Submitted to Environment and Planning B: Urban Analytics and City Science
XX(X):1–17

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], Kari Watkins[1], and Professor WhogaveTweets[3]

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

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

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

diseases, chronic respiratory diseases, diabetes, stroke, joint and bone diseases, and cancer (World Health 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 choice 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 tract-level socioeconomic characteristics. We find a positive relationship where the least park-accessible tracts have an expected physical activity participation rate XX percent lower than the most accessible tracts, though this result is only suggestive of correlation from a statistical perspective. However, we also find a strong and significant negative correlation between increased park accessibility and decreased obesity rates, again controlling for spatial correlation, socioeconomic attributes, and physical activity rates.

We also demonstrate how the choice-based accessibility metric could potentially be expanded to account for other elements of a park's attractiveness by introducing the number of tweets emanating from the park during (WE NEED TO DESCRIBE THE PERIOD). The paper concludes with a discussion of opportunities for future research.

Literature Review

Numerous previous studies have found relationships between characteristics of the built environment and aspects of physical, mental, social, and economic health. Many of these findings are summarized in Table

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

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. (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 reported 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 exist in many cities. 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 sufficient greenspace to meet their needs or wants for physical activity. Considering access at the neighborhood level within a single city may also 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.

When evaluating the impacts of parks and greenspaces on health outcomes at a sub-metropolitan level, one of the fundamental challenges researchers face is selecting 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.” This is at least partly because each method has inherent strengths and weaknesses. Distance is an inherently continuous phenomenon, but software tools ease and enable buffer analyses where distance becomes a binary phenomon of somewhat arbitrary extent. Logan et al. (2017) note that these commonly used approaches can also lead researchers to overlook access-poor populations. Infrastructure-based distances can vary widely depending on which modes of travel are considered. a highway network may not adequately capture trails and bike paths. Similarly, gravity-based measures may not always account for surrounding land uses when considering the attractiveness or impedance of a destination; the findings of Pretty et al. (2005) and other suggest unpleasant pedestrian routes attenuate perceived travel costs.

Though there is no shortage of possible accessibility metrics, in this paper we wish to present and apply a technique that is explicitly based in the choices of individuals as they consider how to access parks.

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 weighted 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, non-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 alternatives 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 Application

In this section we describe an experiment with data for New York City, where we compute a holistic logsum-based accessibility to parks metric and model the relationship between this measure and physical activity rates, controlling for spatial effects and socioeconomic factors. We subsequently model the effect of this accessibility on obesity.

Data

This study uses data available to the public from a variety of federal and state data agencies. The datasets, as well as the analysis code, is available on GitHub at `https://github.com/boydnpb/park_access_new.git` 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 (WE NEED A CITATION). 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 (WE NEED A CITATION). 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 parks. It is worth noting that the particular dataset classifies specific facilities within larger parks as independent features: baseball fields within Central Park, for example, are themselves delineated green spaces. We elect to retain these independent features as they may generate additional physical activity participation aside from the containing park.

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 the following specification:

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

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

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
Median tract income	9829.00	54,566.00 (39,090.00, 73,340.75)	250000.00
Share of low income	0.00	23.40 (15.80, 34.60)	96.20
Share of high income	0.00	22.70 (13.50, 34.50)	82.70
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			
Park Size in Acres	0.50	1.38 (0.82, 3.20)	1446.08
Tweets Emanating from Park	0.00	0.00 (0.00, 2.00)	5452.00

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. If only correlation and not dependence exists in the data generating process, it may be more efficient to estimate a restricted spatial autoregressive model 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 the restricted spatial error model (SEM):

$$\begin{aligned} \boldsymbol{y} &= X\boldsymbol{\beta} + \boldsymbol{u}, \boldsymbol{u} = \lambda W \boldsymbol{u} + \boldsymbol{\epsilon} \\ \boldsymbol{y} &= X\boldsymbol{\beta} + (I - \lambda W)^{-1}\boldsymbol{\epsilon} \end{aligned} \tag{7}$$

If the likelihood ratio test rejects that the models are equivalent, then the SDM should be used because its estimates are unbiased. If the test cannot reject equivalents, the SEM is more efficient.

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

$$W_{ij} = \begin{cases} \frac{1/d_{ij}}{\sum_{k=1}^{n}(1/d_{ik})} & \text{for } d_{ij} \leq 1.8 \text{ miles} \\ 0 & \text{for } d_{ij} > 1.8 \text{ miles} \end{cases} \tag{8}$$

This is a similar specification to then one Macfarlane et al. (2015) used in a study of home prices in Atlanta. We selected this specification after comparing overall model likelihoods between it and an adjacent (queen) tract specification.

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 plausible values for β to generate the choice-based accessibility statistic used in this study.

The functional form of the accessibility statistic is adapted from Equation 2 by log-transforming first the distance (Euclidean, in miles) between the park boundaries and population-weighted tract centroids and second the park size (in acres).

$$V_{ij} = \log(size_j)\lambda_s + \log(distance_{ij})\lambda_d \tag{9}$$

We then calculate the consumer surplus C (log-sum of the exponentiated utilities) for each tract and standardize the result,

$$C'_i = \frac{C_i - \bar{\boldsymbol{C}}}{sd(\boldsymbol{C})} \tag{10}$$

We initially estimated potential values of the λ coefficients by applying a limited-memory, box-constrained optimization algorithm maximizing the 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 $\lambda_d = -1.98$ and $\lambda_s = 0.000$. Though the "most likely" parameter value, in this case it was unsatisfactory for two reasons: first, our prior assertion that park size is a valuable attribute of a park, and second the observation that with these maximum likelihood values tracts bordering on the corners of Central Park were measured with poor overall accessibility. We subsequently adjusted the λ_s value to 0.3; the relative ratio of the coefficients implies that all else equal, an individual would travel 6.6 times further to access a park twice as large.

*Using the `nlm` function in R (**?**)

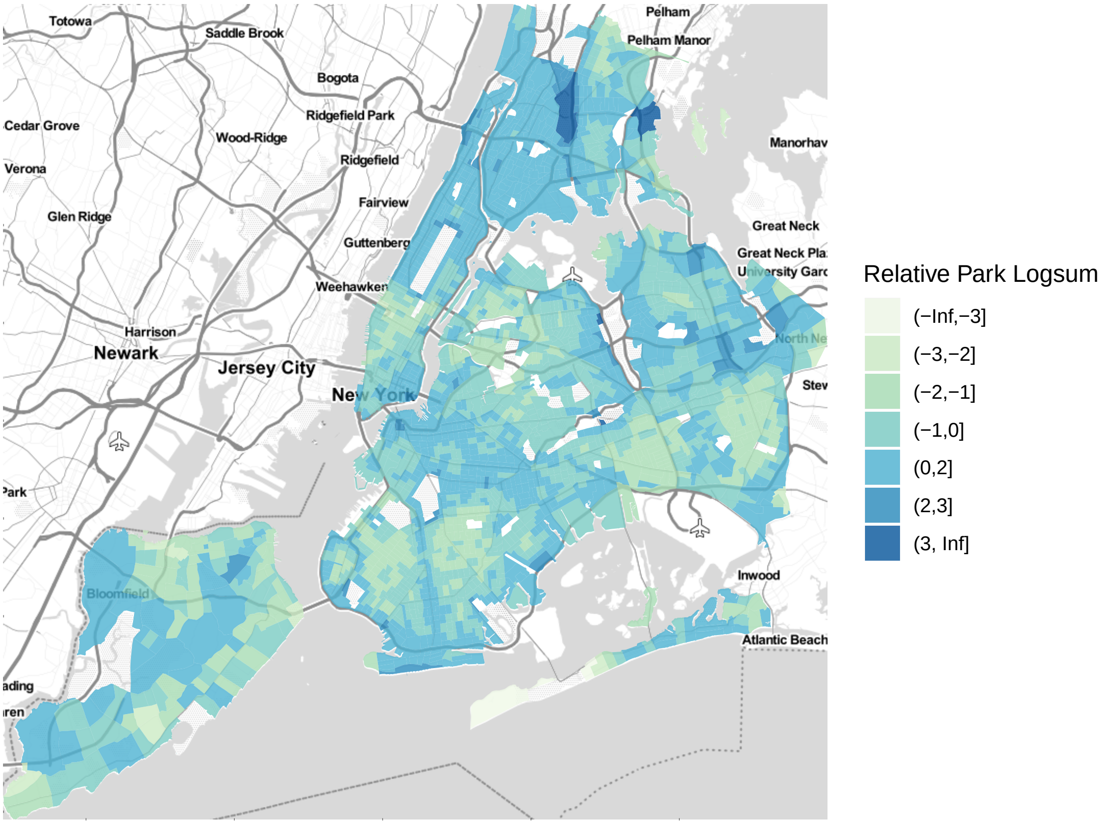

Figure 1. Relative surplus-based accessibility in New York City

Figure 1 shows a map of New York City with each tract highlighted based on its relative accessibility score. The most continuous region of good park access is in upper Manhattan and the Bronx, bracketed by Central Park and the Bronx River. Conversely, some of the poorest-accessibility areas are in Brooklyn tracts not immediately adjacent to Prospect Park. Because the accessibility statistic is normalized, the worst values are slightly below -3, the best somehere above 3.

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 failed to reject that the SEM is equivalent ($\chi^2 = 7.437687$ on 13 degrees of freedom), so we use the SEM specification only going forward. Table 3 presents the estimated coefficients for the base model with income represented in two ways. In the first, the share of the tract population in the bottom three income deciles and the share of the population in the top three income deciles are each included as independent predictors. In the second, the logarithm of the mean tract income is included as a predictor. Notwithstanding the higher model likelihood of the deciles model, comparing coefficients across the models shows that the predictors change magnitude or significance only marginally, if at all. Overall, the base model predictors are significant and of intuitive direction. Tracts with higher shares of full-time workers, college-educated adults, young adults, and high-income households all show a greater share of individuals engaging in regular physical activity. Conversely, tracts with greater population density and a greater share of single adults, children, seniors, minorities of all types, and low income households all report lower rates of physical activity.

The other two models presented in Table 3 include the log-sum accessibility statistic described in Section added as a predictor. In both income specifications, the access logsum is positively correlated with physical activity rates; all else equal, people in tracts with higher accessibility to parks report higher rates of physical activity. The estimated value of this coefficient and its standard error is effectively the same with both representations of income, though in the log representation the estimated t-statistic technically passes the 0.1 level. This is suggestive evidence only from a statistical perspective, but it deserves comment. Taking a compromise value of 0.10 between the two models, the data suggest going from the least-accessible (-3.0) to the most-accessible (3.0) tract implies a 0.6 percentage-point rise in physical activity rates, though with a wide confidence interval around that estimate.

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 accessibility logsum variable. Table **??** presents the estimated coefficients for both income specifications, with and without the accessibility logsum. As in the physical activity models, the coefficients are largely significant and in the expected sign. The accessibility coefficient has practically the same effect as in the physical activity model: moving from poor to excellent parks access lowers the changes the expected obesity rate by around half a percentage point. In this case, however, the parameter p-value is lower than the conventional 0.05 level in each income representation.

Additional Amenities

As discussed when presenting the choice-based accessibility metric above, a key benefit it offers is the natural way in which park amenities beyond simply size can be accomodated. For example, the number of tweets emanating from a park might be seen as a measure of the park's popularity, or a proxy for its use. It may be construed that parks with high popularity or use contain other amenities that encourage residents to use the park space, raising physical activity rates and lowering obesity rates beyond what would be expected with the park's size and proximity alone. It it reasonable to question whether park size and twitter activity are highly correlated, and thus would result in a double-counting of park size in the accessibility statistic. In our data the number of tweets is positively but only loosely correlated with park

Table 3. SEM Coefficients Predicting Physical Activity Rates

	Base	Base: log(Income)	Access	Access: log(Income)
(Intercept)	74.9254^{***}	22.1640^{***}	74.8739^{***}	21.9908^{***}
	(1.0343)	(2.7224)	(1.0339)	(2.7214)
log(density)	-0.3137^{***}	-0.3493^{***}	-0.3052^{***}	-0.3399^{***}
	(0.0710)	(0.0726)	(0.0712)	(0.0727)
Full-time workers	0.0567^{***}	0.0796^{***}	0.0578^{***}	0.0809^{***}
	(0.0075)	(0.0075)	(0.0076)	(0.0075)
College-educated	0.1955^{***}	0.1814^{***}	0.1943^{***}	0.1802^{***}
	(0.0062)	(0.0062)	(0.0063)	(0.0062)
Single Adults	-0.0201^{*}	-0.0248^{**}	-0.0204^{*}	-0.0252^{**}
	(0.0081)	(0.0083)	(0.0081)	(0.0083)
Share of children (0-17)	-0.1050^{***}	-0.1135^{***}	-0.1053^{***}	-0.1138^{***}
	(0.0117)	(0.0120)	(0.0116)	(0.0120)
Share of young adults (18-29)	0.0343^{**}	0.0391^{***}	0.0355^{***}	0.0404^{***}
	(0.0107)	(0.0110)	(0.0107)	(0.0110)
Share of seniors (65+)	-0.0825^{***}	-0.0973^{***}	-0.0840^{***}	-0.0990^{***}
	(0.0114)	(0.0116)	(0.0114)	(0.0117)
Black population share	-0.0380^{***}	-0.0368^{***}	-0.0381^{***}	-0.0369^{***}
	(0.0040)	(0.0041)	(0.0040)	(0.0041)
Asian population share	-0.1121^{***}	-0.1118^{***}	-0.1119^{***}	-0.1116^{***}
	(0.0047)	(0.0048)	(0.0047)	(0.0048)
Hispanic population share	-0.0669^{***}	-0.0690^{***}	-0.0675^{***}	-0.0698^{***}
	(0.0046)	(0.0047)	(0.0046)	(0.0047)
Other Minorities	-0.0488	-0.0385	-0.0480	-0.0375
	(0.0310)	(0.0318)	(0.0309)	(0.0318)
Share high income	0.0119		$0.0127^{\cdot}$	
	(0.0072)		(0.0072)	
Share low income	-0.1493^{***}		-0.1491^{***}	
	(0.0070)		(0.0070)	
λ	0.8288^{***}	0.8278^{***}	0.8277^{***}	0.8268^{***}
	(0.0308)	(0.0310)	(0.0310)	(0.0311)
log(Income)		4.5103^{***}		4.5239^{***}
		(0.2230)		(0.2229)
Access logsum			0.0921	$0.1105^{\cdot}$
			(0.0586)	(0.0601)
Num. obs.	2102	2102	2102	2102
Parameters	16	15	17	16
Log Likelihood	-4405.1949	-4463.9789	-4403.9585	-4462.2882
AIC (Linear model)	9145.1064	9239.9176	9140.9393	9233.9513
AIC (Spatial model)	8842.3898	8957.9579	8841.9171	8956.5763
LR test: statistic	304.7167	283.9597	301.0222	279.3750
LR test: p-value	0.0000	0.0000	0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 4. SEM Coefficients Predicting Obesity Rates

	Base	Base: log(Income)	Access	Access: log(Income)
(Intercept)	25.2568^{***}	55.9024^{***}	25.2541^{***}	55.9516^{***}
	(1.0539)	(2.1855)	(1.0470)	(2.1788)
log(density)	-0.0641	-0.0462	-0.0773	-0.0597
	(0.0549)	(0.0555)	(0.0551)	(0.0556)
Full-time workers	-0.0171^{**}	-0.0297^{***}	-0.0187^{**}	-0.0312^{***}
	(0.0058)	(0.0056)	(0.0058)	(0.0057)
College-educated	-0.0800^{***}	-0.0732^{***}	-0.0786^{***}	-0.0721^{***}
	(0.0049)	(0.0048)	(0.0049)	(0.0048)
Single Adults	0.0252^{***}	0.0273^{***}	0.0256^{***}	0.0277^{***}
	(0.0062)	(0.0063)	(0.0062)	(0.0063)
Share of children (0-17)	0.0568^{***}	0.0612^{***}	0.0571^{***}	0.0615^{***}
	(0.0090)	(0.0091)	(0.0089)	(0.0091)
Share of young adults (18-29)	-0.0721^{***}	-0.0743^{***}	-0.0736^{***}	-0.0758^{***}
	(0.0082)	(0.0083)	(0.0082)	(0.0083)
Share of seniors (65+)	-0.0724^{***}	-0.0642^{***}	-0.0703^{***}	-0.0621^{***}
	(0.0088)	(0.0088)	(0.0088)	(0.0088)
Black population share	0.0913^{***}	0.0909^{***}	0.0913^{***}	0.0909^{***}
	(0.0033)	(0.0033)	(0.0033)	(0.0033)
Asian population share	-0.0781^{***}	-0.0780^{***}	-0.0784^{***}	-0.0783^{***}
	(0.0037)	(0.0037)	(0.0037)	(0.0037)
Hispanic population share	0.0432^{***}	0.0445^{***}	0.0439^{***}	0.0453^{***}
	(0.0036)	(0.0037)	(0.0036)	(0.0037)
Other Minorities	$-0.0407^{\cdot}$	$-0.0468^{\cdot}$	$-0.0417^{\cdot}$	-0.0480^{*}
	(0.0236)	(0.0239)	(0.0236)	(0.0239)
Share high income	$-0.0103^{\cdot}$		-0.0113^{*}	
	(0.0056)		(0.0056)	
Share low income	0.0845^{***}		0.0840^{***}	
	(0.0054)		(0.0054)	
λ	0.9614^{***}	0.9603^{***}	0.9608^{***}	0.9597^{***}
	(0.0100)	(0.0102)	(0.0101)	(0.0103)
log(Income)		-2.6347^{***}		-2.6430^{***}
		(0.1697)		(0.1694)
Access logsum			-0.1228^{**}	-0.1316^{**}
			(0.0457)	(0.0462)
Num. obs.	2102	2102	2102	2102
Parameters	16	15	17	16
Log Likelihood	-3857.6405	-3887.3784	-3854.0369	-3883.3332
AIC (Linear model)	9376.6945	9422.8995	9369.7177	9414.1283
AIC (Spatial model)	7747.2809	7804.7567	7742.0738	7798.6663
LR test: statistic	1631.4136	1620.1428	1629.6439	1617.4620
LR test: p-value	0.0000	0.0000	0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 5. Estimated Relationship between Accessiblity and Outcomes

	Accessibility Components	Estimate	Std. Error	p-value
$y = $Physical Activity				
High and Low Income	Size and Distance	0.0921	0.0586	0.1157
High and Low Income	Tweets, Size and Distance	0.0781	0.0601	0.1937
log(Income)	Size and Distance	0.1105	0.0601	0.0658
log(Income)	Tweets, Size and Distance	0.0856	0.0616	0.1648
$y = $Obesity				
High and Low Income	Size and Distance	-0.0832	0.0372	0.0254
High and Low Income	Tweets, Size and Distance	-0.0815	0.0382	0.0329
log(Income)	Size and Distance	-0.0832	0.0372	0.0254
log(Income)	Tweets, Size and Distance	-0.0805	0.0382	0.0351

size in our data ($\rho = 0.4185386$), indicating that the Twitter data in fact could provide new information or proxy for the desired amenities.

To test this hypothesis, we can use the twitter data described earlier and suggested in Equation 4. In this application, we again log-transform all the components of utility:

$$V_{ij} = \log(size_j)\lambda_s + \log(distance_{ij})\lambda_d + g(tweets_j)\lambda_t \tag{11}$$

where $g(x)$ is a Yeo-Johnson transformation to preserve cases where $log(0)$ would be otherwise undefined. As before, there is little information by which to determine the values of λ in this utility specification; we use the previous values with the addition of $\lambda_t = 0.1$. The ratio between the utility coefficients implies residents would travel almost twenty times as far or use a park three times smaller if it had twice the twitter activity, all else equal.

Figure 2 shows the percent change in the normalized log sum statistic after including the twitter infromation. There is not a recognizable overarching pattern to the changes, though tracts immediately surrounding Prospect Park and Central Park gain even more relative accessibility, and tracts in midtown Manhattan, Queens, and southern Bronx tend to lose some.

Table **??** presents the estimated coefficients relating accessibility to obesity and physical activity rates, with and without the inclusion of Twitter data. In all four cases, including twitter information in the accessibility statistic tempers the strength of the relationship with the dependent variable and modestly widens the standard error without substantively affecting the significance of the test statistics. From this limited example, it appears Twitter activity does not enhance the ability of parks to induce physical activity rates or lower obesity rates, though further investigation of this and other park amenities would certainly be warranted.

Limitations and Future Research Direction

We readily acknowledge limitations in this study. As in any study conducted with areal data, we at risk of falling victim to the ecological inference fallacy, where aggregate statistics mask or contradict disaggregated or individual-level trends. A large-sample survey of individuals in New York City,

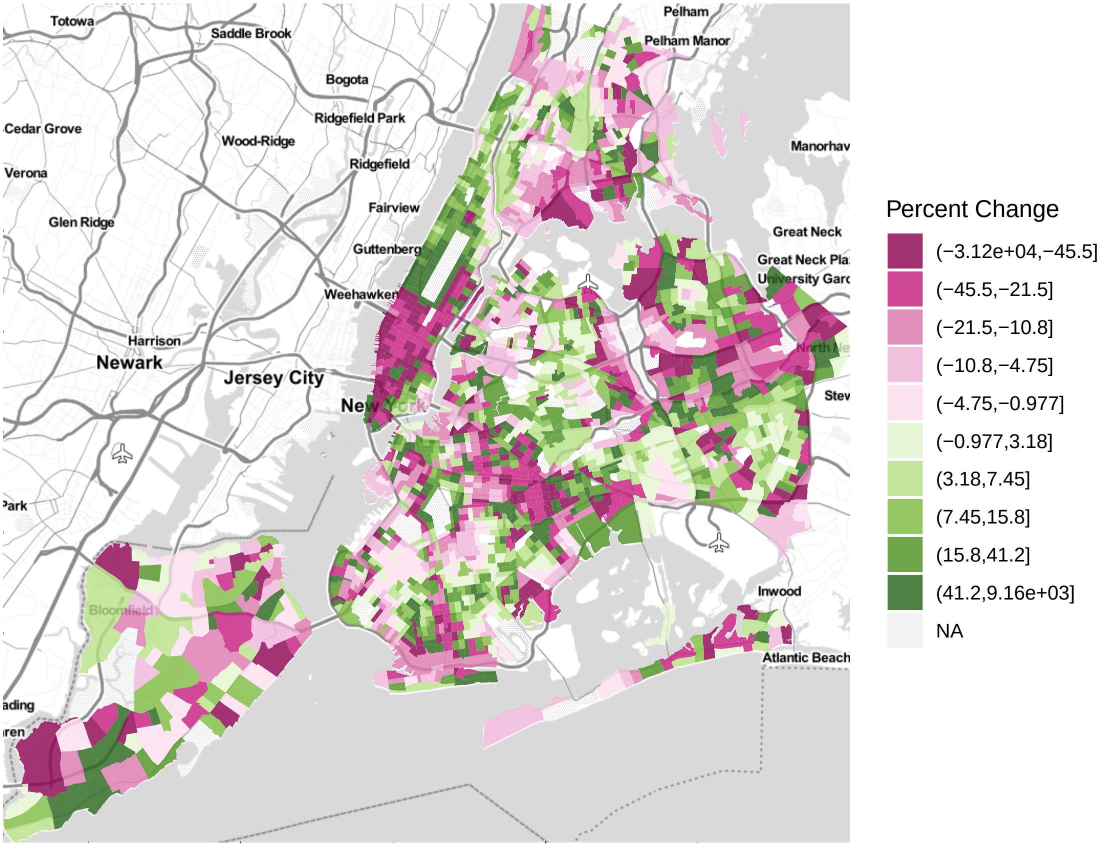

Figure 2. Percent change in relative accessibility when accounting for Twitter activity.

including measured physical activities and obesity would always be prefferable to the tract-level data used in this study. It would be also be preferrable to have obtained the λ accessibility utility coefficients from a high-quality survey of leisure destination choice rather than asserting them to match our prior expectations of what constitutes quality park access. An ideal survey to address the question would incorporate both sets of questions: physical activity and health data on one hand and park use (including which parks were used and how frequently) on the other. As no such dataset exists to our knowledge, this tract-level aggregate analysis with asserted utility coefficients is the possibility that remains.

This paper presents a holistic accessibility statistic that could, in theory, accommodate many attributes of the destination parks as well as the people who might use them. As an illustration: the park-going population could be separated into at least four delineated clusters, each preferring different amenities of a park:

- runners and cyclists: long, interesting trail systems
- sports players: soccer fields, basketball courts, or baseball diamonds
- families with small children: water features and playgrounds
- casual users: water features, gardens, performances, etc.

An analyst could then compute the accessibility logsum for each cluster with different utility values for each park's amenities, and obtain a measure of a neighborhood's accessibility to park features that its residents most care about. In this paper, we proceed only incrementally beyond this theory by adding Twitter activity as an an element of a park's attractiveness above and beyond its size. Exploring additional amenities or market segmentation strategies could provide a more definitive understanding of the relationship between park accessibility and health outcomes.

Finally, this study is primarily focused on the hypothesis that accessibility to parks encourages physical activity, which in turn reduces obesity. There are a multitude of other hypotheses that might be proposed and tested with the basic methodology we have employed here, or competing explanations for the outcomes we have observed. It is distinctly possible, for instance, that individuals who wish to exercise regularly in parks choose to live near them; this could potentially explain why the observed effects of accessibility seem partially dependent on the representation of income in our econometric model. Controlling for this potential self-selection effect would be necessary to isolate the exogenous impacts of park access on obesity or other health outcomes. And regarding these other variables: this study did not consider potential relationships between park access and hospitalization rates, life expectancy, respiratory disease, mental health, or any number of potential beneficial outcomes hypothesized or explored in the existing literature. Exploring these connections and their underlying mechanisms should be a priority as city planners and urban architects attempt to improve the quality of life of urban residents in the future.

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.

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.