Submitted to Environment and Planning B: Urban Analytics and City Science
XX(X):1–13

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

The impacts of park access on health outcomes: a spatial comprehensive approach

Nico Boyd[1], Gregory S. Macfarlane[2], and Kari Watkins[1]

Abstract

This research identifies the correlation between size and access to urban parks and physical activity and obesity outcomes at the neighborhood level. Proximity to parks is associated with increased physical activity and reduced obesity, but little research has been conducted on the relationship between accessibility to parks and health outcomes. Using data for three urban areas, we created a new measure for access to parks called 'park choice accessibility.' Park choice accessibility uses a gravity model to interact distance to parks and the quality of those parks as defined by their size. A small park very close to a neighborhood can have a major impact, but a larger park at a similar distance may have an even larger impact. Similarly, a large park can be further away and still have an impact on health outcomes. Using spatial econometric analysis, we assess whether park choice accessibility is associated with increased physical activity or decreased prevalence of obesity at the neighborhood level. The analysis controls for socioeconomic covariates such as age, marital status, income, and educational attainment.

Introduction

The United States is currently facing an epidemic of obesity and chronic diseases, which are non-communicable diseases of long duration and typically slow development, including cardiovascular diseases, chronic respiratory diseases, diabetes, stroke, joint and bone diseases, and cancer (World Health

[1]School of Civil and Environmental Engineering, Georgia Institute of Technology
[2]Civil and Environmental Engineering Department, Brigham Young University

Corresponding author:
Gregory Macfarlane
430 Engineering Building
Brigham Young University
Provo, UT 84602
USA

Organization 2014) Recent statistics suggest that approximately 300,000 premature deaths each year can be attributed to chronic diseases (World Health Organization 2014). According to Finkelstein et al. (2009), obese Americans spend approximately $1,500 more in healthcare costs annually than Americans of average weight, totaling $147 billion in direct medical expenses nationally.

While a moderate amount of regular physical activity has been established as an effective strategy for reducing and managing obesity and many of the aforementioned chronic diseases Centers for Disease Control and Prevention (2009); **?**, Wolf (2008) notes that "more than 50 percent of U.S. adults do not get enough physical activity to provide health benefits; 24 percent are not active at all in their leisure time. Activity decreases with age and sufficient activity is less common among women than men, and among those with lower incomes and less education" (p. 22). As the epidemiological transition from infectious to chronic diseases is now complete in the developed world, increasing physical activity has become a vital public health task in the 21st century.

Until recently, most large-scale health promotion efforts focused on individual-level interventions intended to educate people about healthy lifestyles and behaviors, touching on topics including diet and exercise. According to Coutts (2008) however, this trend is shifting as professionals "have begun adopting an ecological paradigm, accepting that both individual and environmental determinants play a role in health behavior. This new, arguably revisited, public-health paradigm accounts not only for the compositional (who you are) but also for the contextual (where you are) influences on physical activity." As professionals begin to operate from the assumption that the design and configuration of the built environment can facilitate or inhibit physical activity, they are increasingly looking to public spaces like parks and greenways as key elements of the built environment that can support exercise (Bedimo-Rung et al. 2005).

In this paper, we present a holistic and flexible measurement for park accessibility based in random utility theory. This measure can consider the continuous distance to all parks in the region, weighted by the size of the park and other amenities. We apply this measurement to study the link between park accessibility and attractiveness and Census tract-level aggregate physical activity participation and obesity rates in New York City, controlling for spatial correlation and dependence. We find a positive relationship where the least park-accessible tracts have an expected physical activity participation rate 3.8 percent lower than the most accessible tracts, controlling for socioeconomic characteristics. We also find that increased physical activity rates lead to reduced aggregate obesity rates, but that when controlling for physical activity rates, accessibility to parks does not reduce overall accessibility. This suggests that to be an effective tool for addressing chronic diseases, communities must do more than simply provide parks, but also incentivize more citizens to use them.

Literature Review

Numerous previous studies have found relationships between aspects of the built environment and aspects of physical, mental, social, and economic health. Many of these findings are summarized in Table 1. For the purposes of this study, it is clear that parks and other vegetated areas support outdoor physical activity in ways distinct from other urban environments (Giles-Corti et al. 2005; Wells et al. 2007).

In spite of these generally positive findings, the evidence of a specific link between green space and physical activity is somewhat mixed. Wolf (2008) report residents self-identify as more active and attaining a higher quality of life when a greater portion of the environment is greenspace. Larson et al.

Table 1. Commonly cited benefits of parks and greenspaces

Category	Summary of Benefits
Physical Health	Provide clean drinking water (Benedict and McMahon 2006) Promote faster healing in hospitals (Akbari et al. 2001) Reduce heat-related mortality (Stone and Norman 2006) Reduce incidence of cardiovascular-related mortality (Mitchell and Popham 2008) Improved air quality and related reductions in respiratory-related mortality (Lovasi et al. 2008)
Mental Health	Reduce stress and mental fatigue (Wolch et al. 2014) Reduce aggression (Kuo and Sullivan 2001) Enhance emotional and cognitive development Improve behavioral outcomes in youth
Social Health	Enhanced community aesthetics Reduce crime (**?**Wolfe and Mennis 2012) Increase social interaction (Sullivan et al. 2004)
Economic Health	Provide ecosystem services Increase residential property values and municipal property tax revenues Attract more shoppers and increase economic activity to commercial districts

(2016) used self-reported scores on the Gallup-Healthways Wellbeing Index (citation) to evaluate the relationship between different areas of wellbeing, including physical health, and park quantity, quality, and accessibility in 44 U.S. cities. While the authors found positive associations between wellbeing and park quality and accessibility, these relationships were not statistically significant. A study by West et al. (2012) used park data from the Trust for Public Land's 2010 City Park Facts and public health data from the Behavioral Risk Factor Surveillance System (BRFSS) to examine relationships between the density of parkland, parkland per capita, and levels of physical activity and obesity for 67 metropolitan statistical areas in the U.S. The study found a significant, positive association between park density and physical activity and a significant, negative association between park density and obesity. Conversely, Richardson et al. (2012) examined the relationship between urban greenspace and selected mortality rates, and though the authors did not find a statistically significant relationship between the quantity of urban greenspace and mortality caused by lung cancer, diabetes, heart disease or car accidents, they did find that all-cause mortality was significantly higher in greener cities. In a meta-analysis of 20 peer reviewed journal articles exploring the relationship between parks and objectively measured physical activity, Bancroft et al. (2015) found that five studies reported a significant positive association between the two, six studies produced mixed results, and nine studies found no association at all.

While several of the aforementioned studies yielded results consistent with the hypothesis that park access and use confer health benefits stemming from increased physical activity and attendant decreases in obesity, all of the studies – except some included in the meta-analysis by Bancroft et al. (2015) -– were conducted at the level of the city or metropolitan statistical area. Metropolitan-level analyses do not capture the within-city variation in accessibility to parks that exists. Some cities with large amounts of total greenspace may nevertheless inequally distribute this space throughout the city leading to areas with poor access. Conversely, a city with a smaller overall proportion of greenspace may give all of its citizens better access to significant greenspace. Considering access at the neighborhood level within a single city may eliminate some regional or cultural fixed effects affecting metropolitan-level analyses.

In a study of neighborhoods in New York City, Stark et al. (2014) found that a higher proportion of park space in a neighborhood was associated with a lower BMI for its residents. Pretty et al. (2005) showed the presence of trees and other vegetation in outdoor environments is positively associated with physical activity; further, Suminski et al. (2005) showed that adding vegetation to existing sidewalks and trails increased local motivation to engage in physical activity.

To evaluate the impacts of parks and greenspaces on health outcomes at a sub-metropolitan level, one of the fundamental tasks involved is the selection of a measure of accessibility. In a comprehensive meta-analysis of published studies that measure active accessibility (accessibility using non-motorized travel modes including walking and cycling), Vale et al. (2016) identified four broad categories of studies based on the metric employed:

Distance-based Account only for the Euclidean distance between origins and destinations

Infrastructure-based Explicitly incorporate relevant transportation networks like roads and sidewalks to more accurately measure travel time and distance

Gravity-based Incorporate travel impedance measures to model accessibility as a function of a destination's attractiveness

Walk-score Similar to Gravity-based, with the impedance measure explicitly designed around walking

Importantly, Vale et al. acknowledge that there is not yet a consensus on the most appropriate accessibility metric to use in a given setting, noting that "ways to measure active accessibility are as varied as the number of scholars the measure them." For example, infrastructure-based distances can vary widely depending on which modes of travel infrastructure are considered. Similarly, gravity-based measures may not always account for surrounding land uses when considering the attractiveness or impedance of a destination. In evaluating methodologies of measuring access to urban services, including parks and greenspaces, Logan et al. (2017) note that existing and commonly used approaches "often simplify their measure of proximity by using large areal units and by imposing arbitrary distance thresholds," which often results in access-poor populations being overlooked.

Choice-based Accessibility

Consider that an individual is choosing a park for a recreation activity. If we apply random utility choice theory, specifically the multinomial logit model (McFadden 1974), the expected probability of individual i choosing park j from the set of all regional parks J is

$$P(j|V_{ij}) = \frac{\exp(V_{ij})}{\sum_{p \in J} \exp(V_{ip})} \tag{1}$$

where parks are differentiated from each other by their relative measurable utilities V_{ij}. In principle, V may include any measurable attributes of either the choice maker or the park. In this study we use a linear formulation of

$$V_{ij} = size_j \lambda_s + distance_{ij} \lambda_d \tag{2}$$

incorporating the size of the park in acres and the distance of the park from the census tract i in miles. The coefficients $\boldsymbol{\lambda}$ can be estimated from household surveys, though in the absence of a survey we may assert reasonable values.

A key theoretical outcome of random utility choice models is that the consumer surplus, or the total value of all alternatives of the choice set, can be obtained as the log-sum of the denominator of the choice probability equation (Small and Rosen 1981):

$$CS_i = \ln\left(\sum_{p \in J} \exp(V_{ip})\right) \tag{3}$$

There are several advantages to a log-sum defined metric relative to buffer-based accessibility metrics more commonly found in the literature. First, all individuals are defined as having some access to all parks, rather than an arbitrary cutoff asserted by the researcher. This allows for the fact that some people are more or less sensitive to distances, and that distance is a continuous, and not a binary, phenomenon. Second, the random utility formulation allows the researcher to include any attribute of the park; in this case, we consider the size of the park as an element of accessibility. This suggests that not all parks are equal, and that a large park such as Central Park on Manhattan may provide health and activity benefits over a much larger area than a smaller community square.

Note that this formulation is an extension of the gravity-based accessibility statistics, in the same way that the gravity model is a specific case of a destination-choice model (Daly 1982). But the extension is meaningful, and allows us to consider more attributes of the park or the indivdual beyond size and impedance. For example, we can aggregate the number of tweets emanating from a park as a measure of the park's attractiveness beyond its size, and the utility for the choice maker from Equation 2 then becomes

$$V_{ij} = size_j \lambda_s + distance_{ij} \lambda_d + tweets_j \lambda_t \tag{4}$$

In spite of its advantages, logsum-based accessibilities have not received as much application in the accessibility literature as distance-based or even gravity-based measures. They are commonly used, however, in alternatives analyses of transit infrastructure improvements (de Jong et al. 2007). A reason for this is likely that such analyses are usually conducted in the context of a travel demand model, where calibrated and multimodal logsums are readily available (Geurs et al. 2010).

Empirical Methodology

In this section we describe an experiment with data for New York City, where we compute a holistic logsum-based accessibility component and model the relationship between this measure and physical activity rates, controlling for spatial effects and socioeconomic factors.

Data

This study uses data available to the public from a variety of agencies and data providers.

Table 2. Descriptive Statistics of Tract and Park Variables

	Minimum	Median (IQR)	Maximum
Tract Variables, $N = 2102 $			
Obesity Rate	11.70	26.00 (20.50, 32.30)	48.20
Physical Activity Rate	42.50	71.20 (65.53, 76.50)	89.80
Density: population per square mile	0.06	69.37 (39.89, 107.91)	377.75
Full-time workers	1.90	55.40 (49.10, 60.80)	100.00
College-educated	2.00	27.80 (18.20, 42.08)	92.30
Single Adults	24.40	60.15 (49.30, 69.40)	98.40
Share of children (0-17)	0.00	21.10 (16.80, 25.50)	62.80
Share of young adults (18-29)	0.00	17.90 (14.90, 21.70)	78.30
Share of seniors (65+)	0.00	11.80 (8.80, 15.90)	88.50
Black population share	0.00	11.55 (2.50, 47.25)	99.30
Asian population share	0.00	8.20 (2.90, 21.20)	90.10
Hispanic population share	0.00	18.40 (9.03, 39.68)	95.60
Other Minorities	0.00	0.60 (0.00, 1.50)	12.20
Park Variables, $N = 2777 $			
size	0.50	1.38 (0.82, 3.20)	1446.08
tweets	0.00	0.00 (0.00, 2.00)	5452.00

The primary dataset is a geographic polygons shapefile of Census tracts in New York City. We appended relevant sociodemographic data for each tract level from the American Community Survey 2011-2015 5-year estimates. Data on physical activity participation rate and obesity rates for each tract are available from the Centers for Disease Control and Prevention's 500 Cities Project. After removing tracts with missing population information or outlying and unusual characteristics, we have Sexprnrow(tracts) complete cases. Table 2 presents key descriptive statistics for these data.

In a destination choice framework, the tracts represent the "origins" and "destinations" are parks in New York City. We retrieved a shapefile of public parks and greenspaces within New York City's municipal boundaries and checked it for accuracy and relevance. Upon inspection, we removed several facilities that do not qualify as publicly accessible green space, such as Yankee Stadium and its surrounding parking lots. We also removed parks of less than half an acre in size, as these appear to be predominantly vacant lots or parking strips rather than legitimate public green space. This leaves us with 2,777.

Model

We predict the rate of physical activity (y) in a census tract as a function of the tract's sociodemographic characteristics and choice-based accessibility to parks with a semi-log specification:

$$\log(y_i) = f(\boldsymbol{x}_i, CS_i, g(W, Obesity, \boldsymbol{X})) \tag{5}$$

where $\boldsymbol{x}_i$ is a vector of tract-level sociodemographic attributes, CS_i is a choice-based accessibility measure for the tract computed from Equation 3 and $g(W, \boldsymbol{y}, X)$ represents the impact of spatial neighborhood effects on the outcome.

Using tracts as the unit of observation in the model allows us to consider the availability of parks at a sub-metropolitan level, but it also introduces the problem of spatial correlation of unobserved errors and spatial dependence in the data generating process (Cliff and Ord 1970; Anselin 1980). Spatial correlation occurs when spatially-distributed unobservable or missing variables (school quality, neighborhood prestige, etc.) influence the modeled outcomes. Spatial dependence occurs when the outcome *depends* on the outcome in neighboring areas; for example, seeing neighbors exercising may encourage exercise. The relaxed spatial Durbin model (SDM) proposed by Burridge (1981) controls for both processes:

$$\boldsymbol{y} = \rho W \boldsymbol{y} + X\boldsymbol{\beta} + WX\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{6}$$

where ρ is the estimated spatial correlation between the outcome variables $y_i, y_j, \ldots \in \boldsymbol{y}$, W is a matrix specifying the spatial relationship between observations i and j in the dataset, $\boldsymbol{\beta}$ is a vector of estimable coefficients relating the attributes X of an observation to its outcome, and $\boldsymbol{\gamma}$ is a vector of estimated coefficients relating the attributes of an observation's *neighbors* to its outcome.

The researcher must assert W *a priori*, though there are several common specifications (Dubin 1998). In this study we specify that tracts with centroids located within 1.8 miles of each other are considered neighbors, and the strength of the relationship is the inverse of the distance between them:

$$W_{ij} = \begin{cases} \frac{1/d_{ij}}{\sum_{k=1}^{n}(1/d_{ik})} & \text{for } d_{ij} \leq 1.8 \text{ miles} \\ 0 & \text{for } d_{ij} > 1.8 \text{ miles} \end{cases} \tag{7}$$

```
## Warning: Column `geoid` joining factor and character vector,
coercing into character vector
## Warning in nb2listw(neighbours = tracts.dnn, glist = dists.inv,
zero.policy = TRUE, : zero sum general weights
```

It may be more efficient to estimate a restricted spatial autoregressive model – that controls for only correlation or only dependence – and therefore save degrees of freedom in the estimation. Macfarlane et al. (2015) illustrate a decision algorithm originally proposed by Florax et al. (2003) which uses a likelihood ratio test of the SDM against a restricted spatial error model (SEM): if the likelihood ratio test rejects that the models are equivalent, then the SDM should be used.

A final point to note about spatial autoregressive models is that the partial derivative of the model with respect to a single regressor $\boldsymbol{x_k} \in X$ does not equal the estimated coefficient β_k. Consequently, the estimated coefficients do not represent the *effect* of a variable on the outcome. The direct effects of an observation's attribute on its outcome, the indirect effects of neighbors' attributes on an outcome, and the total effect of own and neighbors' attributes can attained via Monte Carlo simulation. In this study we use the `spdep` package for R (Bivand et al. 2013) for model estimation and impact simulation.

Accessibility

Extant destination choice models used in practice typically use travel time in minutes as a travel impedance term, and employment by sector as a size term or attraction component. Typically such models handle recreational trips together with other non-work and non-school trips, and we therefore can find no previously estimated coefficients using size terms relevant to park acreage or amenities. Given this gap in the literature, we instead assert values for β to generate the choice-based accessibility statistic used in this study.

We initially estimated potential values of the β coefficients by applying a limited-memory, box-constrained optimization algorithm maximizing the SDM model log-likehood. The constraint required that the coefficient on size be positive and that on distance negative; all else equal, we assert people will prefer to use larger parks and parks that are nearer to their residence. This resulted in $\beta_d = -0.42$ and $\beta_s = 0.000$. Given our expectation that park size is a valuable attribute, we subsequently adjusted this value to a value of 1.00. This implies that each mile

Results

We estimated SDM and SEM models regressing the physical activity rate against against the base covariates (without the accessibility logsum). A likelihood ratio test rejects that the SEM is equivalent ($\chi^2 = 157.6637$ on 11 degrees of freedom), so we use the SDM specification only going forward. Table 3 presents the estimated coefficients for the base model as well as models including logsum-based accessibility with and without Twitter activity. A few initial observations can be made from this table. First, adding the logsum measurement significantly improves the overall model fit, using a likelihood ratio test on the model log-likelihood statistics. Second, the coefficient estimates on the controlling variables remain virtually unchanged from the base model.

Recall from the presentation of spatial econometrics above that the total effect of a covariate on the dependent variable is different from the estimated coefficient and involves interaction between direct and indirect effects, or between the β and γ parameters. This is particularly important when the β and γ estimates for a particular covariate show a different sign, such as in the proportion of black residents in a census tract, as interpreting the direction and magnitude of the relationship is not straightforward.

Table 4 shows the simulated marginal impact of the accessibility logsum measurements on the physical activity rate in each tract. Given that the logsum is a relatively continuous measure over space, and that it would be difficult to improve the park accessibility of a tract without also improving the accessibility of the tract's neighbors, the total effect combining the direct effect and the indirect effects of all neighbors is most appropriate. The total effect of tract-level accessiblity on physical activity rates is strongly significant, regardless of whether we consider Twitter activity.

It is of course necessary to also consider the magnitude of the effect. As this is a semi-log model, the coefficient implies a unit increase in the access and tweets logsum leads to a 0.621% increase in the physical activity rate. Referring to the map of logsum measures in Figure 1, the measures span from roughly -3 in heavily industrialized segments of southern Brooklyn to roughly 3 in tracts immediately adjacent to Central Park on the Upper East Side. Thus moving from the least accessible parts of the city to the most accessible will reduce the expected physical activity rate by 3.726%. In other words, if a tract with poor accessibility is expected to achieve the sample mean of 71.2%, the same tract with excellent accessibility is expected to achieve a rate of 73.85%. For an aggregate statistic strongly influenced by other socioeconomic factors, this is not an insubstantial marginal effect.

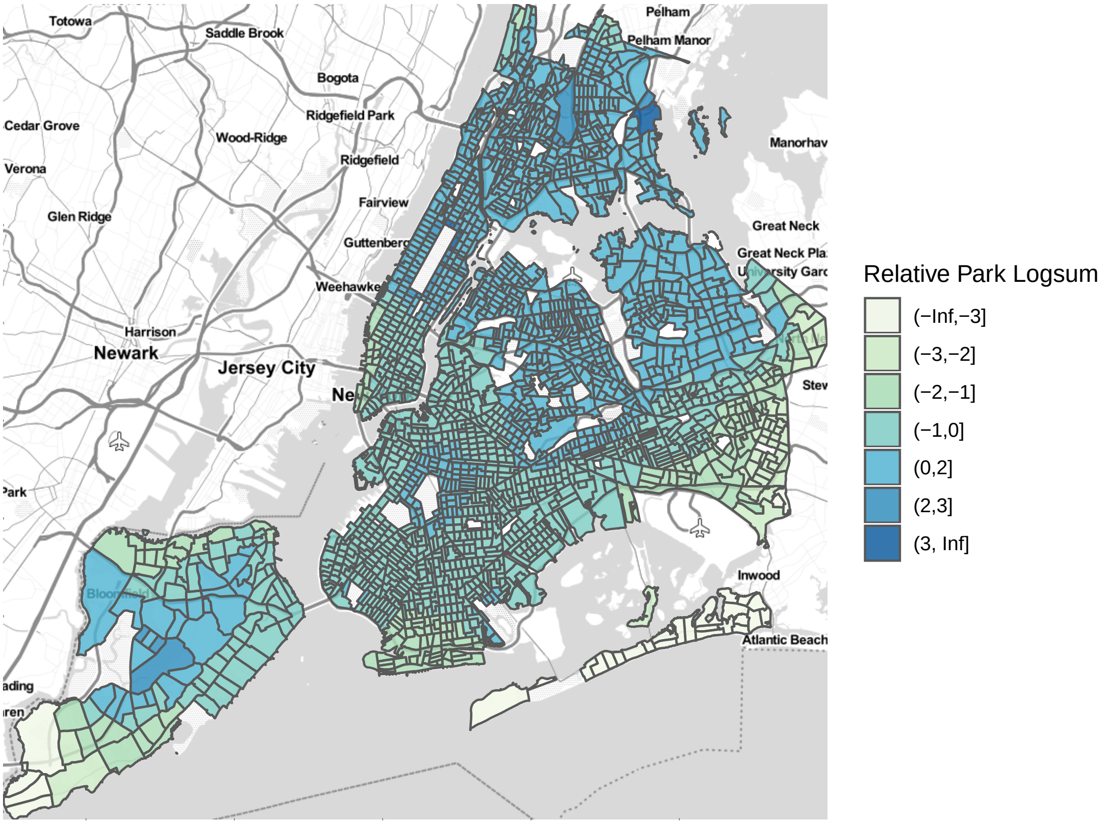

Figure 1. Relative surplus-based accessibility in New York City

We now consider the impacts of a model where the dependent variable is the *obesity* rate, and physical activity becomes an independent covariate alongside the controlling variables and the tweets-augmented logsum variable. Table 5 presents the simulated impacts resulting from this model.

Table 3. Estimated Coefficients Relating Park Access to Physical Activity

	Base	Access	Access + Tweets
Intercept	4.2009 (0.0397)***	4.2127 (0.0405)***	4.2094 (0.0402)***
log(density)	−0.0083 (0.0013)***	−0.0082 (0.0013)***	−0.0082 (0.0013)***
Full-time workers	0.0020 (0.0001)***	0.0020 (0.0001)***	0.0020 (0.0001)***
College-educated	0.0033 (0.0001)***	0.0033 (0.0001)***	0.0033 (0.0001)***
Single Adults	−0.0013 (0.0001)***	−0.0013 (0.0001)***	−0.0013 (0.0001)***
Share of children (0-17)	−0.0022 (0.0002)***	−0.0022 (0.0002)***	−0.0022 (0.0002)***
Share of young adults (18-29)	0.0009 (0.0002)***	0.0010 (0.0002)***	0.0009 (0.0002)***
Share of seniors (65+)	−0.0020 (0.0002)***	−0.0020 (0.0002)***	−0.0020 (0.0002)***
Black population share	−0.0004 (0.0001)***	−0.0004 (0.0001)***	−0.0004 (0.0001)***
Asian population share	−0.0017 (0.0001)***	−0.0017 (0.0001)***	−0.0017 (0.0001)***
Hispanic population share	−0.0010 (0.0001)***	−0.0010 (0.0001)***	−0.0010 (0.0001)***
Other Minorities	−0.0009 (0.0006)	−0.0009 (0.0006)	−0.0009 (0.0006)
lag log(density)	−0.0104 (0.0028)***	−0.0131 (0.0029)***	−0.0143 (0.0030)***
lag Full-time workers	−0.0023 (0.0004)***	−0.0021 (0.0004)***	−0.0020 (0.0004)***
lag College-educated	−0.0002 (0.0002)	−0.0003 (0.0002)	−0.0004 (0.0002)*
lag Single Adults	−0.0044 (0.0004)***	−0.0042 (0.0004)***	−0.0042 (0.0004)***
lag Share of children (0-17)	−0.0050 (0.0007)***	−0.0046 (0.0007)***	−0.0046 (0.0007)***
lag Share of young adults (18-29)	0.0021 (0.0009)*	0.0021 (0.0009)*	0.0021 (0.0009)*
lag Share of seniors (65+)	0.0001 (0.0007)	−0.0001 (0.0007)	−0.0003 (0.0007)
lag Black population share	0.0010 (0.0001)***	0.0010 (0.0001)***	0.0010 (0.0001)***
lag Asian population share	−0.0009 (0.0002)***	−0.0009 (0.0002)***	−0.0009 (0.0002)***
lag Hispanic population share	0.0012 (0.0002)***	0.0009 (0.0002)***	0.0009 (0.0002)***
lag Other Minorities	0.0023 (0.0026)	0.0039 (0.0026)	0.0038 (0.0026)
ρ	0.1232 (0.0146)***	0.1188 (0.0146)***	0.1214 (0.0146)***
Access logsum		0.0031 (0.0029)	
lag Access logsum		0.0012 (0.0034)	
Access + Tweets logsum			0.0023 (0.0029)
lag Access + Tweets logsum			0.0032 (0.0035)
Num. obs.	2102	2102	2102
Parameters	25	27	27
Log Likelihood	4011.7459	4019.3566	4020.8199
AIC (Linear model)	-7905.9204	-7922.2513	-7921.9468
AIC (Spatial model)	-7973.4917	-7984.7132	-7987.6397
LR test: statistic	69.5713	64.4620	67.6929
LR test: p-value	0.0000	0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Discussion

We readily acknowledge limitations in this study. To begin, it would be beneficial to have estimated the logsum coefficients from a high-quality survey rather than asserting them to match *a priori expectations of*

Table 4. Simulated Impact of Accessibility on Physical Activity

Logsum Measure	Estimate	p-value
Access	0.0049558	0.00045 ***
Access + Tweets	0.0062096	9e-05 ***

Note:
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5. Simulated Impacts of Variables on Obesity

	Direct	Indirect	Total
log(density)	-0.0054 **	-0.29297 ***	-0.29838 ***
Full-time workers	2e-05	0.00225	0.00228
College-educated	-0.00186 ***	0.01647 ***	0.01461 ***
Single Adults	0.00072 ***	-0.02557 ***	-0.02485 ***
Share of children (0-17)	0.00014	-0.04293 ***	-0.04279 ***
Share of young adults (18-29)	-0.00321 ***	0.00186	-0.00135
Share of seniors (65+)	-0.00377 ***	-0.02677 ***	-0.03054 ***
Black population share	0.00291 ***	-0.00047	0.00244 *
Asian population share	-0.00471 ***	-0.02114 ***	-0.02585 ***
Hispanic population share	0.00122 ***	-0.00508 **	-0.00387 *
Other Minorities	-3e-04	0.08436 ***	0.08406 ***
Access + Tweets logsum	0.00526	0.06174 ***	0.067 ***
log(Physical Activity Rate)	-0.85142 ***	-6.26845 ***	-7.11987 ***

Note:
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Funding

This project was funded in part via ....

References

Akbari H, Pomerantz M and Taha H (2001) Cool surfaces and shade trees to reduce energy use and improve air quality in urban areas. *Solar Energy* 70(3): 295–310. DOI:10.1016/S0038-092X(00)00089-X.

Anselin L (1980) Estimation methods for spatial autoregressive structures. *Regional Science Dissertation & Monograph Series, Program in Urban and Regional Studies, Cornell University* (8).

Bancroft C, Joshi S, Rundle A, Hutson M, Chong C, Weiss CC, Genkinger J, Neckerman K and Lovasi G (2015) Association of proximity and density of parks and objectively measured physical activity in the United States: A systematic review. *Social Science & Medicine* 138: 22–30. DOI:10.1016/J.SOCSCIMED.2015.05.034.

Bedimo-Rung AL, Mowen AJ and Cohen DA (2005) The significance of parks to physical activity and public health: A conceptual model. *American Journal of Preventive Medicine* 28(2): 159–168. DOI: 10.1016/J.AMEPRE.2004.10.024.

Benedict MA and McMahon ET (2006) *Green Infrastructure: Linking Landscapes and Communities*. Washington: Island Press.

Bivand RS, Pebesma E and Gomez-Rubio V (2013) *Applied spatial data analysis with* $\{R\}$*, Second edition*. Springer, NY.

Burridge P (1981) Testing for a common factor in a spatial autoregression model. *Environment and Planning A* 13(7): 795–800.

Centers for Disease Control and Prevention (2009) CDC - Healthy Places - Physical Activity.

Cliff AD and Ord JK (1970) Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography* 46: 269–292.

Coutts C (2008) Greenway Accessibility and Physical-Activity Behavior. *Environment and Planning B: Planning and Design* 35(3): 552–563. DOI:10.1068/b3406.

Daly A (1982) Estimating choice models containing attraction variables. *Transportation Research Part B: Methodological* 16(1): 5–15. DOI:10.1016/0191-2615(82)90037-6.

de Jong G, Daly A, Pieters M and van der Hoorn T (2007) The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice* 41(9): 874–889. DOI: 10.1016/J.TRA.2006.10.002.

Dubin RA (1998) Spatial autocorrelation: a primer. *Journal of Housing Economics* 7(4): 304–327. DOI: 10.1006/jhec.1998.0236.

Finkelstein EA, Trogdon JG, Cohen JW and Dietz W (2009) Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates. *Health Affairs* 28(5): w822–w831. DOI:10.1377/hlthaff.28.5.w822.

Florax RJGM, Folmer H and Rey SJ (2003) Specification searches in spatial econometrics: The relevance of Hendry's methodology. *Regional Science and Urban Economics* 33: 557–579.

Geurs K, Zondag B, de Jong G and de Bok M (2010) Accessibility appraisal of land-use/transport policy strategies: More than just adding up travel-time savings. *Transportation Research Part D: Transport and Environment* 15(7): 382–393. DOI:10.1016/J.TRD.2010.04.006.

Giles-Corti B, Broomhall MH, Knuiman M, Collins C, Douglas K, Ng K, Lange A and Donovan RJ (2005) Increasing walking: How important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine* 28(2): 169–176. DOI:10.1016/J.AMEPRE.2004.10.018.

Kuo FE and Sullivan WC (2001) Aggression and Violence in the Inner City. *Environment and Behavior* 33(4): 543–571. DOI:10.1177/00139160121973124.

Larson LR, Jennings V and Cloutier SA (2016) Public Parks and Wellbeing in Urban Areas of the United States. *PLOS ONE* 11(4): e0153211. DOI:10.1371/journal.pone.0153211.

Logan T, Williams T, Nisbet A, Liberman K, Zuo C and Guikema S (2017) Evaluating urban accessibility: leveraging open-source data and analytics to overcome existing limitations. *Environment and Planning B: Urban Analytics and City Science* : 239980831773652DOI:10.1177/2399808317736528.

Lovasi GS, Quinn JW, Neckerman KM, Perzanowski MS and Rundle A (2008) Children living in areas with more street trees have lower prevalence of asthma. *Journal of Epidemiology & Community Health* 62(7): 647–649. DOI:10.1136/JECH.2007.071894.

Macfarlane GS, Garrow LA and Moreno-Cruz J (2015) Do Atlanta residents value MARTA? Selecting an autoregressive model to recover willingness to pay. *Transportation Research Part A: Policy and Practice* 78: 214–230.

McFadden DL (1974) Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka P (ed.) *Frontiers in Econometrics*. New York: Academic Press, pp. 105–142.

Mitchell R and Popham F (2008) Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet* 372(9650): 1655–1660. DOI:10.1016/S0140-6736(08)61689-X.

Pretty J, Peacock J, Sellens M and Griffin M (2005) The mental and physical health outcomes of green exercise. *International Journal of Environmental Health Research* 15(5): 319–337. DOI:10.1080/09603120500155963.

Richardson EA, Mitchell R, Hartig T, de Vries S, Astell-Burt T and Frumkin H (2012) Green cities and health: a question of scale? *Journal of Epidemiology and Community Health* 66(2): 160 LP – 165. DOI: 10.1136/jech.2011.137240.

Small KA and Rosen HS (1981) Applied Welfare Economics with Discrete Choice Models. *Econometrica* 49(1): 105. DOI:10.2307/1911129.

Stark JH, Neckerman K, Lovasi GS, Quinn J, Weiss CC, Bader MD, Konty K, Harris TG and 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. DOI:10.1016/J.YPMED.2014.03.026.

Stone B and Norman JM (2006) Land use planning and surface heat island formation: A parcel-based radiation flux approach. *Atmospheric Environment* 40(19): 3561–3573. DOI:10.1016/J.ATMOSENV.2006.01.015.

Sullivan WC, Kuo FE and Depooter SF (2004) The Fruit of Urban Nature. *Environment and Behavior* 36(5): 678–700. DOI:10.1177/0193841X04264945.

Suminski RR, Poston WSC, Petosa RL, Stevens E and Katzenmoyer LM (2005) Features of the neighborhood environment and walking by U.S. adults. *American Journal of Preventive Medicine* 28(2): 149–155. DOI: 10.1016/J.AMEPRE.2004.09.009.

Vale DS, Saraiva M and Pereira M (2016) Active accessibility. *Journal of Transport and Land Use* 9(1): 209–235.

Wells NM, Ashdown SP, Davies EHS, Cowett FD and Yang Y (2007) Environment, Design, and Obesity. *Environment and Behavior* 39(1): 6–33. DOI:10.1177/0013916506295570.

West ST, Shores KA and Mudd LM (2012) Association of Available Parkland, Physical Activity, and Overweight in America's Largest Cities. *Journal of Public Health Management and Practice* 18(5): 423–430. DOI: 10.1097/PHH.0b013e318238ea27.

Wolch JR, Byrne J and Newell JP (2014) Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning* 125: 234–244. DOI: 10.1016/J.LANDURBPLAN.2014.01.017.

Wolf KL (2008) City trees, nature and physical activity: A research review. *Arborist News* 17(1): 22–24.

Wolfe MK and Mennis J (2012) Does vegetation encourage or suppress urban crime? Evidence from Philadelphia, PA. *Landscape and Urban Planning* 108(2-4): 112–122. DOI:10.1016/J.LANDURBPLAN.2012.08.006.

World Health Organization (2014) Noncommunicable diseases country profiles 2014. Technical report, World Health Organization.