talen, E., & Koschinsky, J. (2013). The walkable neighborhood: a literature review. *International Journal of Sustainable Land Use and Urban Planning (IJSLUP), 1*(1). Retrieved from https://www.sciencetarget.com/Journal/index.php/IJSLUP/article/view/211

The CALIS Procedure. (2011). In *SAS/STAT® 9.3 User's Guide*. Cary, NC: SAS Institute Inc.

Thomas, M.-P. (2011). Differentiation in lifestyles and residential choices of families in Switzerland. In *23 rd European Network for Housing Research Conference," Mixité": an urban and housing issue*.

Thorp, A. A., Owen, N., Neuhaus, M., & Dunstan, D. W. (2011). Sedentary Behaviors and Subsequent Health Outcomes in Adults: A Systematic Review of Longitudinal Studies, 1996–2011. *American Journal of Preventive Medicine, 41*(2), 207–215. https://doi.org/10.1016/j.amepre.2011.05.004

Todd, M., Adams, M. A., Kurka, J., Conway, T. L., Cain, K. L., Buman, M. P., … King, A. C. (2016). GIS-measured walkability, transit, and recreation environments in relation to older Adults' physical activity: A latent profile analysis. *Preventive Medicine, 93*, 57–63. https://doi.org/10.1016/j.ypmed.2016.09.019

Tremblay, M. S., Aubert, S., Barnes, J. D., Saunders, T. J., Carson, V., Latimer-Cheung, A. E., … Chinapaw, M. J. M. (2017). Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome. *International Journal of Behavioral Nutrition and Physical Activity, 14*, 75. https://doi.org/10.1186/s12966-017-0525-8

Troiano, R. P., Berrigan, D., Dodd, K. W., MâSse, L. C., Tilert, T., & Mcdowell, M. (2008). Physical Activity in the United States Measured by Accelerometer: *Medicine & Science in Sports & Exercise, 40*(1), 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, *40*(1), 181.

Tuckel, P., & Milczarski, W. (2015). Walk ScoreTM, Perceived Neighborhood Walkability, and Walking in the US. *American Journal of Health Behavior*, *39*(2), 242–256. https://doi.org/10.5993/AJHB.39.2.11

Turner, S., Shunk, G., & Hottenstein, A. (1998). DEVELOPMENT OF A METHODOLOGY TO ESTIMATE BICYCLE AND PEDESTRIAN TRAVEL DEMAND. Retrieved from https://trid.trb.org/view/497047

Urban Land Institute (Ed.). (2015). *America in 2015: a ULI survey of views on housing, transportation, and community*. Washington, D.C.: Urban Land Institute.

U.S. Department of Health and Human Services. (1996). *Physical activity and health: a report of the Surgeon General*. Atlanta, GA: U.S.

U.S. Department of Health and Human Services. (2015). *Step It Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities*. Washington, DC: U.S. Dept of Health and Human Services, Office of the Surgeon General.

U.S. Department of Health and Human Services. (2017). *Annual Update of the HHS Poverty Guidelines*. Retrieved from https://www.federalregister.gov/documents/2017/01/31/2017-02076/annual-update-of-the-hhs-poverty-guidelines

Van Dyck, D., Cardon, G., Deforche, B., Owen, N., & De Bourdeaudhuij, I. (2011). Relationships between neighborhood walkability and adults' physical activity: How important is residential self-selection? *Health & Place*, *17*(4), 1011–1014. https://doi.org/10.1016/j.healthplace.2011.05.005

Van Dyck, D., Cardon, G., Deforche, B., Sallis, J. F., Owen, N., & De Bourdeaudhuij, I. (2010). Neighborhood SES and walkability are related to physical activity behavior in Belgian adults. *Preventive Medicine*, *50, Supplement*, S74–S79. https://doi.org/10.1016/j.ypmed.2009.07.027

Walk Score Methodology. (2016). Retrieved March 20, 2016, from https://www.walkscore.com/methodology.shtml

Walker, J. L., & Li, J. (2006). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, *9*(1), 77–101. https://doi.org/10.1007/s10109-006-0030-0

Warburton, D. E. R., Nicol, C. W., & Bredin, S. S. D. (2006a). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, *174*(6), 801–809. https://doi.org/10.1503/cmaj.051351

Warburton, D. E. R., Nicol, C. W., & Bredin, S. S. D. (2006b). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, *174*(6), 801–809. https://doi.org/10.1503/cmaj.051351

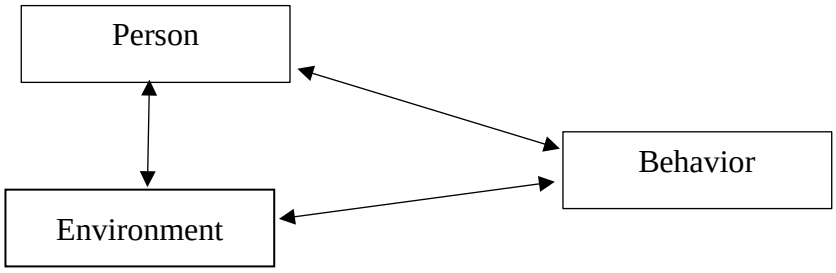
Ward, B. W., Schiller, J. S., & Goodman, R. A. (2014). Multiple Chronic Conditions Among US Adults: A 2012 Update. *Preventing Chronic Disease*, *11*. https://doi.org/10.5888/pcd11.130389

Watson, T. (2009). Inequality and the Measurement of Residential Segregation by Income in American Neighborhoods. *Review of Income and Wealth*, *55*(3), 820–844. https://doi.org/10.1111/j.1475-4991.2009.00346.x

Yu, C.-Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes* (Unpublished doctoral dissertation). University of California, Los Angeles.

APPENDICES

Appendix 1



Social Cognitive Theory (Bandura, 1986)

Social Ecological Model for Health Promotion (McLeroy, Bibeau, Steckler, & Glanz, 1988)

Appendix 2

Reasons for excluding participants living in rural areas:

1) Very limited variance in Walk Score in rural areas (see Table 1a). All participants living in rural areas have a car-dependent Walk Score (see Table 1b). Therefore, keeping rural participants in the sample is not adding additional information to the analysis; it is mainly increasing the sample size.
2) Questions about reasons for moving to current residence might not be as relevant in rural areas. That is, People's lifestyle in rural areas may be fundamentally different from people's lifestyle living in urban areas, and keeping them in the same sample may have undue influences in the analyses.

Table 1a. Descriptive statistics for Walk Score among participants living in rural areas.

Location		**Variability**	
Mean	4.01	**Std Deviation**	9.22
Median	0.00	**Variance**	85.02
Mode	0.00	**Range**	44.00
		Interquartile Range	2.00

Table 1b. Frequency for Walk Score categories among participants living in rural areas.

Walk Score Categories	**Frequency**	**Percent**	**Cumulative**	**Cumulative**
Very car dependent	219	91.63	219	91.63
Car dependent	20	8.37	239	100.00

Table 2a. Descriptive statistics for Walk Score among participants living in urban areas.

Location		Variability	
Mean	40.97	**Std Deviation**	23.62
Median	44.00	**Variance**	558.05
Mode	55.00	**Range**	96.00
		Interquartile Range	33.00

Table 2b. Frequency for Walk Score categories among participants living in urban areas.

Walk Score Categories	Frequency	Percent	Cumulative	Cumulative
Very car dependent	127	26.57	127	26.57
Car dependent	150	31.38	277	57.95
Somewhat walkable	143	29.92	420	87.87
Very walkable	41	8.58	461	96.44
Walkers paradise	17	3.56	478	100.00

Appendix 3

Table 1. Comparison of characteristics of participants with valid and not valid data from the accelerometry section and self-administered questionnaire.

	Total sample[a] N=478	Valid data N=269	Not valid data N=209	*p*-value[b]
Age in years (mean, SD)	47.81 (17.81)	52.63 (16.76)	41.61 (17.22)	<0.0001
Average block group income in $1000 (mean, SD)	25.47 (10.92)	26.58 (10.87)	24.04 (10.84)	0.011
Walk Score (mean, SD)	40.97 (23.62)	39.87 (23.27)	42.39 (24.05)	0.247
Self-reported total physical activity hours/week (mean, SD)	14.46 (19.60)	14.88 (20.90)	13.91 (17.80)	0.582
Kids living in household %	36.40	29.74	44.98	0.001
Male gender %	42.47	42.01	43.06	0.817
White race/ethnicity %	78.66	85.13	70.33	<0.0001
Married %	49.27	55.60	41.15	0.002
Own/lease a car %	78.0	84.01	70.33	0.003
Education %				0.003
High school or less	28.09	25.37	31.58	
Some college	22.64	18.66	27.75	
College graduate	49.27	55.97	40.67	

[a]Total sample= 2016 adult SHOW participants who live in an urban area.
[b]*p*-value is based on chi-square test for categorical variables and t-test for continuous variables.

Appendix 4

Thinking back when you moved to your current residence, at that time, how important were each of the following factors in your decision to move to your current residence?

	Not at all important	Slightly important	Moderately important	Very important
Safe, low-crime area				
Affordable housing				
Attractive, well-kept homes				
Privacy and quiet				
High quality schools				
Lots of trees and other greenery				
Close to friends/family				
Close to work/job				
Friendly neighbors				
Parks and open spaces nearby				
Easy access to health care or other services				
Outdoor recreational opportunities (bike paths, sports fields)				
Amenities such as community center or church				
Lots of things to do (restaurants, movies, shopping, etc.)				
Easy access to public transit				
High racial or ethnic diversity				
Ease of walking or biking to services (such as work, school, grocery stores, restaurants, etc.)				
Opportunities for outdoor activities during free-time (such as walking, rollerblading, biking, playing outdoor games, etc.)				
Easy access to a gym or other work-out facility (such as yoga studio, running track, pool, etc.)				

Appendix 5

The next questions are about the neighborhood you currently live in. Your answers to these questions will help us to understand if it is easy to get around your neighborhood on foot or on a bike, and what kinds of places are nearby.

About how many minutes would it take to <u>walk</u> from your home to the nearest of these facilities?

	0-5 minutes	6-10 minutes	11-20 minutes	21-30 minutes	More than 30 minutes	None within walking distance
Park, playground, or playing field						
Public recreation center						
Trail for walking or biking						
Public swimming pool						
Convenience or small grocery store						
Supermarket						
Post office						
Library						
Elementary school						
Other school						
Fast food restaurant						
Other restaurant						
Pharmacy or drug store						
Salon or barber						
Other type of store						
Place of worship						
Indoor fitness facility						
Golf course						

Appendix 6

Table 1. Reasons for moving to current residence ranked by importance.

Rank	Question	Mean
1	Safe, low-crime area	3.37
2	Affordable housing	3.18
3	Attractive, well-kept homes	3.07
4	Privacy and quiet	2.91
5	Friendly neighbors	2.80
6	Parks and open spaces nearby	2.54
7	Lots of trees and other greenery	2.50
8	High quality schools	2.44
9	Close to work/job	2.41
10	Lots of things to do (restaurants, movies, shopping, etc.)	2.38
11	Opportunities for outdoor activities during free-time (such as walking, rollerblading, biking, playing outdoor games, etc.)	2.32
12	Outdoor recreational opportunities (bike paths, sports fields)	2.25
13	Easy access to health care or other services	2.25
14	Close to friends/family	2.23
15	Amenities such as community center or church	2.16
16	Ease of walking or biking to services (such as work, school, grocery stores, restaurants, etc.)	2.14
17	Easy access to public transit	1.89
18	Easy access to a gym or other work-out facility (such as yoga studio, running track, pool, etc.)	1.75
19	High racial or ethnic diversity	1.66

Table 2. Reasons for moving to current residence relevant to a physically active lifestyle ranked by importance.

Rank	Question	Mean
6	Parks and open spaces nearby	2.54
11	Opportunities for outdoor activities during free-time (such as walking, rollerblading, biking, playing outdoor games, etc.)	2.32
12	Outdoor recreational opportunities (bike paths, sports fields)	2.25
16	Ease of walking or biking to services (such as work, school, grocery stores, restaurants, etc.)	2.14
17	Easy access to public transit	1.89
18	Easy access to a gym or other work-out facility (such as yoga studio, running track, pool, etc.)	1.75