*Environment and Planning B: Planning and Design* 2013, volume 40, pages 489 – 506

doi:10.1068/b38007

# Neighborhood contexts, health, and behavior: understanding the role of scale and residential sorting

**Seth E Spielman**
Department of Geography, University of Colorado at Boulder, 100 Guggenheim Hall/260 UCB, Boulder, CO 80309, USA; e-mail: Seth.Spielman@colorado.edu
**Eun-Hye Yoo**
Department of Geography, University of Buffalo, State University of New York, 105 Wilkeson Quad, Buffalo, NY 14261, USA; e-mail: eunhye@buffalo.edu
**Crystal Linkletter**
Department of Biostatistics, Brown University, Box G-A, Providence, RI 02912, USA; e-mail: Crystal_Linkletter@brown.edu
Received 14 January 2011; in revised form 18 April 2012

**Abstract. Recent reviews in sociology, public health, and urban planning suggest small-scale geographic variations in urban environments are associated with an individual's health, behavior, and well-being. However, estimation of these 'neighborhood effects' are complicated. One complicating factor is residential sorting: individual characteristics such as race, age, and socioeconomic status are associated with behavior and health and these same individual-level factors are often geographically clustered within neighborhoods. This residential sorting leads to a correlation between individual and environmental determinants of behavior and health and poses problems for statistical inference. A second complicating factor is uncertainty about the geographic dimensions of a person's neighborhood. We explore these two potential confounders through a simulation experiment that generates synthetic cityscapes and synthetic behaviors for geolocated individuals. The simulation is used to develop a model of how residential sorting, urban structure, and the geographic definition of an individual's neighborhood affect our understanding of the association between the urban environment and behavior. We find that residential sorting does not systematically affect the magnitude of neighborhood effect estimates, however, the misrepresentation of the geographic dimensions of an individual's neighborhood leads to systematic bias in estimates of neighborhood effects. Unlike previous research on the modifiable areal unit problem we find a systematic relationship between the definition of geographic units of analysis and the magnitude of regression coefficients.**

**Keywords: neighborhood effects, uncertainty, context, residential sorting, geographic scale**

## 1 Introduction

Geographic inequalities are one of the most striking aspects of urban life: health, well-being, and the prevalence of important behaviors like crime and automobile use vary spatially. What accounts for these spatial patterns? Are they simply the result of forces that sort people into neighborhoods on the basis of socioeconomic characteristics? Are they created and/or magnified by neighborhood-based influences? That is, are geographic variations in behavior and well-being due to spatial variation in the composition of populations or are they due to the influence of neighborhood-scale variations in environmental context? Neighborhood-based contextual influences form a significant line of investigation in the social sciences; however, quantifying them is a challenging task. Residential sorting and uncertainty about the definition of neighborhoods may confound our ability to measure neighborhood-based influences on health and behavior (Diez-Roux, 2004; Oakes, 2004).

In this paper we are interested in understanding the influence of the environment on behavior. In the urban context this influence is often called the 'neighborhood effect'. Common questions about neighborhood effects include: Does the food available in a person's neighborhood influence their eating habits? Does the land use near a person's home influence travel behavior? Residential sorting makes it very difficult to answer these questions because observed geographic patterns could be due to prior preferences that shape a person's choice of residence. People who prefer to walk may choose to live in a place where the land-use pattern is conducive to walking; this is an example of residential sorting. On the other hand, geographic patterns could be rooted in the influence of the environment; places that are conducive to walking could induce it: this is an example of the neighborhood effect. Sociospatial patterns are a fact of urban life, but observing these patterns does not indicate which of these mechanisms (sorting or neighborhood effects) is operating. It is likely both the neighborhood effect and sorting simultaneously contribute to geographic patterns in behavior and health. Estimating the relative contribution of these mechanisms to observed patterns is the subject of this paper.

Residential sorting is not the only problem facing the estimation of neighborhood effects; determining the geographic scale of environmental influences on behavior is a persistent 'conundrum' for researchers (Gauvin et al, 2007). The nature of neighborhood is currently the subject of substantial debate (Chaix et al, 2009; Flowerdew et al, 2008; Kwan, 2009; Lebel et al, 2007; Sastry et al, 2002). This uncertainty also complicates efforts to estimate neighborhood effects. Spielman and Yoo (2009) show that the uncertainty about the scale of neighborhoods can systematically bias estimates of neighborhood effects.

Both residential sorting and the influence of the environment on behavior are impractical (or impossible) to observe directly and therefore must be estimated indirectly using statistical models. Because these forces are difficult to measure directly one cannot easily assess the fidelity of statistical models to the 'truth'. In this paper we use a simulation experiment to create cities with known degrees of residential sorting and known neighborhood effects. We use these synthetic cities to assess our ability to accurately estimate the neighborhood effect in the presences of varying degrees of residential sorting and uncertainty about the definition of neighborhoods.

How should we interpret studies of neighborhood effects that occur in the presence of residential sorting but do not control it? When uncertainty about the definition of neighborhoods and residential sorting co-occur, do they interact to exacerbate bias? These are the questions we aim to address in this paper.

### 1.1 Background

In the early 1990s seminal papers in several disciplines kicked off a long-running debate about neighborhood effects. In public health, Haan et al (1987), using data from the Alameda County Study, reported that even after controlling for individual characteristics there is an association between living in a poor neighborhood and lower health status. Brooks-Gunn et al (1993) found neighborhood effects on children's development. Research based on the Project on Human Development in Chicago neighborhoods found that the characteristics of neighborhoods affect crime and collective efficacy (Sampson et al, 1997; 1999). In urban planning both Frank and Pivo (1995) and Cervero and Kockleman (1997) found that character of the built environment around one's residence is associated with travel behavior.

These early papers spawned a large literature on the 'neighborhood' or 'contextual' effects hypothesis. The hypothesis is that, all other things being equal, the residents of some neighborhoods will behave differently because of the characteristics of their neighborhoods. This idea remains controversial in spite of the now substantial literature supporting its basic claims (Chaix, 2009; Clifton et al, 2008; Dietz, 2002; Papas et al, 2007; Sampson et al, 2002).

The controversy arises because correlates of behavior like demographic characteristics, attitudes, and preferences also shape where a person lives (Ewing and Cervero, 2010; Harding, 2003; Oakes, 2004). Since people are not randomly sorted into neighborhoods, patterns emerge. However, these initial geographic patterns may be amplified by environmental factors. People living in similar locations face similar environmental influences and these too may shape behavior.

The 'neighborhood effect' is the portion of a geographic pattern that is due to environmental influences. Understanding the relative contribution of demographic and environmental factors to geographic patterns in behavior (or health) is important because neighborhood-scale environmental factors are often a direct or indirect result of public policies such as policing, land-use regulation, and education. However, research that attempts to isolate the neighborhood effect by controlling for residential sorting has produced mixed results. Several papers found that controlling for residential sorting erases any contextual effects (Eid et al, 2008; Forsyth et al, 2007), while others found that the relationship between the environment and behavior persists after control (Handy et al, 2005; Harding, 2003).

Residential sorting is not the only hurdle facing neighborhood research. Galster (2008) outlines a series of challenges for research on neighborhood effects. First among them is the question of scale. Over what geographic area should the neighborhood (or urban form) be measured? In the study of neighborhood effects, scale is a fundamental and difficult question. Neighborhood is a 'genuinely amorphous' concept (Sastry et al, 2002) and because neighborhoods are hard to define their boundaries are difficult to identify. Some have proposed that measures of neighborhoods should account for spatial and temporal activity patterns (Weber and Kwan, 2003); others advocate the use of street network buffers around a residence (Frank et al, 2004); some call for fuzzy boundaries (Chaix et al, 2009); while others argue that groups of pedestrian-oriented streets are the relevant ecological setting for studying behavior (Grannis, 1998). Determining the appropriate unit of analysis for the study of the relationship between the urban environment and behavior is challenging (Chaix, 2009; Chaix et al, 2009), Spielman and Yoo (2009) showed that it is difficult to understand the environment–behavior relationship unless these measurement questions are sorted out. Given the uncertainty around neighborhood boundaries it is perhaps reasonable to estimate neighborhood effects using multiple geographic units of analysis (as suggested by Flowerdew et al, 2008).

Chaix et al (2006) used multiple geographic units of analysis to study neighborhood effects on mental health in Sweden. They find that the observed neighborhood effects were inversely related to the size of the neighborhood—larger areas resulted in smaller estimates of neighborhood effects. Zhang (2005) studied measures of urban form in Boston at multiple scales and decided the scale that provided the best model fit—1/2 mile—was the best measure of a neighborhood. These multiscale approaches pose some inferential and ontological challenges. Defining the unit of analysis by choosing the model that provides the best fit or most desirable coefficients seems to have little scientific validity (Freedman, 1991). From an ontological perspective, if a person has a 'neighborhood' which exists as a geographically bounded territory only one of the geographic units of analysis is correct. On the other hand, if an individual's neighborhood is a continuous field, and the influence of the field decays with distance, results like Chaix et al's (2006) finding that the influence of neighborhoods declines as scale increases are not surprising.

As a general matter, geographic phenomena can be represented as either objects or fields. The choice of one form of representation over another carries certain ontological commitments (Couclelis, 1992). However, in research on neighborhoods these representational decisions are often constrained by data and made on practical grounds. There has been some experimental work on the definition of neighborhoods. Coulton et al (2001) through a sketch mapping exercise found that people place themselves at the center of a neighborhood, which corroborates Hunter's (1974) findings that people seldom place themselves on the edge of a neighborhood. Galster (1986) develops the concept of a 'negative externality space' as a way to experimentally observe an individual's neighborhood. These experiments suggest that individuals have an egocentric view of their neighborhood, but they provide no clear indication of the nature of neighborhood boundaries. Given the lack of observational evidence regarding neighborhood boundaries, it is not surprising that in statistical models of neighborhood effects it is common to represent neighborhoods as egocentric objects with 'sliding' boundaries (Guo and Bhat, 2007).

Instead of tackling ontological questions about the nature of neighborhood we accept common research practice as the 'truth'. We develop a simulation model that assumes an individual's neighborhood is a discrete territory centered on their home, as is common in a wide variety of disciplines (eg, Clifton et al, 2008; Crane and Boarnet, 2001; Ewing and Cervero, 2001; 2010; Frank et al, 2003; Galster, 2008; Guo and Bhat, 2007; Papas et al, 2007; Transportation Research Board and the Institute of Medicine of the National Academics, 2005). The resulting model, strictly speaking, is not a simulation of neighborhood effects; rather it is a simulation of research on neighborhood effects.

Simulation can illustrate consequences of simple assumptions (Axelrod, 2007). We build a simulation based on the assumptions implicit in the statistical models used to research neighborhood effects. Our goal is to assess the ability of these models to paint an accurate picture of the simplified and abstract reality they posit. Simulation has important advantages, but it also carries a cost. Simulation allows us to design behaviors with a known neighborhood effect and hence model bias. However, the results reported here are only as valid as our abstract system. While this caveat applies generally to statistical models, it is important to note.

### 1.2 Defining neighborhoods: some current debates

We are primarily concerned with how the definition of neighborhoods affects inference about neighborhood effects. There is such a broad range of approaches to the representation of 'neighborhoods' in the social-scientific literature that we need to narrow the concept somewhat. While a full review of the myriad approaches to the definition of neighborhoods is beyond the scope of this paper, the general trends and debates are worth noting. There is broad consensus that neighborhoods exist, but agreement stops there. The major debates in the definition of neighborhoods can be framed as a discussion between those who favor 'place-based' definitions and those who favor 'person-based' definitions (Chaix et al, 2009; Guo and Bhat, 2007; Kwan, 2009; Miller, 2007). Place-based definitions of neighborhoods emphasize the characteristics of places. In contrast, person-based definitions refer to the perception, location, and/or activities of individuals. Both place-based and person-based approaches define a neighborhood as geographic territory that internally differentiates a city; however, the territories are based on very different assumptions.

Place-based measures focus on a person's relative geographic location with respect to the boundaries of a neighborhood. A person lives in a particular place and the boundaries of that place are wholly independent of the individual in question. Location within a neighborhood leads to exposure to social and ecological factors that are geographically embedded (Diez-Roux, 1998; Macintyre et al, 2002). Neighborhood analysis often takes place in the presence of data constraints and as a result researchers often use administrative units such as census tracts to define neighborhood boundaries (Dietz, 2002). However, the use of census tracts and other administrative boundaries is not entirely arbitrary. The actions and interactions of individuals (and institutions) are often constrained by space. These interactions may be important determinants of behavior (Northridge et al, 2003). The Project on Human

Development in Chicago Neighborhoods (PHDCN) defines neighborhoods as clusters of census tracts (Earls and Buka, 1997). Using the PHDCN data Sampson et al (1997) view collective efficacy as a property of neighborhoods that has positive spatial externalities, noting that the willingness of people to intervene on behalf of others can have broad impacts, including the reduction of violence. Browning et al (2006) found that unequal distributions of community resources produced conditions that made residents of some areas more vulnerable than others to the 1995 Chicago Heat Wave. Place-based definitions of neighborhoods emphasize the extra-individual aspects of place such as the positive/negative externalities of communal institutions, cultural norms, and behaviors. Where a person lives generally affects their exposure to these forces; small differences in geographic location are only of significance if they situate a person into a new place: that is, a location with a different set of norms and institutions.

Person-based representations of neighborhoods are wholly dependent upon an individual. They divide space based upon an individual's perception or location in space; in this view, neighborhoods are individual specific. Person-based measures can be static or dynamic. Static measures define a single, stable neighborhood around an individual. Dynamic measures account for an individual's movement in space and time. Static person-based neighborhoods often consider a person's residential/work location as the center of their neighborhood (Chaix, 2009). Static person-based neighborhoods can be constructed at multiple scales: for example, Lee et al (2008) use a series of egocentric concentric rings to measure segregation at multiple scales.

Dynamic person-based measures consider an individual's movement through space and therefore time. As a result such measures are often called time geographic. In this approach individuals construct 'personal' neighborhoods through the conduct of their daily activities (Weber and Kwan, 2003). Hägerstrand (1982, page 324) urged geographers to, "rise up from the flat map, with its static patterns and think in terms of a world on the move, a world of incessant permutations." To this day, this kind of dynamic thinking remains a challenge and constitutes an important subdiscipline of geographic information science. Kwan (2000), Miller (1991; 2005), Laube and Imfeld (2002), and others have made important progress in the visualization, computation, and analysis of dynamic time-geographic patterns. The movement away from traditional place-based conceptualizations of neighborhoods toward person-based neighborhoods is, in part, rooted in what Miller (2007) calls the "place-based fallacy". Miller argues that advances in technology have changed the meaning of place such that where a person resides is a poor measure of a person's life-environment. Kwan (2009) and Chaix et al (2009) similarly stress the importance of individual activities in determining a person's context, arguing that neighborhoods should represent personal exposures to various environments.

The fact that there are multiple ways to define neighborhoods and that these different methods often disagree has significant implications for urban spatial analysis. Neighborhoods, as geographic units of analysis, are inherently modifiable: that is, their boundaries are contingent upon the conceptualization of neighborhoods as 'place' or 'person' based. The modifiability of neighborhoods raises some important statistical and epistemological questions for urban analysis. These different perspectives on 'neighborhoods' are difficult to reconcile. Here, we discuss the nature of this uncertainty but do not attempt to settle questions about the nature of neighborhoods. Ours is a static person-based approach. This approach privileges certain forms of analysis and discounts others. For example, hierarchical statistical models of neighborhood effects are rooted in a place-based approach. When a neighborhood is conceptualized as place based, those living within its boundaries share the same neighborhood, and each neighborhood is shared by multiple people. Hierarchical (or multilevel) models of neighborhood effects exploit this structure. The egocentric approach to neighborhoods is horizontal, not hierarchical. When each person has their own 'personal' neighborhood, hierarchical models do not make sense because there is no hierarchy, each neighborhood contains only one individual.

## 2 Methods

Simulation, like all formal models, requires a substantial dose of reductionism in the form of mathematical formalization. In this section we present our simplified conceptual framework for neighborhood effects. In the opening paragraphs of each part of the methods section we provide a discussion of the components of our model with minimal use of formal notation, then in the following paragraphs we describe our methods using a more formal mathematical language. This structure aims to make the paper both accessible and reproducible.

### 2.1 Overview and conceptual framework

Blalock (1984) and Eid et al (2008) provide a nice conceptual discussion of the regression models widely used to estimate neighborhood effects, here we adopt their basic framework in equation (1).

$$Y = \gamma \boldsymbol{x} + \beta \mathrm{A}_d + \varepsilon \ . \tag{1}$$

Similar to the Blalock–Eid framework, the model that we propose has two components: one describes the influence of individual characteristics ($\gamma \boldsymbol{x}$) and the other describes the influence of the environment ($\beta \mathrm{A}_d$) on the outcome of interest. The individual characteristics that might influence behavior are represented by $\boldsymbol{x}$. Measures of a person's neighborhood are represented with $\mathrm{A}_d$ where the subscript $d$ specifies the geographic dimensions of the neighborhood; larger values of $d$ mean that the person has a larger neighborhood. Finally, the behavior of interest is represented by the $Y$ on the left side of the equation. The coefficients $\beta$ and $\gamma$ are estimated statistically and measure the relative influence of the environment and individual characteristics on behavior. The neighborhood effect, $\beta$, measures the magnitude of a neighborhood's influence on $Y$. The purpose of this model is to estimate the relative contribution of individual characteristics ($\boldsymbol{x}$) and environmental characteristics ($\mathrm{A}_d$) to the outcome of interest ($Y$). Individual characteristics $\boldsymbol{x}$ may include contributors to $Y$, such as age and gender, as well as factors that relate to a person's selection of one neighborhood over another (and hence $\mathrm{A}_d$). The relevant measures of a neighborhood could be any number of variables, including the density of grocery stores, land-use mix, street connectivity, disorder, or collective efficacy to name a few. The important point is that $\mathrm{A}_d$ represents factors that are external to an individual but (potentially) influence behavior.

Considering the significance of neighborhood effects in a variety of disciplines, we leave the outcome $Y$ generic. Commonly $Y$ will be a measure of health or behavior, such as vehicle miles traveled (VMT) or obesity (for example, see Frank et al, 2004). We use an abstract definition of $Y$ because our interest is the statistical estimation of the relative importance of neighborhood ($\beta$) and individual characteristics ($\gamma$) on $Y$ in the presence of residential sorting and uncertainty about the geographic dimensions of neighborhoods. In this paper we use a simulation experiment to assess the accuracy of $\hat{\beta}$, an estimate of the influence of the environment on behavior. We do this assessment by using a simulation model to generate residential sorting and uncertainty about the definition of neighborhoods.

#### 2.1.1 *Simulation*

The simulation proceeds in three steps, and this section is organized accordingly. The first step is to create a series of synthetic cities. For these synthetic cities we specify the level of residential sorting and the structure of neighborhoods using a geostatistical cosimulation algorithm. In the second step we generate a sample from the synthetic population of the city and for the sample we synthesize behavior. This synthetic behavior is conceptualized as a combination of the individual's characteristics and the environment around the individual. We specify how much the individual characteristics ($\boldsymbol{x}$) and the environmental characteristics ($A_d$) contribute to the behavior. In the third step we mimic research practice and estimate equation (1). In practice a researcher would be able to observe behavior, a person's environment (at some scale), and a person's characteristics, but would not be able to directly observe their relative contribution to behavior. Typically, a researcher would use these observations in combination with a statistical model to estimate the relative contribution of individual and environmental characteristics (ie, $\beta$ and $\gamma$) to a behavior. In practice, the accuracy of these estimates would be difficult to assess. Since we generate behavior synthetically by specifying the role of individual and environmental characteristics we can directly assess the accuracy of statistical estimates of the neighborhood effect ($\beta$) and the influence of individual characteristics ($\gamma$) provided by equation (1).

### 2.2 Simulating cities

We simulate cities using a geostatistical cosimulation algorithm called Markov Model 1 (MM1). We encourage readers to see Deutch and Journel (1992) for a full technical discussion of the algorithm; here we provide an overview.

We model cities as pairs of synthetically generated raster maps; each is a 100 × 100 grid of pixels. One member of the pair represents the environment and the other represents the population. We simulate these maps using the software SGeMS (Remy et al, 2008). The pair of maps that together constitute each 'city' are a single realization of two stochastic processes (Chiles and Delfiner, 1999). Some aspects of this random process are controlled. For example, we specify the covariance function that determines the spatial structure of the urban environment. By modifying this covariance function we are able to produce cities where the environment is a dense patchwork of heterogeneous regions and cities with extensive areas of similarity where the environment varies little. The parameter $\theta$ described in table 1 controls the geographic scale of the variability in the urban environment. Technically, $\theta$ corresponds to the range of the spatial autocorrelation in the map of the environment. We generate cities where $\theta$ takes values between 5 and 30 raster cells (hereafter, distance units).

**Table 1.** Simulation parameters.

| Parameter name | Notation | Description |
|---|---|---|
| Residential sorting (or selection) | $\rho$ | This parameter controls the degree of sorting in the city. When high ($\rho \approx 1$) there is high correlation between individual and environmental characteristics: eg, rich people live only in rich neighborhoods and the poor only in poor neighborhoods. When low ($\rho \approx 0$) where a person lives is independent of their characteristics. |
| Urban structure | $\theta$ | This parameter controls the spatial structure of a city. When high ($\theta = 30$) the synthetic urban environments vary little over short distances. This models a city with large homogenous neighborhoods as would be found in the autodependent suburbs of North America. When $\theta$ is small ($\theta = 5$) the synthetic cities are a dense patchwork of small neighborhoods, as one often find in older cities (like New York). |
| Neighborhood (scale) | $d$ | This parameter controls the diameter of a person's hypothesized neighborhood. Large values correspond to large neighborhoods. By varying $d$ we control the area measured by $A_d$ in equation (1). Note the difference between $A_d^*$ and $A_d$ where the * denotes the 'true' neighborhood used to generate behavior in equation (3) and $d$ is the neighborhood used to estimate equation (1). |

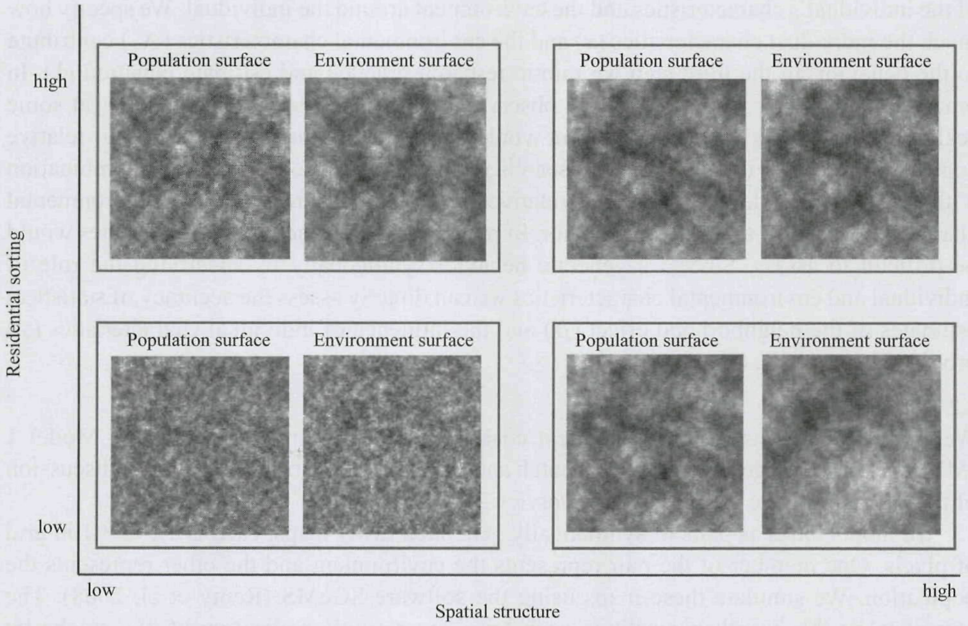


**Figure 1.** Four examples of synthetic cities. The cities in the left-hand column have a small-scale spatial structure, evident by the amount of variation in the surface over small scales. The cities on the right have zones of similarity that are more geographically extensive than those on the left. The cities in the upper row have high residential sorting; the population and environment surface look similar; there is a high correlation between the images at each location. The cities on the lower row have low residential sorting as evident by the significant differences between the population and environment maps.

The bottom row of figure 1 shows the effect of varying $\theta$. When $\theta$ is small, our maps show a dense patchwork of regions. When $\theta$ is large, the maps show extensive areas of similarity.

The second half of the city, the population, is generated using the map of the environment as a starting point. The mechanisms that lead to geographic sorting by race, income, and other demographic factors are openly debated (Clark, 1992; 2009), but the facts are not. In North America, race, ethnicity, and income are powerful predictors of where within a city a person will live. In our simulation model we are interested in recreating the outcome, not the system(s) or structures that created the sorting (as in Bruch and Mare, 2006). In practice, cities with a high degree of residential sorting will exhibit a strong correlation between the environment and individual characteristics.

We model the degree of residential sorting in a city by controlling the correlation at each location on the map pairs. If we denote a location in the city as $\boldsymbol{u}$, a person's home could be described by its location on the environment map $z(\boldsymbol{u})$ and the population map $x(\boldsymbol{u})$. To simulate residential sorting we generate a population map where the correlation with the environment map at each location is controlled by a parameter $\rho$. This sorting parameter $\rho$ varies between 0 and 1. A high value of $\rho \approx 1$ will generate a city where the environment map and the population map are very similar. When $\rho$ is high there is nearly perfect correlation between the values at all locations on the two maps. When $\rho$ approaches zero, the correlation between the two maps will approach zero, recreating a situation where a person's residential location is entirely independent of their characteristics. The extremes of $\rho$ are unlikely to occur in reality, but are useful for the purposes of description. The top row in figure 1 shows simulates cities (map pairs) with high sorting; the bottom row shows cities with low sorting.

The correlation between maps, controlled by $\rho$, can result in a correlation between the $\boldsymbol{x}$ and $A_d$ terms in equation (1). However, the size of a person's neighborhood relative to $\theta$ will also have an impact on model estimation (and the degree of correlation between $\boldsymbol{x}$ and $A_d$). If $\theta$ is small the city is a dense patchwork of regions and if an individual's neighborhood is large it will encompass a variety of different types of regions. However, if $\theta$ is large, as it might be in a postwar suburban environment, a change in the size of person's neighborhood will have little impact on its characteristics. Estimates of neighborhood effects may be more sensitive to the definition of a person's neighborhood when there is high degree of spatial variability in the environment (ie, $\theta$ is set to a low value).

Generally speaking, multivariate stochastic simulation algorithms require a model of auto-covariance and cross-covariance among multiple variables at any pair of two locations in the study domain. Solving such a large cokriging system becomes computationally demanding as the number of variables and/or size of the map increases. The MM1 algorithm is based on a data-screening hypothesis, where the degree of sorting is assumed to screen the adjacent portions of the environment surface at each location. The MM1 model implies that it is sufficient to infer and model the autocorrelation of a single variable to cosimulate two variables whose correlation is controlled by $\rho$.

### 2.3 Conceptualizing behavior

The basic hypothesis of the neighborhood effects literature is that the environment influences behavior (and/or health). Within this simple hypothesis lies enormous complexity (Entwisle, 2007). While we acknowledge this complexity we largely ignore it in our simulation model. One might ask how a model can be valid if it ignores the subtlety of the problem at hand. In response, we remind the reader that our purpose is to model research practice: in a sense ours is a reductionist model of a reductionist practice (quantitative research). As a result we have a very simplified conceptualization of behavior:

$$\text{behavior} = \text{individual characteristics} + \text{environmental characteristics}. \qquad (2)$$

Our synthetic cities consist of a map of population and a map of the environment; these maps provide the inputs into our measure of behavior. To construct the measure, we randomly sample 500 individuals. An individual is a single raster cell with a location $\boldsymbol{u}_i$ ($i = 1, \ldots, 500$); these locations are analogous to a person's home address. The value of the population map at location $x(\boldsymbol{u}_i)$ (the person's home) determines an individual's characteristics. Similarly, an individual's environment is determined by the coincident location $z(\boldsymbol{u}_i)$ on the environment surface. However, an individual's environment is a territory not a discrete location. Therefore, we represent each person's environment as a circle with a diameter of 10 distance units centered on location.

This model of behavior is an abstraction of the Eid–Blalock conceptualization of contextual effects [represented in equation (1)]. It is widely used in research practice. For example, Frank et al (2004) drew a buffer around the homes of 8000 individuals in Atlanta, Georgia. They used both the characteristics of the area contained by the buffer and the characteristics of the homeowner/renter. Specifically, they model the mixture of land uses in the area contained by the buffer. They then estimate the relative contribution of individual characteristics and environmental characteristics to a person's odds of being obese. They found that, as land-use mix increases, a person's odds of being obese decreases. Our model is a direct abstraction of this type of analysis, of which Frank et al (2004) is but one of dozens. For a review of this literature see Papas et al (2007), Ewing and Cervero (2010), Dietz (2002), Transportation Research Board and the Institute of Medicine of the National Academies (2005), among others.

2.3.1 *Synthesis of behavior*

We use the general framework of equation (2) to synthesize behavior. Specifically, the behavior of the person at location $\boldsymbol{u}_i$ is generated as:

$$Y(\boldsymbol{u}_i) = x(\boldsymbol{u}_i) + \omega A^*_{d=10}(\boldsymbol{u}_i) , \qquad (3)$$

where the $x(\boldsymbol{u}_i)$ corresponds to characteristics of the individual at location $\boldsymbol{u}_i$ and $A^*_{d=10}$ corresponds to that person's environment. Specifically, the subscript $d = 10$ denotes the fact that the average of the area within 10 map units of $\boldsymbol{u}_i$ is used as a measure of a person's environment. The asterisk is added to differentiate this measure from $A_d$ in equation (1); $A^*_d$ is used to generate behavior. In research practice the observation of individual characteristics $x(\boldsymbol{u}_i)$ is relatively straightforward, but observation of an individual's neighborhood is not. Simulating behavior in this way allows us to specify both the strength of the neighborhood's influence on behavior using $\omega$ and the size of each person's neighborhood using $A^*_{d=10}$. The parameters $\omega$ and $A^*_{d=10}$ remain fixed throughout the entire experiment, and are used to generate $Y(\boldsymbol{u})$. The result is a sample of 'individuals' and their 'behaviors' based on a "city" with a known level of residential sorting $\rho$ and a known urban structure $\theta$. In this 'city' all individuals are identical in that the geographic scale and magnitude of the environment's influence on behavior is the same for each person; the parameters used to generate behavior do not vary from person to person.

**2.4 Conceptualizing neighborhood effects**

Having generated a synthetic behavior for a synthetic population of a synthetic city we return to the problem of estimating neighborhood effects. For our synthetic population we know exactly how the environment influences behavior. In an observational study, one would be able to observe behavior, but the relative contributions of individual and contextual factors to behavior would have to be estimated statistically. Disentangling these components of behavior is not a trivial task, especially when individual characteristics and environmental characteristics are highly correlated, as happens in the presence of residential sorting (Oakes, 2004).

We replicate research practice by disregarding our prior knowledge of the components of behavior. As would occur in applied settings, we use only the location of each person, his or her behavior, and characteristics to estimate the influence of their neighborhood environment on behavior. In practice, the geographic dimensions of a person's neighborhood would be uncertain and we model this uncertainty by using many definitions of a person's neighborhood in our statistical models. Sometimes we deliberately use a neighborhood that is too small; sometimes we deliberately use a neighborhood that is too large; and sometimes we use the correct neighborhood. The 'correct' neighborhood is one that matches the geographic dimensions of the zone used to generate behavior ($d = 10$). We estimate neighborhood effects using various neighborhood definitions which range in size from 1 to 20 distance units, twice as big as the 'real' neighborhood. This practice of deliberately using the wrong neighborhood allows us to explore the potential implications of uncertainty around the true definition of a person's neighborhood.

In equation (1) neighborhood effects are represented with $\beta$. Typically, it would be difficult to assess the accuracy of this coefficient because the influence of the environment cannot be directly observed. However, we simulated behavior using $\omega$ which determined the net contribution of the environment to behavior. Therefore we can measure the bias by comparing our estimates of neighborhood effects [$\hat{\beta}$ from equation (1)] with the 'real' neighborhood effect ($\omega$) used to generate the synthetic behavior. Specifically, we measure bias: the difference between our estimates of the neighborhood effect and the true effect used to generate behavior. Bias is written as:

$$\text{bias} = \hat{\beta} - \omega \ . \tag{4}$$

This sort of direct measurement of bias is possible only within the confines of a simulation model. We repeat these measurements of bias for a wide range of cities: cities with low sorting, high sorting, a heterogeneous spatial structure, and large areas of similarity.

2.4.1 *Modeling bias in neighborhood effects*

To get a full understanding of how well an egocentric model of neighborhood effects can be estimated under uncertainty about the scale of neighborhoods and residential selection, we design an experiment to explore these factors and their interactions. We generate fifty cities. The parameters that shape the city ($\rho$ and $\theta$) were selected using a space filling Latin hypercube design (Neter et al, 2004). The sampling design was three dimensional; for convenience, we notate the *j*th choice for the three factors as $\phi_j = (\phi_{j1}, \phi_{j2}, \phi_{j3})$, $\theta$, $\rho$ and *d*, respectively.

This space-filling design is an important element of our simulation model. With a relatively small sample of fifty cities it allows the assessment of error in the estimates of neighborhood effects under a wide range of conditions. Latin hypercube sampling (LHS), in this instance, involved creating a three dimensional parameter space: $\theta \in [5, 30]$, $d \in [1, 20]$, and $\rho \in [0, 1]$ and randomly selected combinations of parameters to fill the parameter space in both one-dimensional and two-dimensional projections.

An experimental run consists of choosing a particular combination of the model parameters $\phi_j$. A city is synthesized on the basis of the amount of sorting and urban structure $(\phi_{j1}, \phi_{j2})$. Then 500 people are randomly sampled from the map. Each individual's location ($\boldsymbol{u}$) is used to generate a behavior. Following equation (3), the behavior is generated using a fixed neighborhood with a diameter of 10 units ($A_d^*$). In practice, $A_d^*$ would be unknown and estimated, mimicking this $\phi_{j3}$ is used to construct a neighborhood of various dimensions around each person such that $A_{d=\phi_{j3}}$. The neighborhood based on $\phi_{j3}$ is used in the estimation of equation (1). Often we will use an incorrect neighborhood. The misspecification of an individual's neighborhood is, $A_{d=10}^* - A_{d=\phi_{j3}}$, the difference between the size of the neighborhood used to create the simulated behavior and the size of the neighborhood used to estimate equation (1).

For a given factor combination $\phi_j$ we repeat this process many times, placing 100 sets of 500 people in order to obtain an estimate of the expected value of the neighborhood effect $E\{\hat{\beta}\}$. Estimating equation (1) with the simulated behavioral data we obtain $\text{bias}_j = \frac{\omega - E\{\hat{\beta}\}}{\omega}$, the bias given a particular combination of parameters $\phi_j$. The difference between $\omega$ and $\hat{\beta}$ is the amount of error in the estimation of neighborhood effects; this quantity is dependent upon the choice of $\omega$, to scale the measure we report the error as a proportion.

Using runs $\phi_1, \ldots, \phi_n$ and their corresponding outcomes $\text{bias}_1, \ldots, \text{bias}_n$, we establish the relationship $\text{bias}_j = \mathrm{f}(\phi_j) + e$. We approximate $\mathrm{f}(\phi_j)$ using semiparametric estimation. A third-order polynomial estimation captures the relationship sufficiently well [equation (5)]. That is,

$$\text{bias}_j = \alpha + \sum_i^3 \alpha_i \phi_i + \sum_i^3 \alpha_{ii} \phi_i^2 + \sum_i^3 \alpha_{iii} \phi_i^3 + \sum_{i<j}\sum \alpha_{ij} \phi_i \phi_j + \alpha_{ijk} \phi_i \phi_j \phi_k \ , \tag{5}$$

where $\phi_1, \phi_2, \phi_3$ correspond to $\theta$, $\rho$, and *d*, respectively. Equation (5) models the effect of each model parameter and all parameter interactions on bias. The $\alpha$'s represent the relative contribution of each parameter to bias.

In conclusion, we ran a series of 100 regressions for each of the fifty 'cities' that were randomly sampled from a large set of possible parameter combination. For each 'kind' of city we systematically varied the definition of a person's neighborhood. The series of regressions we ran for each combination of parameters $\phi_j$ provides us with 100 estimates of $\beta$ for each type of city. We then calculate bias for each combination of parameters. Equation (5) allows us to model the independent and joint contributions of residential selection, urban structure, and the definition of an individual's neighborhood to bias.

## 3 Results

In spite of the complexity of the simulation procedure, the results of the experiments can be summarized using two simple and intuitive graphical tools: main effect and interaction plots. The main effect of each parameter is its relationship with bias while the other two parameters are held fixed. For example, the main effect plot of the neighborhood size parameter $d$ in figure 2(b) shows changes in bias as the size of the neighborhood changes from 1 to 20 distance units. The main effect plots in figure 2 display the independent contribution of each simulation parameter to bias. Bias is expressed in proportional terms so that a value of 0.5 on the vertical axis means that for a given parameter setting the neighborhood effect ($\beta$) was overestimated by 50%.

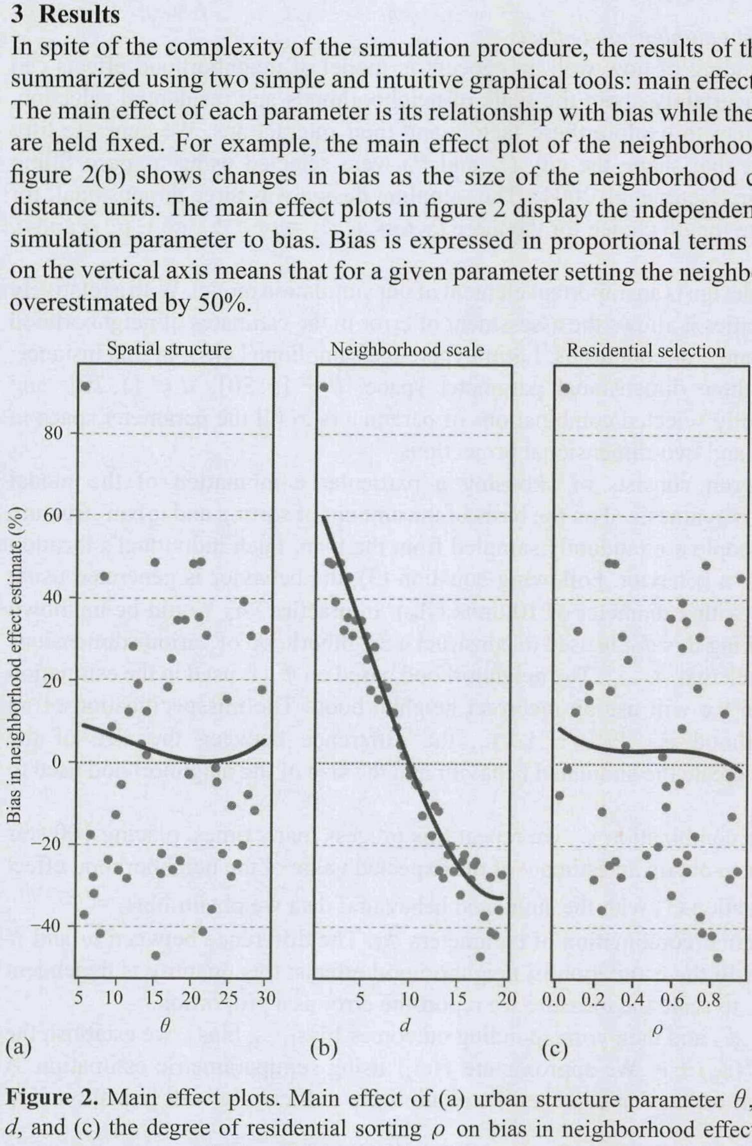


**Figure 2.** Main effect plots. Main effect of (a) urban structure parameter $\theta$, (b) neighborhood size $d$, and (c) the degree of residential sorting $\rho$ on bias in neighborhood effects. Bias is expressed in percentage terms such that a bias of 20% indicates that the strength of the association between the environment and the outcome is overestimated by 20%. Negative values indicate underestimation.

Bias systematically changes as the spatial scale of an individual's neighborhood changes [as shown by the good fit of the curve in figure 2(b)]. It is no surprise that the correct specification of an individual's neighborhood $d = 10$ (ie, $A^*_{d=10} = A_{d=\phi_{j3}}$) leads to an unbiased estimator, $\omega - E\{\hat{\beta}\} = 0$ (this is illustrated by the curve in figure 2 crossing zero when $d = 10$). However, when the definition of a person's neighborhood is wrong, estimates of the association between the environment and behavior are wrong. Small neighborhoods ($d < 10$) result in the overestimation of neighborhood effects and large neighborhoods yield underestimated effects. Of course 'large' and 'small' are relative terms defined by reference to the 'true' neighborhood $A_d^*$. In applied settings, since there is no clear way to assess the actual dimensions of an individual's neighborhood, it is difficult to say if a study has overestimated or underestimated the size of a person's neighborhood. Therefore, it is hard to determine if the association between the environment and behavior is overestated or understated. These results indicate that erring toward overestimation of the size of person's neighborhood is more conservative than underestimating its size.

Unlike the neighborhood size parameter *d*, there is no systematic relationship between residential sorting, the structure of the city, and the bias in $\hat{\beta}$ [figures 2(a) and 2(c)]. The magnitude of the bias due to residential sorting $\rho$ and urban structure $\theta$ is as high as that of *d* (reaching a maximum value 0.9 and minimum $-0.4$), but setting the urban structure parameter to a lower value ($\theta = 5$) is associated with the overestimation (bias $\approx 0.9$) as well as underestimation (bias $\approx -0.4$). The average bias is close to zero in both figures 2(a) and 2(c) (see fitted curve), indicating that, while individual studies are biased by residential sorting across repeated studies of similar populations, this bias tends to be canceled out.

These plots indicate that bias due to the mispecification of a person's neighborhood is predictable whereas bias due to residential sorting is not. Figure 2 shows that the estimates of neighborhood effects from individual studies that do not account for residential sorting are highly unreliable. We have in effect done many independent synthetic studies of neighborhood effects. In the real world these studies are expensive and time consuming. Our results indicate that in a meta-analysis bias caused by residential selection may, on average, cancel out.

### 3.1 Parameter interactions

The interactions between simulation parameters are visualized using a contour plot, where bias in neighborhood effects estimates are displayed as a function of a pair of any two simulation parameters (holding the third parameter constant). In figure 3, for example, we are looking at the neighborhood size–urban structure parameter plane; contour lines connect all points with a similar amount of bias. Bias is expressed in proportional terms, a 0.15 contour line represents an overestimation of the neighborhood effect ($\beta$) by 15%. The contour plots in figures 3–5 suggest complex interactions among variables in that the contours are not parallel straight lines. For example, when one underestimates the size of person's neighborhood (*d*) bias tends to be positive; however, the magnitude of the bias is associated with the degree of sorting (figure 4). Underestimating the size of a person's neighborhood in the presence of a high degree of residential sorting is worse than using the same neighborhood when sorting is low.

Similar to the conclusion drawn from the main effect plot in figure 2(b), the correct specification of the neighborhood parameter $d = 10$ leads to a near zero bias regardless of the specification of the other parameters. One striking feature of the interaction plots is that the magnitudes of the bias in the plots that do not include *d* are relatively low in comparison with those that do. Again, these contour lines represent the average bias for a given combination of parameters; the low values are deceiving as they mask the significant (random) variability in bias that occurs between experimental runs with similar parameter settings. We see variability within a given parameter regime because of the stochastic nature of the cities and the random sampling procedure used to create the population.

In figure 3, the interaction between $\theta$ and *d*, we assume that the degree of residential sorting $\rho$ is held constant. As shown in the main effect plot, the bias due to overspecification and underspecification of the spatial scale parameter *d* results in underestimation and overestimation of the neighborhood effect. The urban structure parameter $\theta$, however, alters

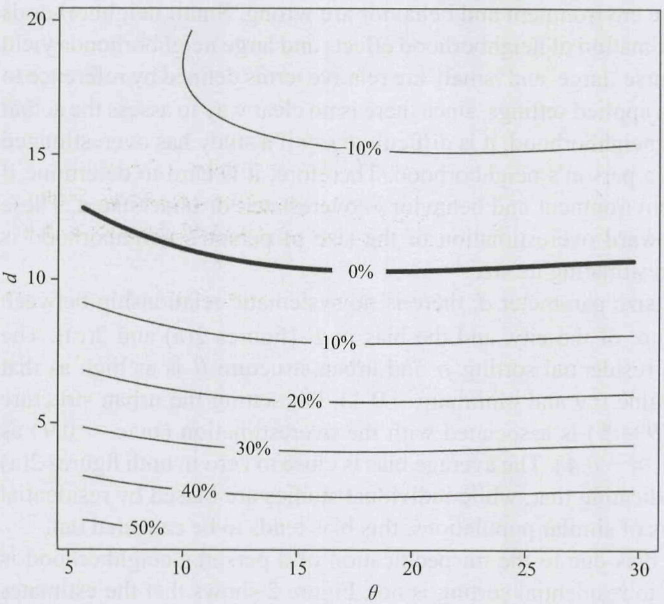


**Figure 3.** Interaction between $\theta$ and $d$. Contour lines show bias $[100(\omega - \beta)/\omega]$ in the neighborhood effect estimate. Bias is expressed in percentage terms such that a bias of 20% indicates that the strength of the association between the environment and the outcome is overestimated by 20%

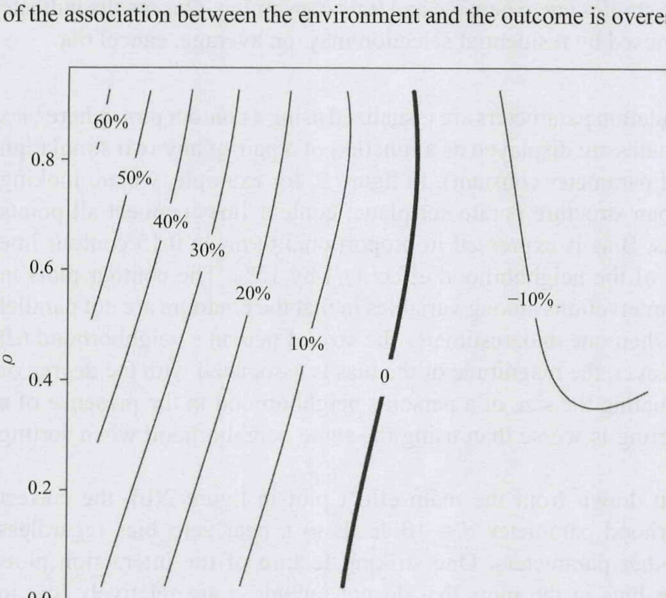


**Figure 4.** Interaction between $\rho$ and $d$. Contour lines show bias $[100(\omega - \beta)/\omega]$ in the neighborhood effect estimate. Bias is expressed in percentage terms such that a bias of 20% indicates that the strength of the association between the environment and the outcome is overestimated by 20%.

the magnitude of the distance parameter's effect on the bias. Generally speaking, a city comprised of smaller communities $\theta < 10$ contributes to higher bias, particularly when $d$ is small. However, once the size of the communities in a city exceed the size of person's neighborhood, changes in $\theta$ have little impact on bias regardless of the scale. This suggests that, if neighborhood effects are estimated using a small homogeneous community around an individual's home but dissimilar surrounding areas contribute to behavior, there is an elevated risk of misestimating neighborhood effects.

## 4 Discussion and conclusions

What is the effect of the urban environment on behavior? Questions about the relationship between the environment and behavior while easy to ask are difficult to answer empirically. In this paper we outline two reasons why this question is so hard to answer. First, we explore the confounding role of residential sorting and, second, we explore how uncertainty about an individual's neighborhood impacts statistical estimates of the association between the environment and behavior.

We conducted an experimental study using simulated cities. Specifically, in the experiment we systematically varied three parameters: the degree of residential sorting, the size of an individual's neighborhoods, and the spatial structure of the urban environment. We modeled bias in neighborhood effect estimates as a function of these three parameters. The advantage of using simulation is that it allows the measurement and modeling of bias through a large number of independent experiments. In order to explore the range of possible combinations of parameter values we used Latin hypercube sampling. Each combination of parameters mimics a different type of city where the interaction between individuals and the environment occurs at a fixed scale.

The simulation study showed that neighborhood effect estimates are strongly influenced by the definition of neighborhoods. Neighborhood effect estimates depend upon the size of the area used to approximate a person's neighborhood relative to their true neighborhood ($A_d - A_d^*$). Of course, in practice it is very difficult to say if a particular conceptualization of a person's neighborhood is 'too big' or 'too small'—we believe that this is a major concern for the social and health sciences. One cannot understand how the environment affects individual behavior until we understand how individuals interact with their environment.

Our experiment uses synthetic data, but is motivated by very real policies. Our results indicate that the evidence base supporting policies which target issues, like obesity and/or auto dependence, through small-scale changes to the built environment may not be as effective as anticipated. For example, in transportation and urban planning Ewing and Cervero (2010) have argued that there is a clear consensus that the built environment relates to travel behavior. This consensus is based in part on literature that uses a unit of analysis like $A_d$ to measure the built environment; if the research tends to underestimate the scale of people's 'neighborhood', the consensus may overstate the truth. The American Center for Disease Control has a Healthy Community Design Initiative which argues that neighborhood design can increase physical activity and decrease stress; the evidence in support of these claims (see Frumkin, 2002) is in part based on studies that use units of analysis like $A_d$. Our work does not tell us if these bodies of research correctly estimated the magnitude of the relationship between the environment and behavior, but it does raise questions about the assumptions that are used to operationalize environmental concepts like 'neighborhood', 'community design', or 'built environment'. We hope this experiment sparks reflection on the assumptions implicit in the evidence base supporting policies that aim to manipulate behavior (or health) by altering the urban environment. To understand the relationship between the environment and behavior we need better conceptual models. Reducing uncertainty about the definition(s) of a person's environment and questioning the use of arbitrary zones (such as 0.25 or 0.5 miles) may be an important step forward. In addition, we need to understand whether the scale of a person's interaction with the built environment is conditioned by age, gender, or socioeconomic status; we need to understand if the relationship varies culturally or regionally. Rich, new sources of information, such as activity traces generated by mobile phone-users, may provide important information about the scale of individual environments (Gonzalez et al, 2008).

Rather than trying to model the complex relationship between individuals and the urban environment directly, we model the assumptions implicit in widely used statistical models. Our simulation model is relatively simple in light of the complexity of the problems that we are dealing with and we have boiled down a wide swath of important questions about the effect of the urban environment on human behavior/health to a series of parameters. This abstraction of reality provides a level of experimental control that would be impossible in an observational study. However, we report on the relationships and outcomes that exist within the confines of the simulation, the degree to which these same relationships exist in the 'real world' is open to debate.

In summary, we found that the definition of neighborhoods plays an important role in the estimation of neighborhood effects. This relationship is systematic and predictable. The problem of residential sorting is amplified when the spatial scale of a person's neighborhood is underspecified; residential sorting's impact is minimal when the neighborhood is measured over an appropriate geographic unit.

**References**


Axelrod R, 2007, "Simulation in the social sciences", in *Handbook of Research on Nature Inspired Computing for Economy and Management* Ed. J-P Rennard (Idea Group, Hershey, PA) pp 90–100

Blalock H M, 1984, "Contextual-effects models: theoretical and methodological issues" *Annual Review of Sociology* **10** 353–372

Brooks-Gunn J, Duncan G, Klebanov P, Sealand N, 1993, "Do neighborhoods influence child and adolescent development?" *American Journal of Sociology* **99** 353–395

Browning C, Wallace D, Feinberg S, Cagney K, 2006, "Neighborhood social processes, physical conditions, and disaster-related mortality: the case of the 1995 Chicago heat wave" *American Sociological Review* **71** 661–678

Bruch E, Mare R, 2006, "Neighborhood choice and neighborhood change" *American Journal of Sociology* **112** 667–709

Cervero R, Kockelman K, 1997, "Travel demand and the 3ds: density, diversity, and demand" *Transportation Research D* **2**(3) 199–219

Chaix B, 2009, "Geographical life environments and coronary heart disease: a literature review, theoretical contributions, methodological updates, and a research agenda" *Annual Review of Public Health* **30** 81–105

Chaix B, Leyland A, Sabel C, Chauvin P, Råstam L, Kristersson H, Merlo J, 2006, "Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmö, Sweden in 2001" *Journal of Epidemiology and Community Health* **60** 427–435

Chaix B, Merlo J, Evans D, Leal C, Havard S, 2009, "Neighbourhoods in eco-epidemiologic research: delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur" *Social Science and Medicine* **69** 1306–1310

Chiles J, Delfiner P, 1999 *Geostatistics: Modeling Spatial Uncertainty* (John Wiley, New York)

Clark W, 1992, "Residential preferences and residential choices in multiethnic context" *Demography* **29** 451–466

Clark W, 2009, "Changing residential preferences across income, education, and age" *Urban Affairs Review* **44** 334–355

Clifton K, Ewing R, Knaap G, Song Y, 2008, "Quantitative analysis of urban form: a multidisciplinary review" *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* **1**(1) 17–45

Couclelis H, 1992, "People manipulate objects (but cultivate fields): beyond the raster vector debate in GIS", in *Theories and Methods of Spatio-temporal Reasoning in Geographic Space* Eds A Frank, I Campari, U Formentini (Springer, Berlin) pp 65–77

Coulton C J, Korbin J, Chan T, Su M, 2001, "Mapping residents' perception of neighborhood boundaries: a methodological note" *American Journal of Community Psychology* **29** 371–383

Crane R, Boarnet M, 2001 *Travel by Design: The Influence of Urban Form on Travel* (Oxford University Press, New York)
Deutsch C, Journel A, 1992 *GSLIB: Geostatistical Software Library and User's Guide* (Oxford University Press, Oxford)
Dietz R D, 2002, "The estimation of neighborhood effects in social sciences: an interdisciplinary approach" *Social Science Research* **31** 539–575
Diez-Roux A, 1998, "Bringing context back into epidemiology: variables and fallacies in multilevel analysis" *American Journal of Public Health* **88** 216–222
Diez-Roux A, 2004, "Estimating neighborhood health effects: the challenges of causal inference in a complex world" *Social Science and Medicine* **58** 1953–1960
Earls F, Buka S, 1997 *Project on Human Development in Chicago Neighborhoods* (National Institute of Justice, Washington, DC)
Eid J, Overman H, Puga D, Turner M, 2008, "Fat city: questioning the relationship between urban sprawl and obesity" *Journal of Urban Economics* **63** 385–404
Entwisle B, 2007, "Putting people into place" *Demography* **44** 687–703
Ewing R, Cervero R, 2001, "Travel and the built environment: a synthesis" *Transportation Research Record* number 1780, 87–113
Ewing R, Cervero R, 2010, "Travel and the built environment—a meta-analysis" *Journal of the American Planning Association* **76** 265–294
Flowerdew R, Manley D, Sabel C, 2008, "Neighbourhood effects on health: does it matter where you draw the boundaries?" *Social Science and Medicine* **66** 1241–1255
Forsyth A, Oakes J M, Schmitz K, Hearst M, 2007, "Does residential density increase walking and other physical activity?" *Urban Studies* **44** 679–697
Frank L D, Pivo G, 1995, "Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking" *Transportation Research Record* number 1466, 44–52
Frank L D, Engelke P, Schmid T L, 2003 *Health and Community Design: The Impacts of Community Design on Physical Activity* (Island Press, Washington, DC)
Frank L, Andresen M, Schmid T, 2004, "Obesity relationships with community design, physical activity, and time spent in cars" *American Journal of Preventive Medicine* **27**(2) 87–96
Freedman D, 1991, "Statistical models and shoe leather" *Sociological Methodology* **21** 291–313
Frumkin H, 2002, "Urban sprawl and public health" *Public Health Reports* **117** 201–217
Galster G, 1986, "What is neighborhood? An externality-space approach" *International Journal of Urban and Regional Research* **10** 243–263
Galster G, 2008, "Quantifying the effect of neighbourhood on individuals: challenges, alternative approaches, and promising directions" *Schmollers Jahrbuch* **128**(1) 7–48
Gauvin L, Robitaille E, Rive M, McLaren L, Dassa C, Potvin L, 2007, "Conceptualizing and operationalizing neighbourhoods: the conundrum of identifying territorial units" *Canadian Journal of Public Health* **98** 18–26
González M C, Hidalgo C A, Barabási A-L, 2008, "Understanding individual human mobility patterns" *Nature* **453** 779–782
Grannis R, 1998, "The importance of trivial streets: residential streets and residential segregation" *American Journal of Sociology* **103** 1530–1564
Guo J, Bhat C, 2007, "Operationalizing the concept of neighborhood: application to residential location choice analysis" *Journal of Transport Geography* **15** 31–45
Haan M, Kaplan G, Camacho T, 1987, "Poverty and health prospective evidence from the Alameda county study" *American Journal of Epidemiology* **125** 989–998
Hägerstrand T, 1982, "Diorama, path and project" *Tijdschrift voor Economische en Sociale Geografie* **73** 323–399
Handy S, Cao X, Mokhtarian P, 2005, "Correlation or causality between the built environment and travel behavior? Evidence from northern California" *Transportation Research Part D: Transport and Environment* **10** 427–444
Harding D, 2003, "Counterfactual models of neighborhood effects: the effect of neighborhood poverty on dropping out and teenage pregnancy"*American Journal of Sociology* **109** 676–719

Hunter A, 1974 *Symbolic Communities* (University of Chicago Press, Chicago, IL)
Kwan M-P, 2000, "Interactive geovisualization of activity–travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set" *Transportation Research C* **8** 185–203
Kwan M-P, 2009, "From place-based to people-based exposure measures" *Social Science and Medicine* **69** 1311–1313
Laube P, Imfeld S, 2002, "Analyzing relative motion within groups of trackable moving point objects", in *Geographic Information Science, GIScience 2002* Eds M J Egenhofer, D M Mark (Springer, Heidelberg) pp 132–144
Lebel A, Pampalon R, Villeneuve P, 2007, "A multi-perspective approach for defining neighbourhood units in the context of a study on health inequalities in the Quebec city region" *International Journal of Health Geographics* **6** article 27
Lee B A, Reardon S F, Firebaugh G, Farrell C R, Matthews S A, O'Sullivan D, 2008, "Beyond the census tract: patterns and determinants of racial residential segregation at multiple geographic scales" *American Sociological Review* **73** 766–791
Macintyre S, Ellaway A, Cummins S, 2002, "Place effects on health: how can we conceptualise, operationalise and measure them?" *Social Science and Medicine* **55** 125–139
Miller H, 1991, "Modeling accessibility using space–time prism concepts within geographical information systems" *International Journal of Geographical Information Systems* **5** 287–301
Miller H, 2005, "A measurement theory for time geography" *Geographical Analysis* **37** 17–45
Miller H, 2007, "Place-based versus people-based geographic information science" *Geography Compass* **1** 503–535
Neter J, Wasserman W, Kutner M, Li W, 2004 *Applied Linear Statistical Models* 5th edition (McGraw-Hill, New York)
Northridge M E, Sclar E D, Biswas P, 2003, "Sorting out the connections between the built environment and health: a conceptual framework for navigating pathways and planning healthy cities" *Journal of Urban Health* **80** 556–568
Oakes J M, 2004, "The (mis)estimation of neighborhood effects: causal inference for a practical social epidemiology" *Social Science and Medicine* **58** 1929–1952
Papas M, Alberg A, Ewing R, Helzlsouer K, Gary T, Klassen A, 2007, "The built environment and obesity" *Epidemiologic Reviews* **29**(1) 129–143
Remy N, Boucher A, Wu J, 2008 *Applied Geostatistics with SGeMS: A User's Guide* (Cambridge University Press, Cambridge)
Sampson R, Raudenbush S, Earls F, 1997, "Neighborhoods and violent crime: a multilevel study of collective efficacy" *Science* **277**(5328) 918–924
Sampson R, Morenoff J, Earls F, 1999, "Beyond social capital: spatial dynamics of collective efficacy for children" *American Sociological Review* **64** 633–660
Sampson R J, Morenoff J D, Gannon-Rowley T, 2002, "Assessing 'neighborhood effects': social processes and new directions in research" *Annual Review of Sociology* **28** 443–478
Sastry N, Pebley A, Zonta M, 2002, "Neighborhood definitions and the spatial dimension of daily life in Los Angeles", technical report, Rand Corporation, Santa Monica, CA
Spielman S, Yoo E, 2009, "The spatial dimensions of neighborhood effects" *Social Science and Medicine* **68** 1098–1105
Transportation Research Board and the Institute of Medicine of the National Academies, 2005, "Does the built environment influence physical activity? Examining the evidence", technical report, National Academy of Sciences, Washington, DC
Weber J, Kwan M, 2003, "Evaluating the effects of geographic contexts on individual accessibility: a multilevel approach" *Urban Geography* **24** 647–671
Zhang M, 2005, "Metrics of urban form and the modifiable areal unit problem" *Transportation Research Record* number 1902, 71–79