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

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

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

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. The World

[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

Health Organization (2014) estimates approximately 300,000 premature deaths in the United States each year can be attributed to chronic diseases. Finkelstein et al. (2009) find an obese American spends nearly $1,500 more per year in health care costs than an American of normal weight. As obesity and chronic disease have become rampant, it is no surprise that healthcare costs have risen to nearly one-fifth of the United States' gross domestic product (Harnik and Welle 2011).

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 (Gordon et al. 2013). Wolf 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" (Wolf 2008, 22). Similarly, the United States Department of Health and Human Services notes that less than fifty-percent of Americans meet established recommendations for moderate to vigorous physical activity, or MVPA (U.S. DHHS 2010). 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.

Research has shown that the design of the built environment influences a range of behaviors, including those related to physical activity. The presence of trees and other vegetation in outdoor environments have been shown to be positively associated with physical activity (Pretty et al. 2005); one finding suggests that after sidewalks and trails have been constructed, the introduction of natural elements positively impacts motivation to engage in physical activity (Suminski et al. 2005). Additional evidence indicates that commonly vegetated areas, such as parks and open space, support outdoor physical activity Giles-Corti et al. (2005); Wells et al. (2007). Perhaps most telling is the finding that "people in large cities perceive themselves to be generally healthier if a greater percentage of the living environment is greenspace, are inclined to be more active, and claim the ability to relax faster" (Wolf 2008, 24). Thus, by providing space for active recreation, public parks and greenspaces may increase the likelihood of engaging in more physical activity. As such, public investment in parks can be thought of as a public health intervention for chronic diseases and conditions, as well as general population health.

To date, the literature exploring the relationship between parks and health outcomes, specifically those related to physical activity and obesity, has yielded mixed results. Importantly, there is considerable variation in the design of past studies, including the spatial scale of analysis, the population of interest, and the measure of proximity and/or accessibility. While studies of neighborhood level health impacts do exist, they typically focus on discrete populations of interest within individual parks instead of examining the impacts of larger park networks (Bancroft et al. 2015; Coutts 2008). Other studies have used the city and metropolitan statistical area as the spatial scale of analysis, which in some cases has resulted in the discovery of positive associations between increased park space and positive health impacts but may conceal more nuanced relationships that exist at smaller spatial scales (Larson et al. 2016). Evaluating the health impacts of parks and greenspaces using a smaller unit of analysis is an important research gap to fill, especially considering that park use and physical activity within parks varies considerably according to residential proximity to parks and park facilities as well as a number of sociodemographic factors (Kaczynski et al. 2014). There are important urban design considerations here; is it better to build a single large park with many different amenities, or to build a series of smaller parks nearer people's homes. Similarly, do parks placed near low-income households affect the behavior of those households?

This study attempts to fill that gap using New York City as a case study example. Using a new measure of park access called 'Park Choice Accessibility,' we find

Existing Literature

Urbanization and Public Health Outcomes

Across the world, the population of cities is growing at an unprecedented rate. With a current estimated global urban population of over 3.5 billion people, the United Nations Population Fund estimates that by 2030, 5 billion people will inhabit cities worldwide. By 2050, an additional 3 billion people will live in cities – a sixty percent increase over twenty years – with much of this growth expected to take place in the developing world (United Nations Population Fund, 2016). Despite the greater expected concentration of urban growth in other parts of the globe, the trend of urbanization is highly visible in the United States as well. According to the U.S. Census Bureau, "The nation's urban population increased by 12.1 percent from 2000 to 2010, outpacing the nation's overall growth rate of 9.7 percent for the same period." Nowak and Walton (2005) estimate that urban land area as a percentage of total land area in the U.S. will increase from 3.1% in 2000 to approximately 8.1% by the year 2050, with urbanized places collectively comprising a land area larger than the state of Montana. More than eighty percent of the U.S. population now lives in urban areas, compared to sixty-four percent in 1950 (United States Census Bureau 2007).

The global trend of urbanization has profound implications for population health, both positive and negative. Because urbanization corresponds with an increase in population density compared to rural and suburban settlement patterns, residents of dense urban areas throughout the world have better access on average to many health services and health-promoting amenities simply because of their closer proximity to such resources (Larson et al. 2016). However, despite this benefit of increased density and the ability of cities to provide many opportunities for innovation, economic growth and social progress, historical and current evidence suggests that rapid urban growth often leads to congestion and numerous negative environmental and human health outcomes. Interactions between growing urban populations and their environment, marked by intensive resource consumption and ecologically harmful patterns of development, lead to numerous undesired consequences including pollution and sanitation issues as well as racial and socioeconomic disparities (Larson et al. 2016).

Because of the considerable growth of urbanization throughout the world, many city planning and public health professionals have begun to pay more attention to the role of the built environment in promoting or discouraging health behaviors. Until recently, most large-scale health promotion efforts focused on individual-level interventions intended to educate people about health 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 Coutts (2008); Bedimo-Rung et al. (2005).

Review of the Health Benefits of Parks

A large body of literature exists documenting the many benefits of urban parks and greenspaces to human health and wellbeing. Importantly, these benefits stem from the provision of ecosystem services, which occur when the natural environment supplies something that people demand, improving quality of life and well-being (Larson et al. 2016). These services can include the provisioning of goods such as fresh water and agricultural products; regulatory functions including protection of drinking water quality, heat mitigation, air purification, and stormwater management; and cultural functions, such as improving aesthetics, providing opportunities for recreation, tourism, and physical and mental health, and promoting biodiversity (**?**). According to Wolch et al. (2014), parks and greenspaces provide ”. . . a wide range of ecosystem services that could help combat many urban ills and improve life for city dwellers—especially their health. . . Ecosystem services provided by urban greenspace not only support the ecological integrity of cities, but can also protect the public health of urban populations.” Table 1 summarizes commonly cited benefits of parks and greenspaces.

Table 1. Commonly cited benefits of parks and greenspaces

Category	Summary of Benefits
Physical Health	Provide clean drinking water (Benedict and McMahon 2006) Foster increased physical activity Promote faster healing in hospitals (Akbari et al. 2001) Reduce of heat-related mortality Reduce incidence of cardiovascular-related mortality Improved air quality and related reductions in respiratory-related mortality (Lovasi et al. 2008) Reduce incidence of childhood asthma (Wolf 2008) Improve birth outcomes
Mental Health	Reduce stress and mental fatigue Reduce aggression Enhance emotional and cognitive development Improve behavioral outcomes in youth
Social Health	Enhanced community aesthetics Reduce crime Increase social interaction
Economic Health	Provide of ecosystem services Increase residential property values and municipal property tax revenues Attract more shoppers and increase economic activity to commercial districts

Given the general benefits of parks and greenspaces in Table 1, several studies have attempted to quantify the impacts of these spaces on different facets of health and wellbeing at the city level, yielding mixed results. Larson et al. (2016) used self-reported scores on the Gallup-Healthways Wellbeing Index to evaluate the relationship between different areas of wellbeing, including physical health, and park quantity, quality, and accessibility in 44 U.S. cities. The authors found that “Park quantity (measured as the percentage of city area covered by public parks) was among the strongest predictors of overall wellbeing, and the strength of this relationship appeared to be driven by parks' contributions to physical and community wellbeing" (Larson, Jennings & Cloutier 2016). 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. In a study of New York City, Stark et al. (2014) found that the "proportion of neighborhoods that was large or small park space and park cleanliness were associated with lower BMI among NYC adults after adjusting for other neighborhood features such as homicides and walkability, characteristics that could influence park usage."

Interestingly, in a study by Richardson et al. (2012) that examined the relationship between urban greenspace and selected mortality rates, 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, but found 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 that were included in the meta-analysis by Bancroft et al. (2015) -– were conducted at the level of the city or metropolitan statistical area, potentially concealing variation in health outcomes at more fine-grained levels of analysis (Larson et al. 2016).

Accessibility to Parks

To evaluate the impacts of parks and greenspaces on health outcomes at any scale, 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 active accessibility metric employed: "distance-based, gravity-based or potential, topological or infrastructure-based, and walkability and walk score-type measures." Distance-based measures account only for the Euclidean distance between origins and destinations, while infrastructure-based measures explicitly incorporate relevant transportation networks like roads and sidewalks to more accurately measure travel time and distance. Gravity-based measures incorporate cost measures to model accessibility as a function of a destination's attractiveness (i.e., size, commercial activity, etc.) and the cost of traveling to that destination from a given origin (i.e., travel time or distance).

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." However, their meta-analysis highlights some of the limitations associated with each of the types of metrics described. For example, in describing distance-based accessibility measures, the authors note that such measures are ”extremely sensitive to the way in which travel impedance (i.e., distance) is measured. Accordingly, four types of distance can be identified: Euclidean distance, Manhattan distance, shortest network distance, and shortest network time”. The appropriateness of one distance-based measure over another can vary significantly depending on the topography of the environment and the travel mode that is being employed. In describing gravity-based accessibility measures, which ”assume that travel is a derived demand and there is a tradeoff between the benefit of the opportunity and the cost to reach it from a given origin,” the authors note that such measures do not always explicitly account for land use characteristics near origins and destinations, which may impact that true accessibility of those places.

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. The authors concede that many existing approaches have long been necessary because the computational power required to use higher-resolution analytical techniques was unavailable. However, due to recent advances in computation power and the advent of municipal open-data policies, Logan et al. recommend that future analyses of accessibility disaggregate population data to the building or parcel level and use network distance instead of Euclidean distance to measure proximity.

Choice-based Accessibility

Consider that an individual is choosing a park for a recreation activity. According to basic choice theory (McFadden 1974), the probability of choosing park p from the set of all regional parks J is:

$$P(p|V_p) = \frac{\exp(V_p)}{\sum_{j \in J} \exp(V_j)}$$

where parks are differentiated from each other by their relative measurable utilities V. 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

$$V_{ij} = size_j \beta_s + distance_{ij} \beta_d$$

incorporating the size of the park in acres and the distance of the park from the census tract i in miles. The coefficients β are typically estimated from surveys, though in the absence of a survey we apply a maximum likelihood technique described below.

A key theoretical understanding of random utility choice models is that the consumer surplus of the choice set can be otained as the log-sum of the denominator of the choice probability equation. In plainer terms, the *total value* of an individual’s park accessibility is defined to be:

$$CS_i = \ln \left(\sum_{j \in J} \exp(V_{ij}) \right)$$

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 limit of 1/2 mile or so. 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 New York City's Central Park 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 (**?**). The extension

Empirical Application

Data

```
tracts <- geojson_read("data/nyc_tracts.geojson", what = "sp")

# identify complete cases
complete_tracts_index <- tracts@data %>%
  tbl_df() %>%
  complete.cases()
tracts <- tracts[complete_tracts_index, ]


tracts@data <-  tbl_df(tracts@data) %>%
  dplyr::transmute(
    geoid = GEOID,
    county_fips = substr(GEOID, 3, 5),
    borough = case_when(
      county_fips == "081" ~ "Queens",
      county_fips == "047" ~ "Brooklyn",
      county_fips == "061" ~ "Manhattan",
      county_fips == "005" ~ "Bronx",
      county_fips == "085" ~ "Staten Island",
      TRUE ~ as.character(NA)
    ),
    obesity = OBESITY,
    physact = 100 - Phys_Act,
    log_obesity = log(OBESITY),
    log_physact = log(100 - Phys_Act),
    density = Pop_Density,
    fulltime = FulltimeWork, college = CollegeDeg,
    single = Single_Percent,
    black = PctBlack, asian = PctAsian, hispanic = PctHispanic,
    other = PctNative + PctPacific,
```

```
    Pct0to17, Pct18to29, Pct65plus,
    Income1, Income2, Income3, Income4, Income5, Income6,
    Income7, Income8, Income9, Income10
  )
```

This study used data available to the public from a variety of agencies and data providers. We obtained data on aggregate health outcomes from the Centers for Disease Control and Prevention (CDC) for census tracts in five boroughs of New York City. After removing tracts with missing population information, we have Sexprnrow(tracts) complete cases. Table 2 presents key descriptive statistics for these data.

```
parks <- sf::st_read("data/nyc_parks.geojson", quiet = TRUE) %>%
  transmute(
    size = Park_Acres,
    log_size = log(Park_Acres),
    tweets = TWEET_COUNT,
    log_tweets = yeo.johnson(TWEET_COUNT, 0)
  )
```

We also collected data on park geometry.

Model

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

$$\log(Obesity)_i = f(\boldsymbol{X}_i, A_i, g(W, (Obesity), \boldsymbol{X})) \tag{1}$$

```
base_formula <- formula(
  ~ log(density) +
    fulltime + college + single +
    Pct0to17 + Pct18to29 + Pct65plus + # need to leave out a category for coll
    black + asian + hispanic + other)
```

where $\boldsymbol{T}_i$ is a vector of tract-level attributes, A_i is a choice-based accessibility measure for the tract, and $g(W, (Obesity), \boldsymbol{T})$ 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 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:

$$\boldsymbol{y} = \rho W \boldsymbol{y} + X\boldsymbol{\beta} + WX\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{2}$$

Table 2. Summary Statistics of Tract Data

```
our_summary <- list(
  "obesity"     = tab_summary(tracts$obesity),
  "physact"     = tab_summary(tracts$physact),
  "pop_density" = tab_summary(tracts$density),
  "fulltime"    = tab_summary(tracts$fulltime),
  "college"     = tab_summary(tracts$college)
)

summary_table(tracts@data, our_summary) %>%
  kable("latex", booktabs = T, escape = FALSE) %>%
  group_rows("Obese: percent of tract which is obese", 1, 4) %>%
  group_rows("Physical Activity: Percent of tract reporting physical activity",
  group_rows("Population Density: individuals per square mile", 9, 12)  %>%
  group_rows("Full Time: percent of adults who work full time", 13, 16) %>%
  group_rows("College: percent of adults with a college degree", 17, 20)
```

	tracts@data (N = 2102)
Obese: percent of tract which is obese	
min	11.7
median (IQR)	26.00 (20.50, 32.30)
mean (sd)	26.51 ± 7.21
max	48.2
Physical Activity: Percent of tract reporting physical activity	
min	42.5
median (IQR)	71.20 (65.53, 76.50)
mean (sd)	71.19 ± 8.06
max	89.8
Population Density: individuals per square mile	
min	0.05951505
median (IQR)	69.37 (39.89, 107.91)
mean (sd)	81.50 ± 55.92
max	377.7478
Full Time: percent of adults who work full time	
min	1.9
median (IQR)	55.40 (49.10, 60.80)
mean (sd)	54.91 ± 9.28
max	100
College: percent of adults with a college degree	
min	2
median (IQR)	27.80 (18.20, 42.08)
mean (sd)	33.25 ± 20.40
max	92.3

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:

$$W_{ij} = \begin{cases} \frac{1/d_{ij}}{\sum_{k=1}^{n}(1/d_{ik})} & \text{for } d_{ij} \leq 1.8 \text{ miles} \\ 0 & \text{for } d_{ij} > 1.8 \text{ miles} \end{cases} \tag{3}$$

```
pop_weighted_centroids <- read_csv("data/CenPop2010_Mean_TR36.csv") %>%
  mutate(geoid = str_c(STATEFP, COUNTYFP, TRACTCE))

## Parsed with column specification:
## cols(
##   STATEFP = col_double(),
##   COUNTYFP = col_character(),
##   TRACTCE = col_character(),
##   POPULATION = col_double(),
##   LATITUDE = col_double(),
##   LONGITUDE = col_double()
## )

pop_weighted_centroids <- left_join(tracts@data, pop_weighted_centroids, by = '
  dplyr::select(geoid, LONGITUDE, LATITUDE) %>%
  st_as_sf(crs = 4326, coords = c("LONGITUDE", "LATITUDE")) %>%
  st_transform(2263)

## Warning: Column `geoid` joining factor and character vector,
coercing into character vector

tracts.dnn <- dnearneigh(gCentroid(as(pop_weighted_centroids, "Spatial"), byid
                         0, 1.8 * 5280)
dists <- nbdists(tracts.dnn, gCentroid(tracts, byid = TRUE))
dists.inv <- lapply(dists, function(x) 1 / x)
W <- nb2listw(neighbours = tracts.dnn, glist = dists.inv,
              zero.policy = TRUE, style = "W")

## Warning in nb2listw(neighbours = tracts.dnn, glist = dists.inv,
zero.policy = TRUE, : zero sum general weights

trMC <- trW(as(W, "CsparseMatrix"), type="MC") # trace used in montecarlo impac
```

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.

```
#' Calculate destination choice logsums from a distance matrix
#'
#' @param d An $n\times p$ matrix with the distance from all tracts to
#'   all parks
#' @param sizes A p-length vector of park sizes
#' @param tweets A p-length vector of tweets at parks
#' @param betas A vector containing the size, distance, and tweet coefficients
#'   If we submit two variables the tweets are ignored.
#'
#' @return An n-length vector containing the weighted log-sum based
#'   accessibility between a tract and all parks.
#' @details If we have n tracts and p parks, distances needs to be a
#'
calculate_park_logsums <- function(d, sizes, tweets = NULL,
                                   betas = c(-0.4,0.2,0.001)){

  # A is n x p
  a <- betas[1] * d

  # B is p x 1
  b <- betas[2] * sizes

  if(!is.null(tweets))
```

```
    b <- b + betas[3] * tweets

  # calculate observed utility by adding the weighted park-level attributes
  # to the columns of the matrix
  # V is n x p, with b added by-column to each element in a
  V <- sweep(a, 2, b, `+`)

  # log-sum of exponentiated utility, Output is n-length vector
  log(rowSums(exp(V)))

}
```

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

```
# calculate centroids of tracts and parks

parks <- parks %>% st_transform(st_crs(pop_weighted_centroids))

# distances from all tracts to all zones
distances <- st_distance(pop_weighted_centroids, parks, byid = TRUE) %>%
  units::set_units(miles) %>%
  units::drop_units()

# assert that a tract must be at least 1/10 mile from a park
distances <- pmax(distances, 0.1)
```

```
tracts@data <- tracts@data %>%
  mutate(
    access_ls = calculate_park_logsums(log(distances), parks$log_size,
                                       betas = c(-0.42, 1.0)),
    tweets_ls = calculate_park_logsums(log(distances), parks$log_size, parks$lo
                                       betas = c(-0.42, 1.0, 0.1)),
    access_ls = (access_ls - mean(access_ls) )/ sd(access_ls),
    tweets_ls = (tweets_ls - mean(tweets_ls) )/ sd(tweets_ls)
  )

tracts_sf <- tracts %>% st_as_sf() %>% st_transform(4326)
```

```r
# this interactive map is much easier to create and explore
pal <- colorQuantile("Spectral", tracts_sf$access_ls, n = 5)
leaflet(tracts_sf) %>%
  addProviderTiles(providers$CartoDB.Positron) %>%
  addPolygons(group = "access", color = ~ pal(access_ls),
              label = ~as.character(round(access_ls,2))) %>%
  addPolygons(group = "tweets", color = ~ pal(tweets_ls),
              label = ~as.character(round(tweets_ls,2))) %>%
  addLayersControl(baseGroups = c("access", "tweets"))
```

Results

We estimate

```r
obese_base_sem <- errorsarlm(update(base_formula, log_obesity ~ .),
                             data = tracts@data, listw = W, zero.policy = TRUE)
obese_base_sdm <- lagsarlm(update(base_formula, log_obesity ~ .),
                           data = tracts@data, listw = W, zero.policy = TRUE,
                           type = "mixed")
test_sdmsem <- lmtest::lrtest(obese_base_sem, obese_base_sdm)
```

A likelihood ratio test reveals that the SEM is not preferred, so we use the SDM only going forward.

```r
# estimate models
access_sdm <- update(obese_base_sdm, .~ . + access_ls)
tweets_sdm <- update(obese_base_sdm, .~ . + tweets_ls)
```

Conclusion

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.

Table 3. Estimated Coefficients on Obesity

```
library(texreg)
obesity_models <- list(Base = obese_base_sdm, `Access` = access_sdm,
                       `Access + Tweets` = tweets_sdm)
texreg(obesity_models, digits = 4, table = FALSE, single.row = TRUE,
       booktabs = TRUE, use.packages = FALSE)
```

	Base	Access	Access + Tweets
(Intercept)	2.8025 (0.0705)***	2.8074 (0.0717)***	2.8233 (0.0712)***
log(density)	0.0058 (0.0024)*	0.0059 (0.0024)*	0.0060 (0.0024)*
fulltime	−0.0013 (0.0002)***	−0.0013 (0.0002)***	−0.0013 (0.0002)***
college	−0.0044 (0.0002)***	−0.0044 (0.0002)***	−0.0044 (0.0002)***
single	0.0024 (0.0003)***	0.0024 (0.0003)***	0.0024 (0.0003)***
Pct0to17	0.0036 (0.0004)***	0.0035 (0.0004)***	0.0035 (0.0004)***
Pct18to29	−0.0035 (0.0004)***	−0.0035 (0.0004)***	−0.0035 (0.0004)***
Pct65plus	−0.0011 (0.0004)**	−0.0011 (0.0004)**	−0.0011 (0.0004)**
black	0.0033 (0.0002)***	0.0033 (0.0002)***	0.0033 (0.0002)***
asian	−0.0030 (0.0002)***	−0.0030 (0.0002)***	−0.0030 (0.0002)***
hispanic	0.0020 (0.0002)***	0.0021 (0.0002)***	0.0021 (0.0002)***
other	−0.0004 (0.0010)	−0.0004 (0.0010)	−0.0003 (0.0010)
lag.log(density)	−0.0497 (0.0055)***	−0.0562 (0.0057)***	−0.0591 (0.0058)***
lag.fulltime	−0.0065 (0.0006)***	−0.0059 (0.0006)***	−0.0057 (0.0006)***
lag.college	0.0018 (0.0004)***	0.0015 (0.0004)***	0.0013 (0.0004)**
lag.single	−0.0017 (0.0008)*	−0.0011 (0.0008)	−0.0011 (0.0008)
lag.Pct0to17	−0.0139 (0.0011)***	−0.0126 (0.0011)***	−0.0125 (0.0011)***
lag.Pct18to29	−0.0099 (0.0016)***	−0.0097 (0.0016)***	−0.0097 (0.0016)***
lag.Pct65plus	−0.0106 (0.0011)***	−0.0108 (0.0011)***	−0.0111 (0.0011)***
lag.black	−0.0022 (0.0003)***	−0.0023 (0.0003)***	−0.0023 (0.0003)***
lag.asian	−0.0008 (0.0003)**	−0.0009 (0.0003)**	−0.0008 (0.0003)**
lag.hispanic	−0.0006 (0.0003)	−0.0012 (0.0004)**	−0.0012 (0.0004)**
lag.other	0.0096 (0.0047)*	0.0134 (0.0048)**	0.0133 (0.0048)**
ρ	0.5224 (0.0231)***	0.5073 (0.0233)***	0.5061 (0.0232)***
access_ls		−0.0052 (0.0052)	
lag.access_ls		0.0155 (0.0062)*	
tweets_ls			−0.0022 (0.0053)
lag.tweets_ls			0.0146 (0.0064)*
Num. obs.	2102	2102	2102
Parameters	25	27	27
Log Likelihood	2738.0669	2747.6239	2749.7162
AIC (Linear model)	-4999.8214	-5047.4899	-5050.1421
AIC (Spatial model)	-5426.1338	-5441.2478	-5445.4325
LR test: statistic	428.3124	395.7579	397.2904
LR test: p-value	0.0000	0.0000	0.0000

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

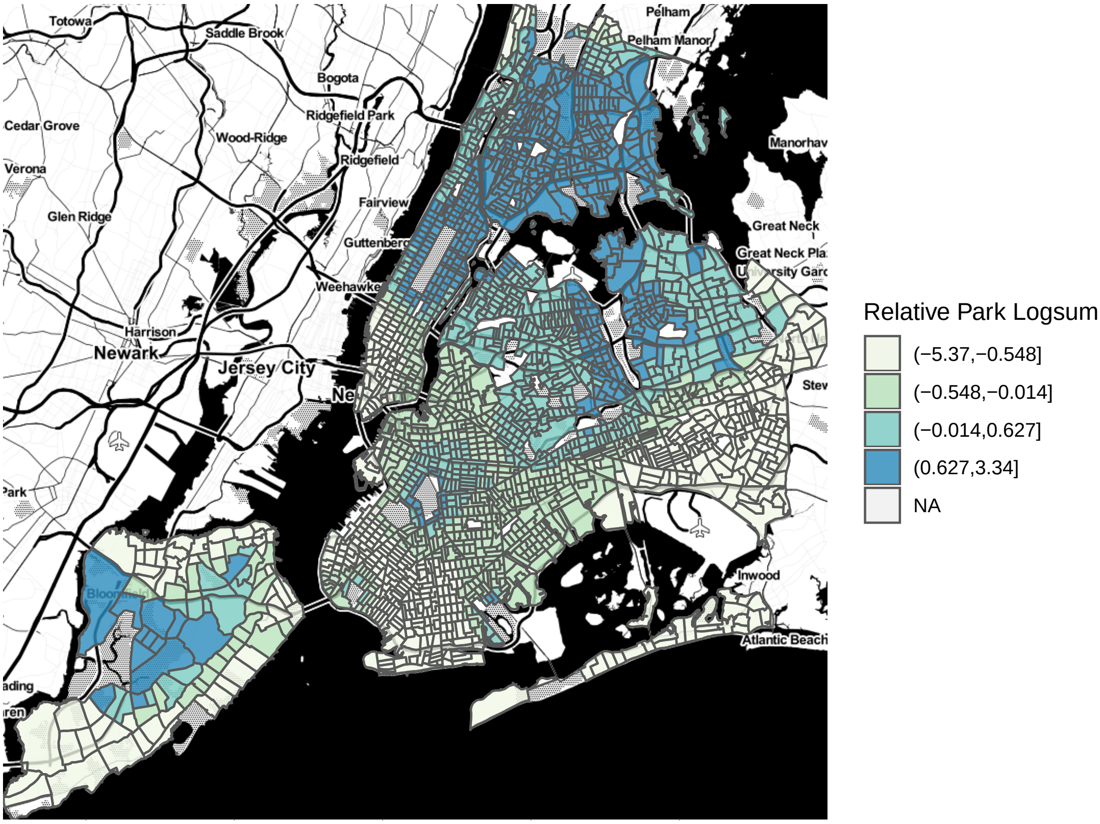

Figure 1. Relative surplus-based accessibility in New York City

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.

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.

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.

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.

Gordon B, Wang Z and Luo X (2013) Chronic disease and the link to physical activity. *Journal of Sport and Health Science* 2(1): 3–11. DOI:10.1016/J.JSHS.2012.07.009.

Harnik P and Welle B (2011) *From fitness zones to the medical mile: How urban park systems can best promote health and wellness*. Trust for Public Land.

Kaczynski AT, Besenyi GM, Stanis SAW, Koohsari MJ, Oestman KB, Bergstrom R, Potwarka LR and Reis RS (2014) Are park proximity and park features related to park use and park-based physical activity among adults? Variations by multiple socio-demographic characteristics. *International Journal of Behavioral Nutrition and Physical Activity* 11(1): 146. DOI:10.1186/s12966-014-0146-4.

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.

Nowak DJ and Walton JT (2005) Projected Urban Growth (2000–2050) and Its Estimated Impact on the US Forest Resource. *Journal of Forestry* 103(8): 383–389. DOI:10.1093/jof/103.8.383.

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.

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.

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.

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