# Capturing the Relationship Between Spatial Structure and Individual Outcomes: Variation in the Concept of "Access to Parks" and its Association with BMI

ELIJAH KNAAP, University of California-Riverside
CHANDRA A. REYNOLDS, University of California-Riverside
SERGIO REY, University of California-Riverside
ROBIN P. CORLEY, University of Colorado Boulder
SALLY J. WADSWORTH, University of Colorado Boulder

In this study, we examine how the choice of methods and data sources affect observed relationships between park accessibility and physical health, as measured through Body Mass Index (BMI). With a longitudinal study of the effects of neighborhoods on human development and cognitive aging as our backdrop, we examine how choices of spatial representation, sources of spatial data, and methods of spatial analysis yield a variety of different conclusions regarding the fundamental research questions about the associations between neighborhoods and their residents' life. Our results help clarify the amount of variance that subjective decisions like these introduce into quantitative studies. We use these results to provide guidance on how certain decisions should be made, and when researchers' omitted discussions about such choices should raise red flags; in so doing, we set the stage for a broader discussion about social science's replication crises in the special context of spatial data.

*Keywords*: accessibility, BMI, neighborhood effects, externality space, uncertain geographic context problem

## INTRODUCTION

The relationship between the built environment and physical activity has been a locus of research for both public health and urban planning for decades (Ewing et al., 2003). While many questions remain, a large share of "public health research has shown that neighborhood conditions are associated with health behaviors and outcomes" (Bader et al., 2015). Among this scholarship a growing body of work examines the relationship between the availability of parks and recreational space in a neighborhood and the physical fitness levels of neighborhood residents. Uncovering the existence and nature of this relationship assumes an elevated importance because it lies at the nexus of several disciplines including public health, sociology, geography, and developmental psychology. Further, because of its potential impact on critical life outcomes, access to park space is a longstanding social equity issue in urban policy and planning (Barbosa et al., 2007; Ewing et al., 2003; Talen, 2010; Talen & Anselin, 1998), and prior work demonstrates inequitable distribution typically in favor of white affluent residents (Wolch et al., 2014).

As the vast and growing body of research on neighborhood effects emanating from both environmental conditions (both social and physical) has grown, it has benefitted greatly from the increasingly wide variety of open geospatial data available on the web, as well as new statistical methods, estimation algorithms, and analytical techniques, notwithstanding the dissemination, discovery, and reuse of computer code through version control systems and other social platforms such as GitHub, Reddit, and StackOverflow. But while these new methods and data sources have become more popular and made research quicker and more sophisticated, so too have they made the research process more complex. With these new data and methods comes a new potential for entropy to enter the research process: which methods will the research team adopt, and which data sources will they choose? When neighborhood effects scholarship defines "the neighborhood," how are spatial relationships specified? Does the relationship change depending on the effect under study? Together these questions imply serious consequences for reproducibility (Kane & Kim, 2018).

Similarly, geospatial data can be collected from local governments, downloaded from public resources like OpenStreetMap, or mined from external resources like Craigslist, each of which has particular strengths and idiosyncrasies. When selecting among these choices, most scholars in the contemporary literature make reasonable justifications for the choices at hand. Rarely, however, are the choices of methods and data explored in depth, their tradeoffs thoroughly explored, and the impacts of these choices analyzed with respect to the paper's ultimate outcomes. Together, these issues contribute to continued uncertainty regarding the relationship between neighborhood context and health. There are several reasons that uncertainty persists: (1) potential selection bias inhibits attributing directionality to any observed association, (2) imprecise and coarsely-measured data on both neighborhood conditions and individual outcomes, and (3) inconsistent and untested definitions of neighborhood spatial context (i.e. the concepts and operationalization of a neighborhood effect mechanism) (Galster, 2003).

In this paper we focus on issues 2 and 3, examining how subjective decision-making during the research process may lead to a wide variety of conclusions about the relationship between spatial structure and public health. Motivating our discussion is the association between physical health and activity space. Specifically, we measure the association between physical health, measured by individual-level BMI, and "access" to "parks and recreation space," where we vary systematically the concepts of (a) accessibility (ranging from simple euclidean distance to the nearest park, to a variety of accessibility measures that apply exponential weighting to parkland accessible along the pedestrian travel network, discounted for nearby competition) and (b) data considered to represent physical activity space. Our results make clear that the estimated relationship between parks and health varies dramatically depending on the specification of both parkland and accessibility, and may lead researchers to defend a wide variety of conflicting results based on which versions of each are chosen, or which data are available.

With a longitudinal study of the effects of neighborhoods on human development and cognitive aging as our backdrop, we examine how choices of spatial representation, sources of spatial data, and methods of spatial analysis yield a variety of different conclusions regarding the fundamental research questions about the relationship between spatial context and individual outcomes. Our results help clarify the amount of variance that subjective decisions like these introduce into quantitative research. We use these results to provide guidance on how certain decisions should be made, and when researchers' omitted discussions about such choices should raise red flags; in so doing, we set the stage for a broader discussion about social science's replication crises in the special context of spatial data.

## NEIGHBORHOODS AND PHYSICAL HEALTH

Capturing the relationship between the built environment and individual-level health is a sincere challenge, given that ecological models of health behavior recognize multiple pathways of influence including individual behaviors, community-level behaviors, and physical infrastructure (Casey et al., 2014; Cohen et al., 2008; Diez Roux, 2001; Ellen et al., 2001; Minh et al., 2017; Sallis et al., 2008). The theoretical nexus between health and environment, in the present context, is that shorter distances and better facilities for recreation space will lead to increased physical activity and thus, improved physical health. But uncovering this relationship empirically has assumed different forms, depending on available data and research questions involved. Joseph & Maddock (2016) for example, review park-based physical activity literature, finding that a majority of park users engaged in "moderately vigorous physical activity" (MVPA). Results elsewhere are mixed, however. For example after controlling for individual and neighborhood level socioeconomic variables, Witten et al. (2008) find no association between BMI, sedentary behaviour, or physical activity and neighborhood access to parks. In addition to mixed results, the directionality of the relationship between park access and physical health remains unclear: it may be that park access increases physical health, but it may also be that people who are more physically fit prefer to live closer to parks. Although they do not address it directly, Lin et al. (2014) hint at these notions of selection bias, demonstrating that "orientation" toward park use is a greater predictor of park usage than simple accessibility.

### Park Access and Physical Activity

The relationship between recreation space and physical health articulated above relies on two critical pathways: first, that increased availability of parks and recreation space will lead to increased physical activity, and second that the increased activity induced by ease of recreation will manifest in healthier body measurements. If either of these pathways fails, then the connection between local environment and individual health fails to manifest. Thus, we briefly examine the literature focusing on each of these connections. In general, prior work is supportive of the notion that "living closer to parks and open space is generally related to increased physical activity levels," but even among early reviews of the work, scholars are careful to highlight that "research on proximity to parks and physical activity has been limited by several shortcomings, including a lack of detail in the measurement of park proximity and in the measurement of physical activity" (Kaczynski et al., 2009).

Another confounding issue in the literature is that neighborhoods are complex entities in which multiple pathways influence individual behavior, and simple proximity alone may not be enough to stimulate park usage. In particular, the social environment that characterizes a neighborhood can exert a strong influence on resident activities. For example, Moore et al. (2010) finds that the instability of a neighborhood may influence older adults disuse of a nearby park given safety concerns; moreover, the age composition of neighborhood may promote or dissuade older adults to use a nearby park if they perceive more or less social connectedness.

Examining this question in greater detail, Wang et al. (2015) find that social variables are statistically significant predictors of perceived park accessibility but substantively less important than "physical and locational features such as proximity to the park, a pleasant walking experience, and a sufficient number of parks in the neighbourhood." This observation raises two important issues beyond park accessibility that may mediate park usage (and, by extension physical activity), including supportive infrastructure, such the presence and quality of sidewalks and footpaths that facilitate pleasant transportation, and heterogeneity in services and amenities provided by local parks such as the presence of playgrounds, gardens, scenery, or athletic equipment that invite different types of recreation (potentially from different types of users). And there is some evidence that these features may have an important role, as prior work has found evidence that such features, particularly paved paths, can have an important influence (Kaczynski et al., 2008).

### Neighborhoods and Body Mass

A small but growing number of studies have examined directly the relationship between park availability and BMI. In general, the evidence of a relationship between park access and body mass is mixed. Early work by Tilt et al. (2007) shows that objective measures of the physical environment such as greenness, walkability, and accessibility can vary widely from subjective measures, and that measures like greenness increased walking trips and lowered BMI. Further, Wolch et al. (2011) find a significant negative relationship between park acres within a 500-meter distance of children's homes and their observed BMI at age 18, and Wen & Kowaleski-Jones (2012) find that even while controlling for individual and neighborhood sociodemographic variables, lower access to parks as indexed via walkability, density and distance metrics, are associated with risk of obesity.

In a recent review, however, Casey et al. (2014) examine associations between youth weight status and objective environmental conditions, finding consistent support for the notion that increased walkability is associated with lower weight, but that evidence for other sources (e.g. parks) remains unclear. This discrepancy likely stems from the fact that measures of access to parks and other environmental features are highly variable and more research is necessary to disentangle the relationships. Dony et al. (2015) make similar observations, arguing that heterogeneity in the ways that "accessibility" is conceived underlies the divergence in research results. This sentiment is shared by Carthy et al. (2020) who examine the role of transport-network mediation in the relationship between green space accessibility and BMI, finding that "the relationship between urban green spaces and BMI among older adults is highly sensitive to the characterization of local green space."

Together these findings raise questions about the reliability of research that relies on coarse specifications of park accessibility, such as Stark et al. (2014) who find small park space within zip codes were each associated with lower BMI, adjusting for sociodemographic variables at the individual and zip-code levels. While encouraging and consistent with theory, zip codes are large and poor units of analysis in most cases (Grubesic, 2008) suggesting that when used to define spatial contexts in neighborhood effects scholarship, they may be capturing more noise than signal. Put simply, the mixed results of neighborhood-health associations may arise from the varied definitions of the neighborhood features (Kwan, 2012).

## QUANTIFYING NEIGHBORHOOD CONTEXT

Since the neighborhood effects literature began to burgeon during the 1990s, inspired largely by the work of Wilson (1987), one of the most critical yet unanswered questions remains: what *is* a neighborhood? And how do we measure it? "Obtaining neighborhood-level measures that approximate the theoretical constructs of interest" (Duncan & Raudenbush, 2001) have been among the chief methodological considerations in the neighborhood effects literature since its inception, and there has been considerable methodological attention devoted to capturing theoretically meaningful and empirically valid measures of neighborhood conditions (Raudenbush, 2003; Raudenbush & Sampson, 1999; Sampson et al., 2002). Yet despite longtime conceptual and empirical attention, these questions remain largely unanswered today (Galster, 2001; Galster, 2019). Most canonical studies in the neighborhood effects literature adopt units such as census tracts for convenience and data availability (or occasionally bespoke neighborhood boundaries in the case of Chicago, which are subject to the same criticisms), effectively operationalizing neighborhoods as containers, and ignoring any potential influence from outside the published boundaries (Sampson et al., 2008, 2005, 2002). In many of these cases, the relevant externality space may be captured well by census tract boundaries, but in others they may not.

Adopting census tracts as estimators of neighborhoods is exceedingly common since they comport loosely with notions of "neighborhood" and are delimited according to major physical landmarks such as highways and railroads. Further, in some instances, the only geographic information a researcher has available is the census tract in which their subjects reside so adopting Census enumeration units is a choice of necessity. Upon stricter scrutiny, however, the notion that Census geographies can serve as proxies for discrete neighborhood delimiters begins to unravel. In reality, residents of a given geographic area (even two neighbors) define the same "neighborhood" location in significantly (and systematically) different ways, and the boundaries they articulate rarely coincide with administrative designations (Hwang, 2016). Furthermore, a person living at the edge of a census tract boundary (e.g. on the street that demarcates one tract from another) is likely to describe their neighborhood as continuous in all directions, rather than demarcated with a hard boundary on one edge.

### Conceptual and Definitional Questions

In his recent book reflecting on decades of neighborhood scholarship, Galster (2019) makes a strong argument that the boundaries of neighborhoods are socially construed and not necessarily consistent among residents of the same location. Furthermore, the meaning and extent of one's neighborhood depends critically on the context in which it is invoked. More specifically, when considering neighborhood effects, the meaning of "neighborhood" depends upon the mechanism through which the effect operates (e.g. participation in a high-quality education system, or cumulative exposure to violence) and the externality space over which that mechanism transmits, where "externality space" corresponds to the effective area over which the effect is detectable (Galster, 2019). For an education system, an externality space would naturally extend to the school district or attendance zone boundaries that sort students into schools, whereas for exposure to violence, the relevant externality space for an adolescent resident may be the walkable area surrounding her home (a much smaller region than the school district boundaries).

Galster's externality space concept contends that neighborhoods take on continuous, overlapping, and inconsistent boundaries. Facially, this implies that studies that reduce neighborhoods to a universal definition based on administrative boundaries (like Census tracts) necessarily misspecify the appropriate externality spaces for one or more geographic resources, and scholarship relying upon these definitions may mistake a statistical artifact as a substantive relationship (or conversely, the lack of an observed association may be taken as evidence that none exists). Instead, however, a true effect may exist but manifest at a different geographic scale. Kwan (2012) refers to this issue as the "uncertain geographic context problem" (UGCoP), and what Galster distinguishes is that the UGCoP applies separately to each potential neighborhood influence.

The externality space concept shares some parallels with accessibility indices developed in the context of urban economics (Hansen, 1959; Huff, 1963; Levinson, 1998) and extended for applications in healthcare (Luo & Wang, 2003; Saxon & Snow, 2020; Wan et al., 2012; Wang & Luo, 2005). These measures assume a set of suppliers like jobs or healthcare providers establish locations in a given metropolitan area, and a set of consumers (e.g. workers or the sick) compete to consume these resources in space, ultimately resulting in an accessibility surface as a function of transportation cost and congestion. Where the concepts of accessibility and externality spaces converge is the notion of distance decay, and the diminishing utility a consumer gains from a resource the further it is located from her. Classic work in job accessibility uses gravity as a model for attractiveness and a reasonable approximation of a decay coefficient (Hansen, 1959), but the externality space concept also recognizes that distance thresholds and decay functions also vary according to the pathway of neighborhood effect and that nonlinear effects may exist (Galster et al., 2000).

### Capturing Activity Space

As the discussion above makes clear, in some cases the extent of "neighborhood" or externality space can be reasonably intuited, such as when it corresponds to officially designated boundaries like school districts. In other cases, however, such as the physical environment that induces individual activity and recreation, the geographic extent of an externality space is ambiguous. Ceteris Paribus, we would expect that shorter distances to the nearest park (i.e. a reduced transportation cost) would result in greater park usage. But parks differ in the amenities they offer, their relative size, their popularity with other neighborhood residents, and perceptions regarding their attractiveness, safety, accessibility, etc (Wang et al., 2015). Finally, these issues may interact with individual characteristics to influence recreation patterns. Wealthier residents with access to a car may be more drawn to larger regional parks with a broader diversity of amenities, whereas younger residents may prefer local, walkable parks where other youths congregate around sports and social activities (and these same characteristics may dissuade older residents who prefer privacy and quiet during their recreation time) (Goličnik & Ward Thompson, 2010). It is also possible that wealthier residents with access to cars may also live close to local parks, and their park activity is larger since they frequent both the local and regional parks. Ignoring the latter may bias upwards the effect of being close to the local park.

Beyond conceptual definitions, the operationalization of adequate externality spaces (for parks, in the context of the present study) requires formalizing a set of geographic locations that provide an amenity (recreation space) and capturing availability to such spaces within each participant's neighborhood. Among the first critical decisions facing a researcher is how to quantify park land. For policy analyses, researchers can rely on officially designated park land, but in behavioral studies where the focus is on recreational space more than designated land use, the question is murkier, and researchers might consider official parks in addition to open space, undeveloped space, or school yards.

After a park has been defined, the next question is whether it should be treated as a point (such as its centroid) or as a polygon, the answer to which depends on two additional considerations. The first is researcher's hypothesis about the effect of parks, (e.g. "an increase in the number of parks is associated with a decrease in BMI," versus "an increase in the amount of parkland is associated with a decrease in BMI"). The second consideration is the analyst's geoprocessing strategy, for example summing the acres of parkland that intersect the participant's buffer, versus summing all acres of land in the park if its centroid falls within the participant's buffer. In the case of network-based analysis, this choice is analogous to whether a participant can access the edge of a park within some specified travel distance, or whether she can reach the center of the park within some travel distance–and depending on the size and shape of the park, these two locations may be quite distant from one another.

The images in Figure 1 and Figure 2, both taken from Metropolitan Denver, help illustrate these issues; park land in the images is shown in dashed lines, tract boundaries are shown in solid (blue) outlines, and each tract's centroid is plotted as a blue circle. In particular, Figure 1 helps show that treating census tracts as containers can be misleading, as sometimes the residents of streets have excellent access to large parks in neighboring tract (literally across the street in some cases) that are ignored in such cases. On the other hand, Figure 2 helps demonstrate that sometimes the tract centroid is difficult to reach by either network or euclidean distance metrics, but neighborhoods inside the tract can have very good access to parks. The networks in Figure 2 also helps show how estimates of the shortest distance to the nearest park can vary substantially if routing along a network or traveling "as the crow flies."

An alternative to measuring park (or park area) density inside a buffer zone, a natural extension includes the use of accessibility indices discussed above (Dony et al., 2015). Accessibility indices are designed to capture the spatial interplay of supply and demand, thus many of the continued innovations in accessibility indices are designed to capture *competition* more adequately. In the context of parks and activity space, however, this may be adjusting in the opposite direction; unlike a job, which can hold only one employee, parks are often social features that draw out additional recreators, such as the case described by Goličnik & Ward Thompson (2010) above. Thus, the appropriate analogue, at least in some cases, is growth or agglomeration as opposed to competition. What this discussion makes clear is that operationalizing "access to parks" in studies focused on its relationship with individual health is riddled with critical, subjective decision-making throughout the analytical pipeline, no guidelines for which have yet been established for most neighborhood influences. As a consequence, we argue there is an obvious need to understand how much this subjectivity can influence reported research results and how much variation we observe in the association between the built environment and individual health outcomes. In the following sections, we present such an analysis.

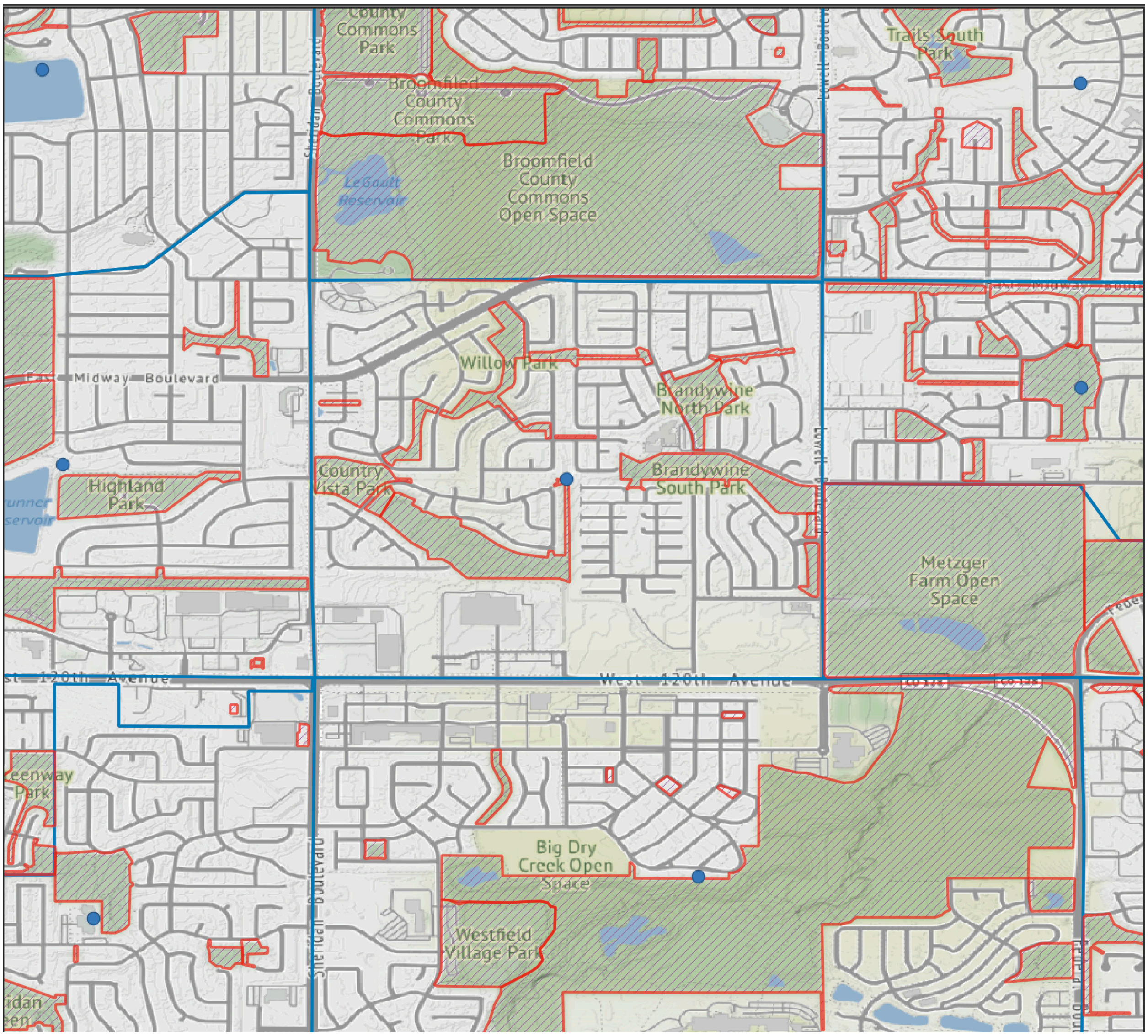


Figure 1: Suburban Park Area (diagonal lines), Tract Boundaries (blue outline) and Tract Centroids (blue point)

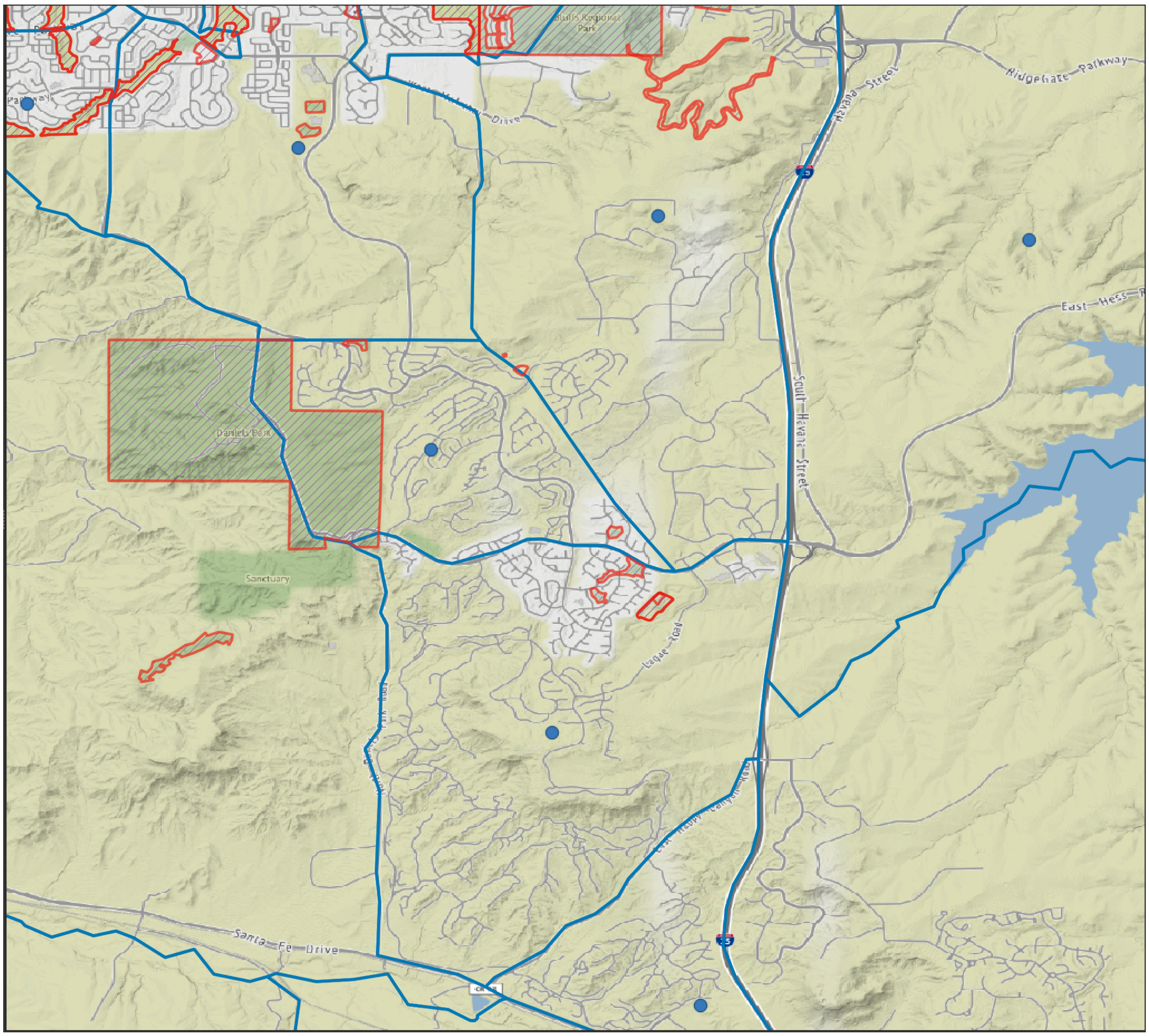

Figure 2: Rural Park Area (diagonal lines), Tract Boundaries (blue outline) and Tract Centroids (blue point)

## SPECIFYING SPATIAL INFLUENCES ON INDIVIDUAL OUTCOMES

We argue there is a clear need to explore how much variation may result from studies that operationalize park access in different ways. Our work is guided by theoretical foundations for the ecological model specification of neighborhood effects, notably Galster (2008) and Galster (2012), the first of which describes the challenges of quantifying neighborhood effects on individual outcomes and the second of which provides a conceptual model specification for doing so. In this paper, our goal is to estimate a model in which individual-level Body Mass Index is associated with the availability of parks in a person's local neighborhood, controlling for additional personal and neighborhood-level covariates, and assess the degree to which variation in the coefficient for "park availability" results from different representations of distance, accessibility, and measurement techniques. We leverage unique data on siblings that participated in a longitudinal study on cognitive and behavioral health, and we measure the way that park access is associated with their BMI outcomes as adults approaching midlife.

### The CATSLife Project

The present study includes the "year 5 sample" of 1278 CATSLife participants tested between July 2015 and February 2020. Participants were recruited from two parent studies, the Colorado Adoption Project (CAP) and the Longitudinal Twin Study (LTS) who have been followed for over 30 years (see Wadsworth et al. (2019)). Multiple outcomes were measured at CATSLife including cognitive and physical functioning and well-being. In the current study, we included participants who have body mass index (BMI) via measured height and weight (N=1224), US address information at the current assessment[1] (N=1251), and sociodeomgraphics (race, ethnicity, and educational attainment; N=1267) for a total analysis sample of 1178 individuals. A summary of variables describing CATSLife participants and the characteristics of their neighborhoods is provided in Table 2. We use the address information for each participant to construct a series of accessibility measures ranging from simple to complex, the first of which are based on a participants street address, the second set of which area based on the census tract in which their address is situated. A summary of park variable measures is provided in the appendix.

### Parks and Recreation Space

For our measures of parks and recreational space, we collect data from OpenStreetMap (OSM), a worldwide, crowdsourced dataset containing spatial and attribute information about physical, social and administrative features across the globe. As a crowdsourced dataset, OSM depends on contributions from interested parties to fill out its data archive, and while this feature means it does not constitute an "official" inventory of parkland from, e.g. local government datasets, OSM is nonetheless a thoroughly comprehensive and robust data platform, particularly for capturing elements of the physical environment (Crooks et al., 2016). OpenStreetMap uses a data structure based on tags and we test two subsets of parkland, the smaller of which contains OSM polygons whose amenity tags include "park,"

[1] We omitted seven individuals for whom geocoding information was not accurate at the address level

Table 1: Summary Statistics

| Variable | N | Mean | Std. Dev. | Min | Pctl. 25 | Pctl. 75 | Max |
|---|---|---|---|---|---|---|---|
| Family Type | 1235 | | | | | | |
| ... Adoptive | 255 | 20.6% | | | | | |
| ... Control | 293 | 23.7% | | | | | |
| ... Dizygotic (Fraternal Twins), | 331 | 26.8% | | | | | |
| ... Monozygotic (Identical Twins) | 356 | 28.8% | | | | | |
| White | 1234 | | | | | | |
| ... No | 97 | 7.9% | | | | | |
| ... Yes | 1137 | 92.1% | | | | | |
| Hispanic | 1234 | | | | | | |
| ... No | 1139 | 92.3% | | | | | |
| ... Yes | 95 | 7.7% | | | | | |
| Sex | 1235 | | | | | | |
| ... Male | 579 | 46.9% | | | | | |
| ... Female | 656 | 53.1% | | | | | |
| Project | 1235 | | | | | | |
| ... CAP | 548 | 44.4% | | | | | |
| ... LTS | 687 | 55.6% | | | | | |
| Participant Age | 1235 | 33.28 | 4.966 | 28.052 | 28.658 | 37.613 | 49.333 |
| Highest Year of Education | 1225 | 16.882 | 2.937 | 11 | 14 | 18 | 22 |
| Neighborhood Population Density (Persons per Square Meter) | 1235 | 0.002 | 0.005 | 0 | 0 | 0.002 | 0.071 |
| Neighborhood % Hispanic/Latino | 1235 | 15.555 | 14.45 | 0 | 6.053 | 19.577 | 83.795 |
| Neighborhood % Black | 1235 | 4.518 | 8.249 | 0 | 0.321 | 4.976 | 85.914 |
| Share of Developed Land in Neighborhood | 1235 | 0.681 | 0.341 | 0 | 0.421 | 0.994 | 1.00 |

Table 2: Descriptive Statistics for Input Variables (Note: All variables Z-transformed in models)

"recreation ground," or "meadow," and the larger contains the prior in addition to "forest" and "nature reserve."

#### Measuring Available Parks

To capture the availability of parks and recreation space available to each participant, we construct 32 separate measures. For each of our definitions of parkland, we first measure the euclidean distance to the nearest park from each participant's last-known address. We then construct euclidean buffers originating from each participant's home in distances of quarter, half, and one mile increments, and count both the number of parks in each buffer and the total acres of parkland within each (we also separately log transform each of these to capture potential nonlinear effects). Following, we construct accessibility variables based on classic urban economic concepts of gravity and attraction potential (El-Geneidy & Levinson, 2006; Hansen, 1959; Kwan et al., 2003; Levinson, 1998) using data from both the street address of each participant and the centroid of the census tract in which they lived, varying both the decay function applied in the access metric and the effective threshold that defines the accessibility surface. A description of the accessibility metrics included in the analysis is shown in Table 3 and a complete description of the results with all combinations we varied is available in Table 5. To compute our measures we use the open-source Python package `access` from the PySAL family of spatial analysis software (Rey et al., 2020; Rey & Anselin, 2010), and we use the open-source `pandana` Python package for calculating shortest path routes through the pedestrian transportation network (Foti et al., 2012).

### Modeling Strategy

To assess the sensitivity of the relationship between BMI and park availability (generally defined), we construct a series of multilevel models using the R package `nlme` (Pinheiro et al., 2018), in which we vary the concept of park availability among our 32 measures described above, holding constant all other terms in the model. Accordingly, we fit the models using the equation

$$Y_{ij} = \alpha + \gamma P_{ij} + \theta N_{ij} + \mu a_{ij} + \eta a_{ij} ses_{ij} + \varepsilon_j + \varepsilon_{ij} \tag{1}$$

with

$$\varepsilon_j \sim \mathcal{N}(0, \mathbf{G})$$

$$\varepsilon_{ij} \sim \mathcal{N}(0, \mathbf{R})$$

and

$$\mathbf{G} = \begin{bmatrix} \sigma^2_{betweenA} & 0 & 0 & 0 \\ 0 & \sigma^2_{betweenC} & 0 & 0 \\ 0 & 0 & \sigma^2_{betweenDZ} & 0 \\ 0 & 0 & 0 & \sigma^2_{betweenMZ} \end{bmatrix}$$

,

$$\mathbf{R} = \begin{bmatrix} \sigma^2_{withinA} & 0 & 0 & 0 \\ 0 & \sigma^2_{withinC} & 0 & 0 \\ 0 & 0 & \sigma^2_{withinDZ} & 0 \\ 0 & 0 & 0 & \sigma^2_{withinMZ} \end{bmatrix}$$

Where $Y_{ij}$ is the BMI for individual $i$ in the $j$ sibling set; $P_{ij}$ is a vector of personal characteristics (e.g., educational attainment, sex, ethnicity); $N_{ij}$ is a vector of characteristics of neighborhood where an individual resides (e.g., tract SES); $a_{ij}$ is a measure of park accessibility in neighborhood where an individual resides; $\eta a_{ij} ses_{ij}$ is a term capturing the interaction between neighborhood park accessibility and neighborhood SES; $\varepsilon_j$ is a random error of between-sibling random effects; $\varepsilon_{ij}$ is a random error of within-sibling random effects; $i$ and $j$ are subscripts for individuals and siblings, respectively.

Our $P$ variables include participant age, sex, race, ethnicity, and educational attainment, and our $N$ variables include race, ethnicity, a factor variable representing socioeconomic status (SES)[2], population density, and the share of the local area considered "developed" by the National Land Cover Database (Wickham et al., 2020) each of which is measured at the census tract level. When we include measures of park accessibility in the model, we include both a main effect and an interaction with SES to capture the potential for perceptions and/or social effects. That is, in higher SES communities proximity to a park may be associated with a recreation amenity, whereas in lower SES communities a park may be associated with nuisance or perceptions of inadequate safety. Furthermore, wealthier neighborhoods may have more better transportation infrastructure or more nuanced architecture that change the attractiveness of local transportation. By using network analysis implicit in

[2]Variables in the factor include median household income, median home value, median contract rent, the share of adults with greater than a Bachelor's degree, poverty rate and unemployment rate

the accessibility metrics, we also capture additional elements of the urban morphology, such as the connectivity and routability of local streets (Boeing, 2018).

| Name | Description | Reference |
|---|---|---|
| Weighted Catchment | Sum of resources within a catchment, weighted by resource provider distance | Hansen 1959 |
| Floating Catchment Area (FCA) | Ratio of providers to clients within a given travel time to the provider | Huff 1963; Wang and Luo 2004 |
| Two-Step FCAs (2SFCA) | Sum of provider-to-client ratio for each provider for each point of origin | Wang and Luo 2005; Luo and Wang 2003 |
| Enhanced 2SFCA (E2SFCA) | 2SFCA with distance decay applied within the catchment area | Luo and Qi 2009 |
| Three-Step FCA (3SFCA) | E2SFCA with distance-based allocation function | Wan, Zhan, Lu et al 2012 |
| Rational Agent Access Model (RAAM) | Weighted minimum travel and congestion cost to all providers within catchment | Saxon and Snow 2020 |
| Join Count | Sum of distinct parks within a specified radius | N/A |
| Distance to Nearest | Distance along the shortest path from an origin (e.g. tract centroid or home address) to a destination | N/A |

Table 3: Description of Park Accessibility Measures

We fit each model adjusting for sibling dependencies by estimating random effects by family/sibling type given variation in genetic relatedness: i.e., siblings from adoptive (A) or non-adoptive control (C) families, (DZ) fraternal or dizygotic twins, or (MZ) identical or monozygotic twins. The variance of the between-sibling random effects represents similarity among siblings in BMI by sibling type $\sigma^2_{betweenA}$ $\sigma^2_{betweenC}$, $\sigma^2_{betweenDZ}$, $\sigma^2_{betweenMZ}$ The corresponding variance of within-sibling random effects represents differences in BMI among siblings $\sigma^2_{withinA}$ ,$\sigma^2_{withinC}$, $\sigma^2_{withinDZ}$, $\sigma^2_{withinMZ}$). We place no constraints on the magnitudes of random effects estimated by sibling types. Notably, the model we estimate in Equation 1 differs from the ideal specification provided by Galster (2008) in a few ways. First, we do not include time-varying personal characteristics or metropolitan-wide characteristics, and we do not attempt to account for unobserved personal characteristics (i.e. we omit the $UP_t$, $P_t$ and $M_t$ terms). In the present context, the omission of these variables is not a serious concern, since our goal is not to estimate the ideal model of neighborhood effects, but simply to explore variance in the association of a single category of neighborhood influences, and the variables we include in the model are sufficient controls for doing so.

## RESULTS & DISCUSSION

A summary of the base model is provided in Table 4 and coefficients for our battery of park accessibility measures is shown in Table 5 with all coefficients expressed in $z-$standardized units. At the individual level, our model shows that being female and additional years of education are associated with lower BMIs in early adulthood. At the neighborhood level, *none* of the variables are significant in the base model. As we iterate through concepts of park access, however, an interesting picture emerges. The only two park access variables statistically significant at the conventional $p = 0.05$ level are distance to the nearest park (at the address level) for both large and small sets of "park" land. All coefficients are in the expected direction with greater access to recreation space associated with decreases in BMI. Perhaps more importantly in the context of the present study, there is wide variation in both the magnitude of the estimated coefficient and the statistical significance thereof across the range of park accessibility metrics (despite the fact that the signs for every single metric are consistent with extant theory). As we describe at the outset, each of the metrics provided in Table 5 either has been used previously in the literature or could be justified on a rational basis, but selecting a single index among the subset could have a dramatic influence on the associated inference.

Prime among these findings is that we have significant variance across nearly every one of the dimensions we tested. This speaks to the importance of model specification and careful attention to the hypothesized pathway through which a neighborhood effect is expected to operate, as well as the importance of spatial scale in both data collection and analysis. According to our results, researchers lacking access to address-level geographic information about their participants will be unlikely to identify an association between park land and individual health, even when one exists because every one of our tract-based measures failed to meet significance.

A second important finding is the best evidence of an association between recreation space and personal health we find comes from shorter distances to the nearest park location. This speaks to both the way parks may be conceived and operationalized in scholarship on associations with physical health, as well as the rapidly diminishing utility of greater access to park land in one's neighborhood. Our models suggest that, after a person can quickly reach their nearest park, having *more* parks at their disposal or more total acres of park land within their catchment zone, there is no additional association with BMI. This suggests that people gain just as much benefit from a small, local "pocket" park as a large regional nature preserve and could have important implications for urban planning practice. Further, it means that, in the context of parkland specifically, the additional computational cost and labor burden required to construct more complex accessibility indices may provide little additional benefit since they were found to have essentially no association with BMI in this study[3].

For those models whose main effects of distance-based park access are significant, the interaction with our neighborhood SES factor variable is also significant and positive for each model, highlighting the potential importance of high-quality amenities at park locations or the social circumscription of park usage. Further, in some models where the main effect of park access was not significant, the interaction between neighborhood SES and park access did remain significant. Put differently, the benefits of having a nearby park (with respect to BMI) are amplified if the park is located in a well-to-do neighborhood. Because we lack an individual-level variable measuring socioeconomic status, and neighborhoods tend to be economically homogenous, this could also be capturing a greater willingness among higher SES participants to use local amenities. The fact that the main effect of SES is never significant in the models suggests that it is not simply a matter of higher status individuals or higher quality parks that lead to lower BMI.

[3] We note this finding with a large grain of salt, since our main finding is that important associations can go missed in this context if an inapplicable metric is applied, so we maintain that good strategy is to test several alternative specifications.

Table 4: Base Model Results

| | Base Model Results |
|---|---|
| Intercept | 3.322 *** (0.034) |
| Project LTS | -0.018 (0.026) |
| White | -0.017 (0.031) |
| Hispanic/Latino | 0.013 (0.033) |
| Years of Education | -0.027 *** (0.006) |
| Age | 0.025 * (0.013) |
| Sex (Female) | -0.044 *** (0.013) |
| Share of Hispanic/Latino Residents | 0.001 (0.006) |
| Share of Black (non-Hispanic/Latino) Residents | 0.007 (0.005) |
| Population Density | -0.005 (0.006) |
| Share of Developed Land | -0.009 (0.007) |
| Socioeconomic Status | -0.005 (0.006) |
| $\sigma^2_{withinA}$ | 0.042 (NA) |
| $\sigma^2_{withinC}$ | 0.031 (NA) |
| $\sigma^2_{withinDZ}$ | 0.031 (NA) |
| $\sigma^2_{withinMZ}$ | 0.013 (NA) |
| $\sigma^2_{betweenA}$ | 0.060 (NA) |
| $\sigma^2_{betweenC}$ | 0.088 (NA) |
| $\sigma^2_{betweenDZ}$ | 0.100 (NA) |
| $\sigma^2_{betweenMZ}$ | 0.167 (NA) |
| $N$ | 1178 |
| $logLikelihood$ | 259.340 |
| $AIC$ | -466.681 |

*** p < 0.001; ** p < 0.01; * p < 0.05.

| Measure | Quantity | Distance Metric | Travel Distance | Origin | Decay Function | Park Set | Estimate | Std Error | Statistic | p-Value |
|---|---|---|---|---|---|---|---|---|---|---|
| Distance | Miles to nearest | Euclidian | Minimum | Address | None | Small | 0.0185 | 0.0083 | 2.2326 | 0.026 |
| Distance | Miles to nearest | Euclidian | Minimum | Address | None | Large | 0.0201 | 0.0088 | 2.2917 | 0.0224 |
| Distance | Log miles to nearest | Euclidian | Minimum | Address | None | Small | 0.0089 | 0.0069 | 1.2958 | 0.1957 |
| Distance | Log miles to nearest | Euclidian | Minimum | Address | None | Large | 0.0122 | 0.0067 | 1.8223 | 0.069 |
| FCA | Park Count | Euclidian | One Mile | Address | None | Small | -0.0036 | 0.0061 | -0.584 | 0.5595 |
| FCA | Park Count | Euclidian | Half Mile | Address | None | Small | -0.0081 | 0.006 | -1.36 | 0.1745 |
| FCA | Park Count | Euclidian | Quarter Mile | Address | None | Small | -0.0055 | 0.0057 | -0.964 | 0.3355 |
| FCA | Park Count | Euclidian | One Mile | Address | None | Large | 0.0009 | 0.0062 | 0.14 | 0.8887 |
| FCA | Park Count | Euclidian | Half Mile | Address | None | Large | -0.0038 | 0.0059 | -0.6402 | 0.5223 |
| FCA | Park Count | Euclidian | Quarter Mile | Address | None | Large | -0.0026 | 0.0057 | -0.4661 | 0.6414 |
| FCA | Park Count | Euclidian | One Mile | Address | None | Small | -0.0039 | 0.007 | -0.5476 | 0.5842 |
| FCA | Park Count | Euclidian | Half Mile | Address | None | Small | -0.0051 | 0.0063 | -0.8124 | 0.417 |
| FCA | Park Count | Euclidian | Quarter Mile | Address | None | Small | -0.0008 | 0.006 | -0.136 | 0.8919 |
| FCA | Park Count | Euclidian | One Mile | Address | None | Large | -0.0025 | 0.0069 | -0.361 | 0.7182 |
| FCA | Park Count | Euclidian | Half Mile | Address | None | Large | -0.0047 | 0.0062 | -0.7478 | 0.4549 |
| FCA | Park Count | Euclidian | Quarter Mile | Address | None | Large | -0.0002 | 0.0059 | -0.0286 | 0.9772 |
| RAAM | Park Acres | Network | 30 Min | Tract Centroid | None | Large | -0.0137 | 0.007 | -1.955 | 0.0512 |
| 2SFCA | Park Acres | Network | 30 Min | Tract Centroid | Step Function | Large | -0.0001 | 0.0574 | -0.0024 | 0.9981 |
| E2SFCA | Park Acres | Network | 30 Min | Tract Centroid | Gaussian | Large | -0.0001 | 0.0058 | -0.0144 | 0.9885 |
| 3SFCA | Park Acres | Network | 30 Min | Tract Centroid | Step Function | Large | -0.0001 | 0.0574 | -0.0024 | 0.9981 |
| FCA | Park Acres | Network | 30 Min | Tract Centroid | None | Large | -0.0001 | 0.0574 | -0.0024 | 0.9981 |
| FCA | Park Acres | Network | 60 Min | Tract Centroid | None | Large | -0.0001 | 0.0574 | -0.0024 | 0.9981 |
| RAAM | Park Acres | Network | 30 Min | Tract Centroid | None | Small | -0.0051 | 0.0073 | -0.7016 | 0.4833 |
| E2SFCA | Park Acres | Network | 30 Min | Tract Centroid | Step Function | Small | -0.2205 | 0.1411 | -1.5625 | 0.1189 |
| E2SFCA | Park Acres | Network | 30 Min | Tract Centroid | Gaussian | Small | -0.0111 | 0.0076 | -1.4621 | 0.1444 |
| 3SFCA | Park Acres | Network | 30 Min | Tract Centroid | Step Function | Small | -0.2205 | 0.1411 | -1.5625 | 0.1189 |
| FCA Ratio | Park Acres | Network | 30 Min | Tract Centroid | None | Small | -0.2205 | 0.1411 | -1.5625 | 0.1189 |
| FCA Ratio | Park Acres | Network | 60 Min | Tract Centroid | None | Small | -0.2205 | 0.1411 | -1.5625 | 0.1189 |
| Weighted Catchment | Park Acres | Network | 2000km | Address | Linear | Small | -0.0031 | 0.0061 | -0.5104 | 0.61 |
| Weighted Catchment | Park Acres | Network | 2000km | Address | Linear | Small | -0.0056 | 0.0063 | -0.8905 | 0.3736 |
| Weighted Catchment | Park Acres | Network | 2000km | Address | Linear | Large | -0.0086 | 0.006 | -1.4276 | 0.1541 |
| Weighted Catchment | Park Acres | Network | 2000km | Address | Linear | Large | -0.005 | 0.0063 | -0.7905 | 0.4296 |

Table 5: Model Results for Park Access Variables

One possible reason we find no evidence of an association between most accessibility variables and BMI (and evidence on the subject has been mixed in the literature to date) is that most accessibility indices place a heavy emphasis on accounting for nearby competition, since they were developed under the precepts of economic equilibrium. Park land, on the other hand is closer to a public good, in that consumption by one person does not inhibit consumption by another. We say that it is "closer" to a public good because park land is still finite and parks have practical limits on occupancy–but this distinction seems to matter little according to our results. This result might suggest that the externality space for local parks is highly localized, extending only a short distance around a person's residential address, and lends new ways about thinking of quantifying the equitable distribution of urban park land (or, the optimal allocation for influencing public health) Especially given that other features such as population density, a proxy for compact development hypothesized by urban planning scholars and Smart Growth advocates to increase physical activity, had no discernible association with BMI.

To date, the vast majority of the literature on green space (however defined) and physical health reports from observational studies, and ours is no exception. This is understandable, since experimental designs that assign children to place of residence are exceedingly rare, with the best known instances being the Moving to Opportunity study from the 1990s, and other quasi-experimental designs such as discussed by Lucero et al. (2018) and Galster & Santiago (2017). But this condition also limits full understanding of the relationship between park access and physical health. Absent these more rigorous research designs, causal attribution remains impossible, and any observed associations may result from a desire among those with lower BMIs to live nearer to parks (as opposed to the park exhorting its own force to induce recreation and lower BMI) (Burdick-Will et al., 2010; Ludwig et al., 2008). We are cognizant of such limitations here and although we proceed from a plausible theoretical pathway in which access to parks stimulates physical activity, in turn lowering BMI, our work here cannot establish the causality of such a path.

## CONCLUSION

In this paper we cast an exploratory lens on the relationship between neighborhood physical conditions and individual health outcomes measured by BMI. Using data from the CATSLife project, we demonstrate the powerful roles of spatial scale, operationalization, and geographic representation in research examining the association between park access and BMI. Specifically, we show that there is wide variance in both the significance and estimated association between park access and BMI, depending on the particular concept of "park access" used in the model. Our results show that a person's local environment can have a clear, statistically significant relationship with their physical health, but that discovering the association requires careful attention to model specification. Specifically, we find that a person's BMI is associated with accessibility to parks in their local neighborhood, but that the association only manifests at small geographic scales, and does not scale with additional parks or park acreage. Specifically, we find a significant but relatively small association between individual-level BMI and the distance to the nearest park and or recreational space (regardless of whether "parks and recreational space" includes formal park land or includes other types of green space such as forests), but we find no association between BMI and more complicated accessibility metrics, nor does the effect scale with the area or cumulative count of parks or recreation space.

For non-geographers studying associations between neighborhood context and individual outcomes, idiosyncratic decisions about geographic representation, geoprocessing, and spatial analysis can appear inconsequential, but here we demonstrate that they can profoundly influence both the inference and importance associated with features in the local environment. Beyond specification issues, our results also speak to the importance of high-resolution spatial data in the practice of estimating neighborhood effects. Whereas most scholarship to date uses coarse neighborhood data such as census tracts, CATSLife data include detailed information about residential street addresses, a feature that becomes critical, since the only association we observe requires the use of highly-localized data. We recognize that often the best data available to researchers comes at large spatial scales, and for certain variables, these scales may well represent the effective externality space. Regardless, our results make clear that spatial analyses require full transparency, and great caution should be exercised interpreting modeling exercises where the full pipeline of these decisions is not made explicit. Further, our results stress the value and continued importance of interdisciplinary collaboration in scholarship focused on urban sociospatial phenomena.

In future work we plan to pursue two extensions focused on longitudinal dynamics and neighborhood selection, respectively. Toward the first end, we are interested in the ways that *changing* access to recreational space and other neighborhood resources may affect development over the long term. Since the CATSLife project provides individual-level data over several decades, we can examine multiple measurements of BMI for each of our participants over time, but a lingering data challenge is the collection of park land data. OpenStreetMap is an evolving database that improves over time, but it is also designed to reflect the changing conditions on the ground. While it is possible to extract data from a particular snapshot in time, we cannot collect information about neighborhood conditions in the 1980s and 1990s when OSM did not exist. Urban development changes relatively slowly and parks are durable investments, so it is possible that using current data would provide a reasonable approximation. Toward the second end we are interested in exploring the relationship between neighborhood selection and neighborhood effects. Leveraging the longitudinal nature of CATSLife again, we can reasonably extract certain critical elements of the life course (such as a person moving into their own home for the first time) and examine for example whether features like park space influence the residential location they ultimately select.

## REFERENCES

Bader, M. D. M., Mooney, S. J., Lee, Y. J., Sheehan, D., Neckerman, K. M., Rundle, A. G., & Teitler, J. O. (2015). Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health & Place*, *31*, 163–172. https://doi.org/10.1016/j.healthplace.2014.10.012

Barbosa, O., Tratalos, J. A., Armsworth, P. R., Davies, R. G., Fuller, R. A., Johnson, P., & Gaston, K. J. (2007). Who benefits from access to green space? A case study from Sheffield, UK. *Landscape and Urban Planning*, *83*(2-3), 187–195. https://doi.org/10.1016/j.landurbplan.2007.04.004

Boeing, G. (2018). A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science*, 239980831878459. https://doi.org/10.1177/2399808318784595

Burdick-Will, J. A., Ludwig, J., Raudenbush, S. W., Sampson, R. J., Sanbonmatsu, L., & Sharkey, P. T. (2010). Converging Evidence for Neighborhood Effects on Children's Test Scores: An Experimental, Quasi- Experimental, and Observational Comparison. In *Whither opportunity?: Rising inequality, schools, and children's life chances* (pp. 255–276).

Carthy, P., Lyons, S., & Nolan, A. (2020). Characterising urban green space density and footpath-accessibility in models of BMI. *BMC Public Health*, *20*(1), 1–12. https://doi.org/10.1186/s12889-020-08853-9

Casey, R., Oppert, J. M., Weber, C., Charreire, H., Salze, P., Badariotti, D., Banos, A., Fischler, C., Hernandez, C. G., Chaix, B., & Simon, C. (2014). Determinants of childhood obesity: What can we learn from built environment studies? *Food Quality and Preference*, *31*(1), 164–172. https://doi.org/10.1016/j.foodqual.2011.06.003

Cohen, D. A., Inagami, S., & Finch, B. (2008). The built environment and collective efficacy. *Health & Place*, *14*(2), 198–208. https://doi.org/10.1016/j.healthplace.2007.06.001

Crooks, A. T., Croitoru, A., Jenkins, A., Mahabir, R., Agouris, P., & Stefanidis, A. (2016). User-Generated Big Data and Urban Morphology. *Built Environment*, *42*(3), 396–414. https://doi.org/10.2148/benv.42.3.396

Diez Roux, A. V. (2001). Investigating Neighborhood and Area Effects on Health. *American Journal of Public Health*, *91*(11), 1783–1789. https://doi.org/10.2105/AJPH.91.11.1783

Dony, C. C., Delmelle, E. M., & Delmelle, E. C. (2015). Re-conceptualizing accessibility to parks in multimodal cities: A Variable-width Floating Catchment Area (VFCA) method. *Landscape and Urban Planning*, *143*, 90–99. https://doi.org/10.1016/j.landurbplan.2015.06.011

Duncan, G. J., & Raudenbush, S. W. (2001). Neighborhoods and Adolescent Development:How Can We Determine the Links? In A. Booth & N. Crouter (Eds.), *Does it take a village? Community effects on children, adolescents, and families* (pp. 105–136). Pennsylvania State University Press. https://doi.org/10.1.1.591.7549

El-Geneidy, A. M., & Levinson, D. M. (2006). *Access to Destinations: Development of Accessibility Measures* (p. 125). http://www.lrrb.org/PDF/200616.pdf

Ellen, I. G., Mijanovich, T., & Dillman, K.-N. (2001). Neighborhood Effects on Health: Exploring the Links and Assessing the Evidence. *Journal of Urban Affairs*, *23*(3-4), 391–408. https://doi.org/10.1111/0735-2166.00096

Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., & Raudenbush, S. (2003). Relationship between Urban Sprawl and Physical Activity, Obesity, and Morbidity. *American Journal of Health Promotion*, *18*(1), 47–57. https://doi.org/10.4278/0890-1171-18.1.47

Foti, F., Waddell, P., & Luxen, D. (2012). A Generalized Computational Framework for Accessibility : From the Pedestrian to the Metropolitan Scale. *4th Transportation Research Board Conference on Innovations in Travel Modeling (ITM)*, 1–14. http://onlinepubs.trb.org/onlinepubs/conferences/2012/4thITM/Papers-A/0117-000062.pdf

Galster, G. (2001). On the nature of neighbourhood. *Urban Studies*, *38*(12), 2111. https://doi.org/10.1080/00420980120087072

Galster, G. (2003). Investigating behavioural impacts of poor neighbourhoods: towards new data and analytic strategies. *Housing Studies*, *18*(6), 893–914. https://doi.org/10.1080/0267303032000135

492
Galster, G. (2012). The mechanism (s) of neighbourhood effects: Theory, evidence, and policy implications. In M. van Ham, D. Manley, N. Bailey, L. Simpson, & D. Maclennan (Eds.), *Neighbourhood effects research: New perspectives* (pp. 23–56). Springer.
Galster, G. C. (2008). Quantifying the Effect of Neighbourhood on Individuals: Challenges, Alternative Approaches, and Promising Directions. *Schmollers Jahrbuch*, *128*(1), 7–48. https://doi.org/10.3790/schm.128.1.7
Galster, G. C. (2019). *Making our neighborhoods, making our selves*. University of Chicago Press.
Galster, G. C., Quercia, R. G., & Cortes, A. (2000). Identifying neighborhood thresholds: An empirical exploration. *Housing Policy Debate*, *11*(3), 701–732. https://doi.org/10.1080/10511482.2000.9521383
Galster, G., & Santiago, A. M. (2017). Do Neighborhood Effects on Low-Income Minority Children Depend on Their Age? Evidence From a Public Housing Natural Experiment. *Housing Policy Debate*, *27*(4), 584–610. https://doi.org/10.1080/10511482.2016.1254098
Goličnik, B., & Ward Thompson, C. (2010). Emerging relationships between design and use of urban park spaces. *Landscape and Urban Planning*, *94*(1), 38–53. https://doi.org/10.1016/j.landurbplan.2009.07.016
Grubesic, T. H. (2008). Zip codes and spatial analysis: Problems and prospects. *Socio-Economic Planning Sciences*, *42*(2), 129–149. https://doi.org/10.1016/j.seps.2006.09.001
Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, *25*(2), 73–76. https://doi.org/10.1080/01944365908978307
Huff, D. L. (1963). A Probabilistic Analysis of Shopping Center Trade Areas. *Land Economics*, *39*(1), 81. https://doi.org/10.2307/3144521
Hwang, J. (2016). The Social Construction of a Gentrifying Neighborhood. *Urban Affairs Review*, *52*(1), 98–128. https://doi.org/10.1177/1078087415570643
Joseph, R. P., & Maddock, J. E. (2016). Observational Park-based physical activity studies: A systematic review of the literature. *Preventive Medicine*, *89*, 257–277. https://doi.org/10.1016/j.ypmed.2016.06.016
Kaczynski, A. T., Potwarka, L. R., & Saelens, B. E. (2008). Association of Park Size, Distance, and Features With Physical Activity in Neighborhood Parks. *American Journal of Public Health*, *98*(8), 1451–1456. https://doi.org/10.2105/AJPH.2007.129064
Kaczynski, A. T., Potwarka, L. R., Smale, B. J. A., & Havitz, M. E. (2009). Association of Parkland Proximity with Neighborhood and Park-based Physical Activity: Variations by Gender and Age. *Leisure Sciences*, *31*(2), 174–191. https://doi.org/10.1080/01490400802686045
Kane, K., & Kim, Y. (2018). Parcels, points, and proximity: Can exhaustive sources of big data improve measurement in cities? *Environment and Planning B: Urban Analytics and City Science*, 239980831879713. https://doi.org/10.1177/2399808318797135
Kwan, M. P. (2012). The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers*, *102*(5), 958–968. https://doi.org/10.1080/00045608.2012.687349
Kwan, M.-P., Murray, A. T., O'Kelly, M. E., & Tiefelsdorf, M. (2003). Recent advances in accessibility research: Representation, methodology and applications. *Journal of Geographical Systems*, *5*(1), 129–138. https://doi.org/10.1007/s101090300107

Levinson, D. M. (1998). Accessibility and the journey to work. *Journal of Transport Geography*, *6*(1), 11–21. https://doi.org/10.1016/S0966-6923(97)00036-7

Lin, B. B., Fuller, R. A., Bush, R., Gaston, K. J., & Shanahan, D. F. (2014). Opportunity or Orientation? Who Uses Urban Parks and Why. *PLoS ONE*, *9*(1), e87422. https://doi.org/10.1371/journal.pone.0087422

Lucero, J., Santiago, A., & Galster, G. (2018). How Neighborhood Effects Vary: Childbearing and Fathering among Latino and African American Adolescents. *Healthcare*, *6*(1), 7. https://doi.org/10.3390/healthcare6010007

Ludwig, J., Liebman, J. B. B. B., Kling, J. R. R. R., Duncan, G. J. J. J., Katz, L. F. F. F., Kessler, R. C. C. C., & Sanbonmatsu, L. (2008). What Can We Learn about Neighborhood Effects from the Moving to Opportunity Experiment? *American Journal of Sociology*, *114*(1), 144–188. https://doi.org/10.1086/588741

Luo, W. (2004). Using a GIS-based floating catchment method to assess areas with shortage of physicians. *Health & Place*, *10*(1), 1–11. https://doi.org/10.1016/S1353-8292(02)00067-9

Luo, W., & Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, *15*(4), 1100–1107. https://doi.org/10.1016/j.healthplace.2009.06.002

Luo, W., & Wang, F. (2003). Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region. *Environment and Planning B: Planning and Design*, *30*(6), 865–884. https://doi.org/10.1068/b29120

Minh, A., Muhajarine, N., Janus, M., Brownell, M., & Guhn, M. (2017). A review of neighborhood effects and early child development: How, where, and for whom, do neighborhoods matter? *Health & Place*, *46*(November 2016), 155–174. https://doi.org/10.1016/j.healthplace.2017.04.012

Moore, S., Gauvin, L., Daniel, M., Kestens, Y., Bockenholt, U., Dubé, L., & Richard, L. (2010). Associations among Park Use, Age, Social Participation, and Neighborhood Age Composition in Montreal. *Leisure Sciences*, *32*(4), 318–336. https://doi.org/10.1080/01490400.2010.488193

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). *nlme: Linear and Nonlinear Mixed Effects Models*. https://cran.r-project.org/package=nlme

Raudenbush, S. W. (2003). The Quantitative Assessment of Neighborhood Social Environments. In *Neighborhoods and health* (pp. 112–131). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195138382.003.0005

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods. *Sociological Methodology*, *29*(1), 1–41. https://doi.org/10.1111/0081-1750.00059

Rey, S., Anselin, L., Amaral, P., Arribas-Bel, D., Cortes, R., Gaboardi, J., Kang, W., Knaap, E., Li, Z., Lumnitz, S., Oshan, T., Shao, H., & Wolf, L. (2020). The PySAL ecosystem: philosophy and implementation. *Geographical Analysis*.

Rey, S. J., & Anselin, L. (2010). PySAL: A Python Library of Spatial Analytical Methods. In *Handbook of applied spatial analysis* (Vol. 37, pp. 175–193). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03647-7_11

Sallis, J., Owen, N., & Fisher, E. (2008). Ecological Models of Health Behavior In Health Behavior and Health Education. In *Health education & behavior* (pp. 465–486).

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*(1), 443–478. https://doi.org/10.1146/annurev.soc.28.110601.141114

Sampson, R. J., Morenoff, J. D., & Raudenbush, S. (2005). Social anatomy of racial and ethnic disparities in violence. *American Journal of Public Health*, *95*(2), 224–232. https://doi.org/10.2105/AJPH.2004.037705

Sampson, R. J., Sharkey, P., & Raudenbush, S. W. (2008). Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(3), 845–852. https://doi.org/10.1073/pnas.0710189104

Saxon, J., & Snow, D. (2020). A Rational Agent Model for the Spatial Accessibility of Primary Health Care. *Annals of the American Association of Geographers*, *110*(1), 205–222. https://doi.org/10.1080/24694452.2019.1629870

Stark, J. H., Neckerman, K., Lovasi, G. S., Quinn, J., Weiss, C. C., Bader, M. D. M., Konty, K., Harris, T. G., & Rundle, A. (2014). The impact of neighborhood park access and quality on body mass index among adults in New York City. *Preventive Medicine*, *64*, 63–68. https://doi.org/10.1016/j.ypmed.2014.03.026

Talen, E. (2010). The Spatial Logic of Parks. *Journal of Urban Design*, *15*(4), 473–491. https://doi.org/10.1080/13574809.2010.502335

Talen, E., & Anselin, L. (1998). Assessing Spatial Equity: An Evaluation of Measures of Accessibility to Public Playgrounds. *Environment and Planning A*, *30*(4), 595–613. https://doi.org/10.1068/a300595

Tilt, J. H., Unfried, T. M., & Roca, B. (2007). Using objective and subjective measures of neighborhood greenness and accessible destinations for understanding walking trips and BMI in Seattle, Washington. *American Journal of Health Promotion*, *21*(4 SUPPL.), 371–379. https://doi.org/10.4278/0890-1171-21.4s.371

Wadsworth, S. J., Corley, R. P., Munoz, E., Trubenstein, B. P., Knaap, E., DeFries, J. C., Plomin, R., & Reynolds, C. A. (2019). CATSLife: A Study of Lifespan Behavioral Development and Cognitive Functioning – Corrigendum. *Twin Research and Human Genetics*, *22*(6), 829–829. https://doi.org/10.1017/thg.2019.126

Wan, N., Zou, B., & Sternberg, T. (2012). A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science*, *26*(6), 1073–1089. https://doi.org/10.1080/13658816.2011.624987

Wang, D., Brown, G., Liu, Y., & Mateo-Babiano, I. (2015). A comparison of perceived and geographic access to predict urban park use. *Cities*, *42*(PA), 85–96. https://doi.org/10.1016/j.cities.2014.10.003

Wang, F., & Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health and Place*, *11*(2), 131–146. https://doi.org/10.1016/j.healthplace.2004.02.003

Wen, M., & Kowaleski-Jones, L. (2012). The built environment and risk of obesity in the United States: racial-ethnic disparities. *Health & Place*, *18*(6), 1314–1322. https://doi.org/10.1016/j.healthplace.2012.09.002

Wickham, J., Stehman, S. V., Neale, A. C., & Mehaffey, M. (2020). Accuracy assessment of NLCD 2011 percent impervious cover for selected USA metropolitan areas. *International Journal of Applied Earth*

*Observation and Geoinformation*, *84*(December 2018), 101955. https://doi.org/10.1016/j.jag.2019.101955

Wilson, W. J. (1987). *The Truly Disadvantaged: The Inner City, The Underclass, and Public Policy*. University of Chicago Press.

Witten, K., Hiscock, R., Pearce, J., & Blakely, T. (2008). Neighbourhood access to open spaces and the physical activity of residents: A national study. *Preventive Medicine*, *47*(3), 299–303. https://doi.org/10.1016/j.ypmed.2008.04.010

Wolch, J., Jerrett, M., Reynolds, K., McConnell, R., Chang, R., Dahmann, N., Brady, K., Gilliland, F., Su, J. G., & Berhane, K. (2011). Childhood obesity and proximity to urban parks and recreational resources: A longitudinal cohort study. *Health and Place*, *17*(1), 207–214. https://doi.org/10.1016/j.healthplace.2010.10.001

Wolch, J. R., Byrne, J., & Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning*, *125*, 234–244. https://doi.org/10.1016/j.landurbplan.2014.01.017

## APPENDIX

Table 6: Summary Statistics

| Variable | N | Mean | Std. Dev. | Min | Pctl. 25 | Pctl. 75 | Max |
|---|---|---|---|---|---|---|---|
| Parks3f_One_Mile_JCount | 1235 | 13.461 | 16.105 | 0 | 3 | 18 | 142 |
| Parks3f_Half_Mile_JCount | 1235 | 4.275 | 5.644 | 0 | 1 | 6 | 72 |
| Parks3f_Quarter_Mile_JCount | 1235 | 1.437 | 2.038 | 0 | 0 | 2 | 25 |
| Parks5f_One_Mile_JCount | 1235 | 16.606 | 18.901 | 0 | 5 | 22 | 165 |
| Parks5f_Half_Mile_JCount | 1235 | 5.047 | 6.425 | 0 | 1 | 7 | 73 |
| Parks5f_Quarter_Mile_JCount | 1235 | 1.65 | 2.24 | 0 | 0 | 2 | 25 |
| ln_Parks3f_One_JCount | 1235 | 2.131 | 1.129 | 0 | 1.386 | 2.944 | 4.963 |
| ln_Parks3f_Half_JCount | 1235 | 1.27 | 0.886 | 0 | 0.693 | 1.946 | 4.29 |
| ln_Parks3f_Quart_JCount | 1235 | 0.655 | 0.65 | 0 | 0 | 1.099 | 3.258 |
| ln_Parks5f_One_JCount | 1235 | 2.343 | 1.127 | 0 | 1.792 | 3.135 | 5.112 |
| ln_Parks5f_Half_JCount | 1235 | 1.402 | 0.899 | 0 | 0.693 | 2.079 | 4.304 |
| ln_Parks5f_Quart_JCount | 1235 | 0.726 | 0.673 | 0 | 0 | 1.099 | 3.258 |
| Closest_Park3f_mi | 1235 | 0.561 | 1.526 | 0 | 0.091 | 0.42 | 23.95 |
| Closest_Park5f_mi | 1235 | 0.458 | 1.27 | 0 | 0.083 | 0.357 | 22.72 |
| ln_Closest_Park3f_mi | 1235 | 0.314 | 0.402 | 0 | 0.087 | 0.351 | 3.217 |
| ln_Closest_Park5f_mi | 1235 | 0.275 | 0.355 | 0 | 0.08 | 0.305 | 3.166 |
| raam_area_lg | 1219 | 0.488 | 2.375 | 0 | 0.005 | 0.031 | 31.932 |
| X_2sfca_area_lg | 1219 | 7265268.754 | 194294945.328 | 0 | 22.716 | 11520.72 | 6760335209 |
| g2sfca_area_lg | 1219 | 3202886.916 | 20029718.438 | 0 | 0 | 571608.793 | 333965189.3 |
| X_3sfca_area_lg | 1219 | 6989188.541 | 186911737.403 | 0 | 21.853 | 11082.932 | 6503442471 |
| fca60_area_lg | 1219 | 7265268.754 | 194294945.328 | 0 | 22.716 | 11520.72 | 6760335209 |
| fca120_area_lg | 1219 | 7265268.754 | 194294945.328 | 0 | 22.716 | 11520.72 | 6760335209 |
| raam_area_sm | 1219 | 0.427 | 2.314 | 0 | 0.006 | 0.047 | 40.199 |
| X_2sfca_area_sm | 1219 | 2132601.637 | 64595382.922 | 0 | 7.684 | 1307.423 | 2253294615 |
| g2sfca_area_sm | 1219 | 468079.587 | 3392462.444 | 0 | 0 | 68246.216 | 51114395.97 |
| X_3sfca_area_sm | 1219 | 2051562.775 | 62140758.381 | 0 | 7.392 | 1257.741 | 2167669420 |
| fca60_area_sm | 1219 | 2132601.637 | 64595382.922 | 0 | 7.684 | 1307.423 | 2253294615 |
| fca120_area_sm | 1219 | 2132601.637 | 64595382.922 | 0 | 7.684 | 1307.423 | 2253294615 |
| park_area_3fc | 1221 | 819698.893 | 1314240.858 | 0 | 84213.236 | 1081622.845 | 21456744.9 |
| park_count_3fc | 1221 | 17.406 | 22.033 | 0 | 2.932 | 22.953 | 186.701 |
| park_area_5fc | 1221 | 990490.617 | 1448147.337 | 0 | 126889.603 | 1307339.056 | 21560130.44 |
| park_count_5fc | 1221 | 22.166 | 27.614 | 0 | 4.164 | 31.36 | 292.862 |

**Distance Models**

*Model Results – Dependent Variable: Three Feature*

| | (1) |
|---|---|
| Intercept | 3.322 *** (0.034) |
| Project | -0.014 (0.026) |
| White | -0.016 (0.031) |
| Hispanic or Latino | 0.010 (0.033) |
| Years of Education | -0.027 *** (0.006) |
| Age | 0.029 * (0.013) |
| Sex | -0.045 *** (0.013) |
| Share of Hispanic/Latino Residents | 0.001 (0.006) |
| Share of Black (non-Hispanic/Latino) Residents | 0.007 (0.005) |
| Population Density | -0.005 (0.006) |
| Share of Developed Land | -0.002 (0.007) |
| Socioeconomic Status | -0.006 (0.007) |
| Closest_Park3f_mi | 0.018 * (0.008) |
| Closest_Park3f_mi * SES | 0.028 ** (0.009) |
| $N$ | 1178 |
| $loglik$ | 264.283 |
| $AIC$ | -472.566 |

*** p < 0.001; ** p < 0.01; * p < 0.05.

*Model Results – Dependent Variable: Five Feature*

| | (1) |
|---|---|
| Intercept | 3.325 *** (0.034) |
| Project | -0.014 (0.026) |
| White | -0.019 (0.031) |
| Hispanic or Latino | 0.009 (0.033) |
| Years of Education | -0.027 *** (0.006) |
| Age | 0.028 * (0.013) |
| Sex | -0.044 *** (0.013) |
| Share of Hispanic/Latino Residents | 0.001 (0.006) |
| Share of Black (non-Hispanic/Latino) Residents | 0.007 (0.005) |
| Population Density | -0.004 (0.006) |
| Share of Developed Land | -0.002 (0.006) |
| Socioeconomic Status | -0.005 (0.007) |
| Closest_Park5f_mi | 0.020 * (0.009) |
| Closest_Park5f_mi * SES | 0.029 ** (0.009) |
| *N* | 1178 |
| *loglik* | 264.734 |
| *AIC* | -473.468 |

*** p < 0.001; ** p < 0.01; * p < 0.05.

*Model Results – Dependent Variable: Natural Log Three Feature*

| | (1) |
|---|---|
| Intercept | 3.321 *** (0.034) |
| Project | -0.016 (0.026) |
| White | -0.015 (0.031) |
| Hispanic or Latino | 0.011 (0.033) |
| Years of Education | -0.027 *** (0.006) |
| Age | 0.027 * (0.013) |
| Sex | -0.044 *** (0.013) |
| Share of Hispanic/Latino Residents | 0.001 (0.006) |
| Share of Black (non-Hispanic/Latino) Residents | 0.007 (0.005) |
| Population Density | -0.004 (0.006) |
| Share of Developed Land | -0.002 (0.007) |
| Socioeconomic Status | -0.006 (0.007) |
| ln_Closest_Park3f_mi | 0.009 (0.007) |
| ln_Closest_Park3f_mi * SES | 0.019 * (0.008) |
| *N* | 1178 |
| *loglik* | 262.641 |
| *AIC* | -469.282 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

*Model Results – Dependent Variable: Natural Log Five Feature*

| | (1) |
|---|---|
| Intercept | 3.321 *** (0.034) |
| Project | -0.016 (0.026) |
| White | -0.015 (0.031) |
| Hispanic or Latino | 0.011 (0.033) |
| Years of Education | -0.027 *** (0.006) |
| Age | 0.027 * (0.013) |
| Sex | -0.044 *** (0.013) |
| Share of Hispanic/Latino Residents | 0.001 (0.006) |
| Share of Black (non-Hispanic/Latino) Residents | 0.007 (0.005) |
| Population Density | -0.004 (0.006) |
| Share of Developed Land | -0.002 (0.007) |
| Socioeconomic Status | -0.020 (0.008) |
| ln_Closest_Park5f_mi | 0.034 (0.019) |
| ln_Closest_Park5f_mi * SES | 0.056 * (0.020) |
| *N* | 1178 |
| *loglik* | 262.641 |
| *AIC* | -469.282 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.