



Published in final edited form as:

Psychol Sport Exerc. 2011 January 1; 12(1): 54–60. doi:10.1016/j.psychsport.2009.09.003.

Environment and Physical Activity Dynamics: The Role of Residential Self-selection

Janne Boone-Heinonen,

University of North Carolina at Chapel Hill, 123 West Franklin St. CB#8120; Chapel Hill, NC 27516-3997

Penny Gordon-Larsen *

University of North Carolina at Chapel Hill, 123 West Franklin St. CB#8120; Chapel Hill, NC 27516-3997

David K Guilkey,

University of North Carolina at Chapel Hill, 123 West Franklin St. CB#8120; Chapel Hill, NC 27516-3997

David R Jacobs Jr., and

University of Minnesota, Division of Epidemiology and Community Health, 1300 South Second Street, Suite 300; Minneapolis, MN 55454

Barry M Popkin

University of North Carolina at Chapel Hill, 123 West Franklin St. CB#8120; Chapel Hill, NC 27516-3997

Abstract

Objectives—Within the socio-ecologic framework, diet and physical activity are influenced by individual, interpersonal, organizational, community, and public policy factors. A basic principle underlying this framework is that environments can influence an individual's behavior. However, in the vast majority of cross-sectional and even the few longitudinal studies of this relationship, the question of whether individuals select their area of residence based on physical activity-related amenities is ignored.

In this paper, we address a critical methodological issue: self-selection of residential location, which is generally not accounted for, and can significantly compromise research on the relationship between environmental factors and physical activity behaviors.

Method—We define and discuss the problem of residential self-selection in the study of neighborhood influences on health and health behavior, review methods used to control for residential self-selection in the literature, and present our strategy for addressing this potentially important source of bias.

Conclusion—Existing research has built our understanding of residential self-selection bias, but important gaps remain. Our strategy uses data from a longitudinal cohort study linked to contemporaneous environmental measures to create a multi-equation model system to

© 2009 Elsevier Ltd. All rights reserved.

*Please Address Correspondence & Reprint Requests To: Penny Gordon-Larsen, Ph.D., gordon_larsen@unc.edu, Fax: 919-966-9159.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

simultaneously estimate residential choice, environmental influences on physical activity, and downstream health outcomes such as obesity and clinical cardiovascular disease risk factor measures.

Greater focus on environmental factors is common to most recent examinations of the obesity and inactivity epidemic (Ogden et al., 2006; Popkin & Gordon-Larsen, 2004). In light of limited success of individual-level obesity prevention strategies (Sharma, 2006), evidence that the neighborhood environment is related to obesity, physical activity, and disease risk (Papas et al., 2007; Wendel-Vos, Droomers, Kremers, Brug, & van Lenthe, 2007) is appealing to public health practitioners. However, despite calls for population-wide, environmental interventions (Sallis, Bauman, & Pratt, 1998), research on environmental health determinants is in its infancy, warranting elucidation of methodological shortcomings in existing research. In this paper, we provide an overview of environmental factors hypothesized to influence obesity and physical activity within the socio-ecologic framework, with focus on the socioeconomic and built environment. We discuss limitations to existing neighborhood health research, followed by a multidisciplinary review of strategies to adjust for arguably the most serious limitation, residential self-selection. Finally, we present one of our approaches for addressing residential self-selection in a prospective cohort study of black and white adults followed over 20 years.

Key Aspects of the Physical Activity Environment

A rapidly growing body of research examines the role of a vast range of contextual factors in influencing how individuals move throughout the day (Saelens & Handy, 2008; Wendel-Vos et al., 2007). The socio-ecologic framework (Figure 1) is useful for theorizing and testing contextual influences ranging across built and socioeconomic environments.

The Socio-ecologic Framework

The socio-ecologic framework describes five interactive levels of influence on health-related behaviors: intrapersonal, interpersonal, institutional/organizational, community, and public policy. Historically, health behavior interventions have focused on intra-personal factors such as knowledge (Haskell et al., 2007) and motivation (Marcus & Owen, 1992), interpersonal factors such as influence of a spouse's behavior (Gorin et al., 2008), or organizational factors, such as workplace supports like onsite fitness equipment (Pratt et al., 2007).

While neighborhoods include inter-personal influences (e.g. social support, social cues), the majority of neighborhood health research focuses on the socioeconomic and built environments (Ball, Timperio, & Crawford, 2006; Popkin, Duffey, & Gordon-Larsen, 2005). The importance of multiple levels of influence in public health (Diez-Roux, 2000) has been recognized in the past decade, starting with research showing increased coronary heart disease incidence associated with living in a disadvantaged neighborhood, independent of individual socioeconomic position (Diez-Roux et al., 1997). While each level can be influenced by any other level, public policy can potentially influence all levels and is arguably the ultimate goal of most neighborhood health research (Lee & Moudon, 2004; Sallis et al., 2006).

The Built and Socioeconomic Environments

The built environment is a major component of community design comprised of aspects such as buildings, transportation systems, parks, and greenways. It is a particularly promising target for improving population health due to its apparent influence on physical activity and hence chronic disease prevention, as well as existing linkages with local policies.

Researchers have found relatively consistent relationships between physical activity and urban sprawl (Ewing, Brownson, & Berrigan, 2006), alternatively referred to as “walkability” (Frank, Schmid, Sallis, Chapman, & Saelens, 2005), incorporating street connectivity measures, land use mix, housing density, and block lengths or retail floor area. Associations between physical activity and pedestrian and biking infrastructure such as sidewalks and bike lanes have been mixed (Giles-Corti & Donovan, 2002; Krizek & Johnson, 2006). Physical activity is related to access to recreational resources such as parks or physical activity facilities in youth (Gordon-Larsen, Nelson, Page, & Popkin, 2006) and adults (Diez Roux et al., 2007; Giles-Corti et al., 2005). Retail destinations such as shops and restaurants within walking distance are also correlated with walking behaviors (Handy, Cao, & Mokhtarian, 2006; Krizek & Johnson, 2006).

The socioeconomic environment, comprised of economic factors (e.g., poverty rate) and social factors (e.g., racial composition or crime and safety), is related to physical activity, obesity, and disease in existing research (Rundle et al., 2008; Wen, Browning, & Cagney, 2007).

Dynamic Interactions between Levels of Influence

An environment perceived to be rich in activity opportunities can lower the perceived burden of engaging in physical activity (intra-personal factor), act as a visual reminder to be active, or enforce activity as a cultural norm (inter-personal factors) (Sallis et al., 1990). While perceived environment measures are not addressed in our research, we assume that any causal influence of the objective environment on physical activity occurs at least in part through perceptions of the environment (Sallis & Owen, 2002), thus involving both intra- and inter-personal level factors.

Further, the dynamic relationships among socio-ecologic framework levels imply that intra- and inter-personal factors influence community-level exposures, through changes in the environment around stationary residents, residential mobility, or some combination of the two. By way of simple example, the first mechanism would occur if the city opened a new basketball court in response to demand by community members. The second mechanism would occur if those motivated to exercise choose to move into neighborhoods with parks or close access to fitness facilities. The latter example describes one major criticism of neighborhood health research and the focus of our study: residential self-selection (Oakes, 2004). In other words, due to concerns about self-selectivity, understanding of whether additional or improved recreation options will enhance activity patterns in any selected neighborhood is limited.

Limitations of Current Research

While “new urbanist” and Smart Growth principles (Rodriguez, Khattak, & Evenson, 2006) are already applied in community planning, policies and existing practices designed to encourage physical activity through environmental changes will remain without a solid scientific evidence base until key methodological and conceptual research gaps (Diez Roux, 2001; Oakes, 2004) are filled. Most research has been conducted in samples with limited racial and ethnic diversity and derived from confined geographic areas, limiting generalization of findings as well as variability in environmental measures. Heterogeneity of relationships by gender, race/ethnicity, or life stage has been largely ignored. Further, the literature is largely cross-sectional, which is particularly vulnerable to bias due to residential self-selection. Thus, studies of built environment-health relationships in large, demographically diverse samples residing in diverse environmental contexts and followed over time are needed.

Several of these limitations are related to observational study designs (Oakes, 2004). While randomized controlled trials that experimentally assign families or amenities to neighborhoods (Cummins, Petticrew, Higgins, Findlay, & Sparks, 2005) may help to address these limitations, experimental assignment of neighborhoods (or neighborhood characteristics) is not often financially or politically feasible. Therefore, we focus on advancements in statistical adjustment methods, availability of richer, longitudinal environmental datasets, and innovative study designs that have, and can continue to, vastly improve the validity of observational studies.

In the next sections, we discuss residential self-selection, what we consider the most serious limitation to existing research, and approaches used to adjust for it. We refer readers to two excellent reviews on residential self-selection in the context of travel behavior: Bhat and Guo discuss several adjustment methods for control for residential self-selection and present their own simultaneous modeling strategy (Bhat & Guo, 2007), and Mokhtarian and Cao review methodologies used to address “attitude-induced” residential self-selection (Mokhtarian & Cao, 2008). Our discussion builds on these reviews by incorporating perspectives from health research using epidemiologic and econometric methods.

Residential Self-Selection: A Major Research Gap

Potential bias due to residential self-selection has been identified as the primary limitation in built environment research (Diez Roux, 2004) and is of particular concern in cross-sectional studies which predominate the existing research. That is, selection of a neighborhood may be related to both the neighborhood exposure and the health outcome of interest.

A frequently overlooked point is that bias can result if factors driving residential selection are either *directly* or *indirectly* related to the exposure and outcome. In the context of built environment effects on physical activity, bias due to a direct relationship will result if already physically active individuals select neighborhoods based on their activity-supporting amenities. As an example of indirect relationships which may lead to bias, low income families may choose a neighborhood based solely on the affordability of housing; if these neighborhoods also contain inadequate physical activity resources and the families are less physically active (Gordon-Larsen et al., 2006), the built environment–physical activity relationship can be overestimated. Put simply, positive relationships between the built environment and physical activity can be attributed to (1) the effect of the environment on physical activity, (2) the effect of propensity for physical activity, or *characteristics related to physical activity*, on residential choice, or (3) both.

Formally, consider the following model of physical activity (PA) as a function of vectors of environmental exposures of interest (E), sociodemographic characteristics (S), measured residential preferences (P), and unmeasured or unmeasurable characteristics that are related to PA (U). ε is an error term assumed to be random.

$$PA = \beta_0 + \beta_1 E + \beta_2 S + \beta_3 P + \beta_4 U + \varepsilon \quad (1)$$

Typical analysis of associations between the built environment and physical activity include PA, E, and S using traditional multivariate adjustment of common sociodemographic measures. Some studies include P, which capture self-reported residential preferences that may influence residential choice. However, U is unmeasured and is thus omitted from the model, and variability in PA explained by U must be relegated to the error term in the model to form a composite error $\varepsilon^* = \beta_4 U + \varepsilon$. This is permissible if U is unrelated to E, S, and P. However, if ε^* is correlated with the independent variables, standard estimation methods will lead to biased estimates of the built environment variables coefficients (β_1) while also

potentially distorting the estimates of the remaining coefficients in the equation (β_2 and β_3). Understanding these complex inter-relationships is essential for obtaining precise, robust, and unbiased estimates of neighborhood effects (Duncan & Raudenbush, 2001).

Components of U related to selection of a neighborhood may include unmeasured preferences for neighborhood characteristics related to physical activity amenities or other features such as schools, proximity to work or family, or other factors. These components lead to residential self-selection bias, described as unmeasured confounding by epidemiologists and unobserved heterogeneity by economists (Zohoori & Savitz, 1997).

Direct Evidence of Residential Self-Selection

Largely, the literature on environmental determinants of physical activity and obesity treats residential decisions as exogenous factors. However, the strong roles of race/ethnicity and income in residential selection are well documented in migration and residential mobility (Ioannides & Zabel, 2008; Sampson & Sharkey, 2008) and housing selection (Nechyba & Strauss, 1998; Song, 2003) research. Self-selection into neighborhoods comprised of demographically similar households may result in race/ethnic stratification across neighborhoods, even in the situation of equivalent expenditures (and potentially amenities) across these neighborhoods (Epple, 2003). Coupled with socioeconomic and racial disparities in physical activity (Gordon-Larsen, Adair, & Popkin, 2002), the *indirect* residential self-selection bias mechanism is well supported. Further, physical activity-related facilities are inequitably distributed across neighborhoods (Gordon-Larsen et al., 2006).

There is also evidence for the *direct* residential self-selection bias mechanism. Recent market surveys support increasing yet varied preferences for traditionally designed communities (e.g., centrally located retail, alternative transportation infrastructure) (Handy, Sallis, Weber, Maibach & Hollander, 2008). Consumers who prefer such neighborhoods might tend to have higher physical activity levels. Indeed, subjects citing access to transit as an important reason for living in a “transit-oriented development” were almost 20 times more likely to use rail transit than those who did not cite this reason (Lund, 2006). Likewise, physical activity and belief that an activity-friendly community will support active transit are significant predictors of desiring to live in an activity-friendly community (Librett, Yore, Schmid, & Kohl, 2007).

While these studies provide evidence of residential selection bias, they rely on self-reported preference, belief, and behavior data which have important limitations. First, residential choices are determined by a large set of variables such as affordability, convenience, and proximity to social support networks that may not be articulated by respondents, so reported preference for activity-supportive communities has limited meaning (Nechyba & Strauss, 1998; Song & Knaap, 2003). Second, preferences are endogenous (in epidemiologic terminology, confounded by unmeasured variables): unobserved factors such as financial constraints (i.e. affordability) may influence both self-reported preferences and selection of neighborhoods with amenities of interest. Failure to control for the endogeneity of preferences will result in biased estimates for the preference variables.

In contrast, residential selection measured using observed, rather than self-reported data involves examination of the environments that individuals actually move into, regardless of their motivations for doing so. Bhat and Guo use such an approach with cross-sectional data, estimating the utility function for current residential location (a function of built and other environment attributes) as the first step in a two-part model (Bhat & Guo, 2007). With longitudinal data, exogenous factors can more easily be identified by using measures occurring prior to residential selection.

In particular, the transition from young to middle adulthood captures a period of vast changes in lifestyle associated with marriage, and children, and career advancement as well as weight gain and physical activity decline. Longitudinal data across a large number of residentially mobile individuals and repeated measures of physical activity and built environment through young and middle adulthood provide a great opportunity to examine observed residential selection patterns and estimate the magnitude of potential residential self-selection bias.

Strategies to Control for Residential Selection Bias

Associations between built environment factors and behavioral outcomes have been the topic of investigations across several fields including urban planning, transportation, economics, and epidemiology, each with their own methodological norms, culminating in recent and increased interdisciplinary research (Saelens, Sallis, & Frank, 2003; Sloane, 2006). To adjust for residential self-selection bias, these researchers have taken various approaches, including longitudinal designs, adjustment for self-reported residential preferences, adjustment for observed predictors of residential selection, and simultaneous equation methods.

Longitudinal Designs

The most ideal *observational* designs are longitudinal, and assess changes in physical activity in relation to changes in the built environment. For example, using an annual panel survey that followed households who moved within a metropolitan area, Krizek used first difference models to estimate change in travel behavior as a function of change in the built environment resulting from residential relocation. He found that an increase in neighborhood accessibility was associated with a decrease in vehicle miles traveled (Krizek, 2003). This study design controls for endogenous characteristics (e.g., motivation for physical activity) that *remain constant over time* by subtracting out time invariant components of U in the model above. To illustrate, consider an expansion of Model 1, which distinguishes variables that change (time variant) from those that remain constant (time invariant) for individual i over time t:

$$PA_{it} = \beta_0 + \beta_1 E_{it} + \beta_2 S_i + \beta_3 T_{it} + \beta_4 P_i + \beta_5 U_i + \beta_6 V_{it} + \varepsilon_i + v_{it} \quad (2)$$

S and T are vectors of observed time invariant and variant sociodemographic variables, respectively. U and V are vectors of unobserved time invariant and variant variables, respectively. U might include genetic determinants of propensity to exercise, while V might include desire to actively commute, which may change over time. Associated error terms are ε_i (random, person-specific error) and v_{it} (random error for person i at time t). Recall that because U and V are unmeasured, and the variability in PA explained by U and V is captured in composite errors $\varepsilon_i^* = \beta_5 + \varepsilon_i$ and $v_{it}^* = \beta_6 + v_{it}$, respectively. Model 2 at time 1 subtracted from Model 2 at time 2 yields the following model, which is estimated by first difference models:

$$(PA_{i2} - PA_{i1}) = \beta_1 (E_{i2} - E_{i1}) + \beta_3 (T_{i2} - T_{i1}) + (v_{i2}^* - v_{i1}^*) \quad (3)$$

Time invariant sociodemographics (S), measured preferences (P), and person-specific error (ε_i^*) subtract out of Model 3. In particular, ε_i^* captures unmeasured time invariant factors, which will no longer bias the estimates. However, v_{it}^* , which may capture unmeasured time invariant factors, remains in the model. That is, first difference models (and similarly fixed effects models) are vulnerable to endogenous characteristics that change over time. A recent

study suggests that this remaining bias is not problematic: the relationship between sprawl and obesity was completely attenuated when estimated with first difference models (Eid, Overman, Puga, & Turner, 2008). However, given the complex etiology of obesity, we might expect stronger, more robust relationships between sprawl and more proximate measures such as physical activity. While longitudinal analysis in diverse samples and environmental contexts are greatly needed, there is limited availability of longitudinal data, particularly time-varying environment exposure data, necessitating innovative methods to address selection bias in cross-sectional studies.

Control for Attitudes and Preferences

Residential attitude and preference data were used in some of the first studies to address residential self-selection in cross-sectional analyses. As described in greater detail by Mokhtarian and Cao (Mokhtarian & Cao, 2008), preference data have been used as control variables in multivariate models (Handy, Cao, & Mokhtarian, 2005), or to create variables reflecting dissonance between residential preferences and objective neighborhood characteristics (Frank, Saelens, Powell, & Chapman, 2007; Schwanen & Mokhtarian, 2005a, 2005b).

While these methods are innovative, they require the assumption that preference measures capture true preferences (i.e., not influenced by current environment or transportation behaviors). As we have described above, such assumptions may not hold. In studies of built environment determinants of physical activity, preferences and attitudes pose problems beyond those described for residential choice. Many economists view an individual's attitudes and preferences as being determined by many of the same factors that determine physical activity and residential selection. Additionally, reporting errors associated with self-reported preferences and behaviors are probably strongly correlated: those who value public transit might be more likely to over-state both their preference for and use of public transit. That is, unobserved factors (U and V) affect both preferences and physical activity-related behaviors.

For these reasons, self-reported residential preferences are likely endogenous and thus lead to problematic interpretation when used as control variables in the prediction of self-reported physical activity. There is a large literature on the topic of determinants of preference structure which is outside the purview of this review. Ultimately, it is possible that controlling for residential preferences may not only fail to correct for residential self-selection bias but may introduce additional bias due to correlation of errors in reported preferences and behaviors.

Control for Observed Predictors of Residential Selection

An alternative approach is to control for predictors of observed residential selection using methods such as propensity scores (D'Agostino, 1998; Rosenbaum & Rubin, 1983), which attempt to control for non-random selection into a treatment (or exposure) group in experimental or observational studies, more recently in the built environment literature. For example, using cross-sectional data, Boer et al. showed that cross-sectional associations between walkability measures and walking were attenuated after propensity score adjustment, in some cases near or past the null (Boer, Zheng, Overton, Ridgeway, & Cohen, 2007).

Propensity score methods model "treatment", defined in this context as living in a neighborhood with activity-supportive characteristics, as a function of measured covariates. These strategies model the probability of living in a particular environment, given individual characteristics. Resulting probabilities (propensity scores) are subsequently used as

adjustment or matching variables in models predicting physical activity from environment characteristics. Propensity score methods were developed for binary treatments but can be expanded to multiple-level treatments. However, these methods are not always easily implemented. Indeed, Boer et al were forced to conduct a series of analyses comparing adjacent levels of the built environment measures. Alternatively, predicted probabilities can be incorporated into weighting variables (inverse-to-probability-of-treatment weighting), which offer the advantage of accommodating multi-level or continuously scaled “treatments” (Robins, Hernan, & Brumback, 2000).

There are several advantages of propensity score methods over traditional covariate adjustment. The balance of covariates can be explicitly verified, and larger sets of covariates can be included in analysis. In the case of matching and weighting methods, selection bias induced by conditioning on common effects of the outcome and exposure (e.g. residential movement) can be avoided (Hernan, Hernandez-Diaz, & Robins, 2004). However, propensity score methods only control for observed characteristics and assume adequate measurement of included variables. Therefore, they can control for residential selection bias only to the extent that covariates included in the treatment models capture determinants of selection into activity-supportive neighborhoods (Joffe & Rosenbaum, 1999; Oakes & Church, 2007). Unobserved characteristics correlated with the environment and physical activity will bias the obtained estimates. Thus, propensity score methods may not fully address residential self-selection.

Instrumental Variables

Structural equation modeling accommodates endogenous variables by explicitly modeling error common to multiple equations of interest (e.g., determinants of residential selection and physical activity). One example is instrumental variables analysis, a traditional econometric approach to controlling for endogeneity. An instrument is a variable that (i) has a causal effect on the exposure, (ii) affects the outcome only through the exposure, and (iii) does not share common causes with the outcome (unobserved characteristics correlated with the instrument and the outcome) (Hernan & Robins, 2006). While instrumental variables can be powerful in controlling for endogeneity due to unobserved characteristics, their effectiveness depends on the validity of the instrument. Violation of criteria ii or iii will introduce bias to the association between the exposure and outcome, which will be amplified if the instrument is only weakly associated with the exposure. While the Sargon test assesses the validity of the instrument, others note that criteria ii and iii are not empirically verifiable (Hernan & Robins, 2006; Martens, Pestman, de Boer, Belitser, & Klungel, 2006).

Although some may use attitudes and preferences as instrumental variables (Mokhtarian & Cao, 2008), many economists would argue that these measures are inappropriate instruments because they are also self-determined and endogenous. In order to meet criterion ii and iii above, one study used residential preferences (e.g. importance of having a backyard) as instruments, explicitly excluding travel-related preferences (Khattak & Rodriguez, 2005). While promising, such preferences may fail criterion iii, since residential preferences and travel behavior likely share common causes, such as household structure.

Using a different strategy, others use non-transport environmental characteristics such as racial composition as instruments (Boarnet & Sarmiento, 1998). However, neighborhood-level race/ethnicity and socioeconomic indicators are strong predictors of various health measures (Riva, Gauvin, & Barnett, 2007), independent of built environment characteristics (Rundle et al., 2008), thus failing criterion ii. It is also unlikely that such variables are exogenous: selection of neighborhoods of varying racial compositions are driven by complex factors, primarily individual-level race/ethnicity (Sampson & Sharkey, 2008), which is related to a wide range of behaviors (Kimbrow, Bzostek, Goldman, & Rodriguez,

2008). While Boarnet and Sarmiento's study focused on automobile trips rather than physical activity health-related measures, the instruments may be invalid for physical activity outcomes to the extent that automobile trips are related to active transportation modes. Indeed, empirical testing showed that the instruments were in some cases invalid, depending on the specific built environmental feature examined.

In sum, the instrumental variables approach is promising, but introduces the challenge of identifying a valid instrument. Even if a potential instrument is not *explicitly* related to travel behavior (or physical activity), it may not be valid due to complex inter-relationships between various environmental characteristics and preferences, behaviors, and sociodemographics.

Full information maximum likelihood (FIML)

Another form of structural equations modeling is FIML methods such as factor and path analysis (Kline, 2005), which can simultaneously test multiple pathways. For example, cross-sectional studies of the relationships among residential preferences, neighborhood characteristics, and behavior show that attitudes and preferences are the strongest correlates of behavior and neighborhood environment (Bagley & Mokhtarian, 2002; Cao, Mokhtarian, & Handy, 2007). Given the limitations of self-reported preferences, these findings are unsurprising. However, as discussed in the next section, FIML will accommodate longitudinal data and, when theory-driven (Martens, 2005), is a powerful tool for testing dynamic pathways among individual characteristics, mobility, and neighborhood environments at multiple time points.

Summary

Each of the above described methods has made important contributions to understanding residential self-selection bias. Self-reported preference data may capture otherwise unmeasurable preferences yet may exacerbate existing biases or introduce new ones, propensity score methods are empirically verifiable yet only capture observed predictors of residential selection, and instrumental variables can control for unobserved predictors but identifying a valid instrument has proven to be difficult, particularly with cross-sectional data.

Our Approach to Control for Residential Selectivity

We will use data from the Coronary Artery Risk Development in Young Adults Study (CARDIA), a longitudinal study of the antecedents and risk factors for cardiovascular disease in an ethnicity-, age-, and sex-balanced cohort of 5,115 black and white young adults aged 18–30 years at baseline (1985). Using GIS, we linked time-varying respondent residential addresses from four CARDIA study years (1985, 1992, 1995, and 2001) with contemporaneous data on environmental factors derived from a series of federal and commercial data bases. Recognizing the limitations of existing methods for adjusting for residential selection, we will use a behavioral model that focuses on the sequencing of career and residential choice as a major component, given the changes central to this lifecycle stage.

In our models, physical activity will be treated as an endogenous factor in this decision-making process (see review (Bartik & Smith, 1987)). We acknowledge the possible correlation between built environment variables and the error term that may arise through endogenous sorting of households across communities. We assume that individuals choose their place of residence by comparing the indirect utility that they would receive from alternative locations; the indirect utility will be partly a function of the neighborhood environment “prices,” such as recreational facilities, inter-connected street networks, and

cost of living. We will develop a series of equations for location choice, choice variables relevant to the young- to middle-aged adult such as selection of career and marriage, and physical activity. We assume that the error terms across these equations are correlated with each other. Thus, standard estimation methods will yield biased and inconsistent parameter estimates.

Our solution to the estimation problem is joint estimation of the entire system by FIML similar to the econometric method developed by Bhat and Guo (Bhat & Guo, 2007) in their study of the relationship household residential choice and auto ownership levels. Distinguishing features of our model system include its incorporation of longitudinal data and simultaneous estimation of physical activity, diet, and downstream health outcomes including obesity and clinical cardiovascular disease risk factor measures, whereas Bhat and Guo focused on the cross-sectional relationship between land use and one transportation measure. Additionally, FIML methods typically require assumptions about the distribution of unobservables in the equations with multivariate normality being a common assumption. Bhat and Guo's method accommodates multinomial ordered response data, and we plan to use a semi-parametric method that provides more flexibility by not requiring the assumption of a specific distribution. The discrete factor method is a variant of the Heckman and Singer method (Heckman & Singer, 1984), shown to work well in models such as these (Mroz, 1999; Mroz & Guilkey, 1995). In particular, we will approximate the distribution of the unobserved variables associated with the respondent's original residential location as well as time varying unobserved individual level characteristics; we will specify a discrete distribution where the parameters of the discrete distribution will be estimated with the coefficients of interest. The method has been used with similar modeling problems (Blau, 1994; Guilkey & Riphahn, 1998).

Of course, the use of FIML methods requires statistical identification of the choice of residential location (and career and marriage choice) and physical activity equations (e.g., family background variables such as state level employment related variables and other similar state, county, and Standard Metropolitan Statistical Area measures). Exogenous variables unique to the career choice equation may be state level employment related variables. Exogenous variables unique to the residential choice equations may be other characteristics of the locations (e.g., access to larger cities, cultural activities, cost of living), which may not directly affect physical activity. The importance of obtaining more than just "technical" identification has been well documented in recent literature (Bound, Jaeger, & Baker, 1995). It is important that the identifying variables have significant impact on the dependent variables and the explanatory power of the set of identifying variables is particularly crucial. Our structural model will employ tests by Bound et al. to gauge strength of identification. Since the model is over-identified, we will also be able to test some identifying restrictions (e.g., (Davidson & MacKinnon, 1993)).

Conclusion

The rapidly growing field of built environmental determinants of health behaviors and outcomes is at a crucial juncture. Without better understanding of residential self-selection bias and improvement of methods to overcome such bias, research on the effects of the built environment on health will be seriously limited. Our CARDIA Obesity and Environment database is the first large scale GIS to link community- and individual-level data in both space and time in a large ethnically diverse sample followed over 20 years. Using these longitudinal data, we will implement innovative structural modeling that adjusts for residential self-selection bias, a critical gap in this burgeoning area of research. Our methods and substantive findings will address major methodological limitations and ultimately

advance the scientific knowledge base regarding potential environmental interventions to promote physical activity.

References

- Bagley MN, Mokhtarian PL. The impact of residential neighborhood type on travel behavior: a structural equations modeling approach. *Annals of Regional Science*. 2002; 36:279–297.
- Ball K, Timperio AF, Crawford DA. Understanding environmental influences on nutrition and physical activity behaviors: where should we look and what should we count? *The international Journal of Behavioral Nutrition and Physical Activity*. 2006; 3:33. [PubMed: 16999874]
- Bartik, TJ.; Smith, VK. Urban amenities and public policy. In: Mills, ES., editor. *Urban Economics*. Vol. II. Amsterdam: North Holland; 1987. p. 1207-1249.
- Bhat CR, Guo JY. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B-Methodological*. 2007; 41(5):506–526.
- Blau DM. Labor-force dynamics of older men. *Econometrica*. 1994; 62:117–156. [PubMed: 12290260]
- Boarnet MG, Sarmiento S. Can land-use policy really affect travel behaviour? A study of the link between non-work travel and land-use characteristics. *Urban Studies*. 1998; 35(7):1155–1169.
- Boer R, Zheng Y, Overton A, Ridgeway GK, Cohen DA. Neighborhood design and walking trips in ten U.S. metropolitan areas. *American Journal of Preventive Medicine*. 2007; 32(4):298–304. [PubMed: 17383560]
- Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*. 1995; 90(430):443–450.
- Cao X, Mokhtarian PL, Handy SL. Do changes in neighborhood characteristics lead to changes in travel behavior ? A structural equations modeling approach. 2007
- Cummins S, Petticrew M, Higgins C, Findlay A, Sparks L. Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *Journal of Epidemiology and Community Health*. 2005; 59(12):1035–1040. [PubMed: 16286490]
- D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*. 1998; 17(19):2265–2281. [PubMed: 9802183]
- Davidson, R.; MacKinnon, JG. *Estimation and inference in econometrics*. New York: Oxford University Press; 1993.
- Diez-Roux AV. Multilevel analysis in public health research. *Annual Review of Public Health*. 2000; 21:171–192.
- Diez-Roux AV, Nieto FJ, Muntaner C, Tyroler HA, Comstock GW, Shahar E, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. *American Journal of Epidemiology*. 1997; 146(1):48–63. [PubMed: 9215223]
- Diez Roux AV. Investigating neighborhood and area effects on health. *American Journal of Public Health*. 2001; 91(11):1783–1789. [PubMed: 11684601]
- Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Social Science and Medicine*. 2004; 58(10):1953–1960. [PubMed: 15020010]
- Diez Roux AV, Evenson KR, McGinn AP, Brown DG, Moore L, Brines S, et al. Availability of recreational resources and physical activity in adults. *American Journal of Public Health*. 2007; 97(3):493–499. [PubMed: 17267710]
- Duncan, GJ.; Raudenbush, SW. Neighborhoods and adolescent development: how can we determine the links?. In: Booth, A.; Crouter, N., editors. *Does It Take a Village? Community Effects on Children, Adolescents, and Families*. State College, PA: Pennsylvania State University Press; 2001. p. 105-136.
- Eid J, Overman HG, Puga D, Turner MA. Fat city: Questioning the relationship between urban sprawl and obesity. *Journal of Urban Economics*. 2008; 63(2):385–404.
- Epple D. Modeling population stratification across locations: An overview. 2003

- Ewing R, Brownson RC, Berrigan D. Relationship between urban sprawl and weight of United States youth. *American Journal of Preventive Medicine*. 2006; 31(6):464–474. [PubMed: 17169708]
- Frank LD, Saelens BE, Powell KE, Chapman JE. Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Social Science and Medicine*. 2007
- Frank LD, Schmid TL, Sallis JF, Chapman J, Saelens BE. Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. *American Journal of Preventive Medicine*. 2005; 28(2 Suppl 2):117–125. [PubMed: 15694519]
- Giles-Corti B, Broomhall MH, Knuiaman M, Collins C, Douglas K, Ng K, et al. Increasing walking: how important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine*. 2005; 28(2 Suppl 2):169–176. [PubMed: 15694525]
- Giles-Corti B, Donovan RJ. The relative influence of individual, social and physical environment determinants of physical activity. *Social Science and Medicine*. 2002; 54(12):1793–1812. [PubMed: 12113436]
- Gordon-Larsen P, Adair LS, Popkin BM. Ethnic differences in physical activity and inactivity patterns and overweight status. *Obesity Research*. 2002; 10(3):141–149. [PubMed: 11886936]
- Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*. 2006; 117(2):417–424. [PubMed: 16452361]
- Gorin AA, Wing RR, Fava JL, Jakicic JM, Jeffery R, West DS, et al. Weight loss treatment influences untreated spouses and the home environment: evidence of a ripple effect. *International Journal of Obesity (London)*. 2008; 32(11):1678–1684.
- Guilkey DK, Riphahn RT. The determinants of child mortality in the Philippines: estimation of a structural model. *Journal of Development Economics*. 1998; 56(2):281–305. [PubMed: 12348678]
- Handy S, Cao X, Mokhtarian PL. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D*. 2005; 10:427–444.
- Handy S, Cao X, Mokhtarian PL. Self-selection in the relationship between the built environment and walking. *Journal of the American Planning Association*. 2006; 72(1):55–74.
- Handy S, Sallis JF, Weber D, Maibach E, Hollander M. Is Support for Traditionally Designed Communities Growing? Evidence From Two National Surveys. *Journal of the American Planning Association*. 2008; 74(2):209–221.
- Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine and Science in Sports Exercise*. 2007; 39(8):1423–1434.
- Heckman J, Singer B. A method for minimizing the impact of distributional assumptions in econometric-models for duration data. *Econometrica*. 1984; 52(2):271–320.
- Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004; 15(5):615–625. [PubMed: 15308962]
- Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006; 17(4):360–372. [PubMed: 16755261]
- Ioannides YA, Zabel JE. Interactions, neighborhood selection and housing demand. *Journal of Urban Economics*. 2008; 63(1):229–252.
- Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. *American Journal of Epidemiology*. 1999; 150(4):327–333. [PubMed: 10453808]
- Khattak AJ, Rodriguez D. Travel behavior in neo-traditional neighborhood developments: A case study in USA. *Transportation Research Part a-Policy and Practice*. 2005; 39(6):481–500.
- Kimbro RT, Bzostek S, Goldman N, Rodriguez G. Race, ethnicity, and the education gradient in health. *Health Affairs (Millwood)*. 2008; 27(2):361–372.
- Kline, R. Principles and practice of structural equation modeling. 2nd ed.. New York: The Guilford Press; 2005.
- Krizek KJ. Residential relocation and changes in urban travel. *Journal of the American Planning Association*. 2003; 69(3):265–281.

- Krizek KJ, Johnson PJ. Proximity to trails and retail: effects on urban cycling and walking. *Journal of the American Planning Association*. 2006; 72(1):33–42.
- Lee C, Moudon AV. Physical activity and environment research in the health field: Implications for urban and transportation planning practice and research. *Journal of Planning Literature*. 2004; 19(2):147–181.
- Librett JJ, Yore MM, Schmid TL, Kohl HW 3rd. Are self-reported physical activity levels associated with perceived desirability of activity-friendly communities? *Health Place*. 2007; 13(3):767–773. [PubMed: 16935021]
- Lund H. Reasons for living in a transit-oriented development, and associated transit use. *Journal of the American Planning Association*. 2006; 72(3):357–366.
- Marcus BH, Owen N. Motivational readiness, self-efficacy, and decision making for exercise. *Journal of Applied Social Psychology*. 1992; 22(1):3–16.
- Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology*. 2006; 17(3):260–267. [PubMed: 16617274]
- Martens MP. The use of structural equation modeling in counseling psychology research. *Counseling Psychologist*. 2005; 33(3):269–298.
- Mokhtarian PL, Cao X. Examining the impacts of residential selection on travel behavior: a focus on methodologies. *Transportation Research Part B*. 2008; 42:204–228.
- Mroz TA. Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome. *Journal of Econometrics*. 1999; 92(2):233–274. [PubMed: 12322578]
- Mroz, TA.; Guilkey, D. Discrete factor approximations in simultaneous equation models with both continuous and discrete endogenous variables. University of North Carolina at Chapel Hill; 1995.
- Nechyba TJ, Strauss RP. Community choice and local public services: A discrete choice approach. *Regional Science and Urban Economics*. 1998; 28(1):51–73.
- Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social Science and Medicine*. 2004; 58(10):1929–1952. [PubMed: 15020009]
- Oakes JM, Church TR. Invited commentary: advancing propensity score methods in epidemiology. *American Journal of Epidemiology*. 2007; 165(10):1119–1121. discussion 1122–1113. [PubMed: 17395596]
- Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999–2004. *Journal of the American Medical Association*. 2006; 295(13):1549–1555. [PubMed: 16595758]
- Papas MA, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC. The built environment and obesity. *Epidemiological Reviews*. 2007; 29:129–143.
- Popkin BM, Duffey K, Gordon-Larsen P. Environmental influences on food choice, physical activity and energy balance. *Physiology & Behavior*. 2005; 86(5):603–613. [PubMed: 16246381]
- Popkin BM, Gordon-Larsen P. The nutrition transition: worldwide obesity dynamics and their determinants. *International Journal of Obesity & Related Metabolic Disorders*. 2004; 28 Suppl 3:S2–S9. [PubMed: 15543214]
- Pratt CA, Lemon SC, Fernandez ID, Goetzel R, Beresford SA, French SA, et al. Design characteristics of worksite environmental interventions for obesity prevention. *Obesity (Silver Spring)*. 2007; 15(9):2171–2180. [PubMed: 17890484]
- Riva M, Gauvin L, Barnett TA. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology & Community Health*. 2007; 61(10):853–861. [PubMed: 17873220]
- Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11(5):550–560. [PubMed: 10955408]
- Rodriguez DA, Khattak AJ, Evenson KR. Can New Urbanism encourage physical activity? *Journal of the American Planning Association*. 2006; 72(1):43–54.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70(1):41–55.

- Rundle A, Field S, Park Y, Freeman L, Weiss CC, Neckerman K. Personal and neighborhood socioeconomic status and indices of neighborhood walk-ability predict body mass index in New York City. *Social Science and Medicine*. 2008; 67(12):1951–1958. [PubMed: 18954927]
- Saelens BE, Handy SL. Built environment correlates of walking: a review. *Medicine and Science in Sports Exercise*. 2008; 40(7 Suppl):S550–S556.
- Saelens BE, Sallis JF, Frank LD. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavior Medicine*. 2003; 25(2): 80–91.
- Sallis JF, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. *American Journal of Preventive Medicine*. 1998; 15(4):379–397. [PubMed: 9838979]
- Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Annual Review of Public Health*. 2006; 27:297–322.
- Sallis JF, Hovell MF, Hofstetter CR, Elder JP, Hackley M, Caspersen CJ, et al. Distance between homes and exercise facilities related to frequency of exercise among San Diego residents. *Public Health Reports*. 1990; 105(2):179–185. [PubMed: 2108465]
- Sallis, JF.; Owen, N. Ecologic models of health behavior. In: Glanz, K.; Rimer, BK.; Lewis, FM., editors. *Health Behavior and Health Education: Theory, Research, and Practice*. 3rd ed.. San Francisco: Jossey-Bass; 2002. p. 462-484.
- Sampson RJ, Sharkey P. Neighborhood selection and the social reproduction of concentrated racial inequality. *Demography*. 2008; 45(1):1–29. [PubMed: 18390289]
- Schwanen T, Mokhtarian PL. What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography*. 2005a; 13:83–99.
- Schwanen T, Mokhtarian PL. What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled. *Transportation Research Part D*. 2005b; 10:127–151.
- Sharma M. School-based interventions for childhood and adolescent obesity. *Obesity Reviews*. 2006; 7(3):261–269. [PubMed: 16866974]
- Sloane DC. From congestion to sprawl: planning and health in historical context. *Journal of the American Planning Association*. 2006; 72(1):10–18.
- Song Y, Knaap GJ. New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics*. 2003; 54(2):218–238.
- Song Y, Knaap GJ. New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics*. 2003; 54:218–238.
- Wen M, Browning CR, Cagney KA. Neighbourhood deprivation, social capital and regular exercise during adulthood: A multilevel study in Chicago. *Urban Studies*. 2007; 44(13):2651–2671.
- Wendel-Vos W, Droomers M, Kremers S, Brug J, van Lenthe F. Potential environmental determinants of physical activity in adults: a systematic review. *Obesity Reviews*. 2007; 8(5):425–440. [PubMed: 17716300]
- Zohoori N, Savitz DA. Econometric approaches to epidemiologic data: relating endogeneity and unobserved heterogeneity to confounding. *Annals of Epidemiology*. 1997; 7(4):251–257. [PubMed: 9177107]

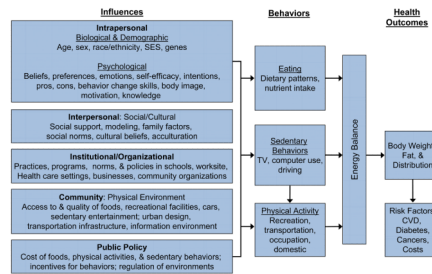


Figure 1. An Ecological Model of Diet, Physical Activity, and Obesity
 Depicts one major pathway linking behaviors to health outcomes; for clarity, direct linkages of behaviors with body size and clinical risk factors and disease are omitted. Adapted from model developed at the NHLBI Workshop: Predictors of Obesity, Weight Gain, Diet, and Physical Activity; August 2004, Bethesda MD