The Impact of Built and Social Environment on Physical Activity among Older Adults

Stephen Mooney

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2016

ABSTRACT

The Impact of Built and Social Environment on Physical Activity among Older Adults

Stephen Mooney

Physical activity, defined as bodily movement produced by skeletal muscles that results in energy expenditure, has many known mental and physical health benefits for older adults. However, as of 2008, only 22.6% of older adults in the United States reported meeting recommended physical activity guidelines. This dissertation examines the role of the built and social environment on physical activity among older adults, with particular focus on physical disorder, or the visual indications of neighborhood deterioration. All empirical analyses use data from the New York City Neighborhood and Mental Health in the Elderly Study (NYCNAMES-II), a three-wave longitudinal study of about 3,500 older adults living in New York City.

We first systematically review the existing literature concerning physical disorder as an influence on physical activity among adults of all ages. We find that most prior studies of disorder and activity have been cross-sectional and that disorder has not consistently been associated with less activity across all studies. However, we also find indications that older adults' activity levels may be more negatively impacted by disorder than younger adults' activity levels.

Next, we use a longitudinal analysis to estimate the association between neighborhood disorder and total physical activity among the NYCNAMES-II cohort. In multivariable mixed regression models accounting for individual and neighborhood factors, for missing data, and for loss to follow-up, we find that each standard deviation increase in neighborhood disorder was associated with an estimated 3.0 units (95% CI: 1.9, 4.2) lower PASE score at baseline, or the equivalent of about 10 minutes of walking per day. There was no significant interaction between physical disorder and changes in PASE score over two years of follow-up.

We next apply a latent transition analysis to identify patterns of types of physical activity the same cohort, identifying seven latent classes of activity. Of these seven classes, three pairs of classes were roughly equivalent except for participation in exercise. About three quarters of subjects remained within each latent class between waves; most transitions that did occur were between classes defined by

exercise to the parallel class without exercise or vice-versa. More neighborhood disorder was modestly associated with moving out of a sports and recreation class (Relative Risk = 1.27, 95% CI = 1.00, 1.61 between waves 1 and 2, Relative Risk = 1.28, 95% CI = 0.85, 1.93 between waves 2 and 3).

Finally, we develop the Neighborhood Environment-Wide Association Study (NE-WAS), an agnostic approach to systematically explore the plethora of neighborhood measures available to modern researchers equipped with geographic information systems (GIS) software. We find that only neighborhood socioeconomic status and disorder measures were associated with total activity and gardening, whereas a broader range of measures was associated with walking.

Substantively, we conclude that more physical disorder was associated with less physical activity, potentially due to decreases in sports and recreation among those living amidst physical disorder, though latent transition analysis estimates were too imprecise to rule out chance. Future longitudinal research on physical disorder as an influence on physical activity would benefit from longer periods of follow-up in which more subjects moved between neighborhoods. Methodologically, the NE-WAS approach appears to be a promising way to systematize neighborhood research as the scale of available spatially located administrative data continues to grow. Future NE-WASes might profitably focus on comparing the spatial scale of neighborhood measures.

Table of Contents

List of Tables and Figures

Acknowledgments

I owe thanks most of all to Andrew Rundle, for his extensive comments, suggestions, and general mentoring not only on this dissertation but throughout my graduate career. Thanks also to the other members of my dissertation committee: Alfred Neugut, Thelma Mielenz, Shaung Wang, and Ryan Demmer, for helpful, thoughtful comments. I also would like to thank the NYCNAMES-II team, especially Magdalena Cerda, Spruha Joshi, and James Quinn, for helping build and clean the dataset I used for the empirical portions of my dissertation, and to John Beard for securing the funding to do the NYCNAMES-II survey in the first place. The Built Environment and Health research group, especially Gina Lovasi, Kathryn Neckerman, James Stark, and Daniel Sheehan, also helped me frame ideas around neighborhood context as a driver of health-related behaviors.

I should also thank the many, many other supportive members of the Mailman community, who were causes of this dissertation in the Rothman sense: without their influence I'd never have gotten to the point of putting it together, at least not how and when I did. On the staff and faculty side, I owe particular thanks to Liliane Zaretsky, Lisa Bates, Sharon Schwartz, Alfredo Morabia, Abdulrahman El-Sayed, Magdalena Cerda, Katherine Keyes, and Charles DiMaggio for friendship, insight, and mentoring. On the student side, the whole cohort who took the methods exam in Fall 2014 have been my loving, tight-knit community. Unofficial class president Stephanie Shiau deserves special credit for organizing us all. More broadly, every epi doctoral student I've interacted with in every cohort has impressed me in some way. Thanks.

Finally, on a more practical note, the time I spent developing and completing this dissertation was supported in part by a fellowship from EHE International, Inc.

Dedication

This dissertation is dedicated to my whole family, for their enduring love and support. But especially, it is dedicated to my nuclear family: to Catherine Williams, for her forbearance in the face of my flaws as a partner, and to Ian Mooney and Zack Mooney, for providing stronger evidence for the social contagion of happiness than Christakis and Fowler could ever hope for.

Chapter 1. Introduction

Physical activity, defined as bodily movement produced by skeletal muscles that results in energy expenditure, has many known mental and physical health benefits for older adults [1, 2]. Physical activity not only prevents coronary heart disease, stroke, colon cancer, and post-menopausal breast cancer, but also helps to maintain a healthy body weight, may prevent or delay dementia, and may minimize both frequency and consequences of falls [3-13]. However, as of 2008, only 22.6% of older US adults reported meeting recommended physical activity guidelines [14]. Nearly a third of Americans age 65 and older reported no leisure-time physical activity in the past month [15]. Physical inactivity was the sixth leading cause of loss of disability-adjusted life years (DALYs) in the United States between 1990-2010, and is an upstream factor in more prominent causes of DALY loss, such as high blood pressure and high body mass index [16].

The gap between the documented need for physical activity and the relatively low activity levels realized among older adults has long been a public health concern [17]. Many early interventions designed to increase activity among older adults focused on behavior change strategies to encourage *exercise*, the subset of physical activity defined as repetitive body movement done to improve or maintain physical fitness [18, 19]. These interventions were frequently successful in the short term, but maintenance of elevated exercise levels after the end of the intervention proved inconsistent [20, 21].

The difficulty of implementing sustainable individually-targeted behavior modifications over the long-term drove physical activity researchers to explore the influence of the social and physical environment in which health behavior choices occur, applying the social ecological theory of behavior change conceptual framework to integrate forces external to individuals with individual decision making processes [22]. This conceptual development, which occurred in parallel with a broader trend of integrating structural forces into individual-centered health research [23-25],holds that built environment, interpersonal relationships, and individual preferences each contribute to individual behavior choices with health consequences, such as whether to walk in one's neighborhood [26, 27]. Figure 1 depicts a social ecological model of influences on physical activity, adapted from prior work [28, 29]. Based on the implications of social ecological models, the United States Centers for Disease Control and Prevention has made it a priority to understand how built environments support physical activity in older adults [15].

One component of social ecology that may be particularly important for physical activity in older adults is physical disorder [30-32]. Physical disorder, defined as the visual indications of neighborhood deterioration, or as the 'broken windows' of the eponymous theory of crime [33, 34], emanates from macro-scale social forces such as informal social control and the economic environment, and manifests in the physical environment experienced by nearby residents. Particularly in the wake of late 20th century de-industrialization, many North American inner cities are characterized by extensive physical disorder [35] and lower activity levels than are observed in the general population [36]. Qualitative evidence from walk-along interviews with older adults suggests physical disorder may incite fear of crime victimization and dissuade walking and other outdoor activity [37].

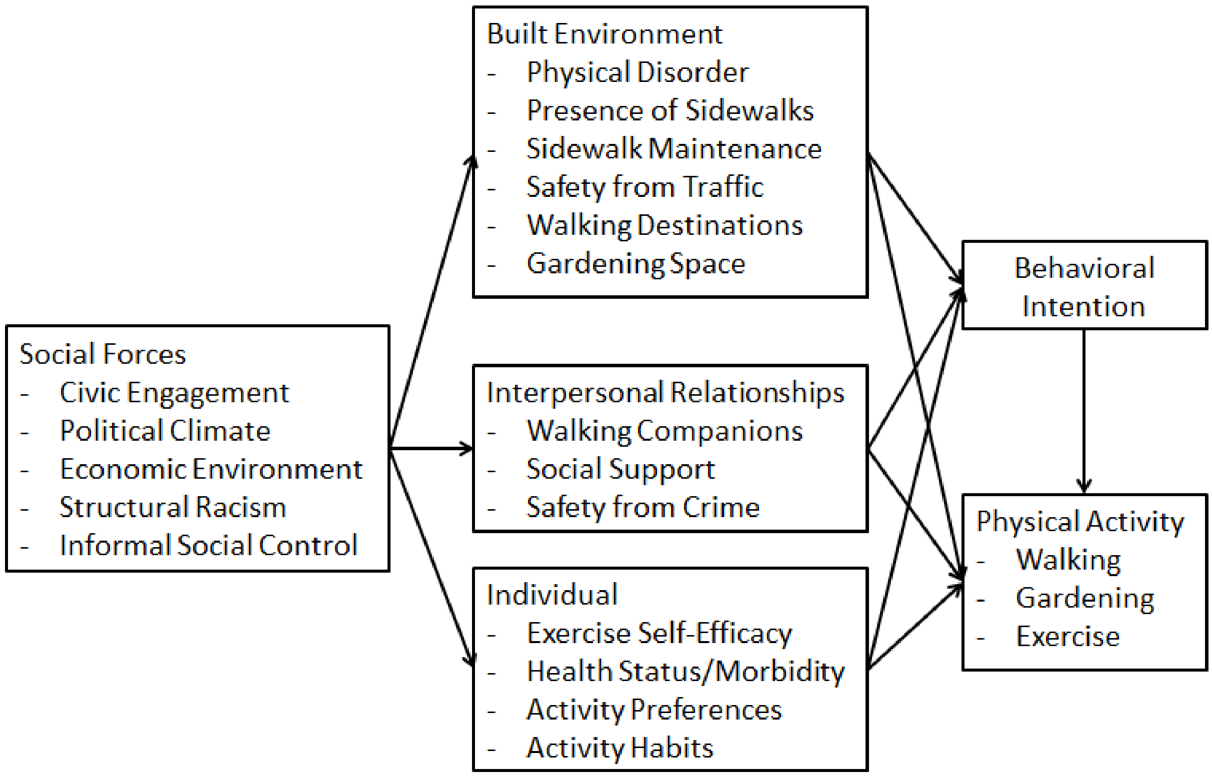
Indeed, in spite of relatively thin quantitative evidence supporting physical disorder (hereafter: disorder) as an impediment to outdoor physical activity [38], the combination of qualitative evidence, the prevalence of less activity and more disorder in more disadvantaged neighborhoods, and an intuitive plausibility that disorder acts as a barrier to activity among older adults has led experts to consider disorder reduction a key component of supporting healthy aging. For example, the American Association of Retired Persons' Livable Communities Evaluation Guide cites well-maintained housing in its section on safety [39], and the Australian Local Government Association's report on age-friendly built environments calls for removing graffiti and rubbish as part of fostering a safe walking environment [40].

Are these calls to focus on maintenance and cleanliness scientifically justified? As noted above, important research gaps surrounding disorder as an influence on physical activity remain. First, little of the research to date regarding disorder has been longitudinal, limiting causal interpretation of identified associations [41-43]. Second, with a few exceptions [44-46], most studies have focused either on overall physical activity or specifically on walking or commuting with active transport [47], failing to account for potentially important differences in individual preferences or social and environmental influences on other types of activities, including gardening, housework, and caring for others [48]. Finally, because neighborhoods are complex, multi-dimensional entities, it is unclear that the most relevant components of the social and environmental context in which physical activity occurs are systematically included in neighborhood studies [42]. If these components were excluded in prior studies, they may result in residual confounding that obscures the true relationship between disorder and physical activity. More generally, inconsistent control for relevant contextual factors in prior studies may explain those studies' inconsistent findings.

This dissertation will target these research gaps, augmenting an analysis of physical disorder as a barrier to activity, as rooted within the social ecological theory of behavior change, with an exploratory 'data science' approach to identify the neighborhood covariates most empirically predictive of physical activity. Specifically, Chapter 2 will synthesize the existing literature concerning physical disorder as an influence on physical activity among adults, with particular focus on older adults. Chapter 3 will work within the social ecological theory of behavior change to investigate the longitudinal relationship between neighborhood physical disorder and physical activity, accounting for individual covariates, in a 3-wave study of older adult residents of New York City. Chapter 4 will expand on this investigation by applying a latent transition analysis to the same cohort, to explore how *types* of activity, such as walking, sports and recreation, or gardening, may be influenced not only by individual preference but also by social and environmental context. Next, Chapter 5 will break from the theoretically rooted approaches to explore the value of data science-style exploratory methods in expanding our knowledge of contextual influences on activity among older adults. Finally, Chapter 6 will summarize the findings from the prior chapters and suggest directions for future research.

The empirical analyses in this dissertation – Chapters 3, 4, and 5 – will use data from NYCNAMES-II, a three-wave survey of about 3,500 New York City residents aged 65-75 at baseline. NYCNAMES-II was initially conducted to assess the effects of neighborhood characteristics, including interventions to improve pedestrian safety, on physical activity and subsequent depression among older adults. The physical context of New York City has previously been characterized in great detail, allowing precise estimates of several hundred characteristics of the neighborhoods surrounding each subject's home address at each wave.

Figure 1. A socio-ecological model of various influences on physical activity



Chapter 2: Neighborhood Physical Disorder and Physical Activity: A Systematic Review

Abstract

Background

Most adults, especially in lower socioeconomic status groups, do not meet physical activity recommendations. Neighborhood disorder – the deterioration of urban landscapes – may be one barrier to physical activity. This systematic review assesses the evidence that neighborhood disorder inhibits physical activity among adults.

Methods

A systematic search of comprehensive databases identified 28 peer-reviewed English language articles on this topic published between 2001 and 2015.

Results

Most studies were based in North America, Europe, and Australia. Broadly speaking, findings were inconsistent. There was some evidence that less severe indicators of disorder, such as litter and graffiti, may also mark areas with more pedestrian activity, negating or reversing the expected inverse association between disorder and physical activity. By contrast, more severe indicators of disorder such as dilapidated or abandoned buildings may discourage discretionary outdoor activity, particularly among vulnerable populations such as women and older adults.

Conclusions

While disorder has been studied in relation to physical activity, substantial variation in disorder measures and study populations have led to disparate results. Future studies of disorder and activity are warranted, particularly studies using longitudinal data, that incorporate validated and internally consistent disorder measures, and that focus on the activity domains and sub-populations that are most likely to be affected by disorder.

Introduction

Physical activity has many known mental and physical health benefits for adults, yet many adults do not meet physical activity guidelines [14, 49]. Many neighborhood contextual factors, including the presence of a structurally safe environment with sidewalks and curb ramps that support walking and the presence of outdoor spaces that support gardening, are thought to influence both the form and the frequency, intensity, and duration of physical activity [45, 50, 51].

One aspect of the neighborhood environment that may affect activity is physical disorder [52]. Physical disorder – the deterioration of urban spaces owing to social forces favoring neglect and abandonment – has long been of interest to social scientists studying domains other than activity [30, 34]. Criminologists and sociologists have debated the controversial 'broken windows' theory that disorder encourages violent crime [53-57]. Exposure to disorder has also been implicated as a potential cause of psychopathology [58, 59], risky sexual practices [60], and obesity [61, 62].

Here, we focus on the potential role disorder may play as a barrrier to physical activity [38]. There are several complementary mechanisms through which disorder may function: at more severe levels, disorder and associated 'incivilities' may induce a fear of crime victimization, which in turn discourages outdoor activity [43, 52]. More modest levels of disorder may also prevent walkers from experiencing aesthetic pleasure they would have encountered in less chaotic environments [63], deterring walking and resulting in less overall physical activity. Finally, disorder's physical manifestations may act as structural impediments to safety, as when cracked sidewalks pose a fall hazard for older adults [37].

However, several prominent studies failed to find support for disorder acting as a barrier to walking [64, 65]. This may be in part because residents of disordered neighborhoods, who are often socioeconomically disadvantaged, are disproportionately 'captive walkers'[66] -- people who lack access to transportation modes other than walking. By definition, captive walkers' transportation mode choice cannot be affected by disorder, and thus a study including a large proportion of captive walkers may find no association between disorder and walking even if disorder inhibits walking among those with alternatives to walking. Conversely, the influence of neighborhood disorder may be particularly strong among older adults, whose mobility may be constrained by declining health [67] or by women, who frequently report stronger perceptions of crime vulnerability [68].

If disorder is to be understood as a barrier to activity, then, it is important to understand the socio-economic contexts and demographic groups in which disorder may have the strongest inhibitory impact. Such contexts and groups may be identified by a critical review of the literature assessing disorder's association with activity. However, while there has been one systematic review nearly a decade ago relating the fear of crime to physical activity that briefly discussed disorder [43] and there have been a host of reviews relating neighborhood influences more broadly to activity and walking [69], to the best of our knowledge, no prior review has examined the influence of neighborhood disorder on physical activity.

Physical Disorder

There is no single definition of physical disorder [70], though most accounts make reference to forms of deterioration or decline in urban landscapes that are abnormal or threatening but not necessarily illegal [31, 71]. A recent review catalogs twenty-six indicators used in disorder measures in the scholarly literature, ranging from immediate threats to safety (e.g. packs of wild dogs) to simple commercial advertising (e.g. alcohol and tobacco advertisements) [70]. In practice, most scales designed to assess disorder include items assessing not only aspects of longer-term abandonment, such as poor building maintenance or vacant lots, but also forms of neglect that may be transient, such as litter or graffiti [58, 72-74]. Finally, neighborhood aesthetics scales designed to assess the overall pleasure of being active in a neighborhood often include disorder subscales, as disorder components such as graffiti and poor building upkeep may make being active in a neighborhood unpleasant [75].

A second key consideration regarding disorder research is the source of the disorder measure [30, 72]. Researchers investigating disorder have traditionally incorporated disorder measures into surveys [70]. However, the growth of 'Big Data' and information technology [76] has made it easier to construct disorder measures from administrative data or remote imagery using geospatial tools.

Neither geospatially developed measures of disorder nor survey measures of disorder are uniformly superior at capturing constructs of researcher interest [64]. Geospatial measures fail to capture between-individual differences in perceptions that may be relevant for physical activity choices [77]. Any given subject's physical activity choices reflect that subject's perceptions of disorder, which are better captured by survey than a geospatial measure. However, self-reported measures can be problematic as well, as social desirability and other response biases may affect measure validity or induce 'same-source bias' (also known as recall bias) when analyzed with respect to other self-reported measures [78-80].

Physical Activity

Physical activity, defined as bodily movement produced by skeletal muscles that results in energy expenditure, has no universally agreed-upon best practice measure [81, 82]. Recreational physical activity taking place in a neighborhood (e.g., going on a walk) may be the form of activity most likely to be affected by neighborhood conditions, but such activity can only be identified if activities' purposes have been recorded. However, self-reported activity data is subject to artifacts due to poor recall and between-individual differences in self-assessment [82]. These artifacts may substantially bias findings [83]. By contrast, accelerometer measures are exempt from biases arising due to self-report, but do not capture the activity's purpose, are vulnerable to measurement error due to improper device placement and external vibration, and fail to capture activity that produce limited torso movement, such as cycling [82, 84]. These limitations in activity assessment may result in biased estimtes of the causes or effects of activity if whether the device is charged properly, worn properly, and returned to the study team in time for analysis is differential by activity undertaken during the study period [85, 86].

This Review

We conducted a systematic review in order to determine the state of the literature regarding neighborhood disorder and physical activity among adults and to make recommendations for future research on this topic. We focused on studies investigating differences in activity or walking levels in adult populations as predicted by disorder in home neighborhood.

Methods

Search strategy

We conducted a systematic search to identify studies published prior to August 2015 in four electronic databases representing the medical and transportation literatures: PubMed, TRID, PsycInfo, and Embase. Figure 1 provides an overview of the search protocol, following the protocol specified by the

PRISMA statement [87]. Following PRISMA guidelines, we customized search terms for each database. Search terms for each database are given in Appendix 1; briefly, each query searched for a phrase indicating walking/physical activity, a phrase indicating disorder/safety, and a phrase indicating a neighborhood or contextual focus.

Eligibility

We selected only quantitative, original research articles studying the potential that neighborhood disorder influences physical activity among adults aged 18 or older. To maximize comparability of studies, included studies used people rather than places as the unit of analysis. For example, a study of 100 people relating each person's neighborhood disorder to each person's activity level would be included, but a study of 100 parks comparing each park's disorder to the number of people walking there would be excluded.

Some researchers have used measures that capture elements of abandonment and untidiness do but do not use the term 'physical disorder' to describe the measure, preferring 'aesthetics' [88] or 'incivilities' [89] instead. For the purposes of the review, we considered measures for which the majority of items reflected visible indicators of neighborhood deterioration or neglect, such as poor housing maintenance, presence of litter, or vacant lots, to constitute disorder.

We emphasize, however, that not all aesthetics measures incorporate indicators of disorder. In particular, many neighborhood audit tools, particularly those focused on the determinants of walking, feature aesthetic subscales assessing pleasing features that may be encountered while walking, such as interesting architecture or beautiful views [90]. In particular, the widely used Neighborhood Environment Walkability Survey (NEWS) audit scale includes an aesthetics subscale [91] that does not include questions assessing deterioration; hence, we excluded many papers whose only measure of neighborhood aesthetics was derived from NEWS.

In addition to a measure of neighborhood disorder, selected papers were required to have a measure of physical activity and to present a quantitative measure of the association of the neighborhood disorder measure and the activity measure. Detailed exclusion criteria, including some examples of excluded papers, for each phase of the review are given in Appendix 2

Results

Selected Studies

Twenty-eight papers met the full inclusion criteria [32, 64, 65, 88, 89, 92-113], including all three papers we had *a priori* anticipated identifying [64, 65, 110]. Characteristics of each study are given in Table 1. Slightly less than half the studies (N=13, 46%) used only study subject reported measures of disorder, half (N=14, 50%) used independently observed measures, mostly neighborhood audits, and one used both [64]. With the exception of one study assessing active commuting [99] and another assessing jogging [106], all studies assessed at least one of leisure-time physical activity (N=7, 25%), total physical activity (N=9, 32%), or walking (N=12, 43%). With two exceptions [93, 113], all activity measures were self-reported. Sample size ranged from 42 to over 31,000; both studies with more than 10,000 subjects were conducted in the Netherlands [104, 108]. Most of the studies were in North America, Australia, or Europe. Figure 2 is a world map identifying the settings in which studies included in this review took place. Two papers reported subgroup analyses of the same intervention study [88, 112], four involved two waves of data collection [89, 97, 101, 103], and the rest were cross-sectional .

Disorder Measures

In the selected studies, there were substantial differences in disorder measures, ranging from administrative reports of street cleanliness [110] to interviewer observations of housing conditions averaged across a district [98]. In general, the studies whose disorder measures were limited to less severe indicators, such as litter-only measures, did not identify associations between neighborhood disorder conditions and activity levels [92, 97, 100, 103, 110, 113]. However, studies whose disorder measures incorporated more severe aspects of deterioration were inconsistent in their findings. For example, Laraia, et al. found that an audit-based disorder measure incorporating building conditions and burned or vacant properties in addition to litter and graffiti was associated with less vigorous physical activity among women in the Raleigh, NC area [89]. However, Hoehner, et al. found an audit-based disorder measure including drug paraphernalia, abandoned cars, and broken windows to be associated with *more* transport-related walking [64].

Some researchers have expressed concerns that using subject-reported measures of both neighborhood exposures and activity may result in 'same-source' bias, wherein correlations in measurement error between exposure and outcome measures results in artificially inflated observed associations [73, 114]. For example, a trend wherein subjects justify their lack of activity by describing their neighborhood as more disordered than it truly is, disorder might appear to be a barrier when it truly is not. However, though about half the studies used both subject-reported measures of disorder and subject-reported measures of activity, there was no evidence that observing an association between disorder and activity was more common when subject-reported measures were used for both constructs (Table 1). While this lack of consistency between measure source and direction of finding does not preclude the possibility of same-source artifacts [79], it does minimize concerns that systematic same-source bias is responsible for all appearance of significant associations.

Activity Measures

In general, studies that measured walking for exercise observed stronger associations than studies measuring physical activity in general or walking for transport. For example, Heinreich, et al.'s study of public housing residents in Kansas City, MO, found that more disorder-related incivilities were associated with less frequent walking for exercise [102], and Mendes de Leon, et. al's study of community-dwelling older adults in Chicago found less walking for exercise where more disorder was present. By contrast, most studies assessing walking for transport, which may be less discretionary, failed to find significant associations [97, 101, 110] or found *more* walking for transport in the presence of more disorder [64, 94, 107]. Several authors whose studies observed the latter association speculated that moderate indictors of disorder such as discarded cigarette butts might themselves be consequences of prevalent walking [64, 94].

Study Populations

The samples used across the identified studies were substantially heterogeneous, ranging from large samples representative of whole countries [98, 104, 105] to selected randomized trial participants [88, 103, 112]. In general, associations between disorder and activity were stronger among women and older adult populations. For example, Cunningham-Myrie et al's study representative of the population of Jamaica, found that disorder was more strongly associated with less activity among women than among men [98]. Similarly, Kwarteng et al. found that disorder was not significantly linked to activity levels in their full sample, but that more disorder was associated with less activity in an analysis limited to older adults [109]. Three of the five studies limited to older adults found significant negative associations between disorder and activity [32, 95, 105, 107].

Analytic Methodology

Nearly all studies acknowledged the potential for geographic clustering of neighborhood measure to bias results. Twelve (43%) used mixed models, three (11%) used cluster robust variance estimators, and two (7%) used generalized estimating equations. Most studies that did not analytically account for clustering explained why clustering was not a relevant source of bias (e.g. non-overlapping subject neighborhoods), suggesting that clustering as a potential threat to validity is well recognized.

Several studies included covariates derived from aggregated measures that, if measured inaccurately and non-differentially at the individual level, may result in biases away from the null for the measures' coefficients [115]; however, none used a disorder measure vulnerable to this source of bias.

Covariates

Every study controlled for individual sex, either statistically or by restrictions intrinsic to study design, and all but one controlled for individual age (Table 2). Many also controlled for education (n=20, 71%) and household income (n=13, 46%). Other common covariates included Race/Ethnicity (n=6), presence of children in the household (n=6), marital status (n=7), co-morbidities (n=6, particularly in studies of older adults), and access to a car (n=6). Only one controlled for disability status, possibly due to disability's uncommonness in the general population. Several studies controlled for environmental covariates as well; however, there were no clear patterns of environmental covariates controlled for. There was not a clear pattern between controlled covariates and estimated relationship between disorder and activity.

Effect modifiers

Not all studies performed subgroup analyses or other tests of effect modification (Table 3). Five studies (18%) explored effect modification by gender, and four (14%) explored effect modification by age. Several other effect modifiers were explored, including presence of children in the household [111], pregnancy status among women of childbearing age [89], and neighborhood poverty [110]. Several studies were restricted to populations explored in subgroup analyses in other studies, including older adults [32, 107] and women [89, 96, 97].

Discussion

Our systematic review criteria identified twenty-eight studies exploring physical disorder in the home neighborhood as an influence on physical activity. The populations studied were mostly urban and suburban residents of the industrialized world, perhaps because disorder is typically conceptualized as a phenomenon of urban decline [31, 54, 70], or perhaps because most research is conducted by research teams based in the industrialized world. All but one of these studies was published in the past decade, including eleven since 2013, suggesting that interest in disorder as a determinant of activity is growing in line with the overall growth of population health research.

Taken as a whole, the evidence base provided by these studies does not indicate that disorder consistently inhibits activity across the general population. However, there were indications that disorder may affect activity in selected contexts, particularly where disorder is more extreme and among women and older adults, suggesting areas for future research. Heterogeneity in measures of disorder and activity in the present studies precluded a quantifying synthesis of findings using meta-analysis or meta-regression.

Methodological Issues in Current Evidence Base and Directions for Future Research

This review identified several methodological issues within research on disorder and activity. First, no single measure of disorder was consistently used across more than a handful of studies. This lack of measure comparability precludes direct comparison between studies or quantitative synthesis of the literature. A recent narrative review catalogs twenty-six indicators that have been used in physical disorder measures [70]. Development and validation of a unified and standardized physical disorder measure for research use would be valuable, not only for research on physical activity, but also for research on mental health, crime, and other outcomes thought to be influenced by disorder.

Second, while theoretical frameworks typically posit that walking is the form of activity most likely to be affected by disorder, many studies used measures of all physical activity. Furthermore, this pattern may grow in the future if accelerometer measures are seen as superior to self-report measures [116] (though self-report measures have defenders [117]). Future research may benefit from use of sophisticated algorithms combining GPS and accelerometer readings to determine modes of activity undertaken [118, 119].

A third methodological issue common to many of the identified studies is the threat of reverse causation artifacts. Cross-sectional studies assessing walking for transport using measures of minor indicators of disorder may be particularly vulnerable to this weakness. Walking for transport is more probable in contexts where errands can more easily be run on foot. Such contexts generally have a larger pedestrian population and in general, more pedestrians create more litter [110]. Thus, while litter is an indicator of disorder, it can also be an indicator of a more walkable location; this may account in part of the repeatedly observed finding that more disorder was associated with more walking, particularly for walking for transport [64, 94, 107]. Similarly, an in-depth ethnography of New York City graffiti artists and taggers revealed that a key motivation for their actions is social status among an audience. They therefore typically choose locations where their works will be observed [120], which suggests that contexts with more pedestrians are themselves better targets for graffiti.

In this light, any conclusions are substantially limited by the lack of longitudinal evidence. None of the identified studies had more than two waves of data, and most were cross-sectional. Though temporal ordering of exposure prior to outcome is generally considered a necessary condition for identifying a causal effect [121], cross-sectional studies are common in studies of contextual influences on activity [43, 122], in part because gathering disorder measures at multiple time points can be expensive. In other research areas, such as the study of the food environment, the use of longitudinal administrative datasets such as the National Establishment Time-Series (NETS) dataset [123] has provided the opportunity to include longitudinal contextual exposures. Future research on disorder might make use of pre-existing administrative records of vacancies or clean streets report cards [110] to obtain longitudinal disorder measures.

Fifth, relatively few of the identified studies included subgroup analyses focused on populations most likely to be affected by disorder. With the exception of gender and age, many factors analyzed as potential confounders in these studies, including presence of children in the household, employment status, and access to alternative transportation, might better be conceptualized as potential effect modifiers. For example, the transport mode choice among ‘captive walkers’ who lack access to a car is typically between walking and transit [124]; such choices may both involve some walking and may be less affected by disorder than the choice by someone with more transport alternatives.

Finally, as in many studies of neighborhood exposures, control for environmental confounding and self-selection (i.e. the choice to live in neighborhoods supporting one’s preferred lifestyle) in these studies might be problematic [125]. One approach to strengthening causal inference in neighborhood studies is to study only those who have moved [126]. Alternately, incorporating evidence from alternate study designs, including discrete choice analysis [127] and mixed-method designs incorporating walk-along interviews or other qualitative evidence [37, 128], may provide important complementary evidence to better understand if disorder acts as a barrier to walking and other outdoor physical activity.

Conclusions

The evidence base systematically compiled here suggests that it is likely that disorder modestly discourages outdoor physical activity in some but not all adults. Further research using longitudinal designs and focusing on those most likely to be vulnerable to inhibitory effects of disorder, including older adults, would provide valuable insight into where and when interventions to decrease disorder may be most likely to be effective at increasing physical activity levels.

Figures

Figure 2.1. PRISMA Flow Diagram representing the selection of articles for this systematic review

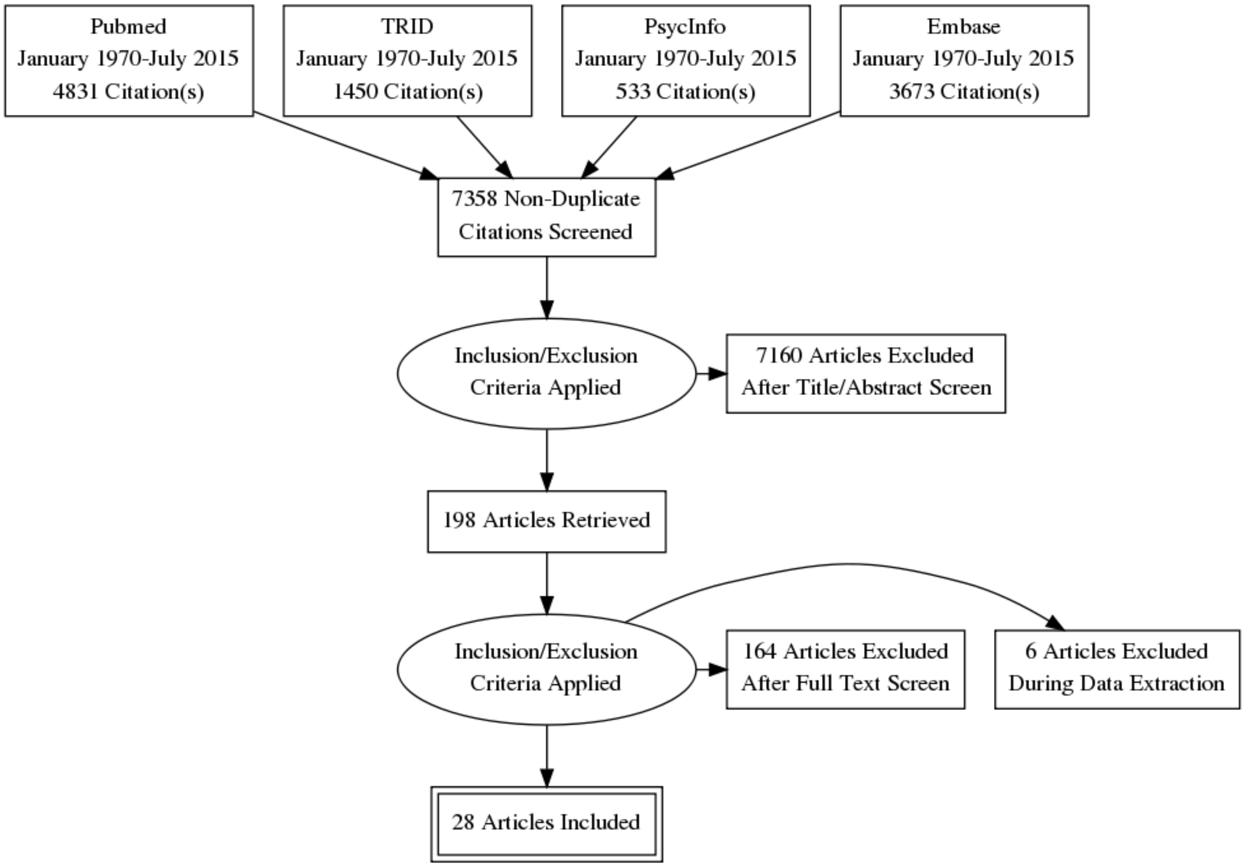


Figure 2.2. Red markers on the map below indicate the locations where studies included in this review were conducted.

Tables

Table 2.1. Descriptive characteristics of the 28 studies identified in this systematic review

First Author	Year	Journal	Title	Exposures	Exposure Detail	Accelerometer	Survey	Outcomes	Analyzed Sample	Geography
Adams, Emma	2013	International Journal of Behavioral Nutrition and Physical Activity	Correlates of walking and cycling for transport and recreation: factor structure, reliability and behavioural associations of the perceptions of the environment in the neighbourhood scale (PENS)	Perceived	Litter	No	IPAQ	Leisure Walking, Walking for Transport	2937	3 cities in Britian
Cain, Kelli L	2014	Social Science & Medicine	Contribution of streetscape audits to explanation of physical activity in four age groups based on the Microscale Audit of Pedestrian Streetscapes (MAPS)	Objective	Aesthetics scale including disorder	Yes	GPAQ/ CHAMPS	LTPA, TA	2022	3 cities in US
Caspi, Caitlin E.	2013	Social Science & Medicine	The social environment and walking behavior among low-income housing residents	Objective	Disorder Audit	No	IPAQ	Leisure Walking, Walking for Transport	729	Boston (low-income housing)
Cerin, Ester	2013	Preventive Medicine	Objectively-measured neighborhood environments and leisure-time physical activity in Chinese urban elders	Objective	Disorder Audit	No	IPAQ	LTPA	484	Hong Kong
Cerin, Ester	2013	Journal of Urban Health	Walking for Recreation and Perceptions of the Neighborhood Environment in Older Chinese Urban Dwellers	Perceived	Hong Kong-specific disorder measaure	No	IPAQ	Leisure Walking	484	Hong Kong

Cleland, Verity	2008	Preventive Medicine	Are Perceptions of the Physical and Social Environment Associated with Mothers' Walking for Leisure and for Transport? A Longitudinal Study	Perceived	Litter	No	Custom	Leisure Walking, Walking for Transport	375	Melbourne, AU
Cleland, Verity	2010	Social Science & Medicine	Individual, social and environmental correlates of physical activity among women living in socioeconomically disadvantaged neighbourhoods	Perceived	Aesthetics as disorder	No	IPAQ-L	LTPA, Transport Activity	3508	Victoria, AU
Cunningham-Myrie, Colette A	2015	Journal of Clinical Epidemiology	Associations between neighborhood effects and physical activity, obesity, and diabetes: The Jamaica Health and Lifestyle Survey 2008	Objective	Aggregated Interviewer Measure	No	Custom	Total Activity	2848	Jamaica
Feuillet, Thierry	2015	International Journal of Health Geographics	Spatial heterogeneity of the relationships between environmental characteristics and active commuting: towards a locally varying social ecological model	Perceived	Aesthetics as disorder	No	Custom	Active Commuting	4164	Paris Metro, FR
Florindo, Alex A	2013	Journal of Physical Activity and Health	Physical activity and its relationship with perceived environment among adults living in a region of low socioeconomic level	Perceived	Litter	No	IPAQ	Total Activity	890	São Paolo, BR
Foster, Sarah	2014	Environment and Behavior	Does fear of crime discourage walkers? A social-ecological exploration of fear as a deterrent to walking.	Perceived	Disorder	No	NPAQ	Leisure Walking, Walking for Transport	485	Perth, AU
Heinrich, Katie M.	2007	International Journal of Behavioral Nutrition and Physical Activity	Associations between the built environment and physical activity in public housing residents	Objective	Disorder Audit	No	Custom	Walking, Total Activity	452	Kansas City, US
Hoehner, Christine M	2005	American Journal of Preventive Medicine	Perceived and objective environmental measures and physical activity among urban adults	Perceived, Objective	Many measures	No	IPAQ	LTPA, TA	1053	St. Louis, MO & Savannah, GA

Jalaludin, Bin	2012	BMC Public Health	A pre-and-post study of an urban renewal program in a socially disadvantaged neighbourhood in Sydney, Australia	Perceived	Litter, Graffiti	No	Custom?	Total Activity	42	Sydney, AU
Jongeneel-Grimen, Birthe	2014	Health & Place	The relationship between physical activity and the living environment: A multi-level analyses focusing on changes over time in environmental factors	Perceived	Aggregated Subject Measure	No	Custom	Total Activity	25309 and 31783	Netherlands
Kamphuis, Carlijn Bm	2009	International Journal of Behavioral Nutrition and Physical Activity	Socioeconomic differences in lack of recreational walking among older adults: the role of neighbourhood and individual factors	Perceived	Disorder	No	Custom	Leisure Walking	1994	Eindhoven, NL
Karusisi, Noëlla	2012	Preventive Medicine	Multiple dimensions of residential environments, neighborhood experiences, and jogging behavior in the RECORD Study	Objective	Maintenance	No	Custom	Jogging	7290	Paris Metro, FR
King, Diane	2008	Journal of Aging and Physical Activity	Neighborhood and individual factors in activity in older adults: results from the neighborhood and senior health study	Objective	Disorder Audit	No	CHAMPS	Total Activity	190	Denver, CO
Kramer, Daniëlle	2013	International Journal of Behavioral Nutrition and Physical Activity	Neighbourhood safety and leisure-time physical activity among Dutch adults: a multilevel perspective	Perceived	Aggregated Subject Measure (same as Jongeneel-Grimen)	No	SQUASH	LTPA	20046	Netherlands
Kwarteng, Jamila L.	2014	Journal of Public Health	Associations between observed neighborhood characteristics and physical activity: findings from a multiethnic urban community	Objective	Disorder Audit	No	IPAQ	Total Activity	919	Detroit, MI
Laraia, Barbara	2007	Journal of Urban Health	Neighborhood Factors Associated with Physical Activity and Adequacy of Weight Gain During Pregnancy	Objective	Disorder Audit	No	Custom	LTPA	703	Wake County, NC

Lovasi, Gina S.	2012	American Journal of Preventive Medicine	Body mass index, safety hazards, and neighborhood attractiveness	Objective	Litter	No	Custom	Walking for Transport	8034	New York, NY
Mendes de Leon, Carlos	2009	Journal of Aging and Health	Neighborhood Social Cohesion and Disorder in Relation to Walking in Community-Dwelling Older Adults:A Multilevel Analysis	Perceived	Aggregated Subject	No	Custom	Walking for Exercise, Other Walking	4317	Chicago, IL
Miles, Rebecca	2008	American Journal of Preventive Medicine	Neighborhood Disorder, Perceived Safety, and Readiness to Encourage Use of Local Playgrounds	Objective	Disorder Audit	No	Custom	LTPA	2123	7 European cities
Oh, April Y.	2010	Journal of Physical Activity and Health	Effects of perceived and objective neighborhood crime on walking frequency among midlife African American women in a home-based walking intervention	Objective	Disorder Crimes	No	Custom	Walking	148	Chicago, IL
Ross, Catherine E.	2001	Journal of Health and Social Behavior	Neighborhood disadvantage, disorder, and health	Perceived	Physical and Social	No	Custom	Walking	2482	Illinois
Strath, Scott J.	2012	International Journal of Behavioral Nutrition and Physical Activity	Measured and perceived environmental characteristics are related to accelerometer defined physical activity in older adults	Objective	Disorder Audit	Yes	--	Total Activity	148	Wisconsin, US
Zenk, Shannon N	2006	Health Education and Behavior	Neighborhood Environment and Adherence to a Walking Intervention in African American Women	Objective	Aesthetics as disorder	No	Custom	Walking	252	Chicago Metro

Table 2.2. Environmental and Individual-level Confounders Controlled for, by study

First Author, Year	Age	Sex	Race/Ethnicity	Urban	Household Size	Children	Marital status	Occupational status	Unemployed	Education	Household Income	Intervention	Smoking	BMI	Disability Status	Neighborhood Tenure	Residential Tenure	Comorbidity	Season	Country of Birth	Current Pregnancy	Renter	Financial Strain	Sunshine During Study Period	Self-efficacy	Access to a car	Access to a Bike	Household Wealth	Recent Crime Victimization	Parking at work	Transit pass
Ross, 2001	X	X	X	X		X	X	X	X	X	X																				
Hoehner, 2005	X	X									X																				
Zenk, 2006	X	R		X						X	X	X																			
Laraia, 2007	X	R	X			X	X			X	X		X	X																	
Heinrich, 2007		X									R																				
Cleland, 2008	X	R				X	X			X																					
Miles, 2008	X	X			X	X	X			X					X	X	X														
King, 2008	X	X									X							X													
Kamphuis, 2009	X	X								X	X																				
Mendes de Leon, 2009	X	X					X			X	X					X		X	X												
Cleland, 2010	X	R				X	X		X	X			X	X				X		X	X										
Oh, 2010	X	R								X	X	X																			
Karusisi, 2012	X	X						X		X												X	X	X	X						
Strath, 2012	X	X												X			X									X					
Jalaludin, 2012																															
Lovasi, 2012	X	X	X			X	X		X	X	X							X		X											
Cerin, 2013a	X	X								X																					
Caspi, 2013	X	X							X									X		X						X					
Adams, 2013	X	X	X	X						X												X				X	X				
Cerin, 2013b	X	X								X																					
Kramer, 2013	X	X			X					X	X																	X			
Florindo, 2013	X	X								X							X									X					
Foster, 2014	X	X								X	X						X									X			X		
Cain, 2014	X	X	X							X								X													
Kwarteng, 2014	X	X	X								X																				
Jongeneel-Grimen, 2014	X	X			X				X	X	X																				
Feuillet, 2015	X	X								X																X	X			X	X
Cunningham-Myrie, 2015	X	X																													
Total	26	22	6	3	3	6	7	2	5	20	13	2	2	3	1	2	4	6	1	3	1	2	1	1	1	6	2	1	1	1	1

X: Controlled for statistically; R: controlled for by restriction.

Table 2.3. Effect modifiers investigated, by study

First Author	Urban-ness	Pregancy Status	Gender	Presence of Children	Ability to walk	Neighborhood poverty	Age	Educational Attainment	Parks
Ross, 2001									
Hoehner, 2005									
Zenk, 2006	X								
Laraia, 2007		X							
Heinrich, 2007									
Cleland, 2008									
Miles, 2008			X	X					
King, 2008									
Kamphuis, 2009									
Mendes de Leon, 2009					X				
Cleland, 2010									
Oh, 2010									
Karusisi, 2012									
Strath, 2012									
Jalaludin, 2012									
Lovasi, 2012			X			X			
Cerin, 2013a			X				X	X	
Caspi, 2013									
Adams, 2013									
Cerin, 2013b									X
Kramer, 2013									
Florindo, 2013									
Foster, 2014									
Cain, 2014							X		
Kwarteng, 2014							X		
Jongeneel-Grimen, 2014			X				X		
Feuillet, 2015									
Cunningham-Myrie, 2015			X						
Total	1	1	5	1	1	1	4	1	1

Chapter 3: Neighborhood Disorder and Physical Activity among Older Adults: A Longitudinal Study

Abstract

Background

Neighborhood physical disorder may inhibit outdoor physical activity, particularly in older adult populations. However, most previous studies of the association between neighborhood disorder and physical activity including older adults have been cross-sectional.

Methods

We examined the relationship between neighborhood disorder and physical activity in a 3-wave longitudinal study using a sample of 3,497 adults aged 65-75 at baseline weighted to be representative of the older adult population of New York City. Our measure of neighborhood disorder was constructed from a virtual street audit using Google Street View imagery. We used longitudinal mixed linear regression controlling for baseline age, sex, race/ethnicity, baseline physical functioning, educational attainment, neighborhood social cohesion, pedestrian injury rate, and walkability and to estimate the effect of disorder on activity score.

Results

In multivariable mixed regression models accounting for individual and neighborhood factors, for missing data, and for loss to follow-up, each standard deviation increase in neighborhood disorder was associated with an estimated 3.0 units (95% CI: 1.9, 4.2) lower PASE score at baseline, or the equivalent of about 10 minutes of walking per day. However, there was no physical disorder was not related to changes in PASE score over two years of follow-up.

Conclusions

In this older adult population, residents of more disordered neighborhoods were on average less active at baseline. Disorder was not associated with decreases in physical activity over time.

Introduction

Physical activity – bodily movement produced by skeletal muscles that results in energy expenditure [129] – prevents or delays onset of many negative physical and mental health outcomes common among older adults, including but not limited to diabetes, cardiovascular disease, breast and colon cancer, arthritis, dementia, disability, falls, loss of independence, and frailty.[1, 2, 9, 11] Despite these protective benefits, fewer than one in four Americans age 65 and over meet recommended physical activity guidelines [14] and nearly a third report engaging in no leisure-time physical activity in the past month.[130]

Physical disorder -- the deterioration of urban landscapes [30] -- may be an important and modifiable barrier to physical activity, particularly walking, among older adults.[67, 131-133] However, the quantitative evidence base that physical disorder acts as a such a barrier is limited. As described in Chapter 2, older adults may be particularly vulnerable to the effects of disorder, but only a few studies have examined disorder in relation to activity among older adults [32, 95, 105, 107, 109, 113, 134]. Furthermore, to the best of our knowledge, all prior studies of neighborhood disorder and physical activity among older adults have been cross-sectional. By assessing neighborhood exposure before changes in activity, longitudinal analyses can establish stronger evidence for causal relationship than are available cross-sectionally [135].

By contrast with cross-sectional data, longitudinal data affords two additional analytic strategies with which to assess effects of an exposure. One strategy is to investigate within-individual change when the exposure of interest changes. A complementary strategy is to analyze between individual differences in change over time, regardless of change[136]. Because people don't change residences often, in many neighborhood studies, the second investigation type (i.e. exploring neighborhood context as a predictor of between-individual differences in change) may often be more appropriate.

In this study, we investigate the longitudinal relationship between neighborhood physical disorder and physical activity in community-dwelling older adults, focusing on the between-individual differences that arise with respect to different disorder levels. We hypothesize that disorder will discourage outdoor activity, such that those living among more disorder will have consistently lower activity levels across all three waves of follow-up.

Methods

Subjects and Setting

We used data from the New York City Neighborhood and Mental Health in the Elderly Study (NYCNAMES-II), a longitudinal study of 3,497 residents of New York City aged 65-75 at baseline. Sampling and recruitment for NYCNAMES-II has been described previously [45]. Subjects were initially recruited by phone, starting in May 2011 and continuing through November 2011. The initial 219 (6.2%) surveys were completed by interviewers at the New York Academy of Medicine; the subsequent 3,281 were completed by Abt-SRBI, a survey research firm. During recruitment, a total of 39,792 telephone numbers were selected from a list of telephone numbers purchased from InfoUSA, a data broker that sells geographically targeted lists of individual's phone numbers and basic demographic characteristics primarily for sales and marketing purposes. The response rate, calculated as screen-outs (i.e. calls that where a household was contacted and the interviewer was able to ascertain that no older adults lived in the household) plus completed interviews divided by all phone numbers selected from the list, was 18%. The co-operation rate, calculated as screenouts plus interviews over screenouts, interviews and declines, was 31%. All surveys were conducted in English or Spanish. Recruitment was terminated when the target 3,500 interviews had been completed; however, during data cleaning, three interviews were determined to be repeats of previously interviewed subjects. These data were discarded, resulting in the final sample of 3,497 subjects. Each subject was followed up by telephone once in summer or fall 2012 and once in summer or fall 2013, resulting in three waves of data collection in all. Each wave's survey, including items not used in this analysis, took approximately 30 minutes. Subjects were paid $10 for participating, either in the form of a CVS gift card as a check.

Phone numbers were selected using a complex survey design whereby census tracts were classified into sixteen strata as follows: first, tracts were cross-classified by three variables: racial-ethnic mix (>75% Non-Hispanic Black, >75% Non-Hispanic White, > 60% Latino, or racial/ethnically mixed), median household income (either above or at/below the median of the median income distribution by tract) and walkability (either above or at/below the median of the walkability distribution). To identify strata, median household income and racial/ethnic mix measures were taken from the United States Census and walkability was taken from a previously validated metric incorporating Census measures and tax lot data [137]. Six tracts were selected from each of the sixteen strata, with probability proportional to number of residents aged 65-75 within the tract. Four additional tracts chosen randomly so that one hundred tracts were selected in all. Subject sample weights were calculated as the probability of a census tract being selected from all tracts in New York City multiplied by the probability of the household being selected from the number of households containing an adult aged 65-75 estimated by the 2010 Census multiplied by the probability of the individual being selected from the number of adults aged 65-75 in the contacted household. Final survey weights were the raked to New York City population estimates from the 2006-2010 American Community Survey for gender and race/ethnicity and from 2010 Census estimates for educational attainment and borough of residence.

Individual Measures

Subjects self-reported sex, age, educational attainment, race/ethnicity, health status, and income. For analysis, we categorized age at baseline as 65-68, 69-71, and 72-75 and categorized household income as <$20,000, $20,000-39,999, $40,000-79,999, and ≥$80,000. Education levels were reported as less than high school graduate, high school graduate, some college, or college graduate; health statuses were reported as excellent, good, fair, or poor. To maintain a balance of individuals in each racial/ethnic group, we categorized race/ethnicity as Non-Hispanic Black, Non-Hispanic White, Hispanic, and Other.

We assessed past-week physical activity using the Physical Activity Scale for the Elderly (PASE).[138] PASE assesses frequency (days/week) and duration (minutes/bout) of engagement in a number of activities older adults report engaging in, such as recreational sports, gardening, housework, and caring for others. PASE survey responses can be compiled into a single score, which is not directly interpretable in terms of energy expenditure [138]. Nonetheless, PASE has been validated in several older adult populations [139-142], and has been shown to have good correlation (r= 0.68) with doubly-labeled water assessment of physical activity [141]. PASE assesses past-week rather than ‘usual’ or ‘typical’ activity not only to minimize survey burden but also to avoid non-differential measurement error that might arise if older adults who are more physically active retain cognitive function longer, as several studies have suggested [8, 143, 144]. One validation study showed PASE to be more strongly correlated with 6-minute test performance than two comparable self-reported older adult physical activity instruments that assessed ‘typical’ activity [140].

Subject physical function was assessed using the Functional Status Questionnaire (FSQ), a 34-item survey instrument designed to monitor for clinically meaningful changes in function [145]. Because NYCNAMES used other instruments to assess mental health and social engagement, only the 9-item physical function sub-scale was administered to NYCNAMES-II subjects. This subscale includes three items that assess basic activities of daily living comprising 1) self-care (e.g. eating and bathing), 2) moving between bed and a chair and 3) walking indoors. The remaining six items assess intermediate activities of daily living, comprising 1) walking several blocks outdoors, 2) walking at least one block or climbing a flight of stairs, 3) doing light housework such as cleaning, 4) running errands such as grocery shopping, 5) driving a car or using public transportation and 6) doing vigorous activities such as running or strenuous sports. Each item has four response categories, comprising 1) usually engaging in the activity with no difficulty, 2) usually engaging with some difficulty, 3) usually engaging with much difficulty 4) not engaging because of health, and 5) not engaging for some other reason. Responses from the nine items were transformed into a score ranging from 0 to 100 for basic activities of daily living and intermediate activities of daily living. At baseline, the sample weighted mean score for basic activities of daily living was 92.8 and 70.7% of subjects had no difficulties performing any basic activity of daily living. The mean score on the intermediate activities of daily living scale score was 80.7 and 40.6% of subjects had no difficulties performing any of the activities.

The subject’s perception of neighborhood social cohesion was assessed using an 8-item scale adapted from an instrument developed by the Project for Human Development in Chicago Neighborhoods [146]. Specifically, subjects were asked about the strength of their agreement with the following statements using a 4-point Likert-type scale: 1) if there are problems around your neighborhood, your neighbors get together to with it, 2) your neighborhood is close-knit, 3) people in your neighborhood generally don’t get along with each other, 4) if you had to borrow $30 in an emergency, you could borrow it from a neighbor, 5) neighbors will keep their eyes open for possible trouble to your place, 6) people in your neighborhood can be trusted, 7) people in your neighborhood don't share the same values, and 8) if you were sick, you could count on your neighbors to shop groceries for you. The overall scale had a Cronbach's alpha of 0.747.

Neighborhood measures

Researchers seeking to study neighborhood disorder while aiming to avoid subject-response and 'same-source' biases (sometimes also termed 'recall bias') often estimate disorder using systematic social observations, wherein trained observers visit neighborhoods and systematically record indicators of physical disorder [109, 147]. These indicators can then be combined to develop an objective measure of neighborhood disorder. For this work, we used a neighborhood disorder measure inspired by systematic social observation but incorporating information technology to improve efficiency. Specifically, we assessed neighborhood using the novel but validated 'virtual street audit' technique in which trained observers record indicators of disorder observed in Google Street View imagery using a computerized system designed to improve reliability and efficiency of virtual street audits [148].

To collect data, trained virtual street auditors used imagery from Google Street View whose initial image capture occurred between August 2007 and October 2011 to assess 532 block faces across New York City for nine indicators of disorder including litter, graffiti, and buildings that appear to be abandoned. Individual items showed kappa scores ranging from 0.34 (for presence of empty alcohol bottles) to 0.80 (for presence of apparently abandoned buildings). Those indicators were then combined using a 2-parameter item response theory model to construct a single disorder scale, which had an internal consistency reliability of 0.93. We used kriging, a geospatial modeling technique that incorporates spatial covariance with distance-weighted measurements [149] to provide an estimate of disorder, with confidence levels, at any point in New York City [150, 151]. We then computed estimates at every vertex of a 100m x 100m grid over the land area of the city and used ArcGIS to compute the mean of the disorder estimates at grid points that fell within each subject's network buffer. Those mean values constituted our estimates of subjects' neighborhood disorder levels.

Neighborhood walkability, defined as the degree to which the neighborhood environment supports pedestrian activity, may confound the relationship between neighborhood disorder and walking. For example, because neighborhoods developed in eras before automobile ownership was widespread have more support for pedestrian activity, yet are also more often disadvantaged, then disorder may impede activity but not be associated with overall activity levels across all neighborhoods. To account for differences in walkability between neighborhoods, we used a validated walkability metric previously described in detail elsewhere [137]. In this measure, the total walkability score is the sum of z-scores of five measures derived from urban planning literature: 1) residential population density, 2) land use mix, 3) intersection density, 4) retail floor area ratio, and 5) subway stop density. This measure has previously been shown to predict BMI [50], engagement in active transport [152], and total physical activity as reported by accelerometer [153].

Finally, our neighborhood pedestrian risk measure was calculated as the density of unique pedestrian-motorist collisions resulting in an injury or fatality to the pedestrian in 2010. These data were abstracted from police records by the New York State Department of Transportation and have been used in prior analyses of pedestrian collisions and influences on physical activity [110, 154, 155].

Each subject reported his or her home address at each of the three waves. We geocoded these addresses to identify the geographic coordinates of the subject's home (96% were geocoded to a rooftop; the remainder were assigned to the age 65-74-population weighted centroid of the reported ZIP code) For each subject, we defined the residential neighborhood as the land area reachable by city streets within ¼ km of the geocoded home location, an area referred to as a ¼ km network buffer and frequently used in neighborhood research [110, 156, 157]. We then assigned a mean disorder, mean walkability, and pedestrian risk score to each participant's residential neighborhood at each wave of follow-up. Relatively few subjects (N=124, 6% of all those followed up) supplied new addresses over the three years of the study.

Missing data and sample weights

Slightly more than seventy percent (n= 2,455) of subjects contacted at baseline were successfully re-contacted at wave 2 and slight more than sixty-seven percent (n=2,355) subjects were re-contacted, including 332 who were not contacted at wave 2. This analysis used data as available from subjects successfully contacted in any of the three waves, such that there were 2,787 (79.7%) subjects with at least two time points of data. For the primary analysis, we used logistic regression to estimate covariate-adjusted probability of inclusion in any given wave, modeling the logit of inclusion in the wave as a linear function of gender, race/ethnicity, educational attainment, borough of residence, neighborhood disorder, neighborhood pedestrian injury rate, and self-reported health status. We then computed inverse probability of observation weights for each subject at each wave as described by Robins, et al. [158]. Censored observations at both wave 2 and wave 3 were more common for subjects with male sex, lower educational attainment and Hispanic ethnicity. Estimated odds of inclusion in Wave 2 and Wave 3 conditional on covariates are given in Appendix 1, Table 1, and kernel density-smoothed histograms of IPCW weights at each wave are given in Appendix 1, Table 2. We weighted our final analyses using the product of IPCW weights and baseline sample weights, such that results using the weights are representative of the population of non-institutionalized New York residents aged 65-75 according to the 2010 US Census (n=571,323). Appendix 2 presents the sample-weighted population demographics in each wave, illustrating that weights were successful in preserving demographic stability.

Relatively few responses were missing for the subjects who were followed up successfully. For example, no more than 1% of data was missing on any PASE component. Nonetheless, to account for possible bias due to missing data, we used five multiple imputations, computed using IVEWARE [159] to model missing covariates from all available covariates, for all missing responses. Following standard practice for multiply imputed data, we used Rubin's rules to calculate combined estimates from statistical models run on each of the 5 imputations individually.

Statistical Analysis

We explored the stability of PASE scores and functional status over three waves of data collection using spaghetti plots and by computing ICCs. To explore the demographic patterning of disorder and functional status, we computed mean disorder levels and median functional status scores, stratified by age, sex, educational attainment, and income.

After plotting disorder and PASE scores to check linearity assumptions, we modeled PASE as a continuous outcome in a longitudinal linear mixed-effects model. Specifically, we first fit a random intercept model predicting PASE score at each wave from neighborhood disorder in that wave, controlling for baseline age, sex, educational attainment and for time-varying perceived social cohesion, neighborhood walkability, and neighborhood pedestrian injury risk. Next, to investigate whether disorder affected the change in PASE score over time (e.g. if older adults living in more disordered neighborhoods encounter a sharper decline in activity) we fit a random intercept/random slope model with an interaction term between baseline disorder and wave. In this model, the interaction term is interpretable as the association between baseline disorder and change in PASE score over time. Finally, we fit a random intercept/random slope model with an interaction term between time-varying disorder and wave. In this model, the interaction term is interpretable as the association between change in disorder over time and change in PASE score.

Exploratory analyses suggested that neighborhood time-varying covariates were unaffected by neighborhood disorder status, so we did not explore complex methods to account for time-varying confounding affected by treatment such as inverse probability weighted marginal structure models [160].

Sensitivity Analyses

We performed four sensitivity analyses to test the robustness of our analysis to various assumptions. First, because we were concerned that past-week activity, as assessed by PASE, might be affected by weather and season, we explored the relationship between PASE and both days since June 1 (to test for seasonal effects) and mean past-week ‘feels like’ temperature using weather data for New York City downloaded from the Weather Underground website [161] and formulae for heat index [162] and wind chill [163] published by the National Weather Service. Second, to test the robustness of our conclusions to our choice of longitudinal modeling strategy, we repeated the primary analysis using generalized estimating equations rather than mixed models [164]. Third, to test the robustness of our results to our model for probability of inclusion in any given wave, we re-ran the main analysis using Abt-SRBI supplied sampling weights for each wave, which were raked to demographic targets as described above but by design could not account for disorder, walkability, or self-reported health status. Finally, since some subjects live in the same larger scale neighborhood areas, here operationalized as NYC Community Districts, we assessed the possibility of non-independence of observations between subjects by fitting a 3-level hierarchical model, clustering on subjects within community districts.

All analyses used R for Windows Version 3.1.0, including the ‘survey’ package to incorporate survey weights to account for sample design. We used the R ‘mitools’ package to combine estimates across imputations using Rubin’s rules [165].

Results

As compared to the older adult population of New York City, the full NYCNAMES-II baseline sample analysis was disproportionately female, well-educated, and non-Hispanic. However, the 2,455 older adults re-contacted in Wave 2 and the 2,355 re-contacted in Wave 3 were roughly comparable to all subjects enrolled at baseline subjects. Table 1 shows selected demographic characteristics of the full study population and the subset who were re-contacted at each wave of follow-up. Relatively few subjects moved during the follow-up period (0.9% at wave 2 and 2.0% at wave 3, n=103 overall).

Neighborhood disorder and functional status at baseline varied between subjects. On average, Hispanics, less educated individuals, and those with lower incomes encountered more disorder. Younger subjects, men, non-Hispanic whites, and those with higher incomes and more education had higher functional status (Table 2). Disorder was not strongly correlated with other neighborhood measures; more broadly, neighborhood measures were only weakly inter-correlated except for pedestrian injury risk and walkability (Table 3).

PASE scores were correlated within people across waves (ICC over 3 waves: 0.67, Figure 1). In spite of the age of the study population, mean PASE at Wave 3 was essentially unchanged from mean PASE at Wave 1 (80.6 vs 81.8), offering little evidence of activity decline across the population over this two-year period. Disorder and PASE score were weakly negatively correlated within each wave analyzed cross-sectionally (Spearman r=-0.13, -0.12, -0.13, Figure 2), though all negative correlations were significantly different from zero ($p<0.001$ for all three).

In a mixed longitudinal random intercept model using IPCW weights to account for censoring and controlling for baseline age, sex, race/ethnicity, educational attainment, and functional status, we observed a statistically significant negative association between disorder and PASE score. In this model, the intercept coefficient for disorder was estimated at -3.1 (95% CI: -4.6, -1.5), which can be interpreted as estimating that each one standard deviation increase in disorder was associated with an average of 3.1 unit lower PASE score, or about 9 minutes of walking/day, at baseline. Adding neighborhood social cohesion, walkability, and pedestrian risk to the models did not substantially alter the random intercept estimate (-3.1, 95% CI: -4.7, 1.5).

In a random slope model including an interaction term between wave and baseline disorder, the estimated coefficient for the disorder/time interaction term was -0.1 (95% CI: -0.9, 0.8), providing no evidence for differences in PASE trajectory by baseline disorder. Figure 3 compares predicted PASE score and 95% prediction intervals for 10,000 bootstrap simulations over three waves for a hypothetical subject living in a neighborhood one standard deviation more disordered than the mean neighborhood above the mean with the estimated PASE score for an equivalent subject living in a neighborhood one standard deviation less disordered than the mean. We observed that the deficit in physical activity that was present at baseline among those living in more disordered neighborhoods remained roughly constant over the following two waves.

Sensitivity Analyses

While there was minor seasonal and temperature variation in PASE score, particularly in the gardening item, past-week temperature was not strongly associated with overall PASE score. Analyses using mean past-week 'feels like' temperature and days since June 1 as covariates are detailed in Appendix 3.

Coefficient estimates computed using a GEE model rather than mixed model were similar to those computed in our primary analysis (Appendix 4, Table 4.1). Similarly, effect estimates computed using a mixed model with Abt-SRBI's sample weights rather than the weights we computed to incorporate health status and other covariates into the model for loss-to-follow up were similar to those computed in our primary analysis (Appendix 5, Table 5.1). Finally, mixed models clustering on community districts were also very similar to the primary analysis (Appendix 6, Table 6.1)

Discussion

In this longitudinal study of older adult residents of New York City, we observed the hypothesized inverse association between neighborhood physical disorder and physical activity. However, while individual subjects' activity levels fluctuated moderately, mean PASE scores for the whole cohort changed little over the two available years of follow-up and we observed no interaction between disorder and change in activity over those two years. Overall, the roughly 3 point PASE score differential per standard deviation of disorder remained constant across all three waves.

PASE scores are pure abstractions and cannot be directly translated in terms of energy expenditure. However, it is possible to conceptualize this 3 PASE points as achievable through roughly 9 minutes/day of walking [138, 142]. That is, if this difference were interpretable as an intervention effect such that removing disorder in a given subject's neighborhood would elevate that subject's activity level, then subjects who currently live in highly disordered neighborhoods and engage in no activity could meet the recommended 30 minutes/day of walking [166] if all litter, graffiti, deteriorated buildings, and bars on the windows in their neighborhoods were removed (equivalent to removing about 3.3 standard deviations of disorder) [150]. We caution, however, that this interpretation is purely a thought experiment to contextualize our estimated 3 PASE points per standard deviation of disorder; our data and study design do not support a causal interpretation of the disorder coefficient estimate. The assumptions that would be required to treat our observed association as an intervention effect are heroic. Not only did too few subjects move for a meaningful estimate of the effect of *changing* disorder exposure in this group [167] but also the causal identifiability assumptions of conditional exchangeability, treatment-variation irrelevance, and lack of interference between units [168] were all likely violated in some degree, and any true causal relationship between disorder and physical activity may be non-linear.

Evidence from walk-along interviews and other qualitative studies of older adults have contributed to the development of theory suggesting that neighborhood disorder may inhibit physical activity among older adults [37, 38, 169]. Several recent cross-sectional quantitative studies generally appear to support this theory, albeit with caveats [32, 107, 134]. Our study provides further support that disorder and activity are inversely associated after controlling for salient factors. We did not find evidence that living amidst disorder led to faster decline in activity levels, though with only two years of follow-up and less than 3% of subjects moving to new neighborhoods, our power to detect such effects was limited.

Unlike most prior studies of neighborhood influence on physical activity, our study used longitudinal data. Within epidemiology broadly, longitudinal studies have a stronger potential for causal interpretation than cross-sectional studies, because cross-sectional studies cannot ensure the exposure of interest precedes the outcome [135]. In particular, with respect to neighborhood influences, cross-sectional studies cannot rule out the limitation that differences in individual preferences may cause those who are more prone to physical activity to select neighborhoods more supportive of those activities, a process known as 'residential self-selection' [170]. Neighborhood researchers relying on cross-sectional data typically control statistically for covariates that might be related to residential self-selection, such as race, income, and education, but also ultimately acknowledge the limitation that self-selection could bias results [42]. In principle, longitudinal data gives investigators the opportunity to assess residential self-selection effects (e.g. by studying those who move) [171], but in practice cohorts not specifically selected from people changing residences (e.g. [126]) often have limited power to assess the effects of change in residential neighborhood characteristics. This was the case in our dataset, where only about three percent of subjects moved during the three years of the study, and as detailed below, it may be that residential self-selection affected the relationship between baseline disorder and physical activity

However, our longitudinal activity measures did allow us to identify changes in physical activity over time. Specifically, the modestly negative relationship we observed between elapsed time and PASE score (activity decreased an average of 0.5 PASE units per year on average, and that estimate that was sufficiently small and imprecise as to be compatible with no change occurring at all) did not appear to be differential by neighborhood disorder level.. Given that we observed no disparity in activity trajectory, there are four complementary explanations for how the presence of the baseline disparity might have arisen. The first is that consistent residual confounding is responsible for the observed consistent association at each wave. Such confounding would need to be independent of covariates in our model including age, educational attainment, income, functional status and race/ethnicity. Nonetheless, we cannot rule this possibility out. A second possible explanation is that residential self-selection is responsible for the emergence of the disparity – that is, on average, subjects selected neighborhoods fitting their activity preferences, and retained their age-specific preferred activity level across all waves of follow-up. A third possibility is that the critical period for neighborhood as an cause of activity norm is prior to age 65, the youngest age in our cohort, such that our subjects had already established physical activity norms suited to their neighborhoods prior to recruitment and continued in these activity behaviors through the duration of the study. Finally, consistent with the socio-ecological model of health behavior, each neighborhood's support for activity was roughly constant over time and the differential in activity between subjects results from the differences in support. Future research might explore these mechanisms in more depth. In particular, the second two possibilities, which interpret the association between disorder and activity as causal, may be particularly promising. For example, if consistent activity levels reflect consistent support for activity, then interventions to decrease disorder are more likely to increase activity.

Strengths

This study has several important strengths. First, as noted above, nearly all prior studies of neighborhood condition and physical activity among older adults have been cross-sectional [122]. The few longitudinal exceptions [172-174] have not examined neighborhood disorder as an influence. Second, this study used a novel low-cost CANVAS/Google Street View measure of neighborhood disorder that can in principle be deployed in other cities, lowering the costs of future replication studies [148, 150]. Third, because this measure of disorder was ascertained independent of survey response, our results are not subject to same-source bias that might arise in survey-only studies [75, 78]. Fourth, we used advanced statistical techniques to account for both missing covariates and loss to follow-up such that missing data would only bias our findings if it was missing not at random conditional on a number of comprehensive covariates [175]. Finally, our results were robust to several sensitivity analyses.

Limitations

However, like most empirical research, the study also has important limitations. First, the low (17%) response rate raises concerns that the sample may not be representative of the older adults in New York City. This low response rate was partially due to a low (57%) contact rate among phone numbers selected from a list of numbers provided by a data vendor; it may be that inaccuracies in address and phone number data included in the list hampered the contact rate, though this hypothesis has not been tested empirically. The cooperation rate among those contacted (31%) was within the 30-40% response rate range typically encountered by New York City Department of Health telephone surveys [176] and in line with response rates reported by a recent test of various survey methodologies conducted in Australia [177]. Concerns about non-response are also somewhat mitigated by the population-based sample design and our use of sample weights in analysis.

A second limitation is that several measures used for this analysis were problematic. Specifically, our social cohesion measure had only mediocre internal consistency in this population, raising the concern that the scale may reflect multiple underlying constructs or may have been interpreted differently by different subjects. Assuming social cohesion independently prevents disorder and encourages physical activity, as has been suggested previously [178], residual confounding due to incomplete control for social cohesion might have biased results away from the null. Similarly, while the PASE questionnaire has been validated in several populations similar to the NYCNAMES-II population,[141, 142] all physical activity questionnaires are subject to imperfect recall and reporting biases, which may be particularly strong among older adult populations. While imperfect recall would be expected to bias our results towards the null, if residents of more disordered neighborhoods simply fail to recall past-week activities, perhaps as a result of stressful neighborhood encounters, the resulting systematic bias would artificially inflate the association between disorder and activity. However, our concerns are tempered by a related analysis, detailed in the next chapter, in which we found that types of activity engaged in were fairly stable across waves, making it unlikely that past-week activity was frequently forgotten in as a consequence of transient events. Finally, our measure of functional status, particularly the basic activities of daily living score, was left-skewed with strong ceiling effects. While functional status was not our primary exposure of interest, if our measure failed to capture functional status variation that was positively correlated with activity and negatively correlated with disorder, then our estimates may be inflated due to residual confounding. More broadly, a more sensitive measure might have resulted in an observable association between neighborhood characteristics, physical activity, and changes in functional status, allowing us to control more completely for time-varying confounding by functional status.

Finally, in this study as in nearly all neighborhood effects studies [125, 179], residential self-selection – the tendency for people to choose neighborhoods that better support their chosen lifestyles -- may have undermined exchangeability. For example, because disorder can act as a barrier to walking only for subjects who would *ever* choose to walk, if those subjects on average choose less disordered neighborhoods, then an estimated effect of disorder failing to account for this difference in walking preferences would be biased. That is, the PASE scores observed among those living in less disordered neighborhoods may not accurately reflect the PASE scores that would have been observed had those who lived in more disordered neighborhoods counterfactually lived in less disordered neighborhoods. We observe, however, that in New York City as in many North American cities, neighborhood disorder is strongly correlated with race/ethnicity and educational attainment of neighborhood residents, as it was for our study participants. Because we controlled for the race/ethnicity and educational attainment of study participants, any confounding introduced by residential self-selection may be somewhat controlled for in our models already.

In conclusion, our study supports prior observations that older adults living in more disordered neighborhoods are on average somewhat less active than those in more ordered neighborhoods. However, we did not find evidence that the presence of disorder induces faster decline in activity levels among older adults. Whether the between-neighborhood disparity in physical activity levels arose as a result of residual confounding, as a result of residential self-selection, as a result of prior neighborhood influence on activity norms, or as a result of unchanging neighborhood support for activity, , is an area for future research.

Tables

Table 3.1. Selected Characteristics of the study population at each wave.

	Interviewed in 2011 (N=3,497)		Interviewed in @012 (N=2,455)		Interviewed in 2013	
Characteristic	**%**	**Mean (SD)**	**%**	**Mean (SD)**	**%**	**Mean (SD)**
Age						
65-68	33		33		34	
69-71	23		23		23	
72-75	44		44		43	
Sex						
Female	60		61		61	
Male	40		39		39	
Race/Ethnicity						
Non-Hispanic White	52		53		54	
Non-Hispanic Black	31		31		30	
Hispanic	11		9		9	
Other	7		7		7	
Educational Attainment						
Less than High School	19		17		17	
Completed High School	27		27		26	
Some College	18		18		17	
Completed College	36		38		40	
Household Income						
Less than $20,000	36		34		33	
$20,000-40,000	25		24		24	
$40,000-80,000	21		22		22	
More than $80,000	18		20		20	
Baseline PASE		80 (46)		81 (46)		81 (46)
Wave 2 PASE				79 (44)		79 (44)
Wave 3 PASE						80 (46)
Baseline functional status		93 (33) [1]		94 (28) [1]		93 (28) [1]
Neighborhood characteristics						
Past-year pedestrian injuries[2]		41 (42)		42 (43)		41 (42)
Walkability		0.00 (2.7)		0.02 (2.8)		-0.02 (2.8)
Disorder		-0.06 (0.25)		-0.07 (0.25)		-0.07 (0.25)
Social cohesion		3.6 (0.9)		3.6 (0.9)		3.6 (0.9)

[1] Functional status score was left-skewed and was reported as median (IQR) rather than mean (SD)
[2] Injuries per km^2 within .25 km network buffer surrounding subject's home address

Table 3.2. Disorder levels, functional status, and PASE score at baseline for 3,497 older adult residents of New York City surveyed in 2011, stratified by demographic and socioeconomic characteristics

Characteristic	%	Mean (SD) Disorder	Median (IQR) Intermediate Activities of Daily Living Score	Mean (SD) PASE Score
Age				
65-68	33	-0.07 (0.24)	94 (67, 100)	82 (47)
69-71	21	-0.07 (0.25)	93 (67, 100)	81 (44)
72-75	46	-0.06 (0.25)	89 (67, 100)	78 (45)
Sex				
Female	60	-0.06 (0.25)	89 (67, 100)	77 (42)
Male	40	-0.06 (0.24)	94 (78, 100)	84 (51)
Race/Ethnicity				
Non-Hispanic White	52	-0.14 (0.26)	94 (78, 100)	85 (47)
Non-Hispanic Black	31	0.02 (0.20)	83 (61, 100)	75 (43)
Hispanic	11	0.07 (0.18)	83 (61, 100)	68 (43)
Other	7	-0.02 (0.22)	89 (67, 100)	82 (48)
Educational Attainment				
Less than High School	19	0.04 (0.20)	80 (56, 100)	67 (44)
Completed High School	27	-0.01 (0.22)	89 (67, 100)	78 (45)
Some College	18	-0.07 (0.24)	93 (67, 100)	83 (44)
Completed College	36	-0.16 (0.26)	100 (83, 100)	87 (46)
Household Income				
Less than $20,000	36	0.03 (0.21)	80 (56, 100)	65 (40)
$20,000-40,000	25	-0.05 (0.23)	94 (72, 100)	86 (47)
$40,000-80,000	21	-0.12 (0.24)	94 (78, 100)	88 (46)
More than $80,000	18	-0.20 (0.26)	100 (83, 100)	91 (47)

Table 3.3. Spearman correlations between selected neighborhood characteristics

Characteristic	Disorder	Social Cohesion	Walkability	Pedestrian Injury Density
Disorder	1.00			
Social Cohesion	-0.05	1.00		
Walkability	0.06	0.06	1.00	
Pedestrian Injury Density	0.03	0.09	0.42	1.00

Table 3.4. Mean Differences in PASE Score at Baseline and Mean Differences in Changes in PASE Score Associated with Baseline Physical Disorder and Changes in Physical Disorder over Time for 3,497 adult residents of New York City Surveyed from 2011-2013.

	Mean Difference (95% Confidence Interval) in PASE Score		
Regression Coefficients	**Model 1[a]**	**Model 2[b]**	**Model 3[b]**
Difference at baseline per one-SD increase in baseline disorder	-3.1 (-4.5, -1.7)	-3.3 (-4.9, -1.7)	-3.1 (-4.7, -1.5)
Change per wave	-0.5 (-1.3, 0.4)	-0.5 (-1.3, 0.4)	-0.5 (-1.4, 0.4)
Difference in change per wave for each one SD increase in baseline disorder		0.1 (-0.8, 0.9)	
Difference in change per wave for each one SD increase in time-varying disorder			0.0 (-0.8, 0.9)

[a] Random intercept model adjusting for baseline age, educational attainment, gender, race/ethnicity, functional status, neighborhood social cohesion, neighborhood pedestrian risk, and neighborhood walkability

[b] Random intercept/random slope model adjusting for baseline age, educational attainment, gender, race/ethnicity, functional status, neighborhood social cohesion, neighborhood pedestrian risk, and neighborhood walkability

Figures

Figure 3.1. Scatterplots of PASE scores across waves showing both correlation and variation across waves

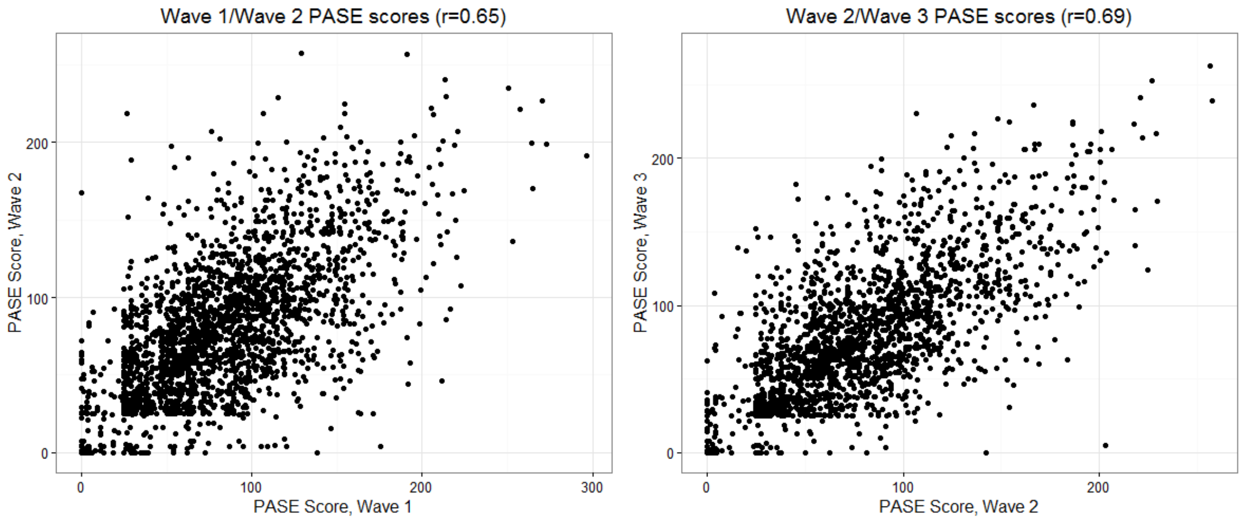


Figure 3.2. Scatterplots of neighborhood disorder score and PASE score at each wave, with an overlaid unadjusted least squares regression line showing the negative correlation at each wave.

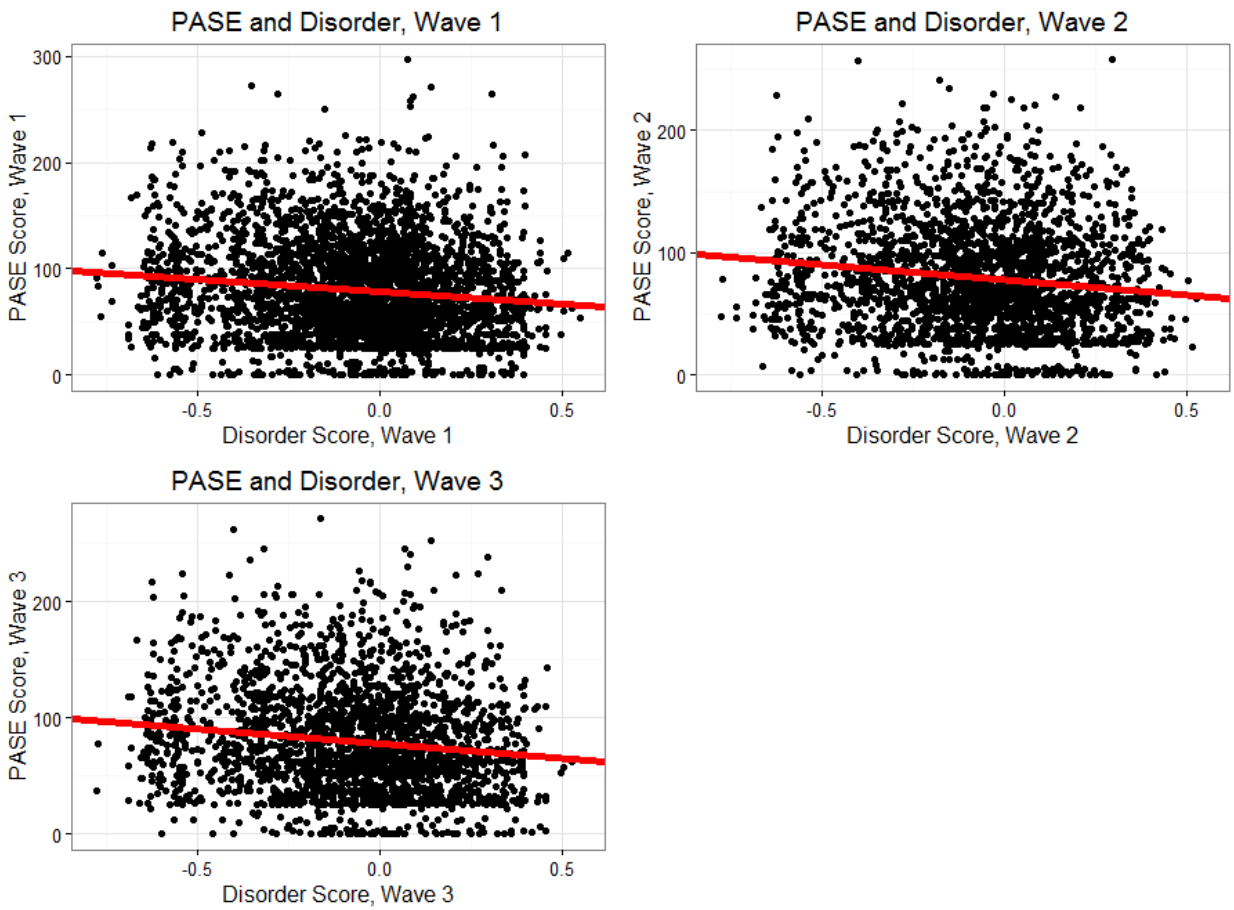
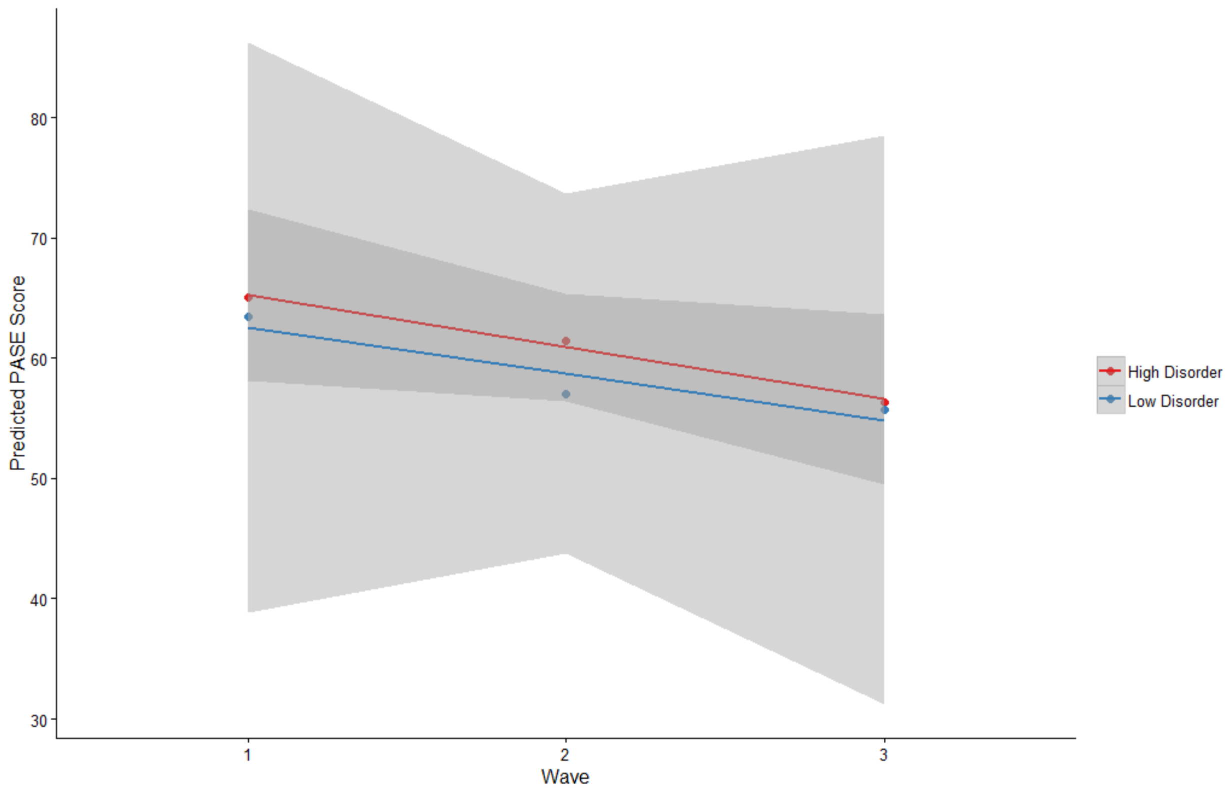


Figure 3.3. Estimated PASE scores and 95% prediction intervals for two hypothetical subjects at each wave

The subject represented by the red dots and lines lives in a high disorder neighborhood whereas the one represented by the blue line lives in a low disorder neighborhood. These predicted values indicate a difference in PASE score by baseline level but high variability in overall score and no interaction between baseline disorder and PASE trajectory over time.



Chapter 4: Longitudinal Physical Activity Patterns among Older Adults: a Latent Transition Analysis

Abstract

Background

Older adults engage in many types of physical activity, including sports and recreation, walking, home repair, and housework. Most studies of the causes of activity focus on overall physical activity or select a single activity to focus on. Both these approaches may fail to capture nuance in older adult activity patterns, particularly as activity is affected by contextual factors such as neighborhood disorder.

Methods

We performed a latent transition analysis to identify patterns of activity types in a cohort of 2,023 older adult residents of New York City and to explore determinants of changes in those activity types over two years of follow-up in relation to individual and contextual factors.

Results

Our analysis identified seven latent classes, which we named Mostly Inactive, Walking, Walking and Exercise, Household Activities, Household Activities and Exercise, Household and Gardening, and Household, Gardening and Exercise. Individual and neighborhood disadvantage was associated with membership in the classes defined by participation in fewer types of activities. At each wave, about 75% of subjects remained in the same class they had been in in the previous wave. Most transitions between classes occurred in or out of classes defined by exercise. There were indications that more neighborhood disorder was associated with moving out of a class defined by exercise (OR = 1.38, 95% CI = 0.96, 2.00 between waves 1 and 2, OR = 1.33, 95% CI = 0.88, 2.15 between waves 2 and 3), but estimates were too imprecise to rule out chance associations.

Conclusions

Patterns of activity types were generally stable, with the notable exception of engagement in exercise. More neighborhood disorder was consistently, but not statistically significantly, associated with ceasing participation in exercise.

Introduction

Researchers studying built environment influences on physical activity, including among older adults, have frequently invoked the need for specificity in both environment measures and activity measures [122, 180-183]. For example, pleasing neighborhood aesthetic characteristics are more plausibly an influence on walking in one's neighborhood than on one's occupational physical activity. A study using a measure of total daily physical activity to assess the impact of neighborhood aesthetics on walking would thus have less power to detect a true effect than one using a measure of walking alone [48].

However, from the standpoint of health promotion among older adults, it may also be problematic to study a single form of activity, such as walking, in isolation. Qualitative evidence suggests that older adults show interest in multiple forms of activity, including not only organized sports or exercise classes but also inexpensive, unstructured activities such as gardening [184]. Neighborhood support for those various forms of activities may vary significantly depending on setting. For example, urban neighborhoods conducive to walking may lack space for gardening, whereas rural settings may lack access to recreational facilities or walkable destinations [185, 186].

Moreover, conventional measures of total physical activity such as metabolic-equivalent units (METs) of energy expenditure [82] may also oversimplify physical activity's effects on health behaviors and health outcomes. Conventionally, researchers have used physical activity measures constructed by combining reports of intensity, frequency, and duration of subjects' activities to estimate total or mean energy expenditure [82]. While this approach provides an integrated measure of total activity, it hides information about which activities a subject engages in, or when within the day or the week that activity occurs. These ignored details, which may be related to cultural, gender, race, or disability related factors, may be highly relevant for determining the factors contributing to activity decisions. For example, two individuals may expend the same amount of energy in a week, but if the first expends the energy through daily gardening tasks and the second attends an intensive bi-weekly exercise class, the impact activity may have on the individuals' physical and mental health may be quite different. [187-189].

However, there are challenges in working with fully disaggregated activity data. For example, analyzing each type of activity (e.g. dishwashing, racquetball, carpentry) separately would require impractically large sample sizes to garner statistical power owing to the small proportion of the population engaging in some activities. As a result, physical activity researchers have recently begun to consider patterns of activity with respect to time or activity type rather than, or in addition to, total energy expenditure. These researchers have used several analytic techniques to date, including: analyses of patterns of activity in minute-by-minute high density accelerometer or fitness tracker data sets; latent class growth models of trajectories of energy expenditure through time; and latent class analyses of questionnaire data to classify study participants into groups with relatively homogeneous activity patterns. These analytic approaches retain theoretically relevant components of the heterogeneity in the underlying data while also collapsing similar groups to increase statistical power as compared to what would be available in fully disaggregated data. Indeed, pattern-oriented approaches may have greater power than energy expenditure approaches. For example, if differing patterns of activity produce similar total energy expenditure, but these differing patterns of activity are associated with health behavior or health outcomes, studies that analyze only total energy expenditure will have low power to detect relevant associations.

We previously applied a latent class analysis to data from the first wave of NYCNAMES-II to classify study participants into 5 groups typified by engagement in differing types of activity. These classes were statistically significant predictors of BMI and depressive symptom count independent of total PASE score [45]. Similarly, a recent analysis of children's activity patterns using minute-by-minute accelerometer data identified a relationship between asthma and timing of physical activity, with decreases in activity seen among children with asthma during the noon to 6 PM time period, when ozone levels are highest [190]. Prior analyses of total accumulated activity using the same dataset over the full monitoring period failed to identify an association between activity and asthma status.[190].

Thus, holistic approaches to physical activity research that identify and analyze patterns of activities may shed light on the contextual supports for and barriers to types of activity above and beyond what could be identified from analyses of energy expenditure alone. In particular, longitudinal mixture modeling techniques such as latent transition analysis may identify characteristic patterns of types of activity [191], wherein subjects may switch between patterns over time. These switches and the variables that predict them may in turn help to identify influences on activity patterns. However, while such mixture models have been applied to types of activities adolescents [192] and younger adults [44] engage in, to the best of our knowledge, no prior study has applies a latent transition analysis to examine patterns of types of activities older adults engage in longitudinally.

In this chapter, we use a latent transition analysis to identify patterns of activity and changes in those patterns of activity over two waves of follow up in the same cohort of older adults we analyzed in Chapter 3.

Methods

Subjects and setting

Like Chapter 3, this analysis used data from NYCNAMES-II, a longitudinal study of residents of New York City aged 65-75 at baseline [45]. Subjects were initially recruited by phone, starting in May 2011 and continuing through November 2011. The final sample size at baseline was 3,497. Abt-SRBI attempted to re-contact each subject by telephone once in summer or fall 2012 and once in summer or fall 2013. Each subject who was successfully recontacted was asked to complete a 30-minute follow-up interview to assesses changes in mental and physical health status, functional status, and neighborhood perceptions. At Wave 2, 2,455 subjects (70.2%) were successfully re-intereviewed, and 2,355 (67.3%) were re-contacted at Wave 3, of which 2,023 had been re-contacted at Wave 2. At each wave, subjects were paid $10 for participating, either in the form of a CVS gift card as a check.

Measures

Demographics and Exposure Measures

All demographic measures, including self-reported sex, age, educational attainment, race/ethnicity, health status, and income were assessed only at baseline. As in Chapter 3, we categorized age at baseline as 65-68, 69-71, and 72-75 and categorized household income as <$20,000, $20,000-39,999, $40,000-79,999, and ≥$80,000. We categorized education levels were as less than high school graduate, high school graduate, some college, or college graduate. We analyzed health statuses as they were reported: excellent, good, fair, or poor. Finally, we categorized race/ethnicity as Non-Hispanic Black, Non-Hispanic

White, Hispanic, and Other.

Physical Activity

All subjects who were followed up successfully were asked at each wave about frequency, duration, and intensity of past week physical activity, using sixteen items derived from the Physical Activity Scale for the Elderly (PASE) [138, 141, 142] and detailed in Appendix Table 1. The full PASE instrument includes an item to assess job-related activity, but that item was not assessed in this mostly retired cohort (12% employed at baseline). PASE items whose response could be inferred from previous responses were not assessed. For example, subjects who had previously indicated during this wave's interview that they walked for zero days in the last week were not asked how much time they spent walking on days they walked.

The PASE instrument was designed to be scored as a continuous measure reflecting total physical activity, a formulation that was used in the previous chapter. Individual items included on the PASE questionnaire do not directly assess independent categories of activity; as such, the questionnaire items required recoding for ease of substantive interpretability and to avoid overfitting latent class models due to item redundancy. Consistent with a previous latent class analysis using PASE items [45], we recoded the sixteen ordinal item responses to twelve dichotomous indicators as follows. There are five physical activity domains where PASE assesses both number of days and average duration which the subject engaged in the activity (walking, light recreational activity, moderate recreational activity, strenuous recreational activity, and muscle strengthening exercises), we computed average duration per week for each subject. Because each level of recreational activity intensity was relatively rare on its own and because the energy expenditure from many recreational activities varies greatly depending on how the activity is performed (e.g. cycling can range from 3.5-16.0 MET units, depending on how vigorous the cyclist is) [193] and hence is subject to misclassification, we did not expect distinguishing recreational activity by vigor to produce interpretable classes. We therefore summed the measures of recreational activity at three intensity levels into one measure reflecting participation in any sport or recreational activity. Finally, for the resulting three domains for which we computed duration per week (walking, sports and recreation, and muscle strengthening), we computed two indicators: an ever- indicator set to true if the subject reported more than 0 minutes/week of the given activity and an often- indicator set to true if the subject reported more than 30 minutes/week of activity. The resulting indicators, their prevalence at each wave, and the maximal MET unit value of any activity associated with that indicator is shown in Table 4.1. This recoding has been used in a previous analysis; several alternate recoding options that were tested in that analysis resulted in less substantively interpretable latent classes [45].

Statistical Analysis

We used a latent class analysis extension to SAS to identify latent classes and assign transition probabilities between classes [194]. For each identified class, we computed median PASE score within the class and the maximal MET score associated with any activity implied by the class to provide an interpretive bridge from these identified classes to more traditional physical activity measures. To increase model interpretability, we imposed measurement invariance across waves, such that the model classifying subjects into a latent class was consistent across waves. Models including covariates in the model fitting step failed to converge. Accordingly, we used a ‘three-step’ modal assignment approach common in latent class analyses [195]. In this approach, step 1 refers to fitting the model, step 2 refers to assigning each subject to the latent class fitting her or his activity pattern best, and step 3 refers to fitting separate multinomial models to assess predictors of class membership and transition between classes. In general, observed associations in three-step models underestimate true associations [196]; we thus expect any bias due to the necessity of using a three-step modal assignment approach to be towards the null.

Whereas in Chapter 2, we used inverse probability of censoring weights to account for loss to follow-up, we were unable to do so for the latent transition analysis because the SAS extension package we used does not at present (version 1.3.1) support weighted analyses [194]. Because latent transition analysis also requires observations at all waves for all subjects, we limited our latent transition analysis to subjects who were observed during all 3 waves. After using assigning each subject a latent activity class at each wave, subsequent analyses used inverse probability of observation weights for each subject at each wave. Specifically, we modeled the logit of being observed in all three waves as a linear function of gender, race/ethnicity, educational attainment, borough of residence, neighborhood disorder, neighborhood pedestrian injury rate, and self-reported health status. We used the inverse of the modeled probability as a weight, as described by Robins, et al. [158]. We then multiplied these weights by the baseline sampling weights to compute final weights for analysis, such that, as in the previous chapter, results are representative of the population of non-institutionalized New York residents aged 65-75 according to the 2010 US Census (n=571,323). As in the previous chapter, all analyses of LTA classes in relation to covariates accounted for missing covariates by combining results across 5 multiply imputed datasets using Rubin's rules.

Because our initial explorations identified relatively stable latent classes wherein the most common transitions were defined by adding or removing recreational sports and muscle strengthening exercises, we focused the third steps of our analyses on the transitions into and out of these exercise-including classes. Specifically, we used weighted Poisson regression with sandwich estimated standard errors to calculate relative risk associated with entering or leaving a latent class incorporating exercise to or from the comparable latent class not incorporating exercise [197].

We used SAS Version 9.3 (Cary, NC) to assign latent classes and performed all subsequent analyses in 64-bit R for Windows version 3.2.3 (Vienna, Austria).

Results

Data were available for all three waves for 2,023 (58%) of the initially recruited 3,497 subjects. As compared to the full cohort, subjects who completed three waves of data collection and accordingly were included in the latent transition analysis were somewhat more likely to be non-Hispanic White, well-educated, and high income (Table 4.2).

Model fit statistics best supported a seven latent class model (Appendix Table 4.2). The seven class model also had acceptable substantive interpretability and a good balance of class prevalence (no class contained less than 10% of the overall cohort). After examining the activity patterns by class, we named the classes Mostly Inactive, Walking, Walking and Exercise, Household Activities, Household Activities and Exercise, Household and Gardening, and Household, Gardening and Exercise, respectively. Table

4.3 describes the latent classes statistically, showing the probability of endorsing each item within each latent class and the prevalence of each latent class at each wave.

Class membership at baseline was strongly associated with individual and neighborhood characteristics. In general, membership in the two less active classes, Mostly Inactive and Walking, were associated with indicators of social disadvantage, such as minority race/ethnicity (35% of non-Hispanic Blacks' activity class was Mostly Inactive or Walking as compared to 23% of non-Hispanic Whites), lower income (38% of those with household incomes below $20,000 were in the Mostly Inactive or Walking classes as compared to 21% of those with household incomes over $80,000) and living in neighborhoods with more disorder (33% of those in neighborhoods above the median disorder level were in Mostly Inactive or Walking, as compared with 23% of those living in neighborhoods below the median disorder level, Wald test $p < 0.01$). Table 4.4 shows the prevalence of latent classes within strata of selected demographic and neighborhood characteristics, and Figure 4.1 shows the distribution of neighborhood disorder measures within each latent class.

Class membership was fairly stable across waves, with about three of every four subjects in each class remaining within their prior latent class across waves and about two of every three in the same class in waves 1 and 3. Transitions between latent classes were most common between latent classes whose defining difference was the addition or subtraction of engaging in recreational sports (Table 4.5, Appendix Figure 4.1). Figure 4.2 depicts the most common transitions between latent classes, including the types of activities defining the differences between classes.

Because most latent class transitions occurred while initiating or ceasing sports and recreation as an activity, we explored predictors of transitioning to and from classes defined by sports and recreation. In bivariate models, older age, less educational attainment, lower income, and above median neighborhood disorder were associated with transitioning from a class including sports and recreation to one that did not (Table 4.6). In multivariable models (Table 4.7), age was strongly associated with transitioning to latent classes not defined by sport and recreational activity between waves 1 and 2 (Relative Risk [RR] = 2.09, 95% CI = 1.21, 3.60 comparing the oldest to the youngest subjects). However, the association did not persist between waves 2 and 3 (RR: 0.78, 95% CI = 0.35, 1.70). There were indications that more disorder was consistently associated with transitioning from a sports-defined class as well, but precision of the weighted estimates was low to rule out chance (RR: 1.27, 95% CI = 1.00, 1.61 between waves 1 and 2, RR: 1.28, 95% CI = 0.85, 1.93 between waves 2 and 3)).

Whereas several covariates were associated with transitioning out of a latent classes defined by sports and recreation, none of the explored variables were strong predictors of transitioning into such a latent class. Multivariable models predicting transition from a class not defined by sports and recreation to one defined by sports and recreation were similarly uninformative (Table 4.8)

Discussion

In this chapter, we used latent transition analysis to identify seven patterns of physical activity performed by a cohort of older adults aged 65-75 in New York City. Based on the types of activity subjects engaged in, we labeled the patterns: 1) Mostly Inactive, 2) Walking, 3) Walking and Exercise, 4) Household Activities, 5) Household Activities and Exercise, 6) Gardening and Household Activities, and 7) Gardening, Household Activities, and Exercise. Transitions between classes over 3 waves of data collection were very uncommon, with the exception of transitioning between classes that were roughly equivalent except for the presence or absence of exercise. There were indications that older age group and more disorder were associated with transitioning out of a class marked by exercise, though neither covariate was consistently statistically significant. Taken together, these findings indicate that though there are between-individual differences in the patterns of activity engaged in by older adults, on average, there is little within-individual difference in activity patterns over the three year period of this study. In short, study participants in general maintained habitual patterns of activity over the study period.
- The finding that exercise activity, specifically muscle strengthening exercises, sports, and recreation, was the type of physical activity most likely to be started or stopped between waves is intriguing. As noted in the sensitivity analysis relating weather to overall activity and types of activity detailed in Chapter 3's appendix, this does not appear to be an artifact of differing weather or season across waves. Broadly speaking, the built and physical environment more strongly affects engagement in active transport than recreational activity, including exercise [47]. It may follow that in a cohort where few subjects moved and the built and physical environment likely didn't change much owing to the relatively short two-year follow-

up period, exercise offers the best prospect for behavior change. Alternately, it may be that recreational sports opportunities vary more week-over-week than other activity types assessed by PASE. For example, a frequent tennis player may require an opponent and a court time to be available in order to schedule a match; the probability that she reports recreational sports in any given week incorporates variability due to this requirement, whereas a frequent gardener may have gardening tasks to accomplish every week. This latter interpretation is consistent with a prior finding that a past-week version of a physical activity survey for older adults had worse test-retest reliability than a past-3 months version of the same survey [198].

Our analysis adds to the literature regarding the changing patterns of activity as adults age. One 2012 study of patterns of physical activity over 12 years among older adult women aged 70 and older at baseline identified four patterns of total activity [46]. Three of the patterns identified in that study, which together accounted for 80% of the women, were defined by roughly constant levels of activity over time, consistent with our finding that activity patterns among older adults, once set, are resistant to change. This finding is further supported by several studies assessing trajectories of total physical activity or leisure-time physical activity in other age groups that have also found activity to be constant over time for the majority of subjects [199-201]. The degree to which this consistency is due to consistent behavioral habits among the subjects as opposed to consistent environmental influence applied to the subjects is an area for future research among older adults who change residences.

Whereas a prior study of patterns of activity in this cohort at baseline had identified five latent classes [45], this analysis, which incorporated activity at three waves, identified seven. Three of the classes identified in this analysis (Mostly inactive, Gardening and Household Activities, and Gardening, Household Activities, and Exercise) are roughly equivalent to three that were identified in the baseline-only analysis. The other four identified here are similar to the remaining two, but the year-over-year changes in sports and recreation and muscle strengthening exercises decrease the covariance between those indicators and the other activities defining the classes. This results in two pairs of classes (Household Activities and Household Activities and Exercise; Walking and Walking and Exercise) that are similar except for participation in exercise.

[202] The patterning of transitions we observed, highlighted in Figure 4.2, is an additional novel contribution of this work. Whereas conventional physical activity research using a single measure of activity, such as MET-Hours of activity per week, is unable to assess the dimensionality of changes in that activity, this latent transition analysis decomposes overall change into more specific between-activity changes. Indeed, that the maximum MET score, a measure of the maximal potential *intensity* of activity [193], does not rank latent classes in the same order as they would be ranked by median PASE score within a class, a measure of total activity [138, 141, 142], is consistent with our understanding that the different patterns of activity reflect both different activity capabilities and different activity intentions among our subjects [203]. The latent classes thus describe distinct patterns of activity, for example whether a subject works in his or her home and garden, rather than an ordered categorization of subjects by energy expenditure.

This analysis had several notable strengths. Though several prior studies have applied latent class growth models or repeated measures latent class analyses to assess trajectories of total activity among adult populations [46, 199, 200], to the best of our knowledge, this was the first analysis to identify longitudinal patterns of activity *types* among older adults. Additionally, though we could not incorporate sampling weights in computing latent classes, our remaining analyses were weighted to allow an estimate of the population prevalence of the latent classes we did identify.

However, our findings should be considered in light of the following limitations. First, because we were unable to conduct the latent transition analysis in the full population, the identified patterns are representative of patterns of activity type among the generally healthier and more privileged groups that were followed up for three waves and may not represent patterns of activity types among older adults more broadly. Our concerns on this front are somewhat mitigated by the similarity between the latent classes identified here and those identified in this cohort at baseline only, as described above, and by our use of weights in analyses of transitions between classes. Second, because NYCNAMES-II had a low response rate, even the full sample may not represent the older adult population of New York City. As in Chapter 3, our use of a population based sample and incorporation of sample weighted analyses somewhat diminishes our concerns on this count. Third, because our only measure of activity types was a survey of past-week activity, there may have been substantial misclassification in our activity measure. We note, however, that the relative stability of latent classes between waves would require the misclassification to be consistent within people over time, and prior studies have generally found PASE to be reliable and valid [138, 139, 142, 204]. Furthermore, though research is evolving on identifying activity types from GPS and accelerometer measures [205-208], survey-based measures remain the core metric for research on activity types [209, 210]. Additionally, in this as in most observational studies of residential context as a determinant of activity, we were unable to fully account for 'residential self-selection', or the tendency for people with a choice of places to live to choose the places that best support their behavioral preferences [211]. However, as in most observational studies of residential context, we did adjust for primary determinants of residential location, including household income and subject race/ethnicity [212].

A final but important limitation to this analysis is that, as in all latent classification analyses, class labels were selected by researchers rather than derived from data, and thus are intrinsically value-laden [202]. As a result, caution must be taken when interpreting classes by name. For example, whereas the "household activities" latent class is defined in opposition to the "household activities with exercise" class by the latter's much more frequent engagement in exercise activities, twenty percent of the members of the "household activities" latent class reported some sports and recreation activities. Indeed, while the the class labels we selected reflect our reading of the types of activities reported by the survey, PASE was not developed with latent class analyses in mind, and we had to recode item responses prior to fitting latent classes models to make the model input more appropriate and interpretable [45]. Future investigation of patterns of activity types might benefit from more open-ended survey options, perhaps taking inspiration from food frequency questionnaires, which have been developed with an expectation of identifying patterns [213]. For example, long-form activity questionnaires that assess the activity engaged in, intensity, frequency, duration, and time of day of each activity bout, though onerous for subjects, might provide the base evidence for a re-evaluation of what should be included on a short-form or grist for a latent class or principal components analysis. A complementary approach, particularly as high-tech devices including mobile phones are incorporated into research, might be the development of dynamic protocols – wherein a subject's mobile phone prompts the subject to identify an activity they'd recently undertaken. Such protocols could be designed with pattern identification in mind. For example, if a subject is active at 8:30 AM every day, it might be especially important to prompt to determine whether that activity represents an active commute.

In summary, this latent transition analysis identified seven patterns of activity types observed over two years in older adult residents of New York City. Patterns of activity types were generally stable, with the notable exception of engagement in exercise. These patterns of activity suggest that barriers to both starting and stopping exercise may be more transient, whereas barriers to other daily life activities may be more constant over time.

Tables

Table 4.1. Prevalence and examples of the categories of past-week physical activity used for this analysis of 2,023 New York City residents aged 65-75, surveyed between June 2011 and November 2011.

Activity Category	Wave 1	Wave 2	Wave 3	Examples[1]	Approximate MET range
Ever-Sports	55.2	54.8	55.2	Golf, dancing, basketball, hiking, jogging	3-14
Often-Sports	27.9	26.6	25.3	Sports for more than 30 minutes a day	
Ever-Exercises	37.6	36.8	38.4	Calisthenics, sit-ups, weight lifting	2-14
Often-Exercises	5.9	5.2	5.5	Exercises for more than 30 minutes a day	
Ever-Walking	91.5	90.6	90.2		2-5[2]
Often-Walking	52.6	50.0	50.0	Walking for more than 30 minutes a day	
Caring for Others	31.1	30.4	31.2	Caring for a person who requires assistance with daily living tasks such as showering	2-4
Outdoor Gardening	21.1	18.7	20.5	Planting plants, pruning, weeding	2-4
Heavy Housework	54.5	54.3	53.0	Vacuuming, sweeping, moving furniture	3-7
Home Repairs	10.7	9.8	10.5	Painting, plumbing, carpentry	2-6
Light Housework	92.0	92.1	92.2	Washing dishes, ironing, laundry	2
Lawn Work or Yard Care	17.5	16.4	16.7	Shoveling snow, lawn mowing	3-6

MET: Metabolic-Equivalent Units
[1] listed activity types are taken from the Physical Activity Scale for the Elderly (PASE) instrument
[2] assuming walking no faster than a ‘very brisk’ pace

Table 4.2. Selected Characteristics of NYCNAMES-II Study Population and the Subset Who Completed All Three Waves of Data Collection

Characteristic	Study Population (N=3,497) %	Completed Three Waves (N=2,023) %
Age		
65-68	34	34
69-71	23	21
72-75	43	45
Sex		
Female	60	61
Male	40	39
Race/Ethnicity		
Non-Hispanic White	52	55
Non-Hispanic Black	31	30
Hispanic	11	9
Other	7	6
Educational Attainment		
Less than High School	19	16
Completed High School	27	26
Some College	18	17
Completed College	36	41
Household Income		
Less than $20,000	36	33
$20,000-40,000	25	24
$40,000-80,000	21	22
More than $80,000	18	21

Table 4.3. Item response probabilities by latent class

Activity	Mostly Inactive	Walking	Walking and Exercise	Household Activities	Household Activities and Exercise	Gardening and Household Activities	Gardening, Household Activities, and Exercise
Ever-Sports	13%	33%	100%	28%	100%	35%	100%
Often-Sports	1%	0%	73%	0%	64%	0%	77%
Ever-Exercises	21%	25%	70%	22%	55%	22%	65%
Often-Exercises	2%	1%	15%	2%	6%	3%	17%
Ever-Walking	36%	100%	98%	95%	99%	96%	99%
Often-Walking	0%	40%	73%	47%	78%	45%	73%
Caring for Others	13%	19%	17%	36%	43%	38%	44%
Outdoor Gardening	1%	2%	6%	3%	4%	77%	85%
Heavy Housework	10%	8%	16%	83%	84%	75%	77%
Home Repairs	2%	3%	2%	6%	11%	25%	33%
Light Housework	68%	90%	85%	100%	99%	96%	96%
Lawn Work or Yard Care	0%	1%	2%	2%	3%	66%	75%
Proportion of Cohort							
Wave 1	11%	16%	11%	24%	16%	11%	12%
Wave 2	11%	16%	12%	24%	16%	12%	10%
Wave 3	12%	16%	10%	23%	17%	12%	10%
Estimated Proportion of Older Adult Residents of New York							
Wave 1	12%	17%	9%	26%	14%	11%	11%
Wave 2	12%	17%	8%	26%	14%	13%	10%
Wave 3	12%	17%	8%	25%	15%	14%	9%
Estimated MET value of most vigorous activity undertaken by majority of class members	2	5	14	7	14	7	14
PASE scores of subjects in this class, median (IQR)	25 (26)	38 (22)	67 (34)	68 (36)	100 (36)	117 (49)	153 (53)

PASE: Physical Activity Scale for the Elderly

Table 4.4. Prevalence of latent classes at baseline by selected individual and neighborhood characteristics

Characteristic	Mostly Inactive	Walking	Walking and Exercise	Household Activities	Household Activities and Exercise	Gardening and Household Activities	Gardening, Household Activities , and Exercise
Age							
65-68	10%	16%	9%	27%	17%	9%	12%
69-71	11%	15%	12%	24%	12%	15%	12%
72-75	13%	19%	9%	26%	14%	9%	10%
Sex							
Female	8%	18%	10%	21%	16%	13%	14%
Male	14%	16%	8%	30%	14%	10%	8%
Race/Ethnicity							
Non-Hispanic White	8%	15%	11%	23%	14%	15%	15%
Non-Hispanic Black	14%	21%	5%	30%	14%	9%	7%
Hispanic	19%	14%	8%	35%	15%	3%	5%
Other	10%	16%	11%	22%	19%	13%	10%
Educational Attainment							
Less than High School	20%	15%	6%	33%	11%	9%	5%
Completed High School	10%	19%	7%	27%	15%	14%	9%
Some College	7%	16%	7%	24%	21%	11%	14%
Completed College	6%	17%	16%	18%	15%	12%	17%
Household Income							
Less than $20,000	21%	18%	6%	30%	13%	8%	4%
$20,000-40,000	8%	16%	6%	28%	21%	9%	12%
$40,000-80,000	4%	16%	9%	24%	12%	17%	17%
More than $80,000	4%	15%	20%	16%	10%	16%	17%
Disorder							
Above Median	14%	19%	8%	28%	16%	9%	7%
Below Median	9%	14%	10%	24%	14%	14%	15%
Unemployment							
Above Median	15%	17%	6%	28%	15%	9%	10%
Below Median	8%	16%	12%	24%	15%	13%	11%

Table 4.5. Probability of Transitioning Between Classes Across Waves, as computed by the Latent Transition Analysis

		Wave 2						
		Mostly Inactive	Walking	Walking and Exercise	Household Activities	Household Activities and Exercise	Gardening and Household Activities	Gardening, Household Activities, and Exercise
Wave 1	**Mostly Inactive**	79%	3%	3%	14%	0%	1%	0%
	Walking	2%	82%	8%	4%	0%	3%	0%
	Walking and Exercise	3%	25%	59%	4%	6%	2%	2%
	Household Activities	7%	0%	1%	74%	12%	5%	1%
	Household Activities and Exercise	0%	0%	2%	24%	70%	0%	3%
	Gardening and Household Activities	1%	3%	3%	2%	0%	78%	13%
	Gardening, Household Activities, and Exercise	0%	0%	4%	1%	4%	20%	71%
		Wave 3						
Wave 2	**Mostly Inactive**	84%	11%	0%	4%	0%	0%	0%
	Walking	5%	79%	9%	6%	0%	1%	0%
	Walking and Exercise	4%	11%	67%	6%	0%	5%	7%
	Household Activities	4%	4%	0%	78%	10%	5%	0%
	Household Activities and Exercise	0%	5%	0%	14%	81%	0%	0%
	Gardening and Household Activities	3%	0%	2%	5%	2%	77%	10%
	Gardening, Household Activities, and Exercise	1%	0%	4%	2%	4%	17%	72%

Table 4.6. Descriptive statistics showing the probability of making a latent class transition related to stopping sports at each wave, weighted to population of adults residents of New York City aged 65-75

	Wave 1 to Wave 2			Wave 2 to Wave 3		
Predictor	**Sports at W1**	**Stopping at W2**	**Starting at W2**	**Sports at W2**	**Stopping at W3**	**Starting at W3**
Baseline Age	**%[a]**	**%[b]**	**%[c]**	**%[a]**	**%[b]**	**%[c]**
65-68	37	14	8	37	12	7
69-71	35	24	11	33	17	4
72-75	33	31	9	30	14	9
Education						
< HS	23	39	10	21	14	7
HS grad	31	24	10	30	20	5
Some College	42	16	7	39	15	9
BA or More	48	16	8	45	8	14
Income						
< 20K	23	28	8	23	14	7
20K-40K	39	24	12	36	21	8
40K-80K	38	19	6	36	9	11
80K +	48	19	11	46	10	8
Disorder						
Above Median	30	26	7	28	17	8
Below Median	39	20	11	38	11	8
Overall	34	23	9	32	14	8

[a]Percent of the total cohort in this stratum at this wave whose activity class included exercise

[b]Percent of the total cohort in this stratum whose activity class included exercise at this wave whose activity class did not include exercise in the next wave

[b]Percent of the total cohort in this stratum whose activity class did not include exercise at this wave whose activity class included exercise in the next wave

Table 4.7. Results of a multivariable Poisson model predicting transitioning from a latent class characterized by exercise to one not characterized by exercise

	Wave 1 to Wave 2			Wave 2 to Wave 3		
	Relative Risk of Stopping Exercise	95% CI		Relative Risk of Stopping Exercise	95% CI	
Baseline Age						
65-68	REF	--	--	REF	--	--
69-71	1.60	0.85	3.02	0.64	0.24	1.69
72-75	2.09	1.21	3.60	0.78	0.35	1.70
Education						
< HS	2.17	1.03	4.55	0.79	0.28	2.24
HS grad	1.34	0.73	2.47	1.40	0.56	3.49
Some College	1.06	0.49	2.27	1.18	0.49	2.84
BA or More	REF	--	--	REF	--	--
Income						
< 20K	0.78	0.32	1.92	1.28	0.49	3.36
20K-40K	0.81	0.36	1.83	2.51	0.92	6.82
40K-80K	0.94	0.45	2.00	0.89	0.31	2.55
80K +	REF	--	--	REF	--	--
Neighborhood Disorder						
Per SD	1.27	1.00	1.61	1.28	0.85	1.93

Table 4.8. Results of a multivariable Poisson model predicting transitioning to a latent class not characterized by exercise to one characterized in part by exercise

	Wave 1 to Wave 2			Wave 2 to Wave 3		
	Relative Risk of Starting Exercise	95% CI		Relative Risk of Starting Exercise	95% CI	
Baseline Age						
65-68	REF	--	--	REF	--	--
69-71	1.46	0.70	3.05	0.57	0.22	1.57
72-75	1.14	0.56	2.32	1.31	0.57	3.00
Education						
< HS	1.28	0.54	3.06	0.80	0.31	2.10
HS grad	1.18	0.57	2.45	0.24	0.08	0.72
Some College	0.90	0.42	1.93	0.95	0.43	2.08
BA or More	REF	--	--	REF	--	--
Income						
< 20K	0.70	0.26	1.86	1.09	0.31	3.79
20K-40K	1.10	0.47	2.59	0.98	0.28	3.49
40K-80K	0.58	0.20	1.71	1.62	0.53	4.95
80K +	REF	--	--	REF	--	--
Neighborhood Disorder						
Per SD	0.90	0.65	1.26	1.05	0.70	1.57

Figures

Figure 4.1. Box plot displaying neighborhood disorder level distribution for each latent class

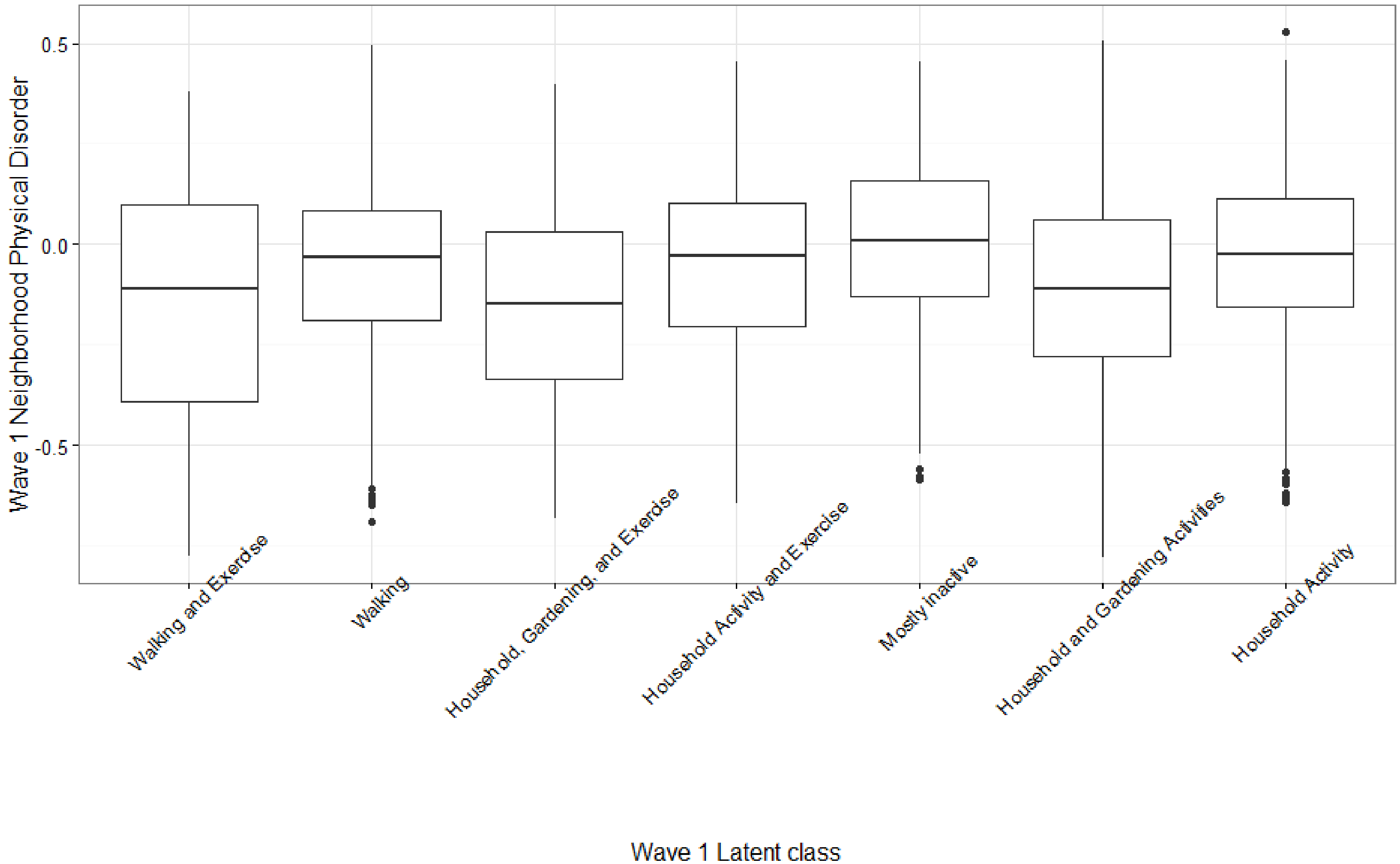
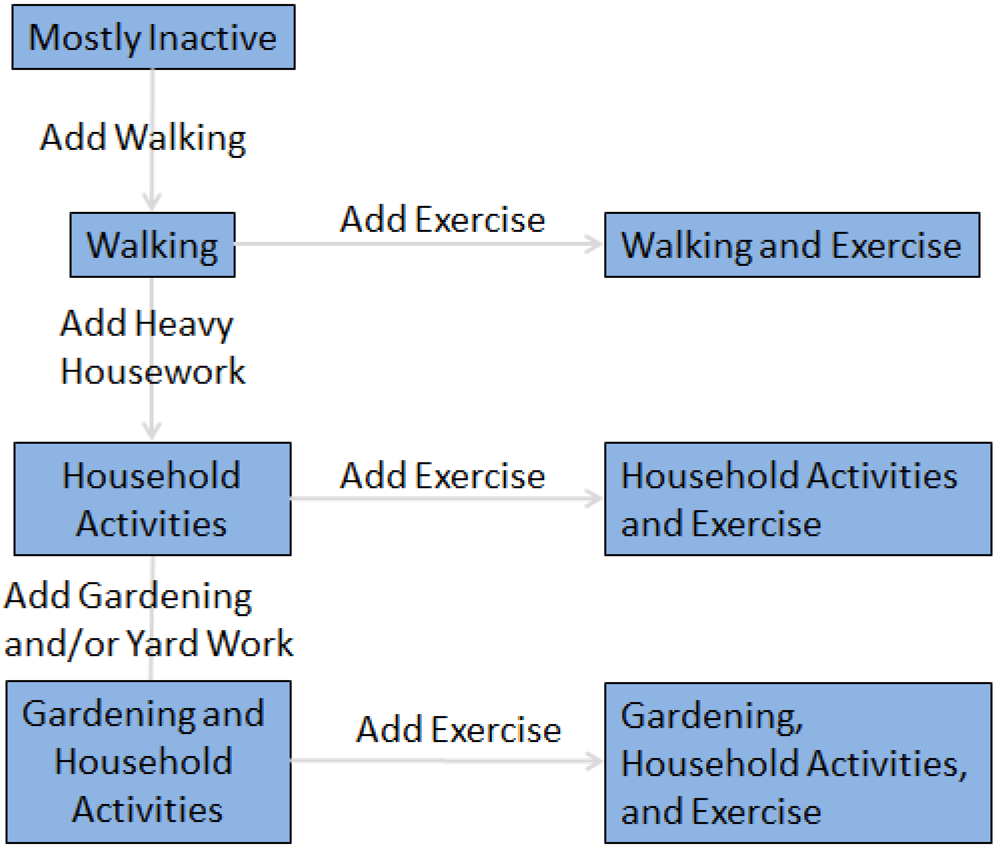


Figure 4.2. Latent classes identified in this analysis, including arrows noting the common transitions between classes and the activity type defining that transition.



Chapter 5: A Neighborhood-environment wide association study (NE-WAS) of physical activity

Abstract

Background

Studies of neighborhood context as a correlate of physical activity typically select a modest number of theoretically informed environmental characteristics to study, analogous, in a genetic context, to a candidate-gene study. We conducted a pilot agnostic ‘Neighborhood Environment-Wide Association Study (NE-WAS)’ approach to studying neighborhood influences on physical activity, analogous, in a genetic context, to a Genome Wide Association Study (GWAS) approach.

Methods

NYCNAMES-II was a telephone survey of 3,497 adult residents of New York City aged 65-75. Using Geographic Information Systems and a variety of previously compiled measures of New York City’s social and physical environment, we constructed 337 measures of neighborhood context for each subject. We explored survey-weighted regression models, LASSO regression, and random forest approaches to select the neighborhood measures most predictive of each of 1) overall physical activity, 2) gardening, 3) walking, and 4) housework.

Results

Of all 337 measures, proportion of residents living in extreme poverty was most strongly associated with total physical activity (estimated decrease of 0.85 Physical Activity Scale for the Elderly units (95% CI: 0.56, 1.14) per 1% increase in proportion of residents living with household incomes less than half the federal poverty line). Only neighborhood socioeconomic status and disorder measures were associated with total activity and gardening, whereas a broader range of measures was associated with walking. As expected, no neighborhood measures were associated with housework after accounting for multiple

comparisons. Regression resulted in more interpretable comparisons between measures than empirical variable selection approaches.

Discussion

A systematic approach to comparing neighborhood measures to activity measures revealed patterns in the domains of neighborhood measures associated with activity. The NE-WAS approach appears promising.

Introduction

Despite considerable qualitative evidence supporting the concept that neighborhoods influence physical activity in older adults [37, 169] and the development of numerous theoretical frameworks [19, 26, 27, 38] exploring these conceptual links, quantitative evidence confirming specific neighborhood factors supporting specific activities has been inconsistent [69].

One reason for inconsistency may be the difficulty of objectively measuring neighborhood constructs described qualitatively. For example, in interviews, older adults frequently indicate that they don't like to walk in their neighborhoods if they feel that they may be targeted for crime [37]. However, measuring neighborhood crime risk is extremely difficult [43]. The area defined by 'neighborhood' is subjective and may vary as older adults lose or recover functional capacity [214-216]. Statistical results from spatial models can be highly sensitive to geographic definition, so problems in neighborhood specification may strongly affect study validity.[217-219] Furthermore, there are many ways to operationalize safety from crime [43]. Whereas one study may operationalize neighborhood crime as reported neighborhood crime in an administrative area such as a county or zip code [220], another may define neighborhood crime by asking subjects to report perceptions of neighborhood safety [186]. Yet it is possible that still another operationalization of neighborhood crime, such as using spatial analysis with techniques such as kriging to estimate crime levels within a buffer around subject homes [149] would more accurately reflect the true impediment to activity. If this were the case, both a zip code-base and a self-report-based study would likely underestimate true effects due to non-differential measurement error [221].

A complementary reason for quantitative inconsistency is the sheer number of neighborhood measures available to modern researchers. Increasingly powerful geographic information systems (GIS) tools and spatial analysis techniques allow researchers to define neighborhoods in creative new ways, such as radial buffers (the area within a given distance 'as the crow flies' from the subject's home address) or network buffers (the area reachable via walking a given distance walk on city streets). Such neighborhood definitions may be more relevant to subjects than definitions used by early neighborhood researchers, such as the Census tract or zip code containing the subjects' residence, but the proliferation of measures creates more opportunities both for artefactual findings [157, 222] and inappropriate statistical control nullifying findings [223]. In the face of so many, often collinear measures, some researchers have called for increased use of neighborhood typologies, as might be identified by a latent class analysis [224-226], rather than specific neighborhood measures. While such techniques validly address researcher concerns regarding multicollinearity of measures, such typology measures may also obscure true effects that might be identified if measures were analyzed individually [181].

An alternate but complementary analytical approach to the multiplicity of measures draws an analogy to the similar conundrum faced by genetic and '-omic' molecular epidemiologists [227]. Along with others, genetic epidemiologists have developed the so-called genome-wide association study (GWAS) approach, wherein 'agnostic' or hypothesis-free empirical analytic approaches are used to search the whole genome for the strongest genetic associations, which then assumed to be the best candidates for subsequent research [228]. Recently, this agnostic paradigm has been replicated, first in the various high-throughput '-omic' fields focused on biomarker discovery, and more recently in the environmental sciences with the advent of 'environment wide association studies' (EWAS) [228-231]. Though such agnostic approaches raise concerns regarding causal interpretation of findings, their systematic nature also enables more straightforward replication [227, 232]. As increasingly sophisticated spatial tools increase the quantity of contextual measures available to neighborhood researchers, neighborhood datasets increasingly resemble GWAS and EWAS datasets. It follows that neighborhood research may similarly benefit by drawing on systematic, agnostic, 'Big Data' research paradigms [233]. For example, empirical variable selection methods such as penalized logistic regression may reveal neighborhood conditions influencing physical activity that conventional approaches have not previously identified [234]. However, with a few exceptions [235, 236], the use of empirical approaches in neighborhood health research has been limited.

In this chapter, we develop the Neighborhood Environment-Wide Association Study (NE-WAS) design, taking explicit inspiration from the EWAS and GWAS approaches. Following the agnostic paradigm, we first test each neighborhood measure as an independent predictor of total physical activity, controlling for individual characteristics. Next, following more recent methodological developments primarily stemming from epigenetic research, we apply machine learning empirical variable selection models to algorithmically identify the most relevant predictors. Finally, we examine the robustness of identified associations to variation in neighborhood definition.

Methods

Subjects and setting

For this exploratory study, we use cross-sectional data from Wave 1 of NYCNAMES-II, a study of residents of New York City aged 65-75. Sampling and recruitment for NYCNAMES-II has been described previously [45]. Telephone recruitment started in May 2011 and was completed in November 2011. The initial two hundred and nineteen (6.2%) surveys were completed by interviewers based at the New York Academy of Medicine. The remaining 3,281 were completed by Abt-SRBI, a survey research firm. During recruitment, a total of 39,792 telephone numbers were selected from a list of telephone numbers purchased from InfoUSA that had previously been geocoded to census tracts. The response rate, calculated as screen-outs (i.e. calls that where a household was contacted and the interviewer was able to ascertain that no older adults lived in the household) plus interviews divided by all numbers selected from the list, was 18%. The co-operation rate, calculated as screen-outs plus interviews divided by screen-outs, interviews and declines, was 31%. Recruitment was terminated when the target 3,500 interviews had been completed; however, the data cleaning process revealed that three wave 1 interviews were with subjects who had previously been interviewed in wave 1. The second interviews for these subjects were discarded, resulting in the final sample of 3,497 adults aged 65-75 living in New York City. Subjects were given a $10 CVS gift card or a $10 check for participating. Surveys were conducted in either English or Spanish.

To maximize power to detect differences due to urban forms and neighborhood racial/ethnic compositions, phone numbers for recruitment were selected using a stratified sampling design. First, all 2,216 census tracts in New York City were classified into one of sixteen strata defined by categories of household income (above or below the 50^{th} percentile of median household income), racial/ethnic mix (more than 75% non-Hispanic Black, more than 75% non-Hispanic White, more than 60% Latino, and ethnically mixed) and walkability (above or below median walkability for New York City). To identify these strata, median household income and racial/ethnic mix were used as reported by the 2006-2010 American Community Survey, and walkability was computed using a previously validated measure incorporating Census measures and tax lot data [237]. For each subject, a sample weights was calculated as the inverse of 1) the probability that the subject's census tract would be randomly selected from among tracts in its stratum times 2) the probability that the subject's household being selected from the number of households containing an adult aged 65-75 estimated by the 2010 Census times 3) the probability of the individual being selected from the number of adults aged 65-75 in the contacted household. Final survey weights were the raked to New York City population estimates from the 2006-2010 American Community Survey for gender and race/ethnicity and from 2010 Census estimates for educational attainment and borough of residence.

Measures

Demographics and Exposure Measures

During the baseline interview, each subject reported his or her sex, age, educational attainment, race/ethnicity, health status, and income. For analysis, we categorized age as 65-68, 69-71, and 72-75 and household income as <$20,000, $20,000-39,999, $40,000-79,999, and ≥$80,000. We categorized educational attainment as less than high school graduate, high school graduate, some college, or college graduate. To maintain a balance of individuals in each racial/ethnic group, we categorized race/ethnicity as non-Hispanic black, non-Hispanic white, Hispanic, and other. We used health status as it was reported: excellent, good, fair, or poor. Because we theorized that the neighborhood environment should only be able to influence physical activity among subjects whose health permitted outdoor physical activity, we excluded those who reported poor health from the primary analysis.

Physical Activity

Physical activity was assessed at baseline using sixteen items derived from the Physical Activity Scale for the Elderly (PASE) [138, 141, 142], a validated survey tool for assessing physical activity among older adults. The PASE instrument assesses past-week physical activity in a number of domains, including strengthening exercises, sports and recreation, walking, gardening, and housework. Exercises, sports and recreation (subdivided as light, moderate, and vigorous), and walking are each assessed with two items, the first assessing frequency of walking bouts and the second assessing average duration walked during each bout; remaining items are assessed as ever- or never-. As detailed in Chapter 3, the PASE score is a linear combination of all sixteen items that reflects total physical activity (r=0.68 with a doubly-labeled water assessment of physical activity in one validation study [141]) but is not directly interpretable in terms of energy expenditure [138, 142].

Neighborhood Measures

During the baseline interview, each subject reported his or her home address. We geocoded these addresses using GeoSupport, a New York City-specific geocoding tool released by New York's Department of City Planning. Ninety-six percent of addresses were successfully geocoded to a rooftop; the remaining four percent were assigned the age 65-74-population weighted centroid of the reported ZIP code as a home location. For each subject, we defined the residential neighborhood as the land area reachable by city streets within a given distance of the geocoded home location, an area referred to as a network buffer and frequently used in neighborhood research [156, 157, 222]. Our primary analysis used ¼ km network buffers; a follow-up analysis described below compared these results to those found using 1 km network buffers.

For each subject, we compiled 337 unique neighborhood measures from a variety of sources. Specifically, demographic and economic characteristics came from the 2006-2010 American Community Survey. Urban form measures were constructed from TIGER/Line shapefiles describing street layout, the New York Metropolitan Transit Authority's ridership reports and a LiDAR scan of the city [137, 238]. Crime and disorder measures were compiled from a measure of crime risk developed by ESRI, Inc., municipal street cleanliness records, an systematic virtual audit using Google Street View imagery, and homicide incident locations as reported by the New York City Police Department to the New York Times [150, 155, 239, 240]. Parks measures, including boundaries and park cleanliness were obtained from The New York City Department of Parks and Recreation [241]. Pedestrian and cyclist injuries counts were compiled from records initially recorded by reporting police officers [242]. Many economic measures make reference to the federal poverty threshold; the threshold was $22,113 for a family of four in 2010 [243].

In order to discern patterns among the forms of support for and barriers to physical activity various neighborhood characteristics provide, we categorized measures into bins according to the aspect of the urban environment each captured (Table 5.1). These bins are analogous to chromosomes in genomic studies, genes in epi-genomic studies, or 'class groupings' in an environment-wide association study

[230].

Some measures, including proportion of residents belonging to certain racial/ethnic minority groups and density of vehicle collisions involving pedestrians, were right-skewed. To be consistent with best practices in agnostic studies and maximize comparability between environmental predictors, we transformed such skewed predictors before analysis [230]. To asses skew for each measure, we visually compared a histogram of the measure and the measure if log-transformed. We retained log-transformed measures for analysis in place of untransformed measures if the log-transformed measure visually appeared closer to a normal distribution than its untransformed analog.

For any pair of measures with spearman correlation coefficients of 1 or -1 (for example, proportion of occupied homes occupied by owners and proportion of occupied homes occupied by renters have a correlation coefficient of -1), we excluded one of the measures. Many measures, particularly as derived from the American Community Survey, are highly but imperfectly correlated with other measures (e.g. because few New York City residents are in the armed forces, proportion of population 16 years and over in the civilian labor force and proportion of the population 16 years and over not in the labor force correlated with an r of -0.99). For this analysis, we excluded only perfectly correlated measures.

Appendix Table 5.1 includes a complete list of all 337 measures used in the final analysis, including their underlying data sources, and whether or not we log-transformed the measure before analysis.

Statistical Analysis

We explored three approaches to identifying the most predictive neighborhood environment variables. First, following analytic techniques used in GWAS studies, we regressed PASE score on each neighborhood environment variable individually to estimate the strength of association between that variable and PASE score. Because we hypothesized that neighborhood environments would affect types of outdoor activity differently, we also used logistic regression for each neighborhood environment variable individually to estimate the strength of association between that variable and engaging in each of three activities: 1) gardening, 2) daily walking, and 3) heavy housework. All regression analyses incorporated survey weights and controlled for individual's age, race/ethnicity, educational attainment, income, and home size.

Second, because many of the predictors used for the regression-based analysis described above were correlated (Figure 5.1), we explored machine learning approaches to empirically select the neighborhood characteristics most predictive of physical activity from combined models. Analogous machine learning approaches are increasingly common in GWAS studies searching for gene-gene interactions [244] and in epigenetic studies [229]. We first explored a least absolute shrinkage and selection operator ('LASSO') regression approach [245]. The LASSO fits a regression model such that the sum of the absolute value of estimated coefficients is less than a constant (the 'penalty'). This penalty shrinks estimates for less relevant coefficients to zero, thereby empirically identifying the remaining variables as a maximally informative subset of variables [246]. All LASSO analyses incorporated sampling weights.

Third, because LASSO regression imposes assumptions about the form of the relationships between variables and outcome (e.g. that predictor effects are additive for continuous outcomes and multiplicative for dichotomous outcomes) we also explored a random forest approach, which allows a more flexible functional form [247]. The random forest algorithm uses randomly selected subsets of observations and randomly selected subsets of variables to build decision trees (i.e. classification trees if the dependent variable is dichotomous, regression trees if the dependent variable is continuous)[248] that best predict the outcome of interest using randomly selected 'bootstrap' samples of the dataset. Decision trees themselves are built using recursive partitioning, an algorithm that repeatedly identifies splits in a continuous covariate that maximizes the Kullback-Leibler divergence (a measure of incremental prediction accuracy) with respect to the outcome variable [248]. Figure 5.2 displays an example of such a decision tree (The original plot, using R and the 'party' package, is given in Appendix Figure A5.2) .For any given set of independent variables, the forest's prediction is the average predicted value of the dependent variable from all trees [249]. This averaging over predictions computed using random selections of observations and variables minimizes model over-fitting [249].

Within the random forest literature, the 'importance' of a given independent variable is defined as the proportion of classification trees including that variable that accurately predicts the observed dependent variable in a cross-validation test set minus the proportion trees including that variable whose value is correct in a subset of the data with the independent variable permuted randomly [250]. Analogously, for regression trees, variable importance is defined by the extent to which trees including that variable minimize observed error minus the error observed when the independent variable is permuted randomly, In effect, the importance measure is an estimate of the extent to which including the variable in the set of variables available for a given tree improves the models predictive accuracy above what would be expected by chance. [249]. Random forests have been used extensively in epigenetic analyses (e.g. [251-253]). In our analysis, random forest models did not incorporate sampling weights owing to lack of software support for weighted analyses.

For both the LASSO and random forest analyses, we used 10-fold cross-validation to tune the algorithmic parameters, such as the penalty constant used for the LASSO, that minimize model predictor error [254]. Such cross-validation is a standard practice in machine learning to minimize model overfitting [255, 256].

For both the regression and empirical variable selection approaches, we identified the five most relevant predictors of each of PASE score as a reflection of overall physical activity. Furthermore, because some types of environmental measures are more plausibly causally linked to some types of activity than others [69, 122], we also used individual items from the PASE measure to measure participation in three specific activity types: daily walking, gardening, and ‘heavy housework’ (vacuuming, sweeping, moving furniture). We hypothesized based on prior literature that daily walking would be associated with measures of urban form [257, 258] and that, because lack of outdoor space poses a barrier to gardening in New York City, gardening would be associated with housing characteristics [259]. We selected heavy housework as a ‘negative control’ [260] – with no *a priori* hypothesis or evidence suggesting that neighborhood conditions should affect participation in heavy housework, finding a large number of correlated neighborhood exposures associated with housework or a pattern of exposures similar to the pattern predictive of other activity measures suggests that residual confounding rather than causality is likely to be responsible for the observed association. We modeled gardening and heavy housework as dichotomous measures as they were included in the original survey. Walking was not assessed with a dichotomous measure in the initial survey; we coded those who reported 5-7 days of past-week walking as daily walkers.

Missing Data

Relatively little data was missing on physical activity (maximum of 1.8% on any PASE item) or demographic covariates (16.2% were missing household income data; no other items were missing for more than 10% of subjects), and no data were missing on a neighborhood covariates. Nonetheless, to address potential non-response biases, we performed all survey-weighted regressions on each of 5 datasets where missing values were imputed using IVEWARE [159]. Following standard practice, we combined the estimates resulting from these models using Rubin's rules [261]. However, because no standard equivalent to Rubin's rules exists to combine variable selection priorities, we used only the first imputed dataset for these steps.

Sensitivity Analyses

To test our regression results' sensitivity to the assumption that neighborhood characteristics were not important for those who reported poor health, we repeated the primary analysis with the full cohort of 3,497 subjects. To test our results' sensitivity to the choice of using the first imputed dataset for the LASSO and random forest analyses, we re-ran all variable selection results with a dataset randomly selected from the other four imputations.

Software

All analyses used 64-bit R for Windows version 3.2.3, including the 'survey' and 'mitools' packages to handle survey weights and combining estimates across imputations [165, 262], the 'glmnet' and 'randomForest' packages for LASSO and random forest models respectively [263, 264], and the 'ggplot2' package to produce Manhattan plots [265].

Results

Two hundred and seventy nine subjects (8%) reported poor health, leaving 3,218 subjects for the primary analysis. PASE scores in the included subjects ranged from 0 to 296 and were slightly right-skewed, with a mean of 84 and a median of 77. Thirty-nine percent of the subjects reported daily walking, twenty-three percent reported gardening, and fifty-seven percent reported doing heavy housework. As compared to the overall cohort, subjects who were excluded from the primary analysis owing to poor health were more likely to be female, to be racial/ethnic minorities, to have lower household incomes, and to be less educated (Table 5.2).

Characteristics Associated with Physical Activity

In linear regression models controlling for individual covariates, measures of neighborhood resident socioeconomic position and disorder were most strongly associated with total physical activity (Figure 5.3). Specifically, the proportion of residents living in households with incomes less than half the poverty level was the most strongly associated with PASE score, with an estimated decrease of 0.85 PASE score units (95% CI: 0.56, 1.14) per 1% increase in proportion of residents living in households with incomes less than half the federal poverty line, equivalent to 10 minutes of daily walking per 4% decrease in proportion of residents living below half the federal poverty line. The remaining four of the top five measures included three other measures of resident socioeconomic position, all showing correlations between more higher-income residents and more physical activity, and one disorder measure, showing well-maintained windows, a marker of building upkeep, to be correlated with more activity (Table 5.3). After Bonferroni correction, no measure of resident demographics, parks, urban form, or pedestrian and cyclist safety were associated with PASE score.

Logistic regression analyses focused only a single type of activity identified many more significant neighborhood correlates than analyses targeting total activity (Figure 5.3, Table 5.3). Measures of high neighborhood socioeconomic status were the strongest predictors of gardening, whereas a wide range of neighborhood characteristics predicted walking 5-7 days in the previous week (Table 5,3). Reassuringly, no neighborhood measures were predictive of heavy housework after Bonferroni correction.

Machine Learning for Neighborhood Variable Selection

As in the regression approach, no neighborhood variables were informative with respect to heavy housework and only a few neighborhood variables remained in the models that best predicted overall PASE score using LASSO regression. Similarly, the LASSO retained many more neighborhood variables when predicting gardening and walking. However, the specific predictors identified by LASSO and random forest approaches were different both from each other and from those identified by unpenalized regression (Table 5.4). In general, predictors that remained non-zero in LASSO models represented different domains of neighborhood influence, consistent with these aspects of neighborhoods operating as independent but minor influences. For example, the top five predictors retained within the LASSO predicting gardening included one measure of household poverty, two measures of urban form, a measure of physical disorder, and a measure of resident employment.

Unlike the LASSO, the neighborhood measures that were most important in in Random Forest models were clustered within domains of neighborhood influences. For example, four of the five most important neighborhood measures included in random forests predicting gardening were measures of housing characteristics (Table 5.5). Random Forest models rank importance of variables rather than select a specific subset of variables as important, so it is not possible to make a direct comparison of the count of variables deemed important in Random Forest analyses compared to LASSO models.

Neighborhood Characteristics and Buffer Size

Using 1km rather than 0.25 km buffers led to more variables being significant after Bonferroni correction, but neither 1 km buffers nor 0.25km buffers were uniformly more strongly correlated with total PASE score (Figure 5.4). Of the 337 neighborhood measures available at both scales, 38 (11%) changed signs between scales, though none of the measures whose coefficient changed signs were nominally significant at a p-value of 0.05 at either scale. There was no clear pattern as to which measures were better correlated at which scales (Table 5.6).

Sensitivity Analyses

The sensitivity analysis conducted using the full cohort identified the same top 5 measures, albeit in a different order (Appendix Table 5.2). The sensitivity analysis applying the LASSO regression to randomly selected subjects resulted in a different set of variables being selected, though the variables that ranked highly in regression analysis were often selected in the sensitivity analysis as well (Appendix Table 5.3)

Discussion

In this analysis, we explored a novel agnostic 'NE-WAS' approach to selecting the neighborhood measures most strongly associated with total physical activity, as well as specifically with walking, gardening, and housework. Identifying such measures can guide both the development of theory and the development of interventions. In our study, the most strongly predictive measure of total physical activity was proportion of residents living in households with incomes below half the federal poverty threshold, equating to $11,056 for a family of four. Neighborhood socioeconomic and disorder measures were most associated with total activity. Socioeconomic measures also strongly predicted gardening, whereas measures of commute distance and commute times were more relevant for walking. As expected, no neighborhood measures significantly predicted housework. Overall, the NE-WAS approach appears promising as a means of identifying neighborhood level contextual factors associated with physical activity. More broadly, NE-WAS may be appropriate for other neighborhood-associated health behaviors and outcomes as well, such as obesity [258] or cardiac arrest [266].

We found that more neighborhood environment measures were significantly associated with walking and gardening than with physical activity as a whole, and no neighborhood measures were associated with housework. Moreover, this pattern was present not only in adjusted bivariate regression models but also in the counts of variables selected by the LASSO approach, This implies that more neighborhood measures predict a non-trivial proportion of the variation in outcomes when outcomes are domain-specific, which is consistent with many prior calls to consider influences on separate domains of activity separately or to assess only activity that might plausibly be influenced by neighborhood when considering neighborhood conditions as a predictor of activity. [42, 43, 69, 183, 267]. Our findings may thus serve as empirical support for an argument frequently made on theoretical grounds alone. However, this finding may also reflect the initial compilation of measures from which we drew our candidates having been selected to study walkability and walking. Future NE-WASes might profitably include more neighborhood measures predictive of alternate sources of activity such as prevalence of hardware stores as a predictor for home repair activity. Modern GIS tools and readily available administrative data make it possible to create novel contextual measures inexpensively.

There are several interpretations for our finding that neighborhood socioeconomic measures were more consistently associated with activity measures than measures with more direct theoretical relevance to specific forms of outdoor activity, such as access to parks. It may be that residents of higher socioeconomic status neighborhoods have used their resources to shape neighborhoods to offer more support for different forms of activity among older adults [268], including dedicated outdoor space that supports gardening and well-maintained sidewalks or amenities such as benches and public restrooms that older adults cite as necessary to supports for walking [269]. A complementary explanation is that the association with neighborhood socioeconomic status is an artifact of residual confounding due to incomplete control for individual socioeconomic position. Higher socioeconomic position older adults are typically more physically active [67, 105], and tend to live in neighborhoods with other high socioeconomic position individuals. We controlled for individual income and educational attainment, but neither fully captures socioeconomic position among older adults [270];statistical control for an imperfectly measured confounder is incomplete [271].

There were substantial differences between the predictors selected by cross-validated empirical variable selection algorithms (‘machine learning’ approaches) and the predictors selected by comparing coefficients from sequentially fit regression models. These differences arise because variable selection algorithms aim to identify the subset of variables with the greatest predictive power when taken together, comparable to stepwise regression, whereas bivariate models adjusted only for individual covariates do not take correlations between neighborhood measures into account [272]. Both approaches are used in agnostic analysis of genes and environment measures, and as agnostic data analysis generally becomes more commonplace, it will be important to explore the inferential and interpretational benefits and drawbacks of each approach. It may be, for example, that sequential regression is more valuable for comparing types of measures (e.g. to ask whether neighborhood socioeconomic characteristics are more associated with activity than park access is) but that empirical variable selection is more valuable for selecting the most informative subset of measures from among many correlated measures [244, 273].

However, we caution that results from empirical variable selection algorithms must be interpreted with full understanding of the algorithms’ workings. In general, LASSO regression resulted in selecting variables representing different domains of neighborhood measurement, whereas the Random Forest models identified multiple measures of the same domain as most informative. The LASSO approach assumes an underlying multivariable linear model [245], which in turn not only implies a linear functional form but also selects a subset of variables that best predict variation in the outcome that is not explained by other covariates. By contrast, the decision trees grown using recursive partitioning in a Random Forest categorize each independent variable independently, and not all predictor variables are available for each tree [249], implying that several inter-correlated neighborhood measures, often drawn from the same domain, may substitute for each other when one is not present or together allow a more flexible functional form. These differences in model fit may explain why the LASSO approach selected variables from different neighborhood measurement domains whereas the Random Forest selected several highly correlated socioeconomic measures as the most important variables.

While the NE-WAS approach explicitly draws an analogy between genetics and neighborhoods, we caution, as others have, that there are vital differences between genomes and modifiable exposures like neighborhoods [232]. Most importantly, unlike genetic SNPs, wherein there are few if any correlations between polymorphisms on separate chromosomes, the correlation structure underlying neighborhood characteristics is strong, complex, and potentially causally circular [125]. However, in proteomics and metabolomics research, wherein measured molecules do show strong and complex inter-correlations, identified molecules are considered to be markers of a process rather than causes of the process and a separate scientific approach, pathway analysis, has developed to integrate knowledge from agnostic analyses to develop and test causal hypotheses [274, 275]. There are analogous systems science-derived integrations of knowledge in neighborhood research (e.g. [276]), though such approaches are still in their infancy. Nonetheless, we anticipate that in this sense the NE-WAS approach is more akin to an –omics approach than a GWAS: the value of the NE-WAS stems not from a precise estimate of the causal effect of some neighborhood characteristic but rather from the ability to systematically identify targets for future exploration and to reveal reproducible patterns in associations across cohorts [227].

We emphasize that while analysis addressed physical activity the principles underlying the NE-WAS approach could be applied to studies of other outcomes that may be related to neighborhood environments. For example, a NE-WAS exploring neighborhood variables associated with crime victimization, psychopathology, respiratory disease, and obesity might be productive [50, 277-279]. Relatedly, NE-WAS may be of value for standardizing neighborhood definitions, as many reviewers of neighborhood effects literature have recommended [69].

This study had several notable strengths. First, the relatively large and population-based sample of older adults residing in a very well-characterized urban environment allowed for relatively precise estimates of associations between neighborhood characteristics and activity outcomes. Second, the use of a survey measure that included items assessing types of activity allowed us to incorporate analyses that target activity measures representing a range of hypothesized susceptibility to neighborhood influence [42, 69]. Third, without a theoretical basis to guide variable selection, agnostic studies are at risk of identifying variables associated with the outcome of interest only due to confounding. In this light, our null finding with respect to the heavy housework outcome serving as our negative control provides some, albeit incomplete, evidence against residual confounding.

However, our results should be viewed in light of limitations as well. First, the 337 neighborhood measures we analyzed were by no means comprehensive or systematic. Rather, befitting such an exploratory approach, they represent a selection of previously developed measures of New York City's urban environment that were available to explore the approach. Future NE-WASes might productively undertake a systematic exploration of neighborhood measures used in the literature to select a more comprehensive set of measures to study, potentially incorporating neighborhood measures of no theoretical relevance as further negative controls. For example, in one prior analysis of childhood obesity, the presence of banks was shown to predict lower BMI in spite of the lack of a plausible mechanism through which banks might cause obesity. This result suggested that residual confounding was responsible for a similar observed association between more neighborhood fast food outlets and lower BMI [280]. Second, we compared only two neighborhood definitions, 0.25 km network buffers and 1.0 km network buffers. It has been repeatedly noted that no single definition captures the construct of a neighborhood [69, 153]; indeed, the meaning of neighborhood may be different for different measures, in different contexts, and for different subgroups [267]. Third, while New York City as a whole comprises a broad range of urban environments, including pockets of sidewalk-free post-war 'sprawl', it nonetheless contains a much more pedestrian-oriented environment than the United States as a whole, and a population at greater extremes of the socioeconomic spectrum. It may be productive to compare results from this NE-WAS to future NE-WASes conducted in environment more representative of the contexts in which most older adults reside. Finally, as in chapters 3 and 4, we were unable to determine whether statistical adjustment for participant race/ethnicity and socioeconomic status fully account for residential self-selection [211].

In conclusion, the NE-WAS is a promising approach to empirically identify neighborhood measures most strongly related to outcomes of interest. In our NE-WAS, neighborhood socioeconomic characteristics were more consistently associated with physical activity than measures of crime, parks, and pedestrian safety. We anticipate performing NE-WASes in other cohorts and other geographic contexts to determine the replicability of the approach and substantive findings [227].

Tables

Table 5.1. Summary of measures used in the NE-WAS

Domain	Number of measures	Data source(s)	Examples
Demographics and housing characteristics	121	American Community Survey	Population Density, % white alone, % boys aged 10-14
Education, Employment, and Income	102	American Community Survey	% College grad, % in labor force, % in food prep sector
Urban Form and Walkability	50	American Community Survey, New York State Accident Location Information Service Line Layer, NYC Transit Authority	% walk to work, density of 4-way intersections, Bus stop density,% of roadbed covered by tree canopy
Crime and Disorder	35	Esri Crime Risk, Google Street View, New York Times Homicide Map, NYC Sanitation Department Report Cards	Weighted average risk of larceny, Mean neighborhood disorder, % filthy streets
Parks	5	New York City Department of Parks and Recreation	% of land area in large parks
Pedestrian Safety	24	New York State Department of Transportation and New York City Police Department	Cyclist injury density in the 1990s, Pedestrian fatality density in the 2000s

Table 5.2. Selected characteristics of the subjects included in this analysis

	Full Cohort (N=3,497)			Fair or better health (N=3,218)		
Characteristic	**N**	**%**	**Sample weighted %**	**N**	**%**	**Sample weighted %**
Age						
65-68	1045	33	34	956	33	33
69-71	664	21	20	608	21	21
72-75	1442	46	46	1335	46	47
Sex						
Female	2094	60	58	1907	59	57
Male	1403	40	42	1311	41	43
Race/Ethnicity						
Non-Hispanic White	1800	51	47	1701	53	48
Non-Hispanic Black	1073	31	26	974	30	26
Hispanic	245	7	14	209	6	13
Other	379	11	12	334	10	12
Educational Attainment						
Less than High School	673	19	32	570	18	30
Completed High School	949	27	29	870	27	29
Some College	627	18	15	570	18	15
Completed College	1248	36	24	1208	38	25
Household Income						
Less than $20,000	1279	37	40	1097	34	37
$20,000-40,000	842	24	24	790	25	24
$40,000-80,000	745	21	21	711	22	22
More than $80,000	631	18	16	620	19	17
Health Status						
Excellent	645	18	17	645	20	18
Good	1523	44	42	1523	47	46
Fair	1050	30	33	1050	33	36
Poor	279	8	9	--	--	--
Activity Measures						
Walked 5-7 days in the last week	1346	38	42	1154	36	39
Gardened in the last week	710	20	23	686	21	24
Performed heavy housework in the last week	1872	54	54	1793	56	57

Table 5.3. Specific neighborhood measures identified as most predictive for several physical activity outcomes using Linear Regression. All analyses control for subject age, race/ethnicity, educational attainment, household income, gender, and home type.

	PASE Score	Gardening	Walking Daily	Heavy Housework
Count of measures that remained significant after Bonferroni correction	5 (1.5%)	33 (9.8%)	49 (14.4%)	0 (0.0%)
Top 5 statistically significant neighborhood measures (by p-value of coefficient)	People living in households with incomes less than half the poverty level (-)	People living in households with incomes less than half the poverty level (-)	Proportion of residents with 60-90 minute travel time to work (-)	--
	People living in households with incomes below the poverty line (-)	Neighborhood Physical Disorder (-)	Broken windows in HVS survey (-)	--
	No problems with windows in HVS survey (+)	People living in households with incomes below the poverty line (-)	Proportion of adult population with at least some college education (+)	--
	People living in households with incomes more than twice the poverty level (+)	People living in households above twice the poverty line (+)	Proportion of working adult population commuting by car, truck, or van (-)	--
	People living in households with incomes between half and three-quarters of the poverty level (-)	People living in households with any interest, dividend, or rental income (+)	Proportion of adult population working in professional or management industries (+)	--

HVS: New York City Housing and Vacancy Survey

Table 5.4. Specific neighborhood measures identified as most predictive for several physical activity outcomes using Least Absolute Shrinkage and Selection Operator ('LASSO') All analyses control for subject age, race/ethnicity, educational attainment, household income, and gender.

	PASE Score	Gardening	Walking Daily	Heavy Housework
Number of neighborhood predictors retained in model minimizing RMSE	3 (0.8%)	45 (12.7%)	22 (6.3%)	0 (0.0%)
Top 5 predictors as ranked by normalized regression coefficient after applying the LASSO penalty	People living in households with incomes less than half the poverty level (-)	Proportion of families that both have incomes below the poverty line and no children under age 18. (-)	Proportion of population living in group quarters (+)	
	Proportion of employed adults working in sales and related occupations (-)	Density of intersections that are cul-de-sacs (+)	Proportion of population that are 5-9 year old males (-)	
	People living in households with incomes between half and three-quarters of the poverty level (-)	No issues with windows reported in HVS survey (+)	Proportion of households with a Hispanic or Latino householder (-)	
		Proportion of employed adults working in farming, fishing, or forestry (-)	Proportion of families that both have incomes below the poverty line and no children under age 18. (-)	
		Proportion of employed adults with commute times of 20 to 29 minutes (-)	Proportion of population that are women aged 35-44. (-)	

HVS: New York City Housing and Vacancy Survey

Table 5.5. Specific neighborhood measures identified as most important by the Random Forest algorithm for several physical activity outcomes.

All analyses incorporate subject age, race/ethnicity, educational attainment, household income, and gender as predictors in the random forest algorithm. Where individual covariates were among the most important as ranked by the algorithm, they were not included in this table. Note that, unlike the LASSO, no direction of association is calculated from the model fit, because random forests predict by recursive partitioning rather than model-based estimation

PASE Score	Gardening	Walking Daily	Heavy Housework
Proportion of housing units that are single-family housing	Proportion of housing units that are single-family housing	Proportion of residents who are males aged 75-84	Proportion of residents who are males aged 10-14
Proportion of housing units that are detached single-family housing	Proportion of housing units that are duplexes	Density of collisions involving bicycles and resulting in an injury or fatality between 2000 and 2009	Proportion of households that are headed by a married couple with no children
Proportion of housing units that are apartment buildings with 50 or more units	Proportion of housing units that are detached single-family housing	Proportion of adults who completed at least some college	Proportion of households with some income other than wages, self-employment income, interest or dividends, social security, supplemental security, public assistance, or retirement income
Proportion of households with incomes between 10K and 15K	Proportion of housing units that are apartment buildings with 50 or more units	Proportion of households with income over 200K	Proportion of population that functions as the head of their household
Proportion of adults whose marital status is separated.	Proportion of housing units that are apartment buildings with 20-49 units	Density of collisions involving bicycles and resulting in an injury between 2000 and 2009	Proportion of households with income over 200K

Table 5.6. The top 5 measures for each neighborhood definition that were more significant predictors of total PASE score than using the alternate neighborhood definition. Plus or minus indicates the direction of association between the neighborhood measure and PASE score.

	Measure	Difference in –$\log_{10}$ p-value
Better at 0.25 km scale	Proportion of population living in households (+)	2.24
	Proportion of population who are naturalized citizens (-)	1.68
	Proportion of population living in households with income below half the poverty level (-)	1.57
	Density of 3-way intersections (+)	1.28
	Proportion of vacant housing units offered for rent (-)	1.27
Better at 1.0 km scale	Proportion of households with incomes between 25K and 30K (-)	2.77
	Proportion of adult residents with a professional degree or more (+)	2.56
	Proportion of households with incomes between 30K and 35K (-)	2.56
	Proportion of family households living below the poverty line with a male householder and no children under age 18 (-)	2.45
	Proportion of total population aged 10 to 14 years	2.42

Figures

Figure 5.1. Histogram of absolute value of pairwise spearman correlation coefficients computed between 337 neighborhood measures analyzed.

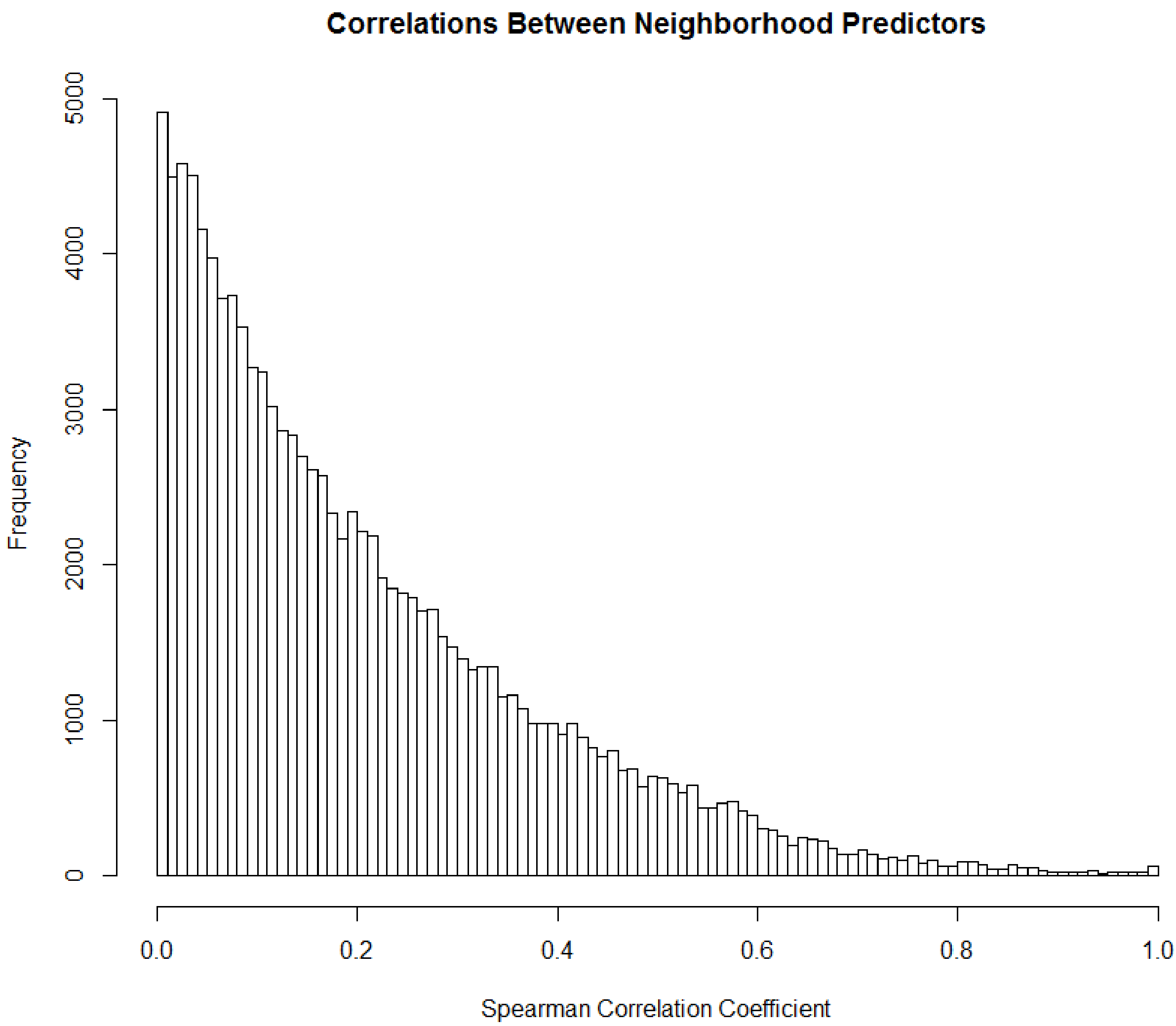


Figure 5.2: Example of a decision tree computed using a recursive partitioning algorithm to predict probability of reporting gardening as a function of neighborhood-level and individual-level covariates among subjects in NYCNAMES-II who reported fair or better health. Probabilities in gray boxes reflect the estimated probability of reporting gardening conditional on observing the covariates in white boxes to have values reflected on the arrows under the boxes. For example, a subject living in a neighborhood wherein more than 51.6% of homes are owner-occupied and more than 14.3% of neighborhood residents live in apartment buildings with 50+ units has an estimated 25.9% probability of reporting gardening. Random forests average estimated probabilities from a group of such decision trees to compute an overall estimated probability.

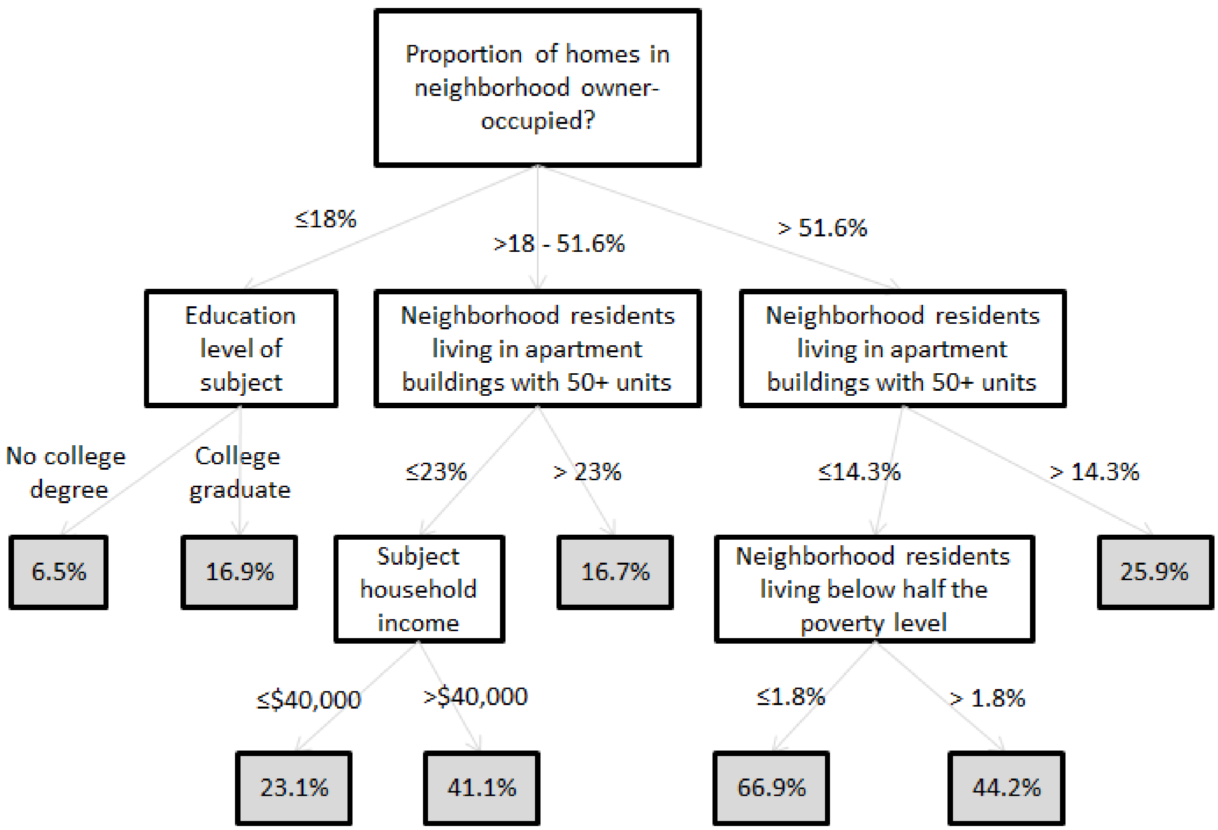


Figure 5.3. Manhattan Plots showing the strengths of correlations between individual neighborhood variables and various physical activity outcomes as measured by the Physical Activity Scale for the Elderly (PASE) after controlling for age, sex, race/ethnicity, educational attainment, income, and housing type.

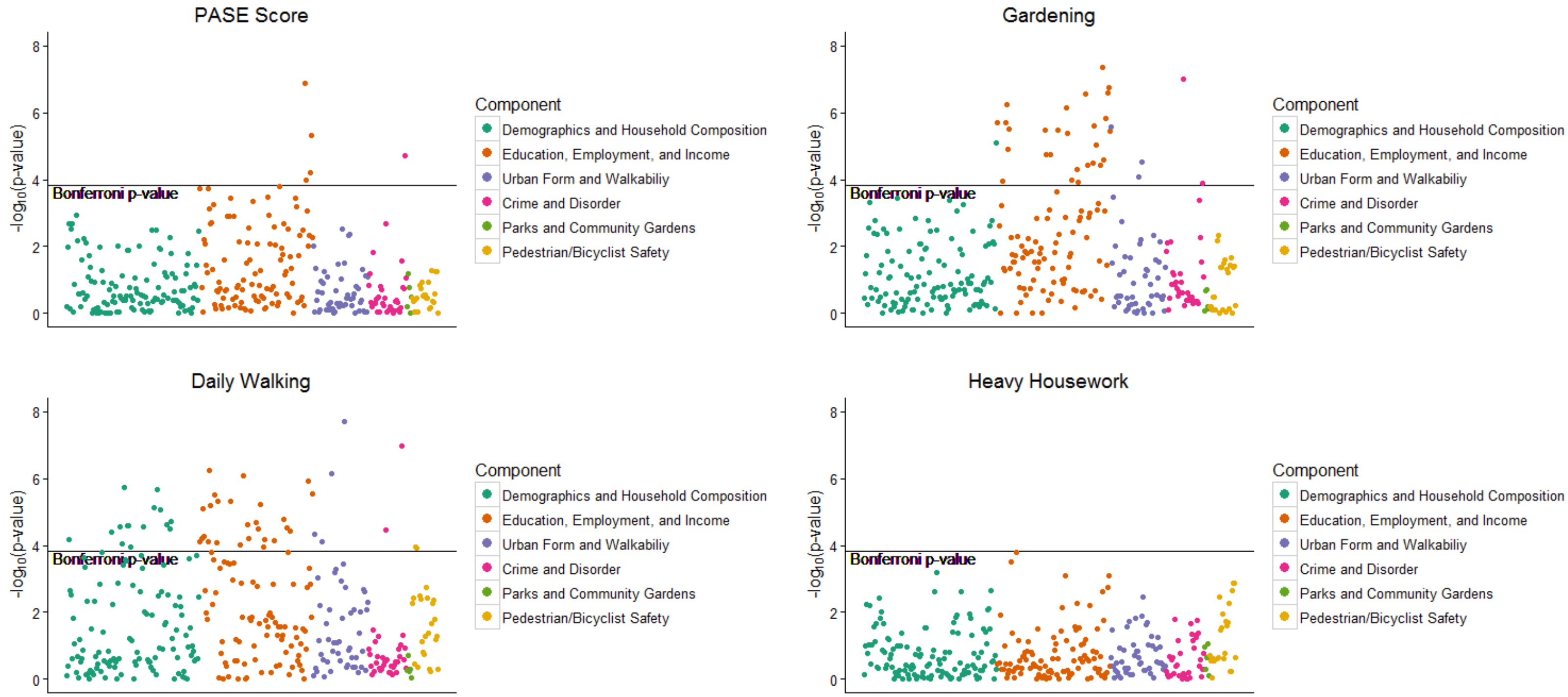
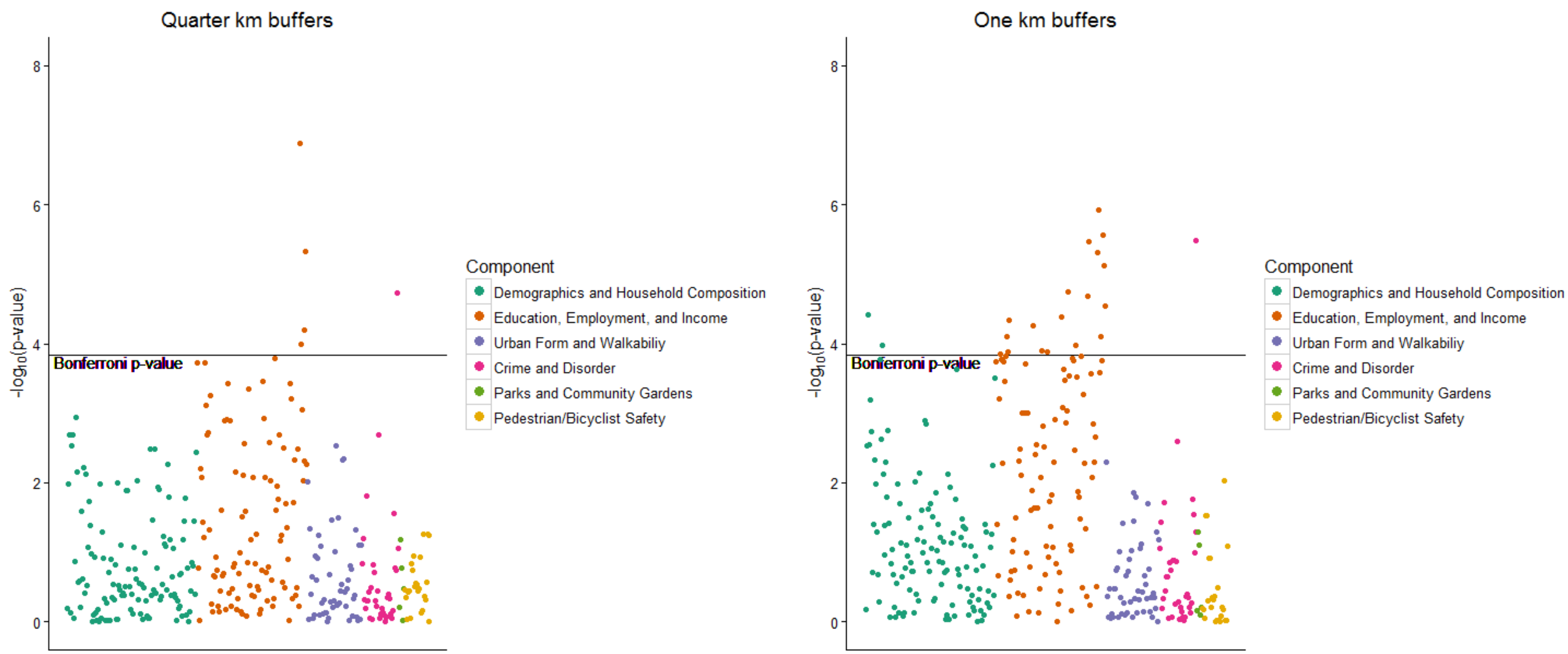


Figure 5.4. Manhattan Plots showing the statistical significance of correlations between individual neighborhood variables total Physical Activity Scale for the Elderly (PASE) score at smaller and larger buffer sizes.



Chapter 6. Conclusion

This overarching goal of this dissertation was to explore neighborhood conditions as an influence on physical activity among older adults. We first reviewed and synthesized the scholarly literature on disorder, a neighborhood condition of particular theoretica; interest, as an influence on activity. We found that a promising direction for neighborhood disorder research was to focus more on vulnerable subgroups and on aspects of activity most likely to be affected by disorder. We then conducted two complementary longitudinal analyses of neighborhood disorder as an influence on activity, one exploring predictors of total physical activity and one identifying patterns of activity and exploring predictors of those patterns. Finally, drawing explicit inspiration from Genome-Wide Association Study (GWAS) and Environment-Wide Association Study (EWAS) approaches, we developed the analogous Neighborhood Environment-Wide Association Study (NE-WAS) design to systematically neighborhood conditions' associations with physical activity. Below, we summarize our core findings and present suggestions for future research on how physical activity is affected by neighborhood conditions, focusing particularly on disorder.

Findings

Our systematic literature search, described in Chapter 2, identified 28 peer-reviewed English language articles on neighborhood disorder as an influence on physical activity. These studies mostly focused on populations in North America, Europe, and Australia. There was some evidence that less severe indicators of disorder, such as litter and graffiti, may also mark areas with more pedestrian activity, negating or reversing the expected inverse association between disorder and physical activity. By contrast, more severe indicators of disorder such as dilapidated or abandoned buildings may discourage discretionary outdoor activity, particularly among vulnerable populations such as women and older adults. We concluded that future studies of disorder and activity are warranted, particularly studies using longitudinal data, that incorporate validated and internally consistent disorder measures, and that focus on the activity domains and sub-populations that are most likely to be affected by disorder.

In Chapter 3, we addressed some of the gaps identified in Chapter 2, using a longitudinal analysis to estimate the association between neighborhood disorder and total physical activity among a cohort of older adult residents of New York City. In multivariable mixed regression models accounting for individual and neighborhood factors, for missing data, and for loss to follow-up, each standard deviation increase in neighborhood disorder was associated with an estimated 3.0 units (95% CI: 1.9, 4.2) lower PASE score at baseline, or the equivalent of about 10 minutes of walking per day. There was no significant interaction between physical disorder and changes in PASE score over two years of follow-up. We concluded that future longitudinal research, particularly with longer periods of follow-up in which more subjects changing residences, might productively explore how these disparities in activity by neighborhood disorder emerge.

Chapter 4 expanded on the definition of physical activity used in Chapter 3 by applying a latent transition analysis to identify patterns of types of physical activity the same cohort of older adults engaged in. Using this method, which to the best of our knowledge has never been used to explore changing patterns of activity types among other adults, we identified seven latent classes of activity. Of these seven classes, three pairs of classes were roughly equivalent except for participation in exercise. About three quarters of subjects remained within each latent class between waves; most transitions that did occur were between classes defined by exercise to the parallel class without exercise or vice-versa . More neighborhood disorder was modestly associated with moving out of a sports and recreation class (Relative Risk = 1.27, 95% CI = 1.00, 1.61 between waves 1 and 2, Relative Risk = 1.28, 95% CI = 0.85, 1.93 between waves 2 and 3) potentially offering a mechanism for the disparity observed in Chapter 3. However, estimates were too imprecise to rule out chance associations.

Finally, in Chapter 5, we developed an agnostic approach to systematically exploring the plethora of neighborhood measures available to modern researchers equipped with geographic information systems (GIS) software. We found that only neighborhood socioeconomic status and disorder measures were associated with total activity and gardening, whereas a broader range of measures was associated with walking. This likely reflected a neighborhood measure set initially collated with walkability in mind. We further found that fitting one survey-weighted regression model for each measure resulted in more interpretable comparisons between measures than empirical variable selection approaches. The NE-WAS approach appears to be a promising way to keep neighborhood research systematic as the scale of available spatially located administrative data continues to grow.

Implications for Research and Practice

These results have several implications for research and policy regarding neighborhood disorder and physical activity. First, all three empirical approaches found inverse associations between disorder and physical activity, consistent with prior findings from the systematic review suggesting that disorder may act as a barrier to activity among older adults. However, our findings lack a strong causal interpretation owing to their observational nature, the risk of residential self-selection, and the low prevalence of changes in disorder these adults experienced over only two years. Nonetheless, these findings justify further research in disorder as a cause of activity, particularly using study designs focused on exogenous *changes* in disorder levels, as after natural disasters or as a result of municipal blight removal policies. The city of Detroit, which suffers from substantial abandonment and has recently invigorated its blight removal program, might be a promising context for such research, as may certain smaller but similarly depopulating post-industrial cities such as Youngstown, OH or Gary, IN. Studies focused on individuals who relocate residences may also be promising, albeit with substantial concerns about residential self-selection.

More broadly, as more municipalities embrace a Health in All Policies approach to governance [281], an understanding that disorder may act as a barrier to physical activity might bring more municipal departments together. Whereas blight removal programs are controversial when understood as undertaken for economic reasons [282], it may be that they would receive broader support if understood as improving health as well. Similarly, broadening the understanding of the role that garbage collection plays in public health from being solely infectious disease vector control to potentially preventing chronic disease by encouraging health-promoting behavior such as physical activity may solidify support for sanitation among policymakers.

Second, whereas the longitudinal analysis provided evidence recapitulating prior cross-sectional findings that disorder is linked to lower levels of activity but limited evidence for the emergence of such disparities, the latent transition analysis offered a potential mechanism. Specifically, the latent transition analysis suggested that residents of more disordered neighborhoods might abandon exercise more frequently without initiating exercise with any greater frequency than residents of ordered neighborhoods. Further analyses, relating neighborhood characteristics to *types* of activity, may plausibly be performed as secondary analyses of already well-characterized cohort studies such as the Black Women's Health Study or the Multi-Ethnic Study of Atherosclerosis. If the necessary data are indeed already collected in these cohorts, such analyses appear to be an inexpensive yet promising direction for future research.

Finally, the NE-WAS yielded intriguing insights about the relative strength of associations between measurable aspects of neighborhoods and physical activity. While, consistent with the GWAS research paradigm, replication in additional contexts is necessary before conclusions can be drawn, the patterns observed here, that socioeconomic and disorder measures were more correlated than parks and pedestrian safety measures, could help to focus a research and policy agenda for activity promotion if they prove replicable. Such replications might require substantial geographic information systems (GIS) work but need not require collecting new human subjects data if they can be applied to existing cohorts; thus, they too ought also to be relatively inexpensive.

In conclusion, we paired two complementary yet distinct theory-driven analyses with an agnostic 'NE-WAS' approach. In these three studies we found some indications but not conclusive evidence that disorder may act as a barrier to physical activity in one cohort of older adults. More research regarding the effects of neighborhood conditions on physical activity among older adults is warranted, perhaps using NE-WAS approaches to best target theory-driven approaches, as has been suggested in the genomics research paradigm..

References

1. Vogel, T., et al., *Health benefits of physical activity in older patients: a review.* International Journal of Clinical Practice, 2009. **63**(2): p. 303-320.
2. Warburton, D.E., C.W. Nicol, and S.S. Bredin, *Health benefits of physical activity: the evidence.* Canadian medical association journal, 2006. **174**(6): p. 801-809.
3. Fried, L.P., et al., *Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study.* Jama, 1998. **279**(8): p. 585-592.
4. Gregg, E.W., M.A. Pereira, and C.J. Caspersen, *Physical activity, falls, and fractures among older adults: a review of the epidemiologic evidence.* Journal of the American Geriatrics Society, 2000. **48**(8): p. 883-893.
5. Iwamoto, J., T. Takeda, and S. Ichimura, *Effect of exercise training and detraining on bone mineral density in postmenopausal women with osteoporosis.* Journal of orthopaedic science, 2001. **6**(2): p. 128-132.
6. Knoops, K.T., et al., *Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project.* Jama, 2004. **292**(12): p. 1433-1439.
7. Landi, F., et al., *Physical activity prevented functional decline among frail community-living elderly subjects in an international observational study.* Journal of Clinical Epidemiology, 2007. **60**(5): p. 518-524.
8. Lautenschlager, N.T., et al., *Effect of physical activity on cognitive function in older adults at risk for Alzheimer disease: a randomized trial.* Jama, 2008. **300**(9): p. 1027-1037.
9. Peterson, M.J., et al., *Physical activity as a preventative factor for frailty: the health, aging, and body composition study.* The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2009. **64**(1): p. 61-68.
10. Rockwood, K. and L. Middleton, *Physical activity and the maintenance of cognitive function.* Alzheimer's & Dementia, 2007. **3**(2): p. S38-S44.
11. Sherrington, C., et al., *Effective exercise for the prevention of falls: a systematic review and meta-analysis.* Journal of the American Geriatrics Society, 2008. **56**(12): p. 2234-2243.
12. Strawbridge, W.J., et al., *Physical activity reduces the risk of subsequent depression for older adults.* American Journal of Epidemiology, 2002. **156**(4): p. 328-334.
13. Sundquist, K., et al., *Frequent and occasional physical activity in the elderly: a 12-year follow-up study of mortality.* American journal of preventive medicine, 2004. **27**(1): p. 22-27.
14. Carlson, S.A., et al., *Trend and Prevalence Estimates Based on the 2008 Physical Activity Guidelines for Americans.* American journal of preventive medicine, 2010. **39**(4): p. 305-313.
15. Centers for Disease Control and Prevention. *Healthy Aging & the Built Environment*. 2012 [cited 2014 April 29].
16. Murray, C.J., et al., *The state of US health, 1990-2010: burden of diseases, injuries, and risk factors.* Jama, 2013. **310**(6): p. 591-606.
17. Heath, G.W., et al., *Evidence-based intervention in physical activity: lessons from around the world.* The lancet, 2012. **380**(9838): p. 272-281.
18. King, A.C., W.J. Rejeski, and D.M. Buchner, *Physical activity interventions targeting older adults: A critical review and recommendations.* American journal of preventive medicine, 1998. **15**(4): p. 316-333.
19. Brownson, R.C., T.K. Boehmer, and D.A. Luke, *Declining rates of physical activity in the United States: what are the contributors?* Annu. Rev. Public Health, 2005. **26**: p. 421-443.
20. Marcus, B.H., et al., *Physical activity behavior change: issues in adoption and maintenance.* Health Psychology, 2000. **19**(1S): p. 32.
21. Sallis, J.F., et al., *An ecological approach to creating active living communities.* Annu. Rev. Public Health, 2006. **27**: p. 297-322.
22. Stokols, D., *Establishing and maintaining healthy environments: toward a social ecology of health promotion.* American Psychologist, 1992. **47**(1): p. 6.
23. Krieger, N., *Epidemiology and the web of causation: has anyone seen the spider?* Social science & medicine, 1994. **39**(7): p. 887-903.
24. Sallis, J.F., N. Owen, and E.B. Fisher, *Ecological models of health behavior.* Health behavior and health education: Theory, research, and practice, 2008. **4**: p. 465-486.

25. Susser, M. and E. Susser, *Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology.* American journal of public health, 1996. **86**(5): p. 674-677.
26. Stokols, D., *Translating social ecological theory into guidelines for community health promotion.* American journal of health promotion, 1996. **10**(4): p. 282-298.
27. Sallis, J., A. Bauman, and M. Pratt, *Environmental and policy interventions to promote physical activity.* American journal of preventive medicine, 1998. **15**(4): p. 379-397.
28. Pikora, T., et al., *Developing a framework for assessment of the environmental determinants of walking and cycling.* Social science & medicine, 2003. **56**(8): p. 1693-1703.
29. Suglia, S.F., et al., *Why the Neighborhood Social Environment Is Critical in Obesity Prevention.* Journal of Urban Health, 2016: p. 1-7.
30. Sampson, R.J. and S.W. Raudenbush, *Seeing disorder: Neighborhood stigma and the social construction of "broken windows".* Social psychology quarterly, 2004. **67**(4): p. 319-342.
31. Sampson, R.J. and S.W. Raudenbush, *Systematic social observation of public spaces: A new look at disorder in urban neighborhoods.* American Journal of Sociology, 1999. **105**(3): p. 603-651.
32. Mendes de Leon, C.F., et al., *Neighborhood Social Cohesion and Disorder in Relation to Walking in Community-Dwelling Older Adults A Multilevel Analysis.* Journal of Aging and Health, 2009. **21**(1): p. 155-171.
33. Wilson, J.Q. and G.L. Kelling, *Broken windows.* Atlantic monthly, 1982. **249**(3): p. 29-38.
34. Keizer, K., S. Lindenberg, and L. Steg, *The spreading of disorder.* Science, 2008. **322**(5908): p. 1681-1685.
35. Latkin, C.A. and A.D. Curry, *Stressful neighborhoods and depression: a prospective study of the impact of neighborhood disorder.* Journal of health and social behavior, 2003: p. 34-44.
36. Lopez, R.P. and H.P. Hynes, *Obesity, physical activity, and the urban environment: public health research needs.* Environmental Health, 2006. **5**(1): p. 25.
37. Van Cauwenberg, J., et al., *Environmental factors influencing older adults' walking for transportation: a study using walk-along interviews.* Int J Behav Nutr Phys Act, 2012. **9**(1): p. 85.
38. Lorenc, T., et al., *Crime, fear of crime, environment, and mental health and wellbeing: Mapping review of theories and causal pathways.* Health & place, 2012. **18**(4): p. 757-765.
39. Kihl, M., *Livable communities: An evaluation guide*. 2005: AARP Public Policy Institute.
40. Association, A.L.G., *Age-friendly built environments: Opportunities for local government*. 2006: Australian Local Government Association.
41. Bancroft, C., et al., *Association of proximity and density of parks and objectively measured physical activity in the United States: a systematic review.* Social Science & Medicine, 2015. **138**: p. 22-30.
42. Lovasi, G.S., S. Grady, and A. Rundle, *Steps forward: review and recommendations for research on walkability, physical activity and cardiovascular health.* Public health reviews, 2012. **33**(4): p. 484.
43. Foster, S. and B. Giles-Corti, *The built environment, neighborhood crime and constrained physical activity: an exploration of inconsistent findings.* Preventive Medicine, 2008. **47**(3): p. 241-251.
44. Silverwood, R.J., et al., *Characterizing longitudinal patterns of physical activity in mid-adulthood using latent class analysis: results from a prospective cohort study.* American journal of epidemiology, 2011: p. kwr266.
45. Mooney, S.J., et al., *Patterns of Physical Activity Among Older Adults in New York City: A Latent Class Approach.* American journal of preventive medicine, 2015. **49**(3): p. e13-e22.
46. Xue, Q.-L., et al., *Patterns of 12-year change in physical activity levels in community-dwelling older women: can modest levels of physical activity help older women live longer?* American journal of epidemiology, 2012. **176**(6): p. 534-543.
47. Van Holle, V., et al., *Relationship between the physical environment and different domains of physical activity in European adults: a systematic review.* BMC public health, 2012. **12**(1): p. 1.
48. Giles-Corti, B., et al., *Understanding physical activity environmental correlates: increased specificity for ecological models.* Exercise and sport sciences reviews, 2005. **33**(4): p. 175-181.
49. Tucker, J.M., G.J. Welk, and N.K. Beyler, *Physical activity in US adults: compliance with the physical activity guidelines for Americans.* American journal of preventive medicine, 2011. **40**(4): p. 454-461.

50. Rundle, A., et al., *Neighborhood food environment and walkability predict obesity in New York City.* Environ Health Perspect, 2009. **117**(3): p. 442-447.
51. Frank, L.D., et al., *Stepping towards causation: do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity?* Social science & medicine, 2007. **65**(9): p. 1898-1914.
52. Skogan, W.G. and M.G. Maxfield, *Coping with crime: Individual and neighborhood reactions*. 1981: Sage Publications Beverly Hills, CA.
53. Kelling, G.L., and James Q. Wilson, in *The Atlantic*. 1982.
54. Kelling, G.L. and C.M. Coles, *Fixing broken windows : restoring order and reducing crime in our communities*. 1996, New York: Martin Kessler Books. xvi, 319 p.
55. Kelling, G.L. and National Institute of Justice (U.S.), *Broken windows and police discretion*. Research report. 1999, Washington, DC: U.S. Dept. of Justice, Office of Justice Programs, National Institute of Justice. vii, 57 p.
56. Taylor, R.B., S.A. Shumaker, and S.D. Gottfredson, *Neighborhood-level links between physical features and local sentiments: Deterioration, fear of crime, and confidence.* Journal of Architectural and Planning Research, 1985. **2**(4): p. 261-275.
57. Braga, A.A., B.C. Welsh, and C. Schnell, *Can policing disorder reduce crime? A systematic review and meta-analysis.* Journal of Research in Crime and Delinquency, 2015. **52**(4): p. 567-588.
58. Keyes, K.M., et al., *Child maltreatment increases sensitivity to adverse social contexts: neighborhood physical disorder and incident binge drinking in Detroit.* Drug and alcohol dependence, 2012. **122**(1): p. 77-85.
59. Leventhal, T. and J. Brooks-Gunn, *Moving to opportunity: an experimental study of neighborhood effects on mental health.* American Journal of Public Health, 2003. **93**(9): p. 1576-1582.
60. Cohen, D., et al., *" Broken windows" and the risk of gonorrhea.* American journal of public health, 2000. **90**(2): p. 230.
61. Burdette, A.M. and T.D. Hill, *An examination of processes linking perceived neighborhood disorder and obesity.* Social science & medicine, 2008. **67**(1): p. 38-46.
62. Franzini, L., et al., *Influences of physical and social neighborhood environments on children's physical activity and obesity.* American Journal of Public Health, 2009. **99**(2): p. 271-278.
63. Alfonzo, M.A., *To walk or not to walk? The hierarchy of walking needs.* Environment and Behavior, 2005. **37**(6): p. 808-836.
64. Hoehner, C.M., et al., *Perceived and objective environmental measures and physical activity among urban adults.* American Journal of Preventive Medicine, 2005. **28**(2): p. 105--116.
65. Ross, C.E. and J. Mirowsky, *Neighborhood disadvantage, disorder, and health.* Journal of Health and Social Behavior, 2001. **42**(3): p. 258--276.
66. Ayvalik, C. and C. Jotin Khisty, *Heuristic analysis of impacts of commuter rail station consolidation on pedestrian access.* Transportation Research Record: Journal of the Transportation Research Board, 2002(1793): p. 47-54.
67. Tucker-Seeley, R.D., et al., *Neighborhood safety, socioeconomic status, and physical activity in older adults.* American journal of preventive medicine, 2009. **37**(3): p. 207-213.
68. Lorenc, T., et al., *Fear of crime and the environment: systematic review of UK qualitative evidence.* BMC public health, 2013. **13**(1): p. 496.
69. Ding, D. and K. Gebel, *Built environment, physical activity, and obesity: what have we learned from reviewing the literature?* Health & place, 2012. **18**(1): p. 100-105.
70. Skogan, W., *Disorder and Decline The State of Research.* Journal of Research in Crime and Delinquency, 2015. **52**(4): p. 464-485.
71. Perkins, D.D., J.W. Meeks, and R.B. Taylor, *The physical environment of street blocks and resident perceptions of crime and disorder: Implications for theory and measurement.* Journal of Environmental Psychology, 1992. **12**(1): p. 21-34.
72. Franzini, L., et al., *Perceptions of disorder: Contributions of neighborhood characteristics to subjective perceptions of disorder.* Journal of Environmental Psychology, 2008. **28**(1): p. 83-93.
73. Mooney, S.J., et al., *Validity of an ecometric neighborhood physical disorder measure constructed by virtual street audit.* American journal of epidemiology, 2014: p. kwu180.

74. Raudenbush, S.W. and R.J. Sampson, *Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods.* Sociological Methodology 1999, Vol 29, 1999. **29**(1): p. 1-41.
75. Echeverria, S.E., A.V. Diez-Roux, and B.G. Link, *Reliability of self-reported neighborhood characteristics.* Journal of Urban Health, 2004. **81**(4): p. 682-701.
76. Mooney, S.J., D.J. Westreich, and A.M. El-Sayed, *Commentary: Epidemiology in the Era of Big Data.* Epidemiology, 2015. **26**(3): p. 390-394.
77. Blacksher, E. and G.S. Lovasi, *Place-focused physical activity research, human agency, and social justice in public health: taking agency seriously in studies of the built environment.* Health Place, 2012. **18**(2): p. 172-9.
78. Duncan, G.J. and S.W. Raudenbush, *Assesing the effects of context in studies of child and youth development.* Educational Psychologist, 1999. **34**(1): p. 29-41.
79. Avolio, B.J., F.J. Yammarino, and B.M. Bass, *Identifying common methods variance with data collected from a single source: An unresolved sticky issue.* Journal of Management, 1991. **17**(3): p. 571-587.
80. Echeverria, M.S.E., A.V. Diez-Roux, and B.G. Link, *Reliability of self-reported neighborhood characteristics.* Journal of Urban Health, 2004. **81**(4): p. 682-701.
81. Pedišić, Ž. and A. Bauman, *Accelerometer-based measures in physical activity surveillance: current practices and issues.* British journal of sports medicine, 2014: p. bjsports-2013-093407.
82. Shephard, R.J. and Y. Aoyagi, *Measurement of human energy expenditure, with particular reference to field studies: an historical perspective.* European journal of applied physiology, 2012. **112**(8): p. 2785-2815.
83. Prince, S.A., et al., *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, 2008. **5**(1): p. 56.
84. Sirard, J.R. and R.R. Pate, *Physical activity assessment in children and adolescents.* Sports medicine, 2001. **31**(6): p. 439-454.
85. Audrey, S., et al., *Adolescent perspectives on wearing accelerometers to measure physical activity in population-based trials.* The European Journal of Public Health, 2012: p. cks081.
86. Van Coevering, P., et al., *Feasibility of using accelerometers to measure physical activity in young adolescents.* Medicine and Science in Sports and Exercise, 2005. **37**(5): p. 867-871.
87. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* Annals of internal medicine, 2009. **151**(4): p. 264-269.
88. Zenk, S.N., et al., *Neighborhood Environment and Adherence to a Walking Intervention in African American Women.* Health Education \& Behavior, 2006. **36**(1): p. 167--181.
89. Laraia, B., et al., *Neighborhood Factors Associated with Physical Activity and Adequacy of Weight Gain During Pregnancy.* Journal of Urban Health, 2007. **84**(6): p. 793--806.
90. Brownson, R.C., et al., *Measuring the built environment for physical activity: state of the science.* American journal of preventive medicine, 2009. **36**(4): p. S99-S123. e12.
91. Saelens, B.E., et al., *Neighborhood-based differences in physical activity: an environment scale evaluation.* American journal of public health, 2003. **93**(9): p. 1552-1558.
92. Adams, E.J., et al., *Correlates of walking and cycling for transport and recreation: factor structure, reliability and behavioural associations of the perceptions of the environment in the neighbourhood scale (PENS).* Int. J. Behav. Nutr. Phys. Act, 2013. **10**.
93. Cain, K.L., et al., *Contribution of streetscape audits to explanation of physical activity in four age groups based on the \ Microscale\ \ Audit\ of \ Pedestrian\ \ Streetscapes\ (\ MAPS\ ).* Social Science \& Medicine (1982), 2014. **116**: p. 82--92.
94. Caspi, C.E., et al., *The social environment and walking behavior among low-income housing residents.* Social Science \& Medicine, 2013. **80**: p. 76--84.
95. Cerin, E.a., *Walking for \ Recreation\ and \ Perceptions\ of the \ Neighborhood\ \ Environment\ in \ Older\ \ Chinese\ \ Urban\ \ Dwellers\.* Journal of Urban Health, 2013. **90**(1): p. pp 56----66.
96. Cleland, V., et al., *Individual, social and environmental correlates of physical activity among women living in socioeconomically disadvantaged neighbourhoods.* Social Science \& Medicine, 2010. **70**(12): p. 2011--2018.

97. Cleland, V.J., A. Timperio, and D. Crawford, *Are perceptions of the physical and social environment associated with mothers' walking for leisure and for transport? A longitudinal study.* Preventive medicine, 2008. **47**(2): p. 188-193.
98. Cunningham-Myrie, C.A., et al., *Associations between neighborhood effects and physical activity, obesity, and diabetes: The Jamaica Health and Lifestyle Survey 2008.* Journal of Clinical Epidemiology, 2015. **68**(9): p. 970--978.
99. Feuillet, T., et al., *Spatial heterogeneity of the relationships between environmental characteristics and active commuting: towards a locally varying social ecological model.* International Journal of Health Geographics, 2015. **14**(1): p. 12.
100. Florindo, A.A., E.P.r. Salvador, and R.S. Reis, *Physical activity and its relationship with perceived environment among adults living in a region of low socioeconomic level.* Journal of Physical Activity \& Health, 2013. **10**(4): p. 563--571.
101. Foster, S., B. Giles-Corti, and M. Knuiman, *Does fear of crime discourage walkers? A social-ecological exploration of fear as a deterrent to walking.* 2014. p. 698--717.
102. Heinrich, K.M., et al., *Associations between the built environment and physical activity in public housing residents.* International Journal of Behavioral Nutrition and Physical Activity, 2007. **4**(1): p. 56.
103. Jalaludin, B., et al., *A pre-and-post study of an urban renewal program in a socially disadvantaged neighbourhood in Sydney, Australia.* BMC Public Health, 2012. **12**(1): p. 521.
104. Jongeneel-Grimen, B., et al., *The relationship between physical activity and the living environment: A multi-level analyses focusing on changes over time in environmental factors.* Health \& Place, 2014. **26**: p. 149--160.
105. Kamphuis, C.B.a., *Socioeconomic differences in lack of recreational walking among older adults: the role of neighbourhood and individual factors.* International Journal of Behavioral Nutrition and Physical Activity, 2009. **6**(1): p. 1.
106. Karusisi, N.l., et al., *Multiple dimensions of residential environments, neighborhood experiences, and jogging behavior in the RECORD Study.* Preventive Medicine, 2012. **55**(1): p. 50--55.
107. King, D., *Neighborhood and individual factors in activity in older adults: results from the neighborhood and senior health study.* Journal of Aging and Physical Activity, 2008. **16**(2): p. 144--170.
108. Kramer, D.l., et al., *Neighbourhood safety and leisure-time physical activity among Dutch adults: a multilevel perspective.* International Journal of Behavioral Nutrition and Physical Activity, 2013. **10**(1): p. 11.
109. Kwarteng, J.L., et al., *Associations between observed neighborhood characteristics and physical activity: findings from a multiethnic urban community.* Journal of Public Health, 2014. **36**(3): p. 358--367.
110. Lovasi, G.S., et al., *Body Mass Index, Safety Hazards, and Neighborhood Attractiveness.* American Journal of Preventive Medicine, 2012. **43**(4): p. 378-384.
111. Miles, R., *Neighborhood Disorder, Perceived Safety, and Readiness to Encourage Use of Local Playgrounds.* American Journal of Preventive Medicine, 2008. **34**(4): p. 275--281.
112. Oh, A.Y., et al., *Effects of perceived and objective neighborhood crime on walking frequency among midlife \ African\ \ American\ women in a home-based walking intervention.* Journal of Physical Activity \& Health, 2010. **7**(4): p. 432--441.
113. Strath, S.J., et al., *Measured and perceived environmental characteristics are related to accelerometer defined physical activity in older adults.* International Journal of Behavioral Nutrition and Physical Activity, 2012. **9**(1): p. 40.
114. Shenassa, E.D., A. Liebhaber, and A. Ezeamama, *Perceived safety of area of residence and exercise: a pan-European study.* American Journal of Epidemiology, 2006. **163**(11): p. 1012-1017.
115. Mooney, S.J., C.A. Richards, and A.G. Rundle, *There goes the neighborhood effect: bias owing to nondifferential measurement error in the construction of neighborhood contextual measures.* Epidemiology, 2014. **25**(4): p. 528-535.
116. STEENE-JOHANNESSEN, J., et al., *Are Self-report Measures Able to Define Individuals as Physically Active or Inactive?* Medicine & Science in Sports & Exercise, 2016. **48**(2): p. 235-244.
117. Haskell, W.L., *Physical activity by self-report: a brief history and future issues.* J Phys Act Health, 2012. **9**(Suppl 1): p. S5-S10.

118. Troped, P.J., et al., *Prediction of activity mode with global positioning system and accelerometer data.* Medicine and science in sports and exercise, 2008. **40**(5): p. 972-978.
119. Duncan, M.J., H.M. Badland, and W.K. Mummery, *Applying GPS to enhance understanding of transport-related physical activity.* Journal of Science and Medicine in Sport, 2009. **12**(5): p. 549-556.
120. Lachmann, R., *Graffiti as career and ideology.* American journal of sociology, 1988: p. 229-250.
121. Hill, A.B., *The environment and disease: association or causation?* Proceedings of the Royal Society of Medicine, 1965. **58**(5): p. 295.
122. Van Cauwenberg, J., et al., *Relationship between the physical environment and physical activity in older adults: a systematic review.* Health & place, 2011. **17**(2): p. 458-469.
123. Kaufman, T.K., et al., *Measuring health-relevant businesses over 21 years: refining the National Establishment Time-Series (NETS), a dynamic longitudinal data set.* BMC research notes, 2015. **8**(1): p. 507.
124. Lovasi, G.S., et al., *Effect of individual or neighborhood disadvantage on the association between neighborhood walkability and body mass index.* American Journal of Public Health, 2009. **99**(2): p. 279.
125. Oakes, J.M., *The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology.* Social science & medicine, 2004. **58**(10): p. 1929-1952.
126. Handy, S.L., X. Cao, and P.L. Mokhtarian, *The causal influence of neighborhood design on physical activity within the neighborhood: evidence from Northern California.* American Journal of Health Promotion, 2008. **22**(5): p. 350-358.
127. Guo, Z. and B.P. Loo, *Pedestrian environment and route choice: evidence from New York City and Hong Kong.* Journal of transport geography, 2013. **28**: p. 124-136.
128. Van Cauwenberg, J., et al., *Physical environmental factors that invite older adults to walk for transportation.* Journal of environmental psychology, 2014. **38**: p. 94-103.
129. Caspersen, C.J., K.E. Powell, and G.M. Christenson, *Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research.* Public health reports, 1985. **100**(2): p. 126.
130. Centers for Disease Control and Prevention. *The State of Aging and Health in America 2013*. [cited 2014 April 29]; Available from: http://www.cdc.gov/aging/pdf/state-aging-health-in-america-2013.pdf.
131. Bird, S., et al., *The influence of the built environment and other factors on the physical activity of older women from different ethnic communities.* Journal of Women & Aging, 2009. **21**(1): p. 33-47.
132. Wilcox, S., et al., *Psychosocial and perceived environmental correlates of physical activity in rural and older African American and white women.* The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 2003. **58**(6): p. P329-P337.
133. Yen, I.H., Y.L. Michael, and L. Perdue, *Neighborhood environment in studies of health of older adults: a systematic review.* American journal of preventive medicine, 2009. **37**(5): p. 455-463.
134. Cerin, E., et al., *Objectively-measured neighborhood environments and leisure-time physical activity in Chinese urban elders.* Preventive Medicine, 2013. **56**(1): p. 86--89.
135. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern epidemiology*. 2008: Lippincott Williams & Wilkins.
136. Singer, J.D. and J.B. Willett, *Applied longitudinal data analysis: Modeling change and event occurrence*. 2003: Oxford university press.
137. Purciel, M., et al., *Creating and validating GIS measures of urban design for health research.* Journal of environmental psychology, 2009. **29**(4): p. 457-466.
138. Washburn, R.A., et al., *The Physical Activity Scale for the Elderly (PASE): development and evaluation.* Journal of clinical epidemiology, 1993. **46**(2): p. 153-162.
139. Bonnefoy, M., et al., *Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study.* Journal of the American Geriatrics Society, 2001. **49**(1): p. 28-35.
140. Harada, N.D., et al., *An evaluation of three self-report physical activity instruments for older adults.* Medicine and science in sports and exercise, 2001. **33**(6): p. 962-970.
141. Schuit, A.J., et al., *Validity of the Physical Activity Scale for the Elderly (PASE): according to energy expenditure assessed by the doubly labeled water method.* Journal of clinical epidemiology, 1997. **50**(5): p. 541-546.

142. Washburn, R.A., et al., *The physical activity scale for the elderly (PASE): evidence for validity.* Journal of clinical epidemiology, 1999. **52**(7): p. 643-651.
143. Laurin, D., et al., *Physical activity and risk of cognitive impairment and dementia in elderly persons.* Archives of neurology, 2001. **58**(3): p. 498-504.
144. Weuve, J., et al., *Physical activity, including walking, and cognitive function in older women.* Jama, 2004. **292**(12): p. 1454-1461.
145. Jette, A.M., et al., *The functional status questionnaire.* Journal of General Internal Medicine, 1986. **1**(3): p. 143-149.
146. Sampson, R.J., S.W. Raudenbush, and F. Earls, *Neighborhoods and violent crime: A multilevel study of collective efficacy.* Science, 1997. **277**(5328): p. 918-924.
147. Sampson, R.J. and S.W. Raudenbush, *Systematic social observation of public spaces: A new look at disorder in urban Neighborhoods 1.* American journal of sociology, 1999. **105**(3): p. 603-651.
148. Bader, M.D.M., et al., *Development and Deployment of the Computer Assisted Neighborhood Visual Audit System (CANVAS).* Health & place, 2015. **31**: p. 163-172.
149. Bader, M.D. and J.A. Ailshire, *Creating Measures of Theoretically Relevant Neighborhood Attributes at Multiple Spatial Scales.* Sociological Methodology, 2014: p. 0081175013516749.
150. Mooney, S.J., et al., *Validity of an Ecometric Neighborhood Physical Disorder Measure Constructed by Virtual Street Audit.* American Journal of Epidemiology, 2014. **In Press**.
151. Quinn, J.W., et al., *Neighborhood physical disorder in New York City.* Journal of Maps, 2014(ahead-of-print): p. 1-8.
152. Freeman, L., et al., *Neighborhood walkability and active travel (walking and cycling) in New York City.* Journal of Urban Health, 2013. **90**(4): p. 575-585.
153. Rundle, A.G., et al., *Using GPS data to study neighborhood walkability and physical activity.* American journal of preventive medicine, 2015.
154. Quinn, J.W., et al., *Neighborhood physical disorder in New York City.* The Journal of Maps, In Press.
155. Lovasi, G.S., et al., *Neighborhood safety and green space as predictors of obesity among preschool children from low-income families in New York City.* Preventive medicine, 2013. **57**(3): p. 189-193.
156. Lovasi, G.S., et al., *Is the environment near home and school associated with physical activity and adiposity of urban preschool children?* Journal of Urban Health, 2011. **88**(6): p. 1143-1157.
157. Rainham, D., et al., *Conceptualizing the healthscape: contributions of time geography, location technologies and spatial ecology to place and health research.* Social science & medicine, 2010. **70**(5): p. 668-676.
158. Robins, J.M. and D.M. Finkelstein, *Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log‐Rank Tests.* Biometrics, 2000. **56**(3): p. 779-788.
159. Raghunathan, T.E., P.W. Solenberger, and J. Van Hoewyk, *IVEware: Imputation and variance estimation software.* Ann Arbor, MI: Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan, 2002.
160. Daniel, R.M., et al., *Methods for dealing with time-dependent confounding.* Statistics in Medicine, 2012.
161. Weather Underground. *Historical Weather*. 2015; Available from: http://www.wunderground.com/history/.
162. Brice, T. and T. Hall. *Heat Index*. 2009; Available from: http://www.srh.noaa.gov/images/epz/wxcalc/heatIndex.pdf.
163. Brice, T. and T. Hall. *Wind Chill*. 2009; Available from: http://www.srh.noaa.gov/images/epz/wxcalc/windChill.pdf.
164. Edwards, L.J., *Modern statistical techniques for the analysis of longitudinal data in biomedical research.* Pediatric pulmonology, 2000. **30**(4): p. 330-344.
165. Lumley, T., *mitools: Tools for Multiple Imputation of Missing Data; 2012.* R package version. **2**.
166. Morabia, A. and M.C. Costanza, *Does walking 15 minutes per day keep the obesity epidemic away? Simulation of the efficacy of a populationwide campaign.* American Journal of Public Health, 2004. **94**(3): p. 437-440.

167. Hernán, M.A., *Invited commentary: hypothetical interventions to define causal effects—afterthought or prerequisite?* American Journal of Epidemiology, 2005. **162**(7): p. 618-620.
168. Hernán, M.A. and J.M. Robins, *Estimating causal effects from epidemiological data.* Journal of epidemiology and community health, 2006. **60**(7): p. 578-586.
169. Moran, M., et al., *Understanding the relationships between the physical environment and physical activity in older adults: a systematic review of qualitative studies.* Int. J. Behav. Nutr. Phys. Act, 2014. **11**: p. 79.
170. Mokhtarian, P.L. and X. Cao, *Examining the impacts of residential self-selection on travel behavior: A focus on methodologies.* Transportation Research Part B: Methodological, 2008. **42**(3): p. 204-228.
171. Boone-Heinonen, J., et al., *Environment and physical activity dynamics: the role of residential self-selection.* Psychology of Sport and Exercise, 2011. **12**(1): p. 54-60.
172. Brown, S.C., et al., *The Relationship of Perceived Neighborhood Social Climate to Walking in Hispanic Older Adults A Longitudinal, Cross-Lagged Panel Analysis.* Journal of aging and health, 2011. **23**(8): p. 1325-1351.
173. Ranchod, Y.K., et al., *Longitudinal associations between neighborhood recreational facilities and change in recreational physical activity in the multi-ethnic study of atherosclerosis, 2000–2007.* American journal of epidemiology, 2014. **179**(3): p. 335-343.
174. Sun, V.K., et al., *How safe is your neighborhood? Perceived neighborhood safety and functional decline in older adults.* Journal of general internal medicine, 2012. **27**(5): p. 541-547.
175. Allison, P.D., *Missing data*. Vol. 136. 2001: Sage publications.
176. Hygiene, N.Y.C.D.o.H.a.M. *Survey data on the health of New Yorkers*. 2013; Available from: http://www.nyc.gov/html/doh/html/data/chs-methods.shtml.
177. Sinclair, M., et al., *Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches.* BMC medical research methodology, 2012. **12**(1): p. 132.
178. Cleland, V., et al., *Individual, social and environmental correlates of physical activity among women living in socioeconomically disadvantaged neighbourhoods.* Social science & medicine, 2010. **70**(12): p. 2011-2018.
179. Diez Roux, A.V., *Estimating neighborhood health effects: the challenges of causal inference in a complex world.* Social science & medicine, 2004. **58**(10): p. 1953-1960.
180. Owen, N., et al., *Understanding environmental influences on walking: review and research agenda.* American journal of preventive medicine, 2004. **27**(1): p. 67-76.
181. Humpel, N., N. Owen, and E. Leslie, *Environmental factors associated with adults' participation in physical activity: a review.* American journal of preventive medicine, 2002. **22**(3): p. 188-199.
182. Ding, D., et al., *Neighborhood environment and physical activity among youth: a review.* American journal of preventive medicine, 2011. **41**(4): p. 442-455.
183. Saelens, B.E. and S.L. Handy, *Built environment correlates of walking: a review.* Medicine and science in sports and exercise, 2008. **40**(7 Suppl): p. S550.
184. Shiu-Thornton, S., et al., *Older adult perspectives on physical activity and exercise: voices from multiple cultures.* 2004.
185. Kegler, M.C., et al., *A qualitative examination of home and neighborhood environments for obesity prevention in rural adults.* International Journal of Behavioral Nutrition and Physical Activity, 2008. **5**(1): p. 65.
186. Wilcox, S., et al., *Determinants of leisure time physical activity in rural compared with urban older and ethnically diverse women in the United States.* Journal of epidemiology and community health, 2000. **54**(9): p. 667-672.
187. Joshi, S., et al., *Beyond METs: Types of Physical Activity and Depression Among Older Adults.* Age and ageing, 2016.
188. Murphy, M.H., et al., *Does doing housework keep you healthy? The contribution of domestic physical activity to meeting current recommendations for health.* BMC public health, 2013. **13**(1): p. 966.
189. Nicklett, E.J., L.A. Anderson, and I.H. Yen, *Gardening Activities and Physical Health Among Older Adults A Review of the Evidence.* Journal of Applied Gerontology, 2014: p. 0733464814563608.

190. Goldsmith, J., et al., *New insights into activity patterns in children, found using functional data analyses.* Medicine & Science in Sports & Exercise, 2016.
191. Collins, L.M. and S.T. Lanza, *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Vol. 718. 2010: John Wiley & Sons.
192. Huh, J., et al., *Identifying patterns of eating and physical activity in children: a latent class analysis of obesity risk.* Obesity, 2011. **19**(3): p. 652-658.
193. Ainsworth, B.E., et al., *Compendium of physical activities: classification of energy costs of human physical activities.* Medicine and science in sports and exercise, 1993. **25**(1): p. 71-80.
194. Lanza, S.T., et al., *PROC LCA: A SAS procedure for latent class analysis.* Structural Equation Modeling, 2007. **14**(4): p. 671-694.
195. Bolck, A., M. Croon, and J. Hagenaars, *Estimating latent structure models with categorical variables: One-step versus three-step estimators.* Political Analysis, 2004. **12**(1): p. 3-27.
196. Bakk, Z., F.B. Tekle, and J.K. Vermunt, *Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches.* Sociological Methodology, 2013. **43**(1): p. 272-311.
197. Lumley, T., R. Kronmal, and S. Ma, *Relative risk regression in medical research: models, contrasts, estimators, and algorithms.* 2006.
198. Delbaere, K., K. Hauer, and S.R. Lord, *Evaluation of the incidental and planned activity questionnaire (IPAQ) for older people.* British journal of sports medicine, 2009.
199. Dishman, R.K., et al., *Using constructs of the transtheoretical model to predict classes of change in regular physical activity: a multi-ethnic longitudinal cohort study.* Annals of Behavioral Medicine, 2010. **40**(2): p. 150-163.
200. Barnett, T.A., et al., *Distinct trajectories of leisure time physical activity and predictors of trajectory class membership: a 22 year cohort study.* International Journal of Behavioral Nutrition and Physical Activity, 2008. **5**(1): p. 57.
201. O'Brien, M., et al., *The ecology of childhood overweight: a 12-year longitudinal analysis.* International journal of obesity, 2007. **31**(9): p. 1469-1478.
202. Shalizi, C.R., *Advanced data analysis from an elementary point of view.* URL: http://www. stat. cmu. edu/cshalizi/ADAfaEPoV/13, 2013. **24**.
203. Rhodes, R.E. and L. Dickau, *Moderators of the intention-behaviour relationship in the physical activity domain: a systematic review.* British Journal of Sports Medicine, 2012: p. bjsports-2011-090411.
204. Liu, R.D., et al., *Psychometric properties of two physical activity questionnaires, the AQuAA and the PASE, in cancer patients.* BMC medical research methodology, 2011. **11**(1): p. 1.
205. Skotte, J., et al., *Detection of physical activity types using triaxial accelerometers.* J Phys Act Health, 2014. **11**(1): p. 76-84.
206. Troiano, R.P., et al., *Evolution of accelerometer methods for physical activity research.* British journal of sports medicine, 2014: p. bjsports-2014-093546.
207. Lim, S., et al., *Measurement error of self-reported physical activity levels in New York City: assessment and correction.* American journal of epidemiology, 2015: p. kwu470.
208. Zhang, S., et al., *Physical activity classification using the GENEA wrist-worn accelerometer.* Medicine and science in sports and exercise, 2012.
209. Arem, H., S.K. Keadle, and C.E. Matthews, *Invited commentary: meta-physical activity and the search for the truth.* American journal of epidemiology, 2015. **181**(9): p. 656-658.
210. Sabia, S., et al., *Association between questionnaire-and accelerometer-assessed physical activity: the role of sociodemographic factors.* American journal of epidemiology, 2014. **179**(6): p. 781-790.
211. Van Dyck, D., et al., *Relationships between neighborhood walkability and adults' physical activity: How important is residential self-selection?* Health & place, 2011. **17**(4): p. 1011-1014.
212. Crowder, K. and S.J. South, *Race, Class, and Changing Patterns of Migration between Poor and Nonpoor Neighborhoods1.* American Journal of Sociology, 2005. **110**(6): p. 1715-1763.
213. Bernstein, M., et al., *Data-based approach for developing a physical activity frequency questionnaire.* American Journal of Epidemiology, 1998. **147**(2): p. 147-154.
214. Basta, L.A., T.S. Richmond, and D.J. Wiebe, *Neighborhoods, daily activities, and measuring health risks experienced in urban environments.* Social Science & Medicine, 2010. **71**(11): p. 1943-1950.

215. Hirsch, J.A., et al., *Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis.* International journal of health geographics, 2014. **13**(1): p. 1.
216. Krieger, N., et al., *Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? the Public Health Disparities Geocoding Project.* American journal of epidemiology, 2002. **156**(5): p. 471-482.
217. Weden, M.M., et al., *Neighborhood archetypes for population health research: Is there no place like home?* Health & place, 2011. **17**(1): p. 289-299.
218. Fotheringham, A.S. and D.W.S. Wong, *The Modifiable Areal Unit Problem in Multivariate Statistical-Analysis.* Environment and Planning A, 1991. **23**(7): p. 1025-1044.
219. Neckerman, K.M., et al., *Body mass index and indices of neighborhood walkability: Heterogeneity of associations across alternative neighborhood definitions.* Under Review.
220. Doyle, S., et al., *Active community environments and health: the relationship of walkable and safe communities to individual health.* Journal of the American Planning Association, 2006. **72**(1): p. 19-31.
221. Jurek, A.M., et al., *Proper interpretation of non-differential misclassification effects: expectations vs observations.* International journal of epidemiology, 2005. **34**(3): p. 680-687.
222. Lovasi, G.S., et al., *Body mass index, safety hazards, and neighborhood attractiveness.* American journal of preventive medicine, 2012. **43**(4): p. 378-384.
223. Glymour, M.M., *Using causal diagrams to understand common problems in social epidemiology.* Methods in social epidemiology, 2006: p. 393-428.
224. Adams, M.A., et al., *Patterns of neighborhood environment attributes related to physical activity across 11 countries: a latent class analysis.* Int J Behav Nutr Phys Act, 2013. **10**(34): p. 10.1186.
225. Adams, M.A., et al., *Neighborhood environment profiles related to physical activity and weight status: A latent profile analysis.* Preventive medicine, 2011. **52**(5): p. 326-331.
226. Kurka, J.M., et al., *Patterns of neighborhood environment attributes in relation to children's physical activity.* Health & place, 2015. **34**: p. 164-170.
227. Ioannidis, J., *Exposure-wide epidemiology: revisiting Bradford Hill.* Statistics in medicine, 2015.
228. Rundle, A., H. Ahsan, and P. Vineis, *Better cancer biomarker discovery through better study design.* European journal of clinical investigation, 2012. **42**(12): p. 1350-1359.
229. Sun, H. and S. Wang, *Penalized logistic regression for high-dimensional DNA methylation data with case-control studies.* Bioinformatics, 2012. **28**(10): p. 1368-1375.
230. Patel, C.J., J. Bhattacharya, and A.J. Butte, *An environment-wide association study (EWAS) on type 2 diabetes mellitus.* PloS one, 2010. **5**(5): p. e10746.
231. Lind, P.M., et al., *An environmental wide association study (EWAS) approach to the metabolic syndrome.* Environment international, 2013. **55**: p. 1-8.
232. Ioannidis, J.P., et al., *Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification.* Science Translational Medicine, 2009. **1**(7): p. 7ps8-7ps8.
233. Mooney, S.J., D.J. Westreich, and A.M. El-Sayed, *Epidemiology in the Big Data Era.* Epidemiology (Cambridge, Mass.), 2015. **In Press**.
234. McKinney, B.A., et al., *Machine learning for detecting gene-gene interactions.* Applied bioinformatics, 2006. **5**(2): p. 77-88.
235. Gale, S.L., et al., *Crime, neighborhood deprivation, and asthma: a GIS approach to define and assess neighborhoods.* Spatial and spatio-temporal epidemiology, 2011. **2**(2): p. 59-67.
236. Pedigo, A. and W. Seaver, *Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches.* PloS one, 2011. **6**(7): p. e22693.
237. Neckerman, K.M., et al., *Disparities in urban neighborhood conditions: evidence from GIS measures and field observation in New York City.* Journal of Public Health Policy, 2009: p. S264-S285.
238. Lovasi, G.S., et al., *Urban Tree Canopy and Asthma, Wheeze, Rhinitis, and Allergic Sensitization to Tree Pollen in a New York City Birth Cohort.* Environmental health perspectives, 2013. **121**(4): p. 494.
239. Lovasi, G.S., et al., *Aesthetic amenities and safety hazards associated with walking and bicycling for transportation in New York City.* Annals of behavioral medicine, 2013. **45**(1): p. 76-85.
240. Times, T.N.Y. *Homicide Map*. Available from: http://projects.nytimes.com/crime/homicides/map/.

241. Stark, J.H., et al., *The impact of neighborhood park access and quality on body mass index among adults in New York City.* Preventive medicine, 2014. **64**: p. 63-68.
242. Mooney, S.J., et al., *Using Google Street View to Assess Environmental Contributions to Pedestrian Injury.* American Journal of Public Health, 2016.
243. Bureau, U.S.C. *Poverty*. 2015; Available from: http://www.census.gov/hhes/www/poverty/data/threshld/index.html.
244. Cantor, R.M., K. Lange, and J.S. Sinsheimer, *Prioritizing GWAS results: a review of statistical methods and recommendations for their application.* The American Journal of Human Genetics, 2010. **86**(1): p. 6-22.
245. Tibshirani, R., *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society. Series B (Methodological), 1996: p. 267-288.
246. Fan, J. and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties.* Journal of the American statistical Association, 2001. **96**(456): p. 1348-1360.
247. Sandri, M. and P. Zuccolotto, *Variable selection using random forests*, in *Data analysis, classification and the forward search*. 2006, Springer. p. 263-270.
248. Quinlan, J.R., *Induction of decision trees.* Machine learning, 1986. **1**(1): p. 81-106.
249. Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
250. Archer, K.J. and R.V. Kimes, *Empirical characterization of random forest variable importance measures.* Computational Statistics & Data Analysis, 2008. **52**(4): p. 2249-2260.
251. Marsit, C.J., et al., *Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma.* Carcinogenesis, 2009. **30**(3): p. 416-422.
252. Christensen, B.C., et al., *Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context.* PLoS Genet, 2009. **5**(8): p. e1000602.
253. Polak, P., et al., *Cell-of-origin chromatin organization shapes the mutational landscape of cancer.* Nature, 2015. **518**(7539): p. 360-364.
254. Hastie, T., et al., *The elements of statistical learning*. Vol. 2. 2009: Springer.
255. Efron, B. and G. Gong, *A leisurely look at the bootstrap, the jackknife, and cross-validation.* The American Statistician, 1983. **37**(1): p. 36-48.
256. Hastie, T., et al., *The elements of statistical learning: data mining, inference and prediction.* The Mathematical Intelligencer, 2005. **27**(2): p. 83-85.
257. Saelens, B.E., J.F. Sallis, and L.D. Frank, *Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures.* Annals of behavioral medicine, 2003. **25**(2): p. 80-91.
258. Rundle, A., et al., *The urban built environment and obesity in New York City: a multilevel analysis.* American Journal of Health Promotion, 2007. **21**(4s): p. 326-334.
259. Echenique, M.H., et al., *Growing cities sustainably: does urban form really matter?* Journal of the American Planning Association, 2012. **78**(2): p. 121-137.
260. Lipsitch, M., E.T. Tchetgen, and T. Cohen, *Negative controls: a tool for detecting confounding and bias in observational studies.* Epidemiology (Cambridge, Mass.), 2010. **21**(3): p. 383.
261. Rubin, D.B., *Multiple imputation for nonresponse in surveys*. Vol. 81. 2004: John Wiley & Sons.
262. Lumley, T., *Analysis of complex survey samples.* Journal of Statistical Software, 2004. **9**(1): p. 1-19.
263. Friedman, J., T. Hastie, and R. Tibshirani, *glmnet: Lasso and elastic-net regularized generalized linear models.* R package version, 2009. **1**.
264. Liaw, A. and M. Wiener, *Classification and regression by randomForest.* R news, 2002. **2**(3): p. 18-22.
265. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2009: Springer Science & Business Media.
266. Mooney, S.J., et al., *In the Wrong Place with the Wrong SNP: Beta-2-Adrenergic Receptor Gene Modifies the Association between Stressful Neighborhood Conditions and Sudden Cardiac Arrest.* Epidemiology, In Press.
267. Feng, J., et al., *The built environment and obesity: a systematic review of the epidemiologic evidence.* Health & place, 2010. **16**(2): p. 175-190.
268. Link, B.G. and J. Phelan, *Social conditions as fundamental causes of disease.* Journal of health and social behavior, 1995: p. 80-94.

269. Mahmood, A., et al., *A photovoice documentation of the role of neighborhood physical and social environments in older adults' physical activity in two metropolitan areas in North America.* Social science & medicine, 2012. **74**(8): p. 1180-1192.
270. Robert, S. and J.S. House, *SES differentials in health by age and alternative indicators of SES.* Journal of Aging and Health, 1996. **8**(3): p. 359-388.
271. Kaufman, J.S., R.S. Cooper, and D.L. McGee, *Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race.* Epidemiology, 1997: p. 621-628.
272. Breiman, L., *Statistical modeling: The two cultures (with comments and a rejoinder by the author).* Statistical Science, 2001. **16**(3): p. 199-231.
273. Patel, C.J. and A.K. Manrai. *Development of exposome correlation globes to map out environment-wide associations*. in *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*. 2014. NIH Public Access.
274. Wu, X., M. Al Hasan, and J.Y. Chen, *Pathway and network analysis in proteomics.* Journal of theoretical biology, 2014. **362**: p. 44-52.
275. Khatri, P., M. Sirota, and A.J. Butte, *Ten years of pathway analysis: current approaches and outstanding challenges.* PLoS Comput Biol, 2012. **8**(2): p. e1002375.
276. Yang, Y., et al., *A spatial agent-based model for the simulation of adults' daily walking within a city.* American journal of preventive medicine, 2011. **40**(3): p. 353-361.
277. Wiebe, D.J., et al., *Mapping activity patterns to quantify risk of violent assault in urban environments.* Epidemiology (Cambridge, Mass.), 2016. **27**(1): p. 32.
278. Jerrett, M., et al., *Traffic-related air pollution and asthma onset in children: a prospective cohort study with individual exposure measurement.* Environmental health perspectives, 2008. **116**(10): p. 1433.
279. Ross, C.E., *Walking, exercising, and smoking: does neighborhood matter?* Social science & medicine, 2000. **51**(2): p. 265-274.
280. Bader, M.D., et al., *More neighborhood retail associated with lower obesity among New York City public high school students.* Health & place, 2013. **23**: p. 104-110.
281. Rudolph, L. and J. Caplan, *Health in All Policies: A guide for state and local governments*.
282. Gordon, C., *Blighting the way: Urban renewal, economic development, and the elusive definition of blight.* Fordham Urb. LJ, 2003. **31**: p. 305.

Appendices

Table A2.1. Search terms used to query each database:

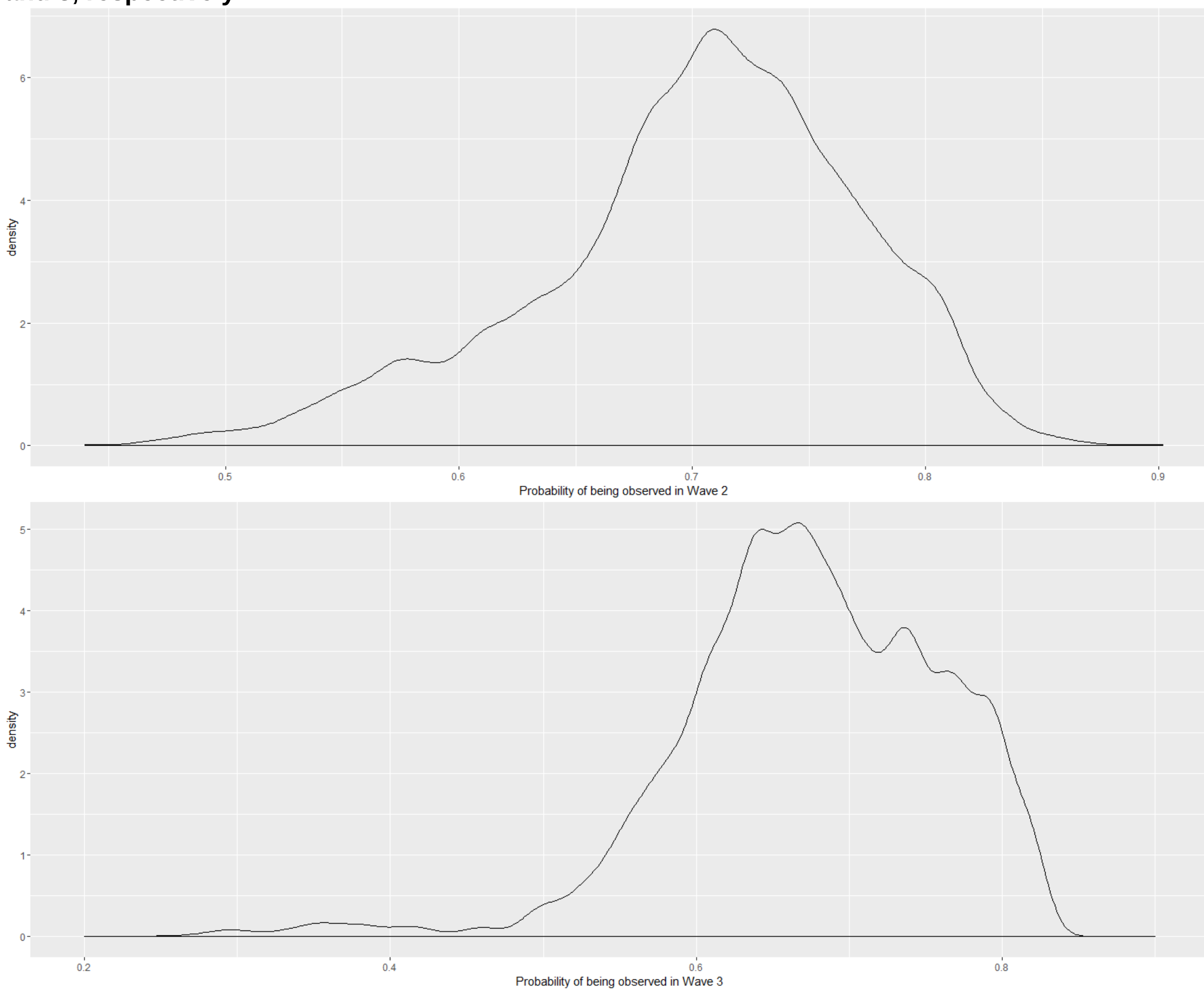
Database	Search Terms
Pubmed	((Walking OR Physical Activity OR Motor Activity[MeSH Terms] OR Exercise) AND (Disorder OR Broken Windows OR Aesthetics OR Incivilities OR Safety[MeSH Terms]) AND (Neighborhood OR Urban OR City OR Residential OR Residence Characteristics[MeSH Terms]))
TRID	((Walking OR Physical Activity OR Motor Activity OR Exercise) AND (Disorder OR Broken Windows OR Aesthetics OR Incivilities OR Safety) AND (Neighborhood OR Urban OR City OR Residential OR Residence Characteristics))
PsycInfo	((Walking OR Physical Activity OR Motor Activity OR Exercise) AND (Disorder OR Broken Windows OR Aesthetics OR Incivilities OR Safety) AND (Neighborhood OR Urban OR City OR Residential OR Residence Characteristics))
Embase	((Walking OR 'Physical Activity' OR 'Motor Activity' OR Exercise) AND (Disorder OR 'Broken Windows' OR Aesthetics OR Incivilities OR Safety) AND (Neighborhood OR Urban OR City OR Residential OR 'Residence Characteristics'))

Appendix 3.1 Tables and Figures Relevant to Computing of Inverse Probability of Censoring Weights.

Table A3.1.1 Parameter estimates from logistic model predicting re-contact in Waves 2 and 3 as a function of demographic characteristics.

	Wave 2			Wave 3		
Covariate	**Odds Ratio**	**95% CI**		**Odds Ratio**	**95% CI**	
Gender						
Female	(Ref)	---	---	(Ref)	---	---
Male	0.83	0.78	0.89	0.87	0.82	0.93
Race/Ethnicity						
Non-Hispanic Black	(Ref)	--	--	(Ref)	---	---
Non-Hispanic White	0.86	0.80	0.94	1.08	1.00	1.17
Other	0.72	0.63	0.82	0.84	0.74	0.96
Hispanic	0.69	0.61	0.77	0.77	0.69	0.87
Educational Attainment						
Less than High School	(Ref)	---	---	(Ref)	---	---
High School Graduate	1.22	1.10	1.34	1.11	1.00	1.22
Some College	1.30	1.17	1.46	0.99	0.89	1.10
College Graduate	1.59	1.43	1.76	1.56	1.41	1.73
Borough						
Unidentified	0.23	0.19	0.28	0.26	0.21	0.31
Manhattan	(Ref)	---	---	(Ref)	---	---
Bronx	0.89	0.79	1.00	0.91	0.81	1.02
Brooklyn	0.89	0.80	0.98	0.83	0.75	0.91
Queens	0.84	0.76	0.93	0.86	0.78	0.95
Staten Island	1.20	0.96	1.51	0.87	0.70	1.08
Health						
Excellent	(Ref)	---	---	(Ref)	---	---
Good	1.06	0.97	1.17	0.87	0.79	0.95
Fair	0.79	0.72	0.88	0.75	0.68	0.83
Poor	0.75	0.65	0.87	0.61	0.53	0.70
Neighborhood Measures						
Disorder (per SD)	1.03	0.98	1.07	0.99	0.95	1.03
Pedestrian Injuries (per SD)	1.07	1.03	1.11	1.01	0.98	1.05

Figure A3.1. Density plots of estimated probability of being re-contacted successfully in Waves 2 and 3, respectively.



Appendix A3.2. Prevalence of demographic characteristics in the sample-weighted study population, showing that post-weighting, the sample retains demographic balance

Characteristic	Wave 1	Wave 2	Wave 3
Age			
65-68	34	33	34
69-71	20	21	21
72-75	46	46	45
Sex			
Female	58	57	59
Male	42	43	41
Race/Ethnicity			
Non-Hispanic White	47	47	46
Non-Hispanic Black	26	27	26
Hispanic	12	13	13
Other	14	14	15
Educational Attainment			
Less than High School	32	31	32
Completed High School	29	29	29
Some College	15	16	16
Completed College	24	24	24
Household Income			
Less than $20,000	40	39	39
$20,000-40,000	24	24	26
$40,000-80,000	21	21	20
More than $80,000	16	17	16

Appendix A3.3. Weather as an influence on past-week physical activity

In order to explore whether our results might be affected by past-week temperatures or season variation in temperature, we downloaded weather data from Weather Underground (http://www.wunderground.com) for New York City over the period in which NYCNAMES-II interviews were conducted.

From the 7 days of weather data immediately prior to each subject's interview date and using formulas from the National Weather Service for heat index (http://www.srh.noaa.gov/images/epz/wxcalc/heatIndex.pdf) and wind chill (http://www.srh.noaa.gov/images/epz/wxcalc/windChill.pdf), we computed a mean past-week 'feels like' temperature for each subject. We then plotted total PASE score and probability of engaging in various components of PASE for each mean temperature overall and stratified by wave. Figures 3.1-3.6, below, display these plots.

After seeing descriptively that the effects of weather appear to be minimal, inconsistent across waves, and without any theory supporting why past-week weather would interact with or be affected by neighborhood disorder, we did not pursue this analysis further or include past-week weather in our final models.

Figure A3.3.1 Total PASE Score and probability of engaging in light housework by mean 'Feels-like' temperature over the past-week.

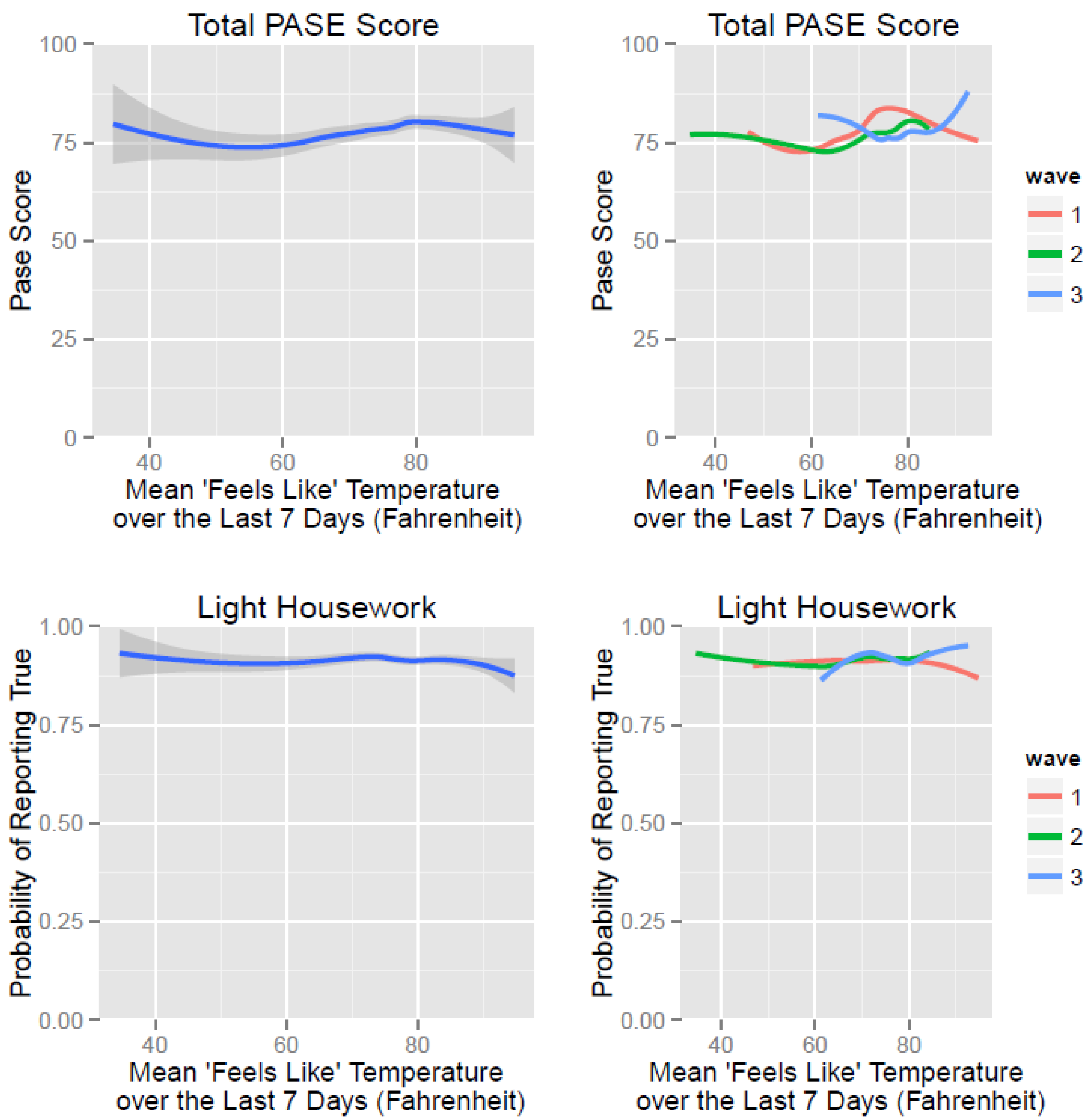


Figure A3.3.2 Probability of engaging in heavy housework and home repair by mean 'Feels-like' temperature over the past-week.

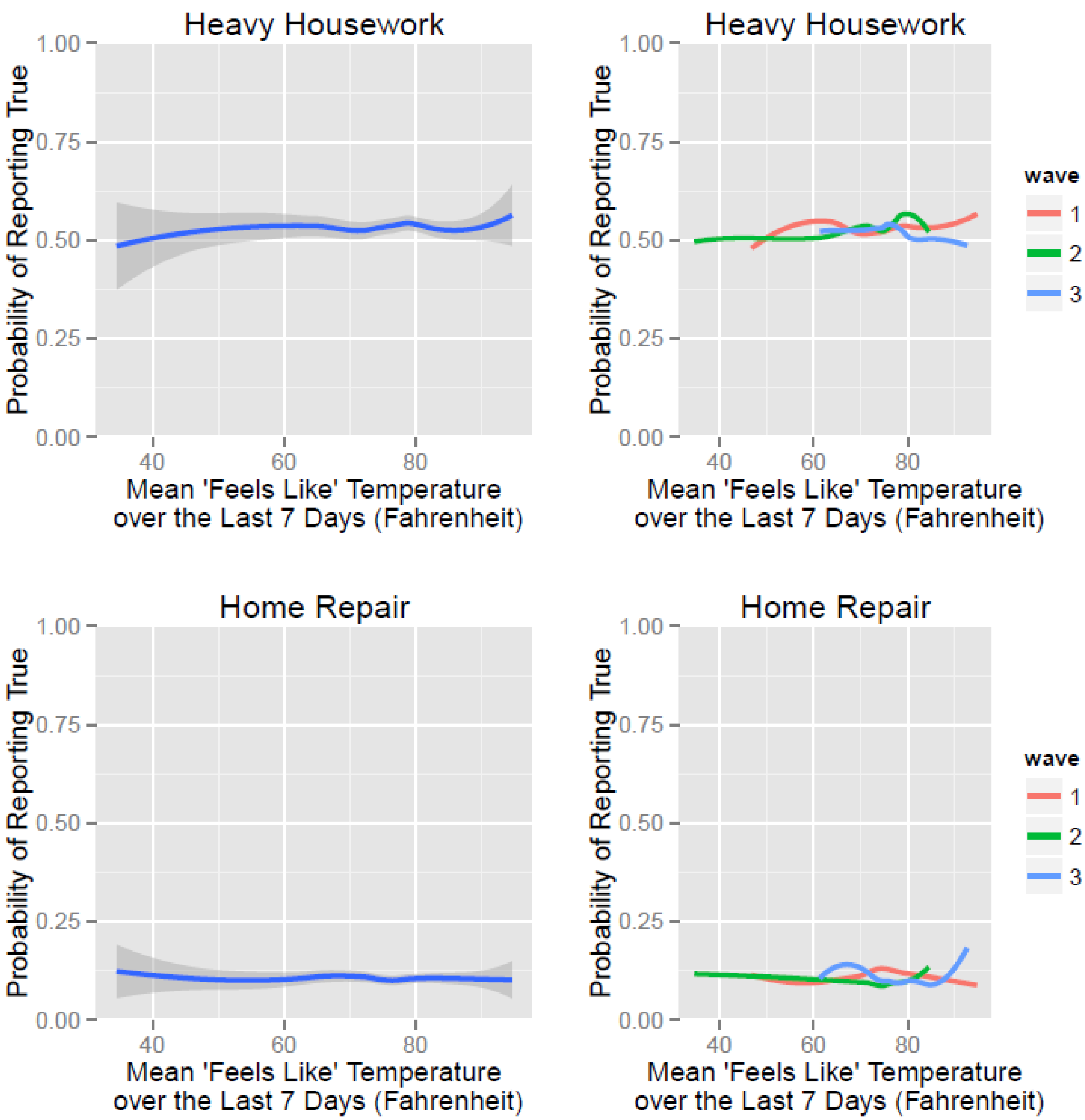


Figure A3.3 Probability of engaging in yard work and gardening by mean ‘Feels-like’ temperature over the past-week.

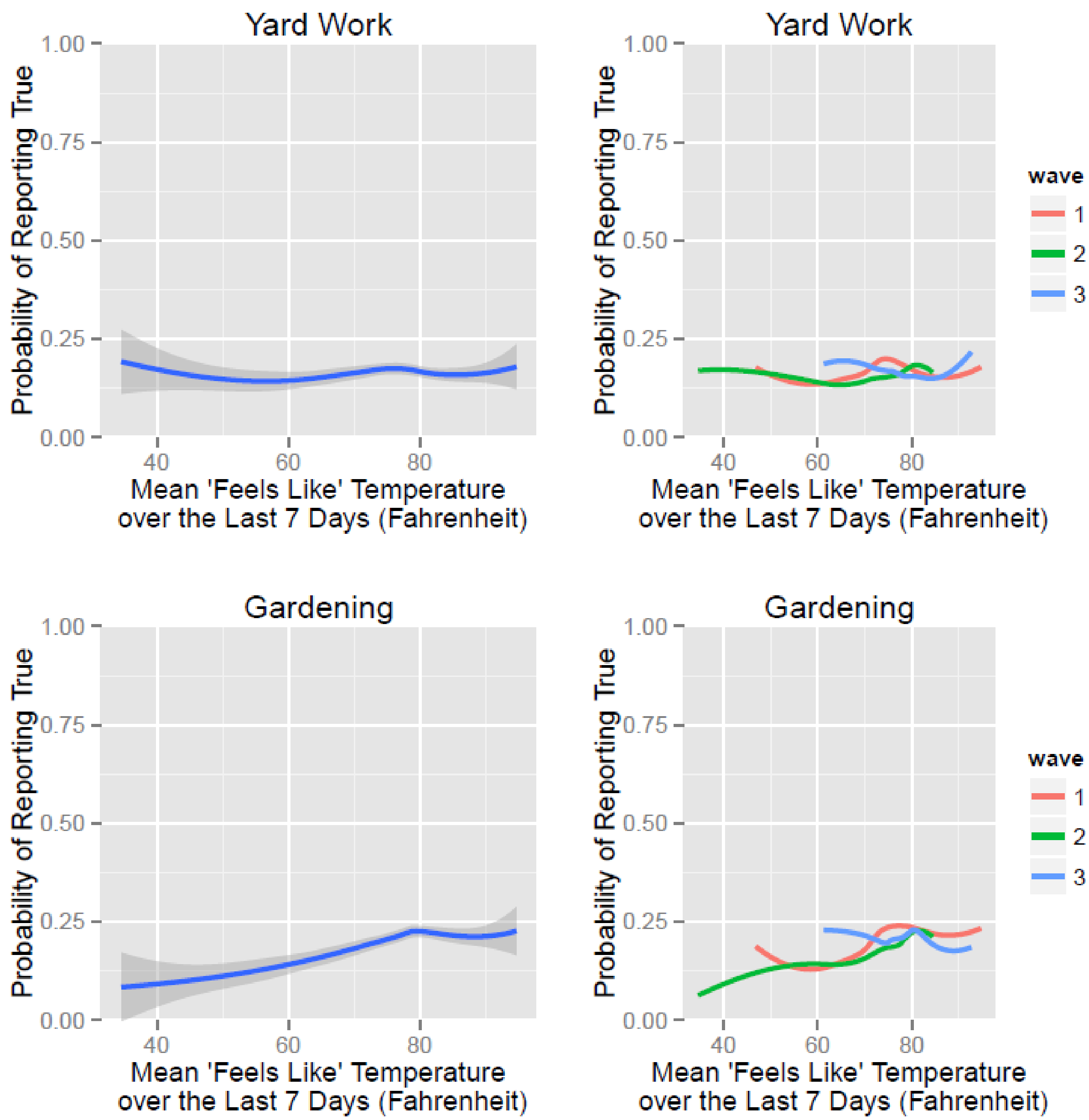


Figure A3.4 Probability of engaging in caring for others and average minutes per day of light sports by mean 'Feels-like' temperature over the past-week.

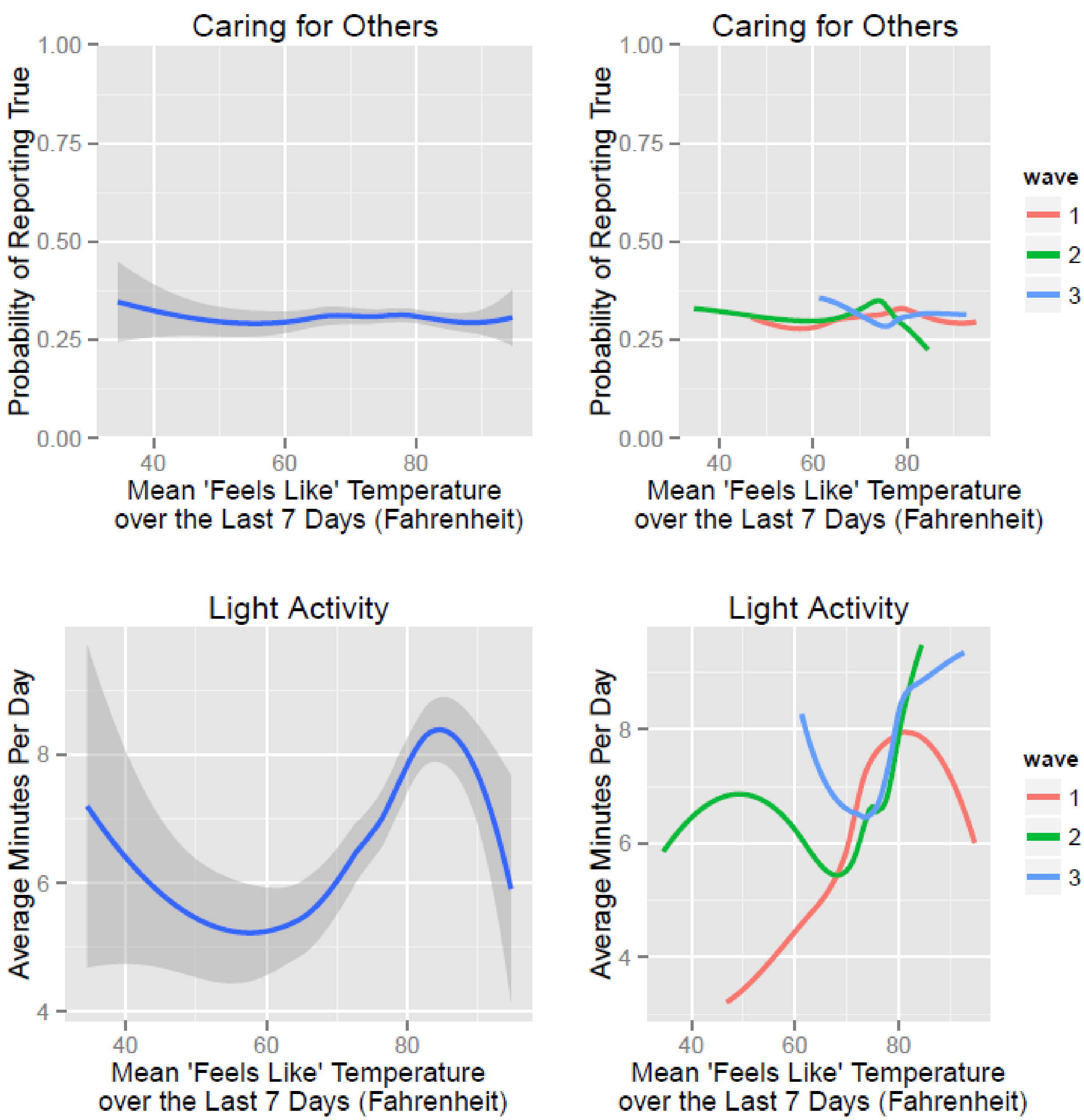


Figure A3.5 Average minutes per day of moderate and strenuous sports by mean 'Feels-like' temperature over the past-week.

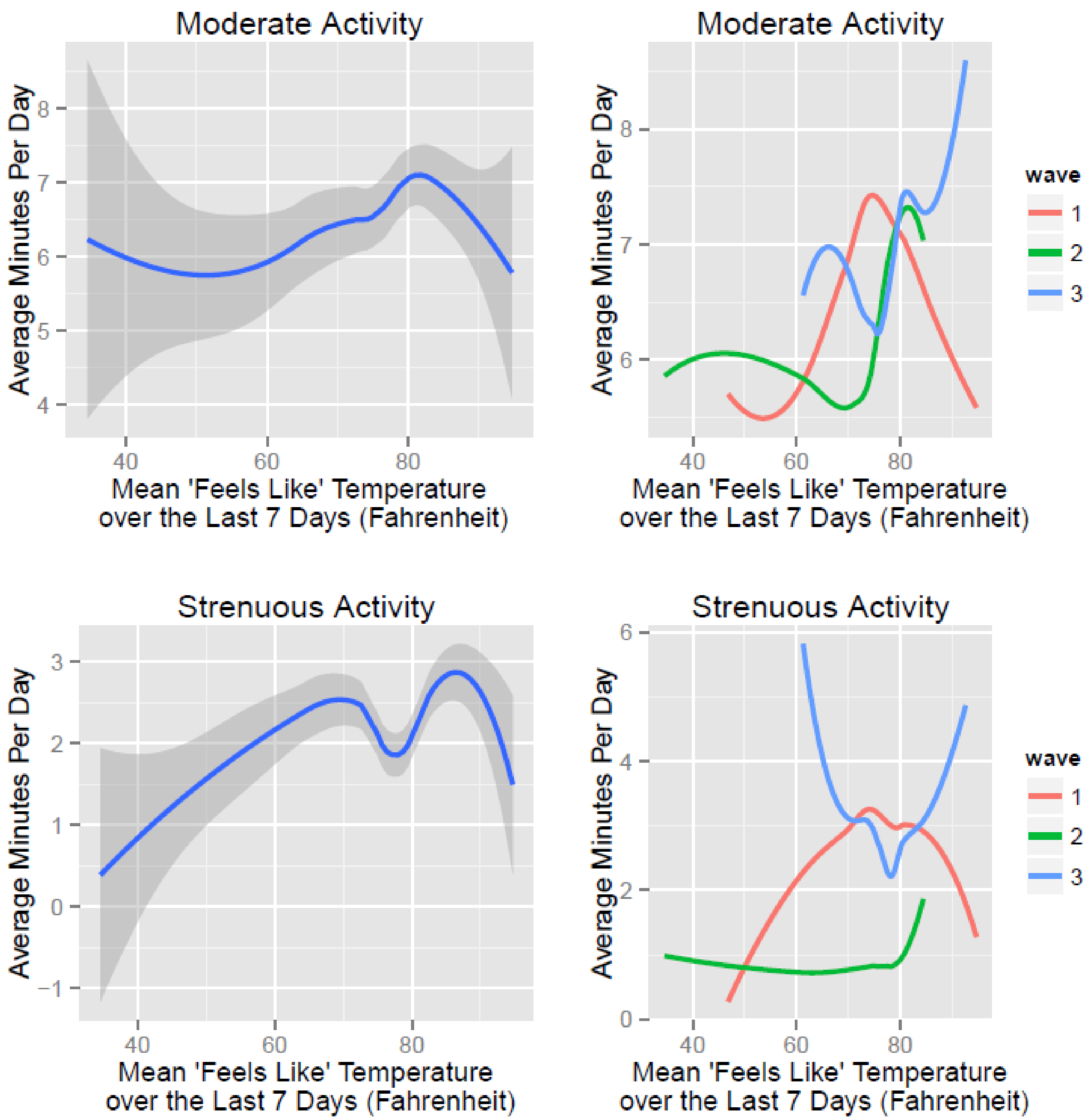
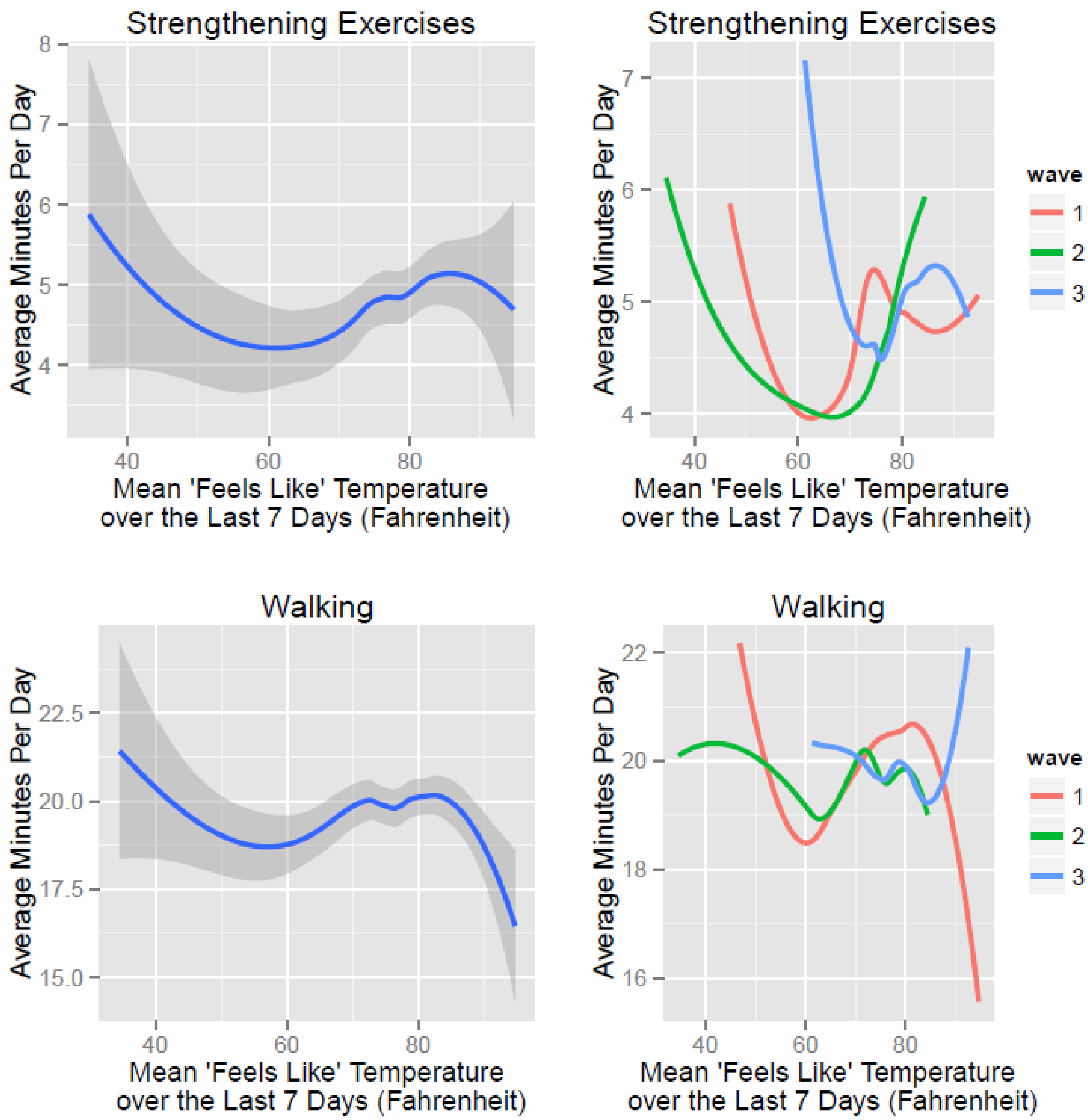


Figure A3.6 Average minutes per day of strengthening exercises and walking by mean 'Feels-like' temperature over the past-week.



Appendix A3.4. Testing the substantive finding's robustness to the choice of mixed rather than GEE models

Table A3.4.1. Mean Differences in PASE Score at Baseline and Mean Differences in Changes in PASE Score Associated with Baseline Physical Disorder and Changes in Physical Disorder over Time for 3,497 adult residents of New York City Surveyed from 2011-2013, using Generalized Estimating Equations (GEE) rather than mixed models.

	Mean Difference (95% Confidence Interval) in PASE Score		
Regression Coefficients	**Model 1[a]**	**Model 2[b]**	**Model 3[b]**
Difference at baseline per one-SD increase in baseline disorder	-4.0 (-7.2, -0.7)	-4.6 (-8.1, -1.1)	-4.4 (-7.9, -1.0)
Change per wave	0.1 (-1.4, 1.1)	-0.1 (-1.4, 1.2)	-0.1 (-1.4, 1.2)
Difference in change per wave for each one SD increase in baseline disorder		0.5 (-0.9, 1.9)	
Difference in change per wave for each one SD increase in time-varying disorder			0.5 (-0.9, 1.8)

[a] GEE model adjusting for baseline age, educational attainment, gender, race/ethnicity, functional status, neighborhood social cohesion, neighborhood pedestrian risk, and neighborhood walkability

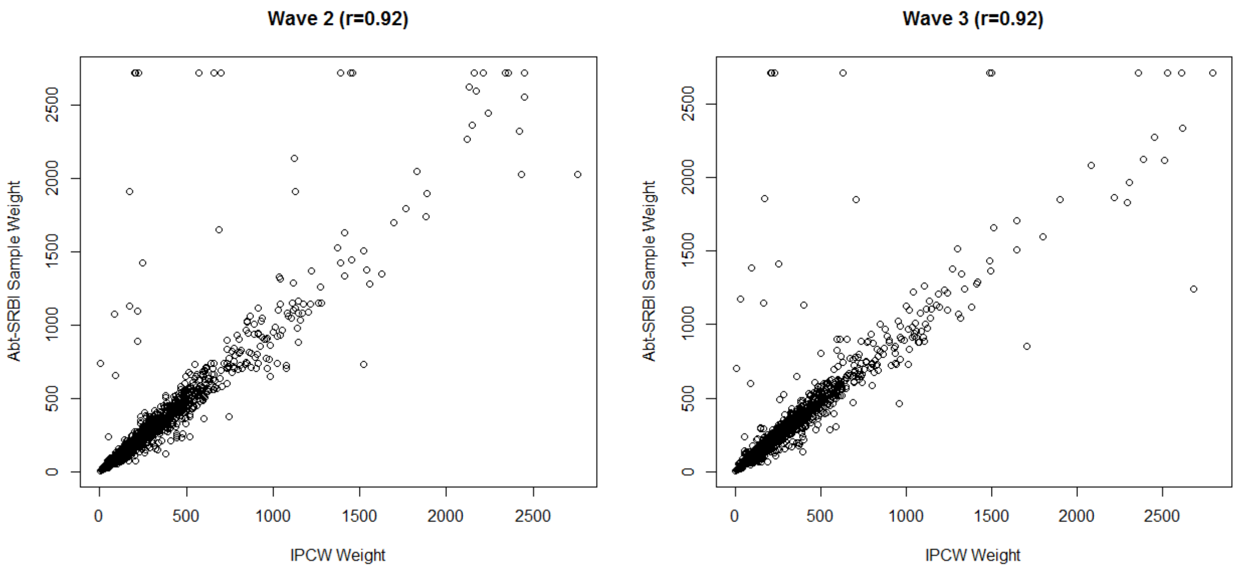
[b] GEE model adjusting for baseline age, educational attainment, gender, race/ethnicity, functional status, neighborhood social cohesion, neighborhood pedestrian risk, and neighborhood walkability, including baseline disorder/wave interaction term

[b] GEE model adjusting for baseline age, educational attainment, gender, race/ethnicity, functional status, neighborhood social cohesion, neighborhood pedestrian risk, and neighborhood walkability, including wave-specific disorder/wave interaction term

Appendix A3.5. Results of investigations of using Abt-SRBI's sample weights for subjects re-contacted at waves 2 and 3 rather than inverse probability of censoring weights incorporating baseline covariates in addition to demographic covariates.

First, we compared Abt-SRBI's sample weights and the IPCW-based weights. As expected, they were highly correlated, and the correlation did not change between waves (r=0.92 at each wave)

Figure A3.5.1 Scatterplots of Abt-SRBI sample weights at waves 2 and 3 and IPCW weights at waves 2 and 3.



As expected, final model results were quite similar (Table 5.2, next page)

Table A3.5.2 Mixed model results for 3,497 older adult residents of New York City, weighted to reflect the 65-75 year old population of New York City using sampling weights computed by Abt-SRBI based on demographic characteristics but not self-reported health or neighborhood characteristics

	Mean Difference (95% Confidence Interval) in PASE Score		
Regression Coefficients	**Model 1[a]**	**Model 2[b]**	**Model 3[b]**
Difference at baseline per one-SD increase in baseline disorder	-3.0 (-4.4, -1.6)	-3.2 (-4.8, -1.7)	-3.0 (-4.6, -1.5)
Change per wave	-0.5 (-1.3, 0.4)	-0.4 (-1.3, 0.4)	-0.5 (-1.3, 0.4)
Difference in change per wave for each one SD increase in baseline disorder		0.1 (-0.8, 0.9)	
Difference in change per wave for each one SD increase in time-varying disorder			0.0 (-0.8, 0.9)

Appendix A3.6. Results of 3-level hierarchical models clustering on subjects within community districts.

Table A3.6.1 3-level Mixed model results for 3,497 older adult residents of New York City, accounting for clustering within community districts

	Mean Difference (95% Confidence Interval) in PASE Score		
Regression Coefficients	**Model 1[a]**	**Model 2[b]**	**Model 3[b]**
Difference at baseline per one-SD increase in baseline disorder	-3.2 (-5.0, 1.5)	-3.3 (-5.2, -1.5)	-3.3 (-5.3, -1.5)
Change per wave	-0.5 (-1.7, 0.4)	-0.5 (-1.6, 0.7)	-0.5 (-1.3, 0.4)
Difference in change per wave for each one SD increase in baseline disorder		0.1 (-1.0, 1.1)	
Difference in change per wave for each one SD increase in time-varying disorder			0.1 (-0.7, 1.0)

Table A4.1. PASE items assessed of 3,497 residents of New York City aged 65-75 at baseline and each of two subsequent waves of follow-up.

Question	Response Options
Over the past week, on how many days did you take a walk outside your home or yard for any reason?	• Never • 1-2 days • 3-4 days • 5-7 days
On the days you walked, on average, how much time per day did you spend walking?	• Less than 30 minutes • 30-60 minutes • More than 60 Minutes
Over the past week, on how many days did you engage in light recreational activities?	• Never • 1-2 days • 3-4 days • 5-7 days
On the days you did them, on average, how much time per day did you engage in these light recreational activities?	• Less than 30 minutes • 30-60 minutes • More than 60 Minutes
Over the past week, on how many days did you engage in moderate recreational activities?	• Never • 1-2 days • 3-4 days • 5-7 days
On the days you did them, on average, how much time per day did you engage in these moderate recreational activities?	• Less than 30 minutes • 30-60 minutes • More than 60 Minutes
Over the past week, on how many days did you engage in strenuous recreational activities?	• Never • 1-2 days • 3-4 days • 5-7 days
On the days you did them, on average, how much time per day did you engage in these strenuous recreational activities?	• Less than 30 minutes • 30-60 minutes • More than 60 Minutes
Over the past week, on how many days did you do any exercises specifically to increase muscle strength and endurance?	• Never • 1-2 days • 3-4 days • 5-7 days
On the days that you did them, on average, how much time per day did you engage in exercises to increase muscle strength and endurance?	• Less than 30 minutes • 30-60 minutes • More than 60 Minutes
Over the past week, did you do any light housework, for example dusting or washing dishes?	• Yes • No
Over the past week, have you done any heavy housework or chores, for example vacuuming, scrubbing floors, washing windows, or carrying wood?	• Yes • No

Over the past week, did you engage in any of the following activities?	
1. Home repairs like painting, wallpapering, electrical work, etc	• Yes • No
2. Lawn work or yard care, including snow or leaf removal, mowing grass, etc.	• Yes • No
3. Outdoor gardening	• Yes • No
4. Caring for another person, such as children, dependent spouse, or another adult	• Yes • No

Table A4.2. Latent Transition Model Fit Statistics

Number of latent classes	AIC	BIC	G^2
3	33251.45	33532.07	33151.45
4	32139.65	32560.58	31989.65
5	30722.59	31306.27	30514.59
6	30299.69	31068.58	30025.69
7	29844.51	30821.06	29496.51
8	29727.84	30934.49	29297.84

Figure A4.1. Heat Map showing probability of transition between latent classes at waves 1 and 2

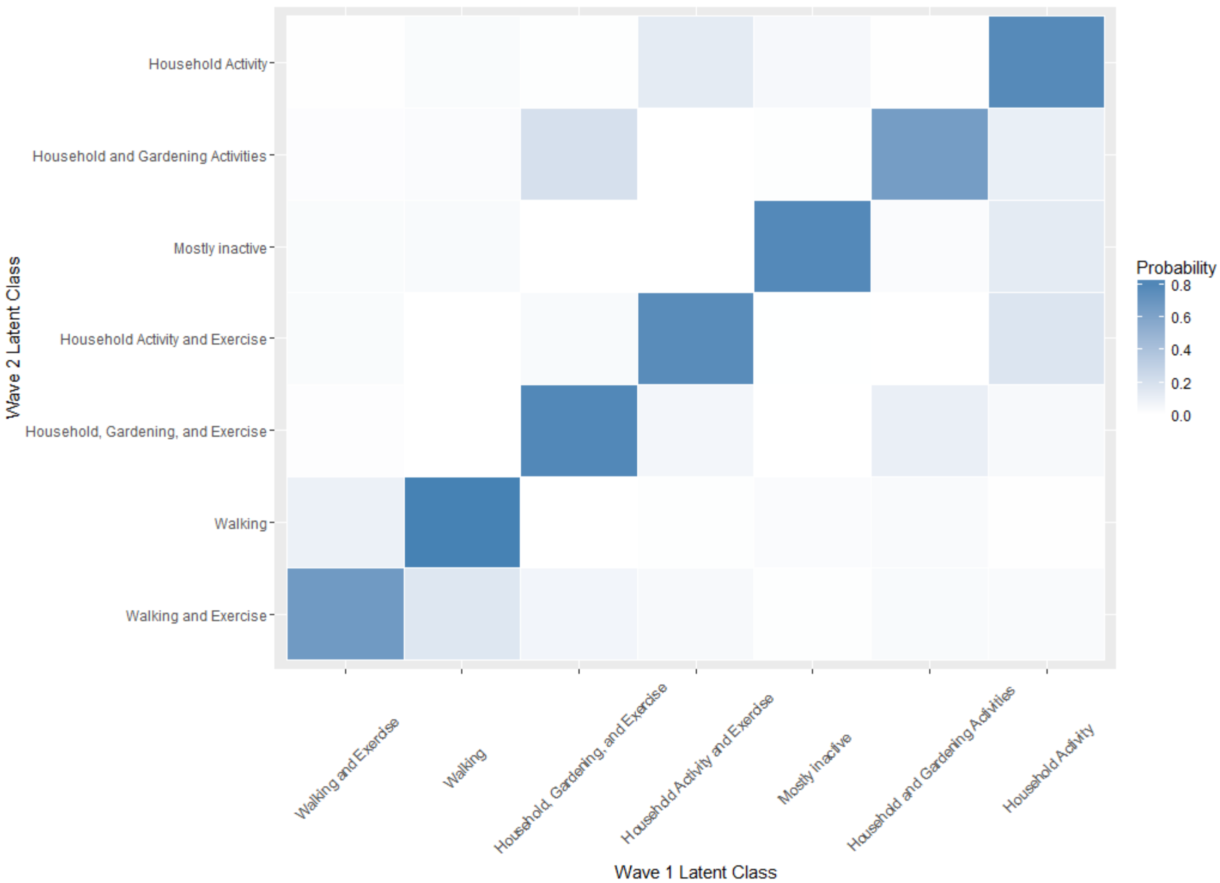


Table A5.1. Measures used in the NE-WAS

Category	Measure	Source	Log-transformed for Analysis
Demographics & Household Characteristics			
	Population Density	ACS	No
	Proportion Male	ACS	No
	Proportion Male Under Age 5	ACS	No
	Proportion Male Age 5-9	ACS	No
	Proportion Male Age 10-14	ACS	No
	Proportion Male Age 15-17	ACS	No
	Proportion Male Age 18-24	ACS	No
	Proportion Male Age 25-34	ACS	No
	Proportion Male Age 35-44	ACS	No
	Proportion Male Age 45-54	ACS	No
	Proportion Male Age 55-64	ACS	Yes
	Proportion Male Age 65-74	ACS	No
	Proportion Male Age 75-84	ACS	No
	Proportion Male Age 85 and Over	ACS	Yes
	Proportion Female Under Age 5	ACS	No
	Proportion Female Age 5-9	ACS	No
	Proportion Female Age 10-14	ACS	No
	Proportion Female Age 15-17	ACS	No
	Proportion Female Age 18-24	ACS	No
	Proportion Female Age 25-34	ACS	No
	Proportion Female Age 35-44	ACS	No
	Proportion Female Age 45-54	ACS	No
	Proportion Female Age 55-64	ACS	No
	Proportion Female Age 65-74	ACS	No
	Proportion Female Age 75-84	ACS	No
	Proportion Female Age 85 and Over	ACS	Yes
	Proportion White Alone	ACS	No
	Proportion Black Alone	ACS	Yes
	Proportion American Indian or Alaska Native Alone	ACS	Yes
	Proportion Native Hawaiian or Pacific Islander Alone	ACS	Yes
	Proportion Some Other Race Alone	ACS	Yes
	Proportion Two or More Races	ACS	Yes
	Proportion Non-Hispanic White Alone	ACS	No
	Proportion Non-Hispanic Black Alone	ACS	Yes
	Proportion Non-Hispanic American Indian or Alaska Native Alone	ACS	Yes
	Proportion Non-Hispanic Native Hawaiian or Pacific Islander Alone	ACS	Yes
	Proportion Non-Hispanic Some Other Race Alone	ACS	Yes
	Proportion Non-Hispanic Two or More Races	ACS	Yes
	Proportion Hispanic White Alone	ACS	Yes
	Proportion Hispanic Black Alone	ACS	Yes
	Proportion Hispanic American Indian or Alaska Native Alone	ACS	Yes
	Proportion Hispanic Native Hawaiian or Pacific Islander Alone	ACS	Yes
	Proportion Hispanic Some Other Race Alone	ACS	Yes

Proportion Hispanic Two or More Races	ACS	Yes
Proportion Family Households	ACS	No
Proportion Married-couple Family Households	ACS	Yes
Proportion Other Family Households	ACS	No
Proportion Other Family with Male Householder Households	ACS	Yes
Proportion Other Family with Female Householder Households	ACS	Yes
Proportion Non-Family Households	ACS	No
Proportion Non-Family with Male Householder Households	ACS	No
Proportion Non-Family with Female Householder Households	ACS	No
Proportion Households with one or more person under age 18	ACS	No
Proportion Family Households with one or more person under age 18	ACS	No
Proportion Married-couple Family Households with one or more person under age 18	ACS	No
Proportion Other Family Households with one or more person under age 18	ACS	Yes
Proportion Other Family with Male Householder Households with one or more person under age 18	ACS	Yes
Proportion Other Family with Female Householder Households with one or more person under age 18	ACS	Yes
Proportion Non-Family Households with one or more person under age 18	ACS	Yes
Proportion Non-Family with Male Householder Households with one or more person under age 18	ACS	Yes
Proportion Non-Family with Female Householder Households with one or more person under age 18	ACS	Yes
Proportion of Households with no people under age 18	ACS	Yes
Proportion Family Households with no people under age 18	ACS	Yes
Proportion Married Couple Households with no people under age 18	ACS	Yes
Proportion Other Family Households with Male Householder and no people under age 18	ACS	No
Proportion Other Family Households with Female Householder and no people under age 18	ACS	Yes
Proportion Non-Family Households with no people under age 18	ACS	No
Proportion Non-Family Households with Male Householder and no people under age 18	ACS	Yes
Proportion Non-Family Households with Female Householder and no people under age 18	ACS	Yes
Proportion of Households with a Householder who is White Alone	ACS	No
Proportion of Households with a Householder who is Black or African American Alone	ACS	Yes
Proportion of Households with a Householder who is American Indian and Alaska Native Alone	ACS	Yes
Proportion of Households with a Householder who is Asian Alone	ACS	Yes
Proportion of Households with a Householder who is	ACS	Yes

Native Hawaiian and Other Pacific Islander Alone		
Proportion of Households with a Householder who is Some Other Race Alone	ACS	Yes
Proportion of Households with a Householder who is two or more races	ACS	Yes
Proportion of Households with a Householder who is Hispanic or Latino	ACS	Yes
Proportion of Households with a Householder who is White Alone, not Hispanic or Latino	ACS	Yes
Proportion of Population living in Households:	ACS	Yes
Proportion of Population living in Family Households:	ACS	No
Proportion of Population living in Family Households and reporting status as Householder	ACS	No
Proportion of Population living in Family Households and reporting status relative to Householder as Spouse	ACS	Yes
Proportion of Population living in Family Households and reporting status relative to Householder as Child	ACS	No
Proportion of Population living in Family Households and reporting status relative to Householder as Grandchild	ACS	Yes
Proportion of Population living in Family Households and reporting status relative to Householder as Brother or sister	ACS	Yes
Proportion of Population living in Family Households and reporting status relative to Householder as Parent	ACS	Yes
Proportion of Population living in Family Households and reporting status relative to Householder as Other relatives	ACS	Yes
Proportion of Population living in Family Households and reporting status relative to Householder as Nonrelative	ACS	Yes
Proportion of Population living in Non-Family Households	ACS	Yes
Proportion of Population Living Alone in Non-Family Households	ACS	Yes
Proportion of Population Not living Alone in Non-Family Households	ACS	Yes
Proportion of Population with Nonrelatives in Non-Family Households	ACS	Yes
Proportion of Population living in group quarters	ACS	Yes
Proportion of population 15 and older that has never been married	ACS	No
Proportion of population 15 and older that is presently married	ACS	No
Proportion of population 15 and older that is presently separated from marital partners	ACS	Yes
Proportion of population 15 and older that is widowed	ACS	Yes
Proportion of population 15 and older that is divorced	ACS	No
Proportion of households with unmarried partners	ACS	No
Proportion of households with unmarried same-sex partners	ACS	Yes
Proportion of households with unmarried male same-sex partners	ACS	Yes
Proportion of households with unmarried female	ACS	Yes

	same-sex partners		
	Proportion of households with unmarried opposite-sex partners	ACS	Yes
	Proportion of households with unmarried opposite-sex partners wherein the male partner is the householder	ACS	Yes
	Proportion of households with unmarried opposite-sex partners wherein the female partner is the householder	ACS	Yes
	Proportion of population age 1 and older who lived at this location 1 year ago	ACS	No
	Proportion of population age 1 and older who have moved within the county in the past year	ACS	Yes
	Proportion of population age 1 and older who have moved between counties but within the state in the past year	ACS	Yes
	Proportion of population age 1 and older who have moved from another state in the past year	ACS	Yes
	Proportion of population age 1 and older who have moved from another country in the past year	ACS	Yes
	Proportion of population born in the United States	ACS	No
	Proportion of population that are naturalized citizens	ACS	Yes
	Proportion of population that are non-citizens	ACS	No
	Proportion of foreign-born population migrating to the US since 2000	ACS	No
	Proportion of foreign-born population migrating to the US between 1990 and 2000	ACS	No
	Proportion of foreign-born population migrating to the US between 1980 and 1990	ACS	No
	Proportion of foreign-born population migrating to the US prior to 1980	ACS	Yes
Education, Employment, and Income		ACS	
	Proportion of the adult population with less than high school education	ACS	No
	Proportion of the adult population who graduated from high school but have no further education	ACS	No
	Proportion of the adult population who have some college education but not a bachelor's degree	ACS	Yes
	Proportion of the adult population who have a bachelor's degree but no post-graduate education	ACS	Yes
	Proportion of the adult population with a master's degree but no doctorate or professional degree	ACS	Yes
	Proportion of the adult population with a degree from a professional school (e.g. law school) but no doctorate	ACS	Yes
	Proportion of the adult population with a doctorate	ACS	Yes
	Proportion of the adult population with a high school education or more	ACS	No
	Proportion of the adult population with some college education or more	ACS	No
	Proportion of the adult population with a bachelor's degree or more	ACS	Yes
	Proportion of the adult population with a master's degree or more	ACS	Yes
	Proportion of the adult population with a professional	ACS	Yes

	school degree or more		
	Proportion of the population aged 3 and over enrolled in school	ACS	No
	Proportion of the population aged 3 and over enrolled in public school	ACS	No
	Proportion of the population aged 3 and over enrolled in public pre-school	ACS	Yes
	Proportion of the population aged 3 and over enrolled in public K-8 school	ACS	No
	Proportion of the population aged 3 and over enrolled in public 9-12 school	ACS	No
	Proportion of the population aged 3 and over enrolled in public college	ACS	No
	Proportion of the population aged 3 and over enrolled in private school	ACS	Yes
	Proportion of the population aged 3 and over enrolled in private pre-school	ACS	Yes
	Proportion of the population aged 3 and over enrolled in private K-8 school	ACS	Yes
	Proportion of the population aged 3 and over enrolled in private 9-12 school	ACS	Yes
	Proportion of the population aged 3 and over enrolled in private college	ACS	Yes
	Proportion of population aged 16-19 not enrolled in high school or high school graduates	ACS	Yes
	Proportion of population aged 16-19 currently enrolled in high school	ACS	No
	Proportion of adult population in the labor force	ACS	No
	Proportion of adult population in the armed forces	ACS	Yes
	Proportion of adult population in the civilian labor force	ACS	No
	Proportion of adult population in the civilian labor force and currently employed	ACS	No
	Proportion of adult population in the civilian labor force and currently unemployed	ACS	Yes
	Proportion of employed adult population working in agriculture, forestry, fishing, or mining	ACS	Yes
	Proportion of employed adult population working in construction	ACS	Yes
	Proportion of employed adult population working in manufacturing	ACS	No
	Proportion of employed adult population working in wholesale	ACS	Yes
	Proportion of employed adult population working in retail	ACS	Yes
	Proportion of employed adult population working in transportation and warehousing	ACS	Yes
	Proportion of employed adult population working in information	ACS	Yes
	Proportion of employed adult population working in finance	ACS	Yes
	Proportion of employed adult population working in professional/scientific industries	ACS	Yes
	Proportion of employed adult population working in arts	ACS	Yes
	Proportion of employed adult population working in	ACS	Yes

Other industries		
Proportion of employed adult population working in public administration	ACS	Yes
Proportion of employed adult population working as management	ACS	Yes
Proportion of employed adult population working as professional	ACS	Yes
Proportion of employed adult population working as healthcare support	ACS	No
Proportion of employed adult population working as protective services	ACS	Yes
Proportion of employed adult population working as food prep worker	ACS	Yes
Proportion of employed adult population working as building and grounds worker	ACS	Yes
Proportion of employed adult population working as personal care and service worker	ACS	Yes
Proportion of employed adult population working as salesperson	ACS	No
Proportion of employed adult population working as office/admin support	ACS	No
Proportion of employed adult population working as farming/fishing/forestry worker	ACS	Yes
Proportion of employed adult population working as construction worker	ACS	Yes
Proportion of employed adult population working as production worker	ACS	Yes
Proportion of employed adult population working as transportation or moving person	ACS	Yes
Proportion of employed adult population working in private sector	ACS	No
Proportion of employed adult population working in public sector	ACS	Yes
Proportion of employed adult population that is self-employed	ACS	No
Proportion of employed adult population working in non-profit sector	ACS	Yes
Proportion of employed adult population working as unpaid family workers	ACS	Yes
Proportion of households with household income less than $10,000	ACS	No
Proportion of households with household income between $10,000 and $14,999	ACS	No
Proportion of households with household income between $15,000 and $19,999	ACS	No
Proportion of households with household income between $20,000 and $24,999	ACS	No
Proportion of households with household income between $25,000 and $29,999	ACS	No
Proportion of households with household income between $30,000 and $34,999	ACS	No
Proportion of households with household income between $35,000 and $39,999	ACS	No
Proportion of households with household income between $40,000 and $44,999	ACS	No
Proportion of households with household income	ACS	No

	between $45,000 and $49,999		
	Proportion of households with household income between $50,000 and $59,999	ACS	No
	Proportion of households with household income between $60,000 and $74,999	ACS	No
	Proportion of households with household income between $75,000 and $99,999	ACS	No
	Proportion of households with household income between $100,000 and $124,999	ACS	No
	Proportion of households with household income between $125,000 and $149,999	ACS	Yes
	Proportion of households with household income between $150,000 and $199,999	ACS	Yes
	Proportion of households with household income of $200,000 or more	ACS	Yes
	Proportion of households with any earned income	ACS	No
	Proportion of households with wage or salary income	ACS	No
	Proportion of households with self-employment income	ACS	No
	Proportion of households with interest, dividend, or rental income	ACS	Yes
	Proportion of households with social security income	ACS	No
	Proportion of households with supplemental security income (SSI)	ACS	Yes
	Proportion of households with public assistance income	ACS	Yes
	Proportion of households with retirement income	ACS	Yes
	Proportion of households with other income	ACS	Yes
	Proportion of families with incomes below the poverty level	ACS	Yes
	Proportion of families that both consist of a married couple and at least one child and are living below the poverty level	ACS	Yes
	Proportion of families that both consist of a married couple no children and are living below the poverty level	ACS	Yes
	Proportion of families that have a male householder with no wife and are living below the poverty level	ACS	Yes
	Proportion of families that have a male householder with no wife and one or more children and are living below the poverty level	ACS	Yes
	Proportion of families that have a male householder with no wife and no children and are living below the poverty level	ACS	Yes
	Proportion of families that have a female householder with no husband and are living below the poverty level	ACS	Yes
	Proportion of families that have a female householder with no husband and one or more children and are living below the poverty level	ACS	Yes
	Proportion of families that have a female householder with no husband and no children and are living below the poverty level	ACS	Yes
	Proportion of the population for whom poverty status is determined living below half the poverty level	ACS	Yes
	Proportion of the population for whom poverty status	ACS	Yes

	is determined living between half and three-quarters the poverty level		
	Proportion of the population for whom poverty status is determined living between three-quarters of the poverty level and the poverty level	ACS	Yes
	Proportion of the population for whom poverty status is determined living between the poverty level and 1.5 times the poverty level	ACS	Yes
	Proportion of the population for whom poverty status is determined living between 1.5 times and 1.99 times the poverty level	ACS	Yes
	Proportion of the population for whom poverty status is determined living above twice the poverty level	ACS	Yes
	Proportion of the population for whom poverty status is determined living below the poverty level	ACS	No
	Proportion of the population for whom poverty status is determined living between the poverty level and 1.99 times the poverty level	ACS	No
Housing			
	Proportion of occupied housing units occupied by owner	ACS	No
	Proportion of housing units that are occupied	ACS	Yes
	Proportion of vacant housing units that are for rent	ACS	Yes
	Proportion of vacant housing units that are for sale	ACS	Yes
	Proportion of vacant housing units that are neither for rent nor for sale	ACS	Yes
	Proportion of housing units that are single-family	ACS	Yes
	Proportion of housing units that are single-family detached	ACS	Yes
	Proportion of housing units that are single-family attached	ACS	Yes
	Proportion of housing units in structures with 2 units	ACS	Yes
	Proportion of housing units in structures with 3-4 units	ACS	Yes
	Proportion of housing units in structures with 5-9 units	ACS	Yes
	Proportion of housing units in structures with 10-19 units	ACS	Yes
	Proportion of housing units in structures with 20-49 units	ACS	Yes
	Proportion of housing units in structures with 50 or more units	ACS	No
	Proportion of housing units that are mobile homes	ACS	Yes
	Proportion of housing units that are boats, RVs, vans, etc.	ACS	Yes
	Proportion of employed adults who commute by car	ACS	Yes
	Proportion of employed adults who commute by transit	ACS	No
	Proportion of employed adults who commute by motorcycle	ACS	Yes
	Proportion of employed adults who commute by bicycle	ACS	Yes
	Proportion of employed adults who commute on foot	ACS	Yes
	Proportion of employed adults who commute by some other mode	ACS	Yes
	Proportion of employed adults who work at home	ACS	Yes
	Proportion of employed adults whose commute takes less than 10 minutes	ACS	Yes

	Proportion of employed adults whose commute takes 10-19 minutes	ACS	Yes
	Proportion of employed adults whose commute takes 20-29 minutes	ACS	Yes
	Proportion of employed adults whose commute takes 30-39 minutes	ACS	Yes
	Proportion of employed adults whose commute takes 40-59 minutes	ACS	Yes
	Proportion of employed adults whose commute takes 60-89 minutes	ACS	Yes
	Proportion of employed adults whose commute takes 90 minutes or more	ACS	Yes
Urban Form			
	Density of unique street intersections per km^2	NYS-ALIS	No
	Density of dead ends and cul-de-sacs per km^2	NYS-ALIS	Yes
	Density of 3-way street intersections per km^2	NYS-ALIS	Yes
	Proportion of intersections that are 3-way intersections	NYS-ALIS	Yes
	Density of 4-way street intersections per km^2	NYS-ALIS	No
	Proportion of intersections that are 4-way intersections	NYS-ALIS	Yes
	Connected node ratio (3+ way intersections/dead ends)	NYS-ALIS	No
	Proportion of land area covered by sidewalks	NYS-ALIS	No
	Proportion of land area covered by street roadbed	NYS-ALIS	Yes
	Proportion of land area covered by tree canopy	NYC Urban Tree Canopy Assessment	Yes
	Proportion of roadbed area covered by tree canopy	NYC Urban Tree Canopy Assessment	Yes
	Proportion of total building area devoted to residential use	NYC Department of City Planning	No
	Land Use Mix		Yes
	Frank Walkability Index		No
	Built Environment and Health Group Walkability Index		No
Transit Access			
	Density of unique bus stops per km^2 in 2004	New York City Metropolitan Transit Authority	Yes
	Density of unique bus stops per km^2 in 2012	New York City Metropolitan Transit Authority	Yes
	Density of bus stops per km^2 allowing transfer between buses (in 2012)	New York City Metropolitan Transit Authority	Yes
	Density of subway stops per km^2 allowing transfer between lines	New York City Metropolitan Transit Authority	Yes
	Density of subway stops per km^2	New York City Metropolitan	Yes

		Transit Authority	
Crime			
	Total Crime Index	ESRI	No
	Personal Crime Index	ESRI	No
	Murder Index	ESRI	No
	Rape Index	ESRI	No
	Robbery Index	ESRI	Yes
	Assault Index	ESRI	Yes
	Property Crime Index	ESRI	Yes
	Burglary Index	ESRI	Yes
	Larceny Index	ESRI	Yes
	Motor vehicle Theft Index	ESRI	Yes
	Homicide Density in 2005	New York Times Homicide Map	Yes
	Homicide Density in 2006	New York Times Homicide Map	Yes
	Homicide Density in 2007	New York Times Homicide Map	Yes
	Homicide Density in 2008	New York Times Homicide Map	Yes
	Homicide Density in 2009	New York Times Homicide Map	Yes
	Homicide Density in 2010	New York Times Homicide Map	Yes
Disorder			
	Street View Disorder	Google Street View/CANVAS	No
	Mean Street Cleanliness in 1999	Project Scorecard	No
	Mean Street Cleanliness in 2000	Project Scorecard	No
	Mean Street Cleanliness in 2001	Project Scorecard	No
	Mean Street Cleanliness in 2002	Project Scorecard	No
	Mean Street Cleanliness in 2003	Project Scorecard	No
	Mean Street Cleanliness in 2004	Project Scorecard	No
	Mean Street Cleanliness in 2005	Project Scorecard	No
	Mean Street Cleanliness in 2006	Project Scorecard	No
	Mean Street Cleanliness in 2007	Project Scorecard	No
	Mean Street Cleanliness in 2008	Project Scorecard	No
	Mean Street Cleanliness in 2009	Project Scorecard	No

	Mean Street Cleanliness in 2010	Project Scorecard	No
	Mean Street Cleanliness in 2011	Project Scorecard	No
	Mean Street Cleanliness in 2012	Project Scorecard	No
	Proportion of houses with broken windows	NYC Housing Vacancy Survey	Yes
	Proportion of houses with rotten or loose windows	NYC Housing Vacancy Survey	Yes
	Proportion of houses with boarded up windows	NYC Housing Vacancy Survey	No
	Proportion of houses with no window problems	NYC Housing Vacancy Survey	No
	Presence of any houses with boarder up or broken windows	NYC Housing Vacancy Survey	Yes
Parks & Greenspace			
	Proportion of land area devoted to community gardens	GrowNYC	Yes
	Proportion of land area devoted to parks		Yes
	Proportion of land area devoted to small parks		Yes
	Proportion of land area devoted to playgrounds		Yes
	Proportion of land area devoted to green streets		Yes
Safety from Traffic			
	Density of collisions involving pedestrians in 2010	New York State Department of Transportation	Yes
	Density of collisions in which pedestrians were injured in 2010	New York State Department of Transportation	Yes
	Density of collisions in which pedestrians were killed in 2010	New York State Department of Transportation	Yes
	Density of collisions involving cyclists in 2010	New York State Department of Transportation	Yes
	Density of collisions in which cyclists were injured in 2010	New York State Department of Transportation	Yes
	Density of collisions in which cyclists were killed in 2010	New York State Department of Transportation	Yes
	Density of collisions involving pedestrians from 1995-	New York	Yes

2010	State Department of Transportation	
Density of collisions in which pedestrians were injured from 1995-2010	New York State Department of Transportation	Yes
Density of collisions in which pedestrians were killed from 1995-2010	New York State Department of Transportation	Yes
Density of collisions involving cyclists from 1995-2010	New York State Department of Transportation	Yes
Density of collisions in which cyclists were injured from 1995-2010	New York State Department of Transportation	Yes
Density of collisions in which cyclists were killed from 1995-2010	New York State Department of Transportation	Yes
Density of collisions involving pedestrians from 1995-1999	New York State Department of Transportation	Yes
Density of collisions in which pedestrians were injured from 1995-1999	New York State Department of Transportation	Yes
Density of collisions in which pedestrians were killed from 1995-1999	New York State Department of Transportation	Yes
Density of collisions involving cyclists from 1995-1999	New York State Department of Transportation	Yes
Density of collisions in which cyclists were injured from 1995-1999	New York State Department of Transportation	Yes
Density of collisions in which cyclists were killed from 1995-1999	New York State Department of Transportation	Yes
Density of collisions involving pedestrians from 2000-2009	New York State Department of Transportation	Yes
Density of collisions in which pedestrians were injured from 2000-2009	New York State Department of Transportation	Yes
Density of collisions in which pedestrians were killed	New York	Yes

	from 2000-2009	State Department of Transportation	
	Density of collisions involving cyclists from 2000-2009	New York State Department of Transportation	Yes
	Density of collisions in which cyclists were injured from 2000-2009	New York State Department of Transportation	Yes
	Density of collisions in which cyclists were killed from 2000-2009	New York State Department of Transportation	Yes

ACS: American Community Survey; NYS-ALIS: New York State Accident Location Information System GIS layer; ESRI: Environmental Systems Research Institute, Inc.

Table A5.2. Specific neighborhood measures identified as most predictive for several physical activity outcomes using regression modes in the full cohort. All analyses control for subject age, race/ethnicity, educational attainment, household income, and gender.

	PASE Score	Gardening	Walking Daily	Heavy Housework
Count of measures that remained significant after Bonferroni correction	7 (2.1%)	32 (9.5%)	40 (11.9%)	2 (0.1%)
Top 5 statistically significant neighborhood measures (by p-value of coefficient)	People living in households with incomes less than half the poverty level (-)	Neighborhood Physical Disorder (-)	Proportion of residents with 60-90 minute travel time to work (-)	Proportion of residents aged 3 and higher enrolled in school (-)
	People living in households with incomes below the poverty line (-)	Proportion of adult population with at least some college education (+)	Proportion of adult population working in professional or management industries (+)	Proportion of residents aged 3 and higher enrolled in public college (+)
	No problems with windows in HVS survey (+)	People living in households above twice the poverty line (+)	Broken windows in HVS survey (-)	
	People living in households with incomes more than twice the poverty level (+)	People living in households with incomes less than half the poverty level (-)	Proportion of working adult population commuting by car, truck, or van (-)	
	People living in households with incomes between half and three-quarters of the poverty level (-)	Proportion of households with incomes between 10K and 15K (-)	Proportion of adult population with at least some college education (+)	

Table A5.3. Specific neighborhood measures identified as most predictive for several physical activity outcomes using LASSO on cohort created by selecting a random imputation for each subject. All analyses control for subject age, race/ethnicity, educational attainment, household income, and gender.

	PASE Score	Gardening	Walking Daily	Heavy Housework
Count of measures that remained significant after Bonferroni correction	23 (6.8%)	9 (2.6%)	23 (6.8%)	21 (6.2%)
Top 5 statistically significant neighborhood measures (by p-value of coefficient)	People living in households with incomes less than half the poverty level (-)	People living in households with incomes between 1 and 1.5 times the poverty level (-)	Broken windows in HVS survey (-)	Proportion of residents aged 3 and older enrolled in any school (-)
	Proportion of intersections that are 3-way (+)	People living in households with incomes less than half the poverty level (-)	Proportion of households with a Hispanic householder (-)	Proportion of residents aged 3 and older enrolled in public college (+)
	People living in households with incomes between half and three-quarters of the poverty level (-)	Neighborhood Physical Disorder (-)	Proportion of residents who commute by car, truck, or van (-)	Density of pedestrian fatalities from 2000-2009 (-)
	Proportion of residents with 30-39 minute commutes (+)	People living in households with incomes between half and three-quarters of the poverty level (-)	Proportion of residents living in non-family households with non-relatives (+)	Proportion of residents whose relationship to the householder is 'parent' (+)
	Proportion of population in non-family households not living alone (+)	Proportion of vacant housing units that are for rent (-)	Proportion of households with incomes between 35K and 40K (-)	Density of cyclist fatalities from 1995-1999 (-)

Figure A5.1. Regularization path graph for the LASSO regression on PASE score. Each line represents the coefficient estimate for one (normalized) parameter as the penalty increases (to the right) The near-horizontal lines represent individual covariates (age, race/ethnicity, educational attainment, gender, household income) that are included in the model unpenalized.

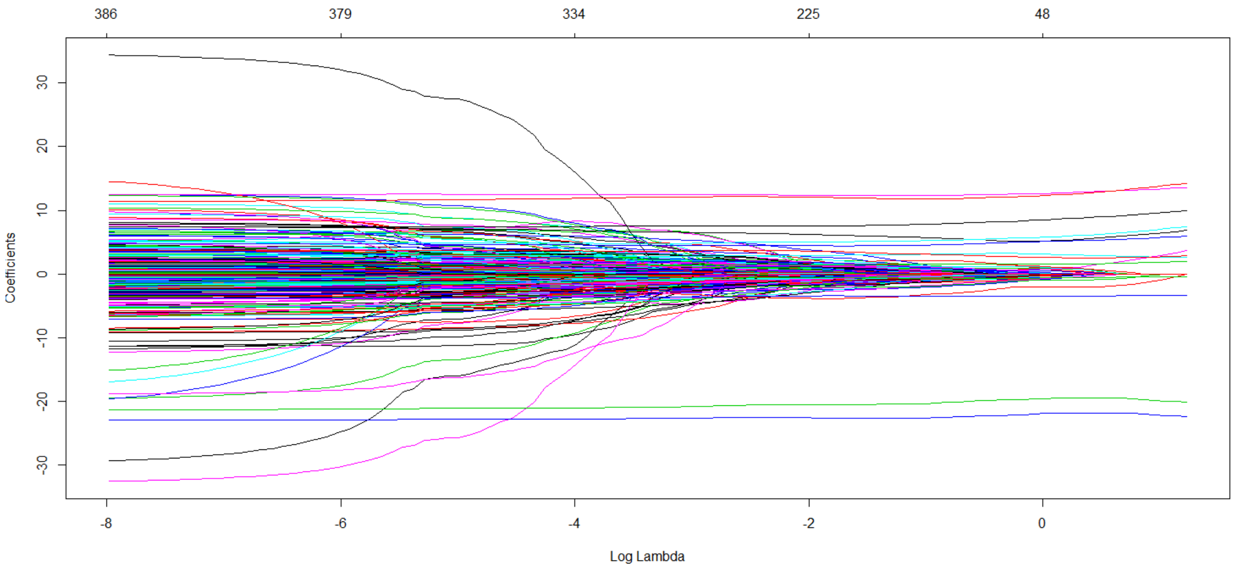


Figure A5.2 Sample output from regression Tree from 'party' package in R predicting probability of gardening as a function of neighborhood and individual covariates. This tree is the raw output corresponding to the tree as that shown in Figure 5.2.