ACSP
Connecting educators, researchers and students

Journal of Planning Education and Research
1–16

DOI: 10.1177/0739456X20915765
journals.sagepub.com/home/jpe
SAGE

Exploring the Nonlinear Relationship between the Built Environment and Active Travel in the Twin Cities

Tao Tao[1], Xinyi Wu[1], Jason Cao[1], Yingling Fan[1], Kirti Das[1], and Anu Ramaswami[2]

Abstract

Active travel is important to public health and the environment. Previous studies substantiate built environment influences active travel, but they seldom assess its overall contribution. Most of the studies assume that built environment characteristics have linear associations with active travel. This study uses Gradient Boosting Decision Trees to explore nonlinear relationships between the built environment and active travel in the Twin Cities. Collectively, the built environment has more predictive power for active travel than demographics, and parks, proximity to downtown, and transit access have important influences. The threshold effects of built environment variables help inform planning practice.

Keywords

machine learning, travel behavior, land use, walking, community design

Introduction

Auto dependence caused by urban sprawl results in many problems, such as traffic congestion, air pollution, and inadequate physical activity, which increase economic costs to society, lead to public health issues and reduce quality of life (Brownson et al. 2009). Urban planners have deployed various policies (such as pedestrian-oriented development and compact development) to mitigate auto dependence and to promote active travel. Active travel (or active transportation) refers to nonmotorized modes of transport, such as walking and cycling (Litman 2003). It is not only a substitute for auto use, but it is also an activity itself. It carries benefits to both society and individuals. To inform planning for active travel, previous studies examined the relationship between the built environment and active travel. Through a meta-analysis, Ewing and Cervero (2010) found that many built environment characteristics, such as density, diversity, design, destination accessibility, and distance to transit, are associated with walking. These findings provide planners supportive evidence to encourage active travel through altering the built environment (Transportation Research Board and Institute of Medicine 2005).

There are two major gaps in the literature (Van Wee and Handy 2016). First, a limited number of studies assess the collective influence of the built environment on active travel compared with other variables, such as demographics. This question is relevant to planners' efficacy of using land-use policies to enhance active travel (Singh et al. 2018). Second, previous studies generally assume a linear relationship between the built environment and active travel (Boarnet, Greenwald, and McMillan 2008) or rely on a predefined nonlinear relationship (Heesch, Giles-Corti, and Turrell 2015). These assumptions disguise the most effective range that a built environment attribute affects active travel (Ding, Cao, and Næss 2018). For instance, as densification increases development costs exponentially, planners wonder how dense is enough to promote active travel. Once they figure out the threshold, they can plan for active travel efficiently.

The purpose of this study is to assess the collective contribution of the built environment to active travel and to scrutinize the potential nonlinear relationship between them. It uses the gradient boosting decision tree (GBDT) approach to the data collected in Minneapolis to answer three research questions:

> **Research Question 1:** Relative to demographics, how much do built environment characteristics collectively contribute to active travel?

Initial submission, December 2018; revised submissions, September 2019; final acceptance, February 2020

[1]University of Minnesota, Minneapolis, MN, USA
[2]Princeton University, Princeton, NJ, USA

Corresponding Author:
Jason Cao, Humphrey School of Public Affairs, University of Minnesota, Twin Cities, 301 19th Avenue South, Minneapolis, MN 55455, USA.
Email: cao@umn.edu

> **Research Question 2:** Among the built environment attributes, which one plays a more important role in affecting active travel?
>
> **Research Question 3:** To what extent should planners change the built environment to efficiently encourage active travel?

This study contributes threefold to the literature. First, we collected active travel data through a self-developed smartphone application. The new data differ from those collected through travel diaries and accelerometers. Short walking trips are often underreported in travel diaries (Neves and Brand 2019). Accelerometers cannot account for activities such as cycling (Strath et al. 2013). Smartphone applications could overcome these disadvantages. It automatically detects respondents' movements and requires them to manually report the mode of each trip. Second, we found that the built environment collectively has a stronger association with active travel than demographic characteristics, challenging the traditional belief that built environment variables are secondary to demographics. Third, this study uses GBDT to validate the nonlinear association between the built environment and active travel. We concluded that built environment variables have prevalent threshold effects on active travel, and the nonlinear patterns vary by variable, challenging the linearity assumption commonly adopted in the literature.

This paper is organized as follows. In the next section, we review the literature on the relationship between the built environment and active travel and identify research gaps. In the third section, we introduce the data and the method used in this study. In the fourth section, we discuss the results. In the last section, we summarize the key results and discuss the associated implications for planning practice.

Literature Review

The Built Environment-Active Travel Connection

A growing number of studies examine the correlation between built environment attributes and active travel and offer great insights for both planning researchers and practitioners. Some studies focused on a single dimension of the built environment, such as land-use types in Oliver et al. (2007), land-use mix in Duncan et al. (2010), and urban design in Giles-Corti et al. (2013). However, they overlooked the effects of other built environment variables. The built environment is a multidimensional concept, consisting of land-use patterns, urban design, and transportation systems (Handy et al. 2002). Land-use policies often bring about changes in several dimensions of the built environment simultaneously. For example, a transit-oriented development may lead to an increase in transit access and density and a reduction in parking supply. Ignoring other built environment dimensions will overstate the effect of the particular dimension included in the analysis and result in omitted variable bias. Furthermore, planners may have to prioritize certain built environment dimensions over others because of budget constraints (Singh et al. 2018). Thus, it is necessary to consider multiple dimensions of the built environment when exploring its relationship with active travel and assess their relative importance.

Many studies explore the effects of multiple built environment dimensions (density, diversity, design, distance to transit, and destination accessibility) on active travel. Cervero and Duncan (2003) investigated the influences of density, diversity, and design on walking and cycling choices, and found that land-use diversity is the strongest predictor for walking. Beenackers et al. (2012) concluded that functional built environment attributes such as density and destination accessibility are more important to transport-related cycling, whereas neighborhood layout, such as street connectivity, is more important to recreational cycling. As the number of empirical studies grows, Ewing and Cervero (2010) conducted a meta-analysis to quantify the effect size of each built environment dimension on walking. They concluded that three built environment attributes are most strongly related to walking: land-use diversity, intersection density, and the number of destinations within walking distance. The elasticity of these three attributes ranges from 0.2 to 0.4. They also argued that the combined effect of several built environment variables on travel behavior can be much higher (Ewing and Cervero 2017).

However, the combined effect of built environment variables is not additive because they are correlated with one another. Some studies start to evaluate the collective impact of built environment variables on travel behavior, as a response to a call from scholars for research (Van Mokhtarian and Cao 2008; Wee and Handy 2016). Singh et al. (2018) quantified the relative contribution of various categories of factors to household vehicle miles traveled (VMT) and concluded that the built environment explains 12 percent of the variation in household VMT. Ding, Cao, and Næss (2018) found that built environment characteristics account for about 63 percent of the predictive power for driving distance on weekdays and about 48 percent for driving distance on weekends. However, a limited number of studies assess the combined influence of the built environment on active travel after controlling for other factors. This study aims to fill the gap by quantifying the extent to which the built environment collectively contributes to active travel and evaluate the relative contribution of multiple built environment dimensions. The results can help planners identify improvement priorities and support the design of effective planning policies.

The Nonlinear Impact of the Built Environment

Although previous studies shed light on the correlation between the built environment and active travel, there is limited discussion on the shape of this relationship. Many studies hypothesized that it is linear (Boarnet, Greenwald, and

McMillan 2008; Kitamura, Mokhtarian, and Laidet 1997; M. Zhang 2004; L. Zhang et al. 2012). However, the built environment could have a nonlinear relationship with travel behavior (Ding, Cao, and Næss 2018; Van Wee and Handy 2016). This nonlinearity indicates that as a built environment attribute changes, its impact on travel behavior also changes. It is in the interest of planners to identify the effective ranges of the built environment variables, where land-use and transportation policies would have the greatest impact.

Some studies recognize the potential nonlinearity and address it mainly in two ways: variable transformation and level classification. When transforming variables, researchers often apply modeling techniques, such as log-linear or Poisson family regression, to examine the data (Schoner and Cao 2014; Wijk et al. 2017). Level classification, a more exploratory approach, is also used in some studies. For example, Heesch, Giles-Corti, and Turrell (2015) divided built environment attributes into four quartiles and ran unordered logistic models. They discovered a U-shaped relationship between neighborhood socioeconomic disadvantages and cycling for transport. Specifically, residents living in both the most advantaged and disadvantaged neighborhoods have higher odds of cycling than residents living in mid-level disadvantaged neighborhoods.

Variable transformation and level classification both involve certain degrees of preassumptions. They test the significance of independent variables but are restrictive in depicting the actual relationship between variables because of the assumptions. Alternatively, some researchers explore the nonlinearity through visualization. By plotting the proportion of each mode against built environment attributes, Frank and Pivo (1994) explored whether mixed land-use and density have nonlinear relationships with mode choices among solo-driving, transit, and walking. Results showed that the number of walking trips remains constant initially and then increases substantially after employment density and population density reach certain thresholds.

Although visualization is capable of illustrating the actual relationship without predefined restrictions, it suffers from confounding effects of other factors, such as socioeconomic and demographic attributes. Recently, scholars have used the GBDT approach to depict the relationship between the built environment and travel behavior, while controlling for other factors. Ding, Cao, and Næss (2018) examined the nonlinear effect of built environment characteristics to conclude that population densification with twelve kilometers of the city center is the most effective in reducing driving distances in Oslo. By analyzing travel diary data in the Washington metropolitan area, Ding, Cao, and Wang (2018) found that population density around residential areas has a negative relationship with car mode choice for commuting, but its effect diminishes when density reaches about thirty-five persons per acre. Furthermore, bus stop density has a limited effect on the probability of car commuting within the range of .20 counts per acre, but reduces the probability dramatically beyond the threshold. These

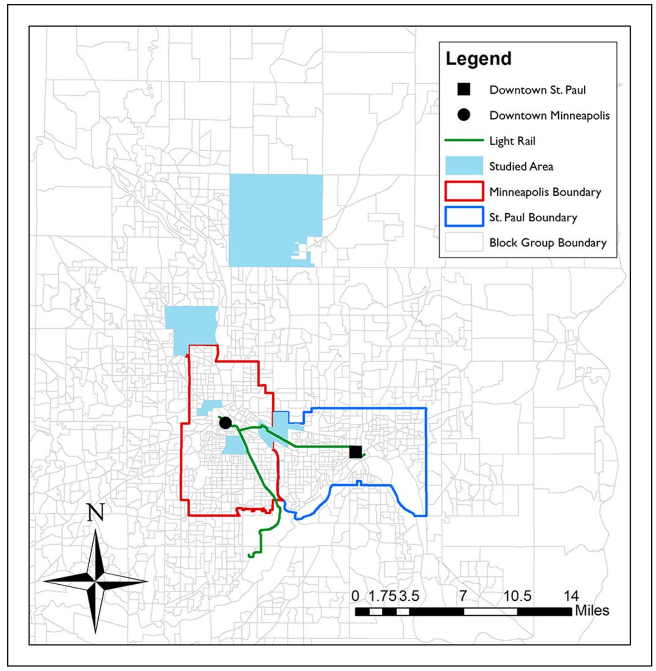

Figure 1. Studied area.

studies suggest that built environment variables may not follow a linear or log-linear relationship with travel behavior. Some attributes may be influential only after reaching a certain threshold, while others may not be more effective beyond a certain threshold. That is, the built environment may have threshold effects on travel behavior. Moreover, threshold effects vary by variable. A preassumed relationship is unable to uncover the varying patterns of relationships.

It is important to scrutinize the nonlinear relationship. Overlooking the nonlinear association between the built environment and active travel may misestimate built environment effects and hence offer erroneous implications for planning practice. Furthermore, planning agencies often provide guidelines for certain planning purposes. For example, the recommended minimum residential density for a bus-based neighborhood transit-oriented development is twelve dwelling units per acre in San Diego (Cervero et al. 2004). An analogy example is how dense is enough to facilitate active travel efficiently. However, the literature provides limited evidence. This study used the GBDT approach to better identify threshold effects. The results offer insights into the ranges of the built environment attributes that shape active travel most effectively.

Data and Method

Data Source

The data came from the *Neighborhood Environment*, *Daily Activities*, and *Well-Being Study*. We administered a survey in the Minneapolis–St Paul Twin Cities area (Figure 1) from

October 17, 2016, to October 25, 2017. We selected six neighborhoods based on urban location versus suburban location, proximity to light rail transit, and income level. For urban areas, we selected one low-income and one middle-income neighborhood *without* light rail access and one low-income and one middle-income neighborhood *with* light rail access. For suburban areas, we chose one low-income and one middle-income neighborhood.

We recruited residents of randomly selected blocks (921 of the 2,443 census blocks in the study neighborhoods) to attain a sample size of four hundred. We conducted a power analysis and found that the sample size should be at least 269. We increased it to four hundred to accommodate for unexpected data issues. We sent a postcard to all homes on the selected blocks with a brief study description and the contact information of the research team. Appendix A presents the content of the postcard. Once contacted, the survey team explained the study in detail to the participants, and those who chose to participate made an appointment with the survey team. Then, a survey team member visited the participants at home to collect their demographic information and provide instructions on how to use a smartphone application called Daynamica™ (previously called SmarTrAC; Fan et al. 2015). Figure 10 in Appendix B presents a brief user guide for this app. The application tracks and automatically breaks down the respondent's day into a series of trips and activities for which they provide additional information (such as companion information and trip purpose). We tracked respondent trip data for seven consecutive days. The recorded information includes trip distances, trip durations, and travel modes. The data collection took a year and sample enrollment ranged from five to fifteen every week. An issue of this long time period is the changing weather throughout the year. We included the average daily temperature to account for temperature variation. Another issue with the data is the continued collection of data after the respondent had completed the study because the phone global positioning system (GPS) was still active. As the survey included an exit interview, we used the exit survey times to truncate the data. The interviewers also went over the data to ensure accuracy.

The dependent variable in this study is the average daily met minutes over the seven days. Met minute is a unit to describe the energy expenditure of an active trip (Physical Activity Guidelines Advisory Committee 2008). One met is defined as 1 kcal/kg/hour. Met minute is the product of met and duration (in minutes) of an activity. If an individual of 75 kg walks at the intensity level of three mets for twenty minutes, the associated energy expenditure is 75 kcal, or 60 met minutes. The Physical Activity Guidelines recommend five hundred to one thousand met minutes per week for health benefits. Met minute is aggregated using equation (1):

$$M = \sum_{j=1}^{q} \sum_{i=1}^{p_j} \frac{D_{ij} R_k}{q}, \tag{1}$$

Table 1. Met for Active Travel.

Speed (mph)	Met for walking
0–2	2
2–2.5	2.8
2.5–2.8	3
2.8–3.2	3.5
3.2–4	4.3
4–4.5	5
4.5–5	7
≥5	8.3
Speed (mph)	**Met for biking**
0–10	4
10–12	6.8
12–14	8
14–16	10
≥16	12

where M is the average daily met minutes of a respondent's active travel, D_{ij} is the duration in minutes of trip i on day j of a respondent, p_j is the number of trips on day j, q is the valid number of days for one respondent, and R_k is the met for the corresponding travel mode of the trip (Table 1), which is based on the research of Ainsworth et al. (2011). After data cleaning, 299 observations were included for further analysis.

Built environment features from publicly accessible data sets (Minnesota Geospatial Commons: http://gisdata.mn.gov) were assembled around the respondents' home locations. To define a neighborhood, this study used the walkability concept by Pushkarev et al. (1975) that an individual's willingness to walk somewhere falls off drastically at about half a mile. The half-mile neighborhood concept is a common practice in the field of land use and transportation (Baker and Lee 2019; Cervero, Sandoval, and Landis 2002; Guo and Xu 2013). Following the literature, we chose a half-mile buffer from respondents' homes to compute area-based built environment variables. Table 2 presents the definitions and descriptive statistics of all variables.

Analysis Method

We applied GBDT to explore the relationships between built environment variables and individual daily active travel. The GBDT method is based on the work of Friedman (2001, 2002). It combines two approaches: decision tree learning and gradient boosting. The decision tree learning method partitions the sample into different subsamples (called terminal leaves) and uses the observations in each subsample to make predictions. In the example in Figure 2, we divided the sample into four subsamples (A, B, C, and D) based on three criteria including whether x_1 is smaller than m, whether x_2 is smaller than p, and whether x_2 is smaller than q. For example, we cluster one observation into subsample D if its x_1 is

Table 2. Variables and Their Definitions.

Variable	Description	*M*	*SD*	Minimum	Maximum
Met minutes	The daily amount of active travel of the respondent	106.64	118.32	2	784.02
Built environment attributes					
Park					
Park area	The area in acre of the parks that have overlaps with the half-mile buffer of the respondent's home	101.06	207.95	0	1,426.76
Distance to the nearest park	The network distance in mile from the respondent's home to the nearest park	0.22	0.14	0	1.06
Transit					
Distance to the nearest LRT station	The network distance in mile from the respondent's home to the nearest LRT station	3.75	4.99	0.03	21.78
Number of bus stops	Number of bus stops within the half-mile buffer of the respondent's home	24.87	15.42	0	60
Local land use					
Land-use entropy	Land-use entropy within the half-mile buffer of the respondent's home	0.54	0.15	0.05	0.95
Population density	Population density within the half-mile buffer of the respondent's home	9.04	6.11	0.76	25.49
Job density	Employment density within the half-mile buffer of the respondent's home	8.44	8.58	0.28	35.58
Design					
Number of dead ends	Number of dead ends within the half-mile buffer of the respondent's home	2.80	2.22	0	11
Regional location					
Distance to downtown Minneapolis	The network distance in mile from the respondent's home to downtown Minneapolis	5.29	4.40	1.08	22.36
Demographic characteristics					
Age	Age in years	49.39	16.51	18	91
Female	A dummy variable indicating whether the respondent is female	0.35	0.48	0	1
Employed	A dummy variable indicating whether the respondent is full or part-time employed	0.69	0.47	0	1
Student	A dummy variable indicating whether the respondent is a full-time student	0.10	0.30	0	1
Household member	Number of household members in the respondent's home	2.68	1.39	1	12
Pet	A dummy variable indicating whether the respondent has a pet	0.50	0.50	0	1
Hispanic	A dummy variable indicating whether the respondent is Hispanic	0.04	0.19	0	1
White	A dummy variable indicating whether the respondent is white	0.79	0.41	0	1
Asian	A dummy variable indicating whether the respondent is Asian	0.04	0.20	0	1
African American	A dummy variable indicating whether the respondent is African American	0.09	0.29	0	1
Less bachelor	A dummy variable indicating whether the respondent has a degree less than bachelor	0.28	0.45	0	1
Bachelor	A dummy variable indicating whether the respondent has a degree of bachelor	0.34	0.47	0	1
Low income	A dummy variable indicating whether the respondent has a low income	0.20	0.40	0	1
Medium income	A dummy variable indicating whether the respondent has a median income	0.34	0.47	0	1
License	A dummy variable indicating whether the respondent has a driver's license	0.89	0.31	0	1
Home owner	A dummy variable indicating whether the respondent owns the house or apartment	0.70	0.46	0	1
Temperature					
Average temperature	Average daily temperature in Fahrenheit.	54.35	17.69	−2.00	86.50

Note: LRT = light rail transit.

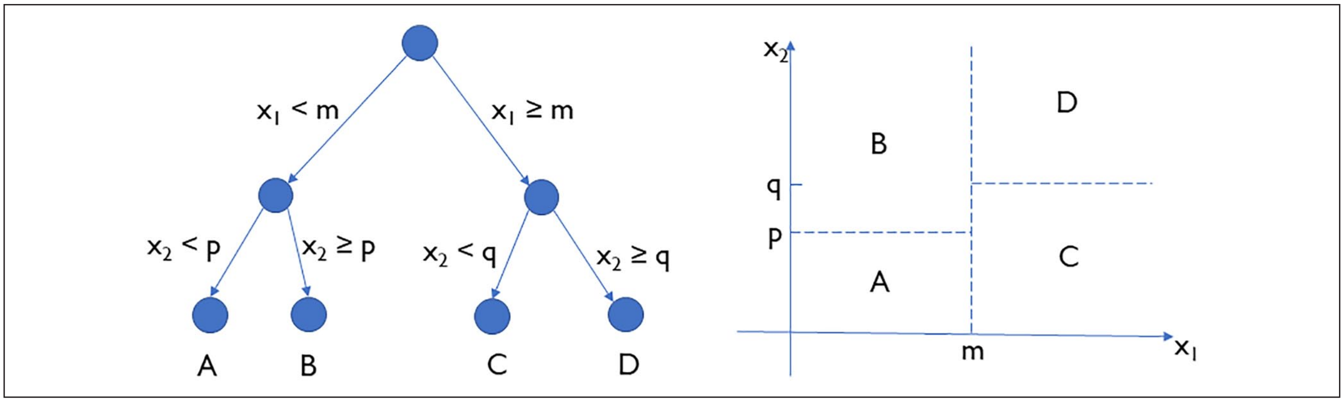

Figure 2. An example of decision tree learning.

larger than m and x_2 is larger than q. Variables x_1 and x_2 are predictors, which could be the built environment variables, demographics, or temperature in our study. The mean of the dependent variable in each subsample is the predicted value of all observations in the subsample. Gradient boosting improves the model performance by combining simple models into one complex model sequentially. In our study, we combined simple trees into one complex tree. We used the generalized boosting method (the package of gbm 2.1.5), developed by Ridgeway (2019), to develop the model in R 3.5.0. Please refer to Appendix C for the algorithm.

Choosing appropriate input parameters is a necessary step for GBDT. In particular, we consider three parameters: tree depth, shrinkage, and the number of trees. Tree depth defines the number of terminal leaves, relevant to the complexity of the trees. Commonly, it is an integer between one and ten. Shrinkage, also known as the learning rate, determines what portion of each simple tree could be combined into the complex one. While smaller shrinkage results in better model performance, it increases computational costs. Ridgeway (2007, 7) recommends a value between 0.001 and 0.01. The number of trees controls how many of the trees will be combined. A higher number of trees could improve model performance, but it may overfit the model. Overfitting happens when the model fits closely to the training data set, but does not fit well to other data sets (e.g., testing data sets). Cross validation could alleviate this problem. This study used a five-fold cross validation. The method partitions the sample into five subsamples, four are training data sets, and one is the testing data set.

In our research, we set shrinkage as 0.001. To find the optimal value of tree depth, we set it from one to ten and estimated ten models. We used root mean square error (RMSE) to assess model performance. We applied the five-fold cross validation to seek the optimal number of trees for each of the models. As shown in Figure 3, RMSE decreases quickly becoming relatively flat after six terminal leaves. We chose six as the tree depth. The corresponding number of trees is 1,613. The pseudo R^2 for the model is .349.

Compared with traditional regression, GBDT has many advantages: its prediction is more accurate; it handles the multicollinearity issue; and it accommodates variables with missing values (Ding, Cao, and Næss 2018; Elith, Leathwick, and Hastie 2008). More importantly, as its ensemble-based boosting approach, GBDT can effectively handle a small sample (Yang et al. 2010). It is evident that applying the GBDT model to a sample with less than one hundred observations produces reliable results in computational biology (Isıkhan, Selen, and Alpar 2016). In transportation, Ding, Cao, and Liu (2019) used GBDT to examine the influence of land-use variables on transit ridership of eighty-six stations in the Washington metropolitan areas and produced reasonable results. Furthermore, sample size is inversely related to p value, which tests whether the corresponding coefficient estimate is statistically significant. A small sample may not have enough statistical power. However, the GBDT approach does not produce p values. Instead, it generates the relative importance of predictors in predicting the response, that is, their practical significance.

The model can produce the relative importance of an independent variable in predicting met minutes, which measures its potential to reduce prediction errors, relative to other independent variables. The relative importance of all predictors adds up to 100 percent. In addition, it can generate a partial dependence plot to demonstrate whether the relationship between an independent variable and daily active travel is nonlinear. The plot is a graphical depiction of the average effect of a variable on the predicted met minutes, while considering its interactions with other independent variables in the model. This is not the independent effect of the variable. In decision trees, the response to a variable in a particular tree relies on the values of other independent variables in the higher level of trees (see Figure 2). Evaluating the travel effect of land-use policies that change one or more dimensions of the built environment should use both the changes and the existing conditions of the unchanged variables as the input for prediction.

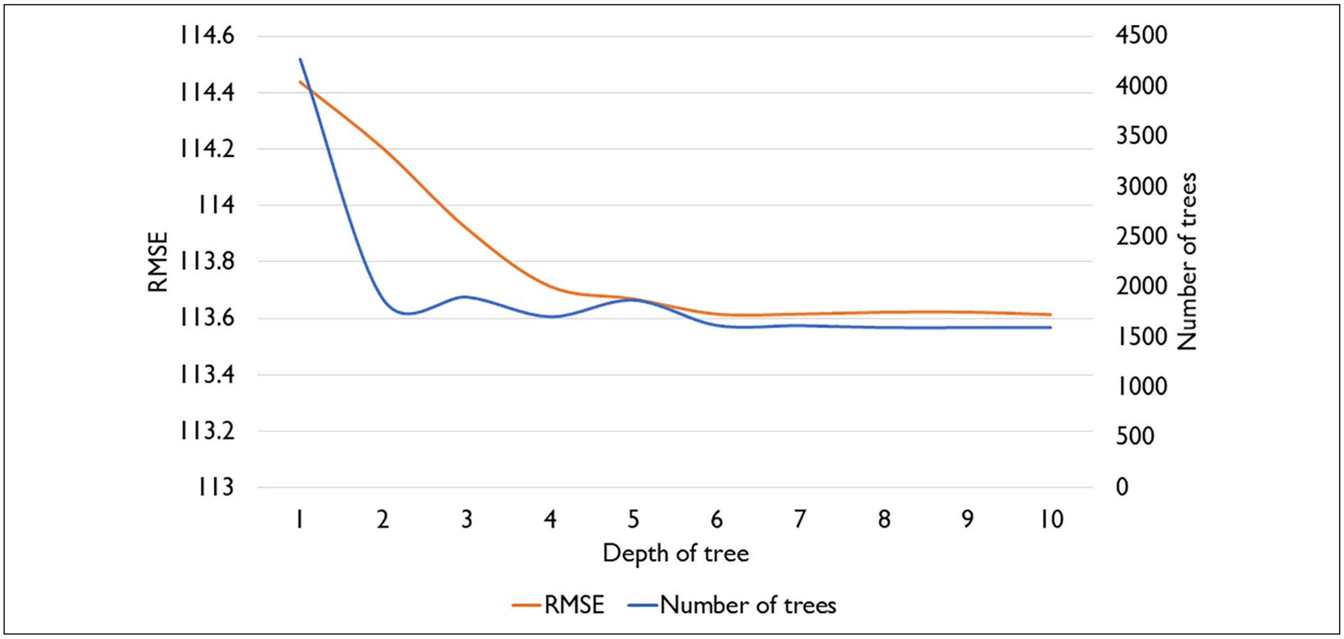

Figure 3. Depth of tree versus RMSE and number of trees.
Note: RMSE = root mean square error.

It is worth noting that although our sample came from six clustered neighborhoods, spatial dependency is not an issue in this study. First, we used individual-based measures (rather than zonal measures) to capture built environment variables. In particular, for distance-based variables, we measured the distances from respondents' homes to the specific locations; for buffer-based variables, we chose a half-mile buffer from the respondent's home. Therefore, respondents living in the same neighborhoods do not have identical measures. Furthermore, we calculated the Moran I index for the residuals of the sample. The index is 0.005, and its p value is .622, so there is no significant spatial dependency in the sample. More importantly, spatial dependency results in an underestimation of standard errors of coefficients, which may make insignificant variables falsely significant. That is, spatial dependency may change the outcomes of hypothesis tests. However, the GBDT approach does not produce p values but emphasizes the practical significance of coefficient estimates.

Results

Relative Importance of Independent Variables

After data analysis, we computed the relative importance of all the independent variables included in our model (Table 3). Built environment characteristics collectively account for more than two-thirds of the predictive power for daily met minutes, suggesting that the built environment plays a dominant role in shaping active travel. Although this finding appears to differ from the common wisdom that demographics are more important to travel behavior than built environment variables (Singh et al. 2018; Stead 2001), it is plausible. First, there are exceptions in the literature. For example, Kitamura, Mokhtarian, and Laidet (1997) found that built environment variables are significantly associated with the frequency of nonmotorized trips in the San Francisco Bay Area while none of demographic characteristics are significant. Furthermore, two studies using the GBDT approach showed the dominant influence of the built environment on driving behavior (Ding, Cao, and Næss 2018; Ding, Cao, and Wang 2018). The GBDT approach does not assume that a built environment variable follows a predefined (mostly linear or log-linear) relationship with travel behavior, and hence can demonstrate the true relationship between these two variables. When it has a threshold effect on active travel (e.g., distance to the closest park as shown later), linear regression overlooks the threshold effect and attempts to fit the data with a regression line. This line obviously misestimates the relationship between the built environment variable and active travel. Moreover, because built environment variables are often correlated with one another, including one variable in the regression may make another originally significant variable insignificant. When researchers manually remove the insignificant variable from the model, the collective effect of built environment variables is underestimated. Combined with previous studies (Ding, Cao, and Næss 2018; Ding, Cao, and Wang 2018), this research challenges common wisdom by depicting the threshold effects of built environment variables on active travel.

Among individual built environment characteristics, distance to the nearest park is the most important, followed by distance to downtown Minneapolis and the number of bus

Table 3. Relative Importance of Variables.

Variable	Relative importance (%)	Rank	Sum (%)	
Built environment attributes				68.8
Park				
Park area	4.6	9	22.1	
Distance to the nearest park	17.6	1		
Transit				
Distance to the nearest LRT station	7.6	7	16.7	
Number of bus stops	9.0	4		
Local land use				
Land-use entropy	7.8	6	16.7	
Population density	3.0	11		
Job density	5.9	8		
Design				
Number of dead ends	2.9	12	2.9	
Regional location				
Distance to downtown Minneapolis	10.4	3	10.4	
Demographic characteristics				
Age	8.2	5		19.8
Female	4.0	10		
Employed	0.4			
Student	0.0			
Household member	1.1			
Pet	0.9			
Hispanic	0.0			
White	0.2			
Asian	0.0			
African American	0.0			
Less bachelor	1.6			
Bachelor	1.2			
Low income	0.2			
Medium income	1.0			
License	0.0			
Homeowner	1.0			
Temperature				
Average temperature	11.5	2		11.5

Note: LRT = light rail transit.

stops with a half-mile buffer from respondents' homes. We further classified built environment attributes into five groups based on their similarities: park, transit, local land use, street network design, and regional location. Park-related variables contribute the most to predicting active travel, with a relative importance of 22.1 percent. Transit and local land use have the same relative importance, at 16.7 percent.

Among demographics, age is the most important, with a relative importance of 8.2 percent. As shown in the third part of this section, it reflects life-cycle effects. The relative importance of gender ranks second. All other demographics contribute to less than 2 percent of the predictive power.

Temperature plays an important role in affecting active travel. Its relative importance is 11.5 percent, ranking second among all the independent variables. This is not surprising because nonmotorized mode users expose directly to the environment and they are less likely to travel and travel shorter during cold days.

The Associations between Built Environment Variables and Daily Active Travel

As shown in Figures 4 to 8, we drew partial dependence plots to illustrate the relationships between built environment characteristics and predicted daily met minutes, while considering the effects of other variables in the model. Because we cannot fully overcome the overfitting issue, small fluctuations appear in the plots. When interpreting the plots, we pay more attention to the general trends of the relationships between variables, than the fluctuations.

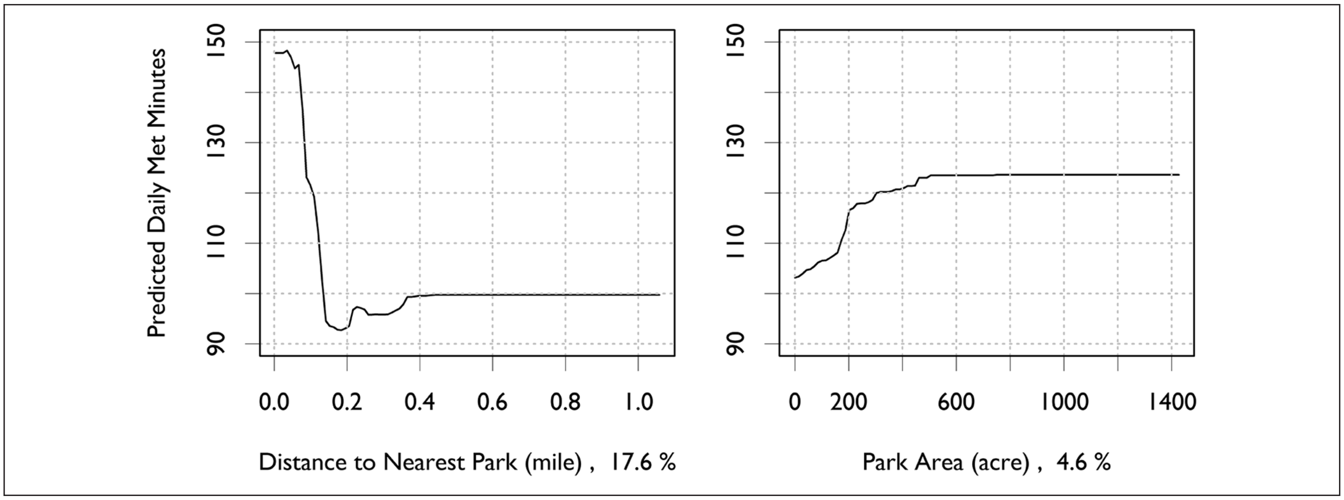

Figure 4. The relationships between active travel and park-related variables.
Note: The relative importance of the variable is presented in the label for the horizontal axis, same for the following figures.

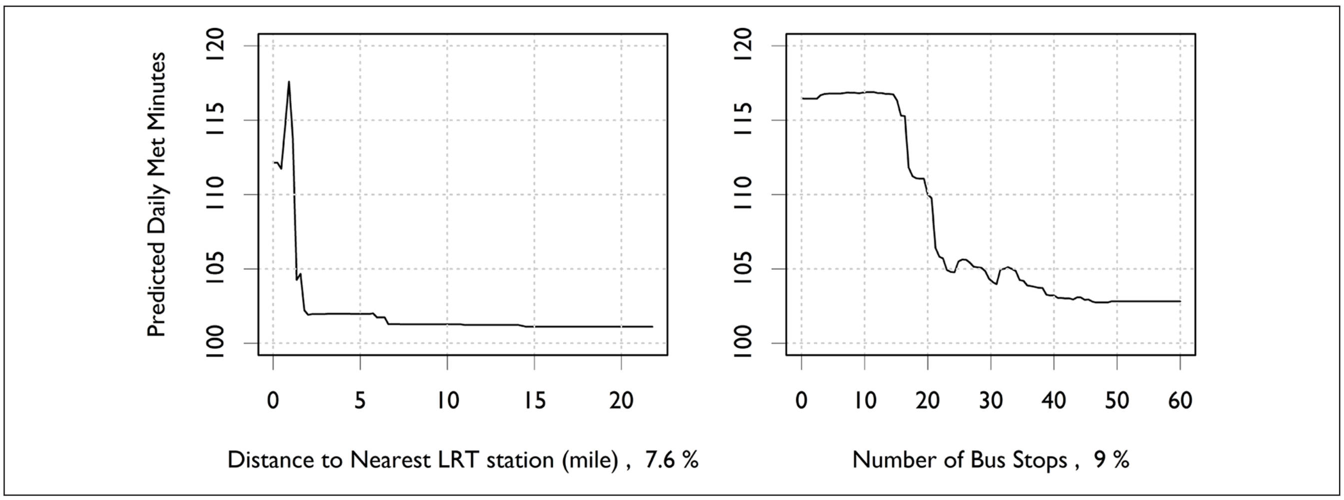

Figure 5. The relationships between active travel and transit-related variables.
Note: LRT = light rail transit.

Park-related variables include the distance to the nearest park and park area (Figure 4). The former has a negative relationship with active travel. Being farther away from parks discourages active travel to parks. This variable shows a clear threshold effect. When the distance increases from 0 to 0.2 miles, active travel decreases from 148 to 93 met minutes. When the nearest park is farther than 0.2 miles, active travel does not change much. Park area has a positive relationship with active travel. The predicted met minutes increase from 103 to 123 when park area increases from 0 to 500 acres. After the threshold, it has no additional influence on active travel. Overall, within their effective ranges, the relationships between park-related variables and active travel are congruent with the gravity-based model: more opportunities (larger parks) attract trips whereas a larger friction (a longer distance) of a park discourages active travel (Cohen, Marsh, et al., 2010; Cohen, McKenzie, et al. 2007). The effect size of distance to the nearest park is a few times larger than that of the park area.

Transit-related variables consist of distance to the nearest LRT station and the number of bus stops, a proxy for bus access (Figure 5). As the distance to LRT station grows from 0 to 0.9 miles, active travel increases by about five met minutes. Within a walkable distance of an LRT station, transit riders walk longer to reach the station when it is farther. Therefore, this positive association makes sense. However, once it is beyond the walkable distance, people may not use LRT and their active travel decreases substantially (Ewing and Cervero 2010). The number of bus stops has a threshold effect on active travel and the relationship is generally negative. With a larger number of bus stops within a walkable distance, it is more likely that a stop is in close proximity to

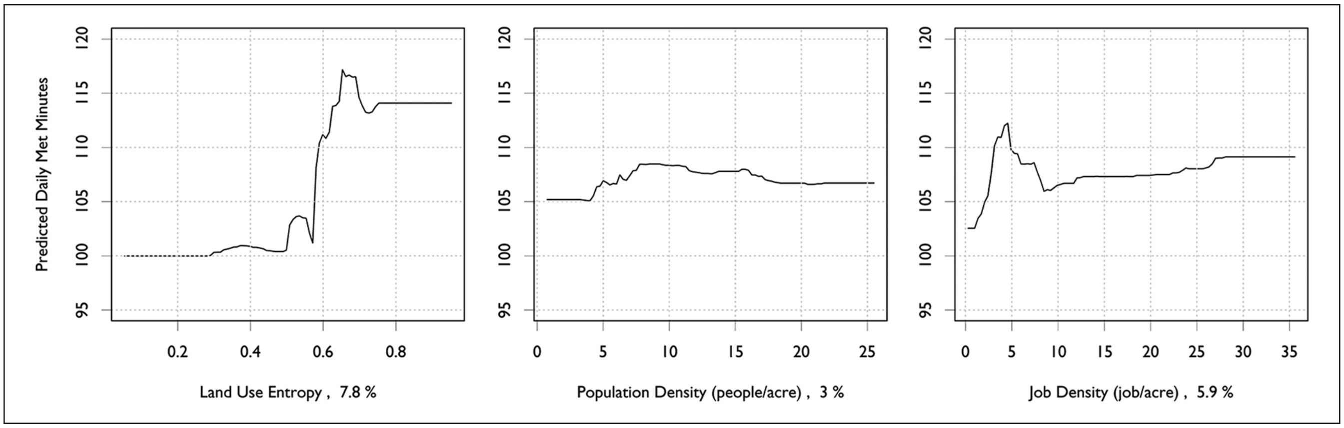

Figure 6. The relationships between active travel and local land-use variables.

individuals' home. They do not need to walk longer to reach the stop. Therefore, the negative association is plausible.

Local land-use-related variables include land-use entropy, population density, and job density (Figure 6). Land-use entropy has a positive relationship with active travel, consistent with the finding of a meta-analysis by Ewing and Cervero (2010). When the entropy index is smaller than 0.5, the predicted active travel is approximately one hundred met minutes. It increases to about 117 met minutes as the index grows to 0.7 and then there is a small fluctuation. Population density also has a positive association with active travel. As population density increases, it seems that the mass of population encourages people to engage in active travel. This is consistent with the literature (Ewing and Cervero 2010). However, once population density exceeds ten people/acre,[1] active travel decreases to some extent. Presumably, population density is positively associated with the number of services and businesses. When they are close to residential areas, people can walk/bike a shorter distance to reach these destinations. This speculation is consistent with the association between job density and active travel. As job density grows, people engage in active travel more. More destinations promote active travel. This is congruent with the finding of Ewing and Cervero's (2010) work. However, when it exceeds five jobs/acre, active travel decreases.

The number of dead end streets, an indicator of street network design, has a negative association with active travel (Figure 7). This finding is somewhat consistent with Schoner and Cao's (2014) work: the number of cul-de-sacs is negatively associated with walking frequencies in Minneapolis. That is, cul-de-sacs are not conducive to active travel. This study also shows that the largest drop in active travel occurs when there is a dead end within a half-mile buffer of respondents' home. Its effect is saturated at about three dead ends.

Figure 8 shows the association between active travel and distance to downtown Minneapolis, a variable indicating regional location. In general, people living within about four

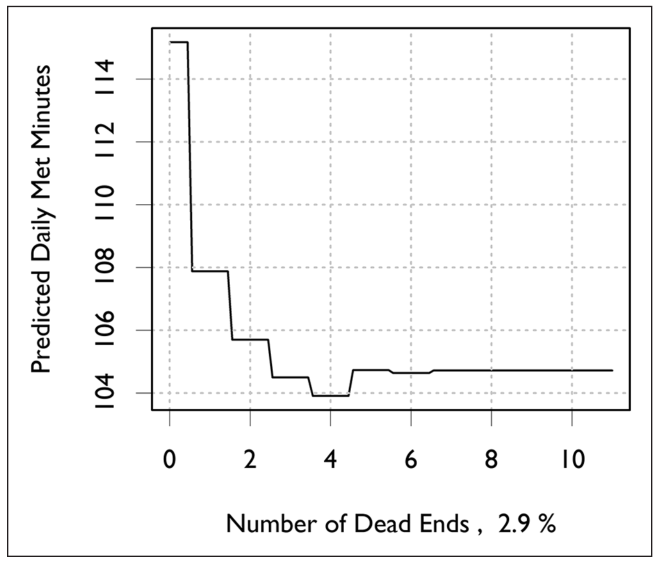

Figure 7. The relationships between active travel and network design variable.

miles from the city center conduct more active travel than those living beyond the distance. Within the threshold, there is a positive association between distance to downtown Minneapolis and active travel. This positive relationship is consistent with the finding of job density: more service and business opportunities enable people to walk/bike a shorter distance to reach their activity destinations. To promote active travel, planners should plan more residential buildings within a few miles from the city center.

The Associations between Active Travel and Other Key Variables

As presented in the first part of this section, average temperature and age have strong associations with active travel. Active travel grows at an exponential rate when the

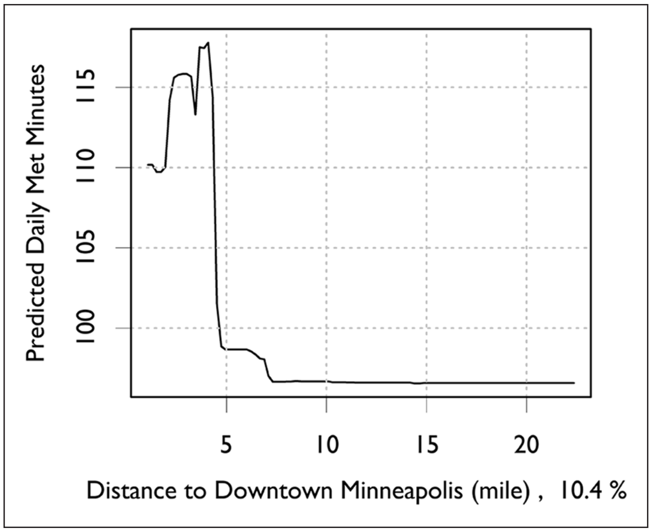

Figure 8. The relationships between active travel and regional location variable.

temperature is above the freezing level (Figure 9). Its marginal effect is saturated at about seventy-seven degrees. In terms of age, individuals reach the peak of active travel when they are at forty-two years old and reach the valley when they are beyond seventy-one years old. These results are consistent with our expectations.

Conclusion

In this research, we used the GBDT approach to examine the influences of built environment variables on individual daily active travel in the Twin Cities. The paper has a few limitations. First, because the survey with Daynamica^{TM} is costly and time-consuming, we sampled only six neighborhoods in urban, inner-inning suburban, and suburban areas and did not recruit respondents from exurban areas in the Twin Cities. This limits the range of built environment variables that can be tested in the model, particularly distance to the city center. With that said, we do not expect suburban residents and exurban residents to show distinct active travel patterns. For example, our model showed that the influence of distance to downtown Minneapolis is saturated after about seven miles. Nevertheless, it is desirable to obtain a random sample of respondents from all types of neighborhoods in the region. Second, this study converted various means of active travel into met minutes and ignored behavioral differences between walking and biking. The implication is that we masked different mechanisms that built environment variables affect walking and biking. On the contrary, if we model walking and biking differently, we will overlook the interaction between walking and biking if people have a budget for physical activity (Rodríguez, Khattak, and Evenson 2006). This is one of the reasons that many studies combine different means of active travel in their analysis (Cao 2015; Sallis et al. 2009). Third, this study controlled for demographic variables but not predispositions toward travel and land use. Thus, residential self-selection may be at work (Guan, Wang, and Cao 2020). For example, people who would like to use parks might choose to live closer to parks. Future studies should include related attitudinal questions in the survey to control for self-selection effects. Fourth, this study considered street connectivity (i.e., number of dead ends), but not the quality of the walking environment. As walking is very sensitive to the pedestrian environment, future research could include additional walkability measures, such as the existence of sidewalks and sidewalk width. Nevertheless, this study makes significant contributions to the literature of the built environment and active travel.

The results showed that the built environment has a critical impact on active travel: the collective contribution of built environment variables to predicting met minutes is about 69 percent. Land-use policies could be effective to promote healthy living. The top three built environment contributors are distance to the nearest park, distance to downtown Minneapolis, and the number of bus stops.

Parks are important to active living. The results further showed that the distance to the nearest park has a more important impact on active travel than park size, which contributes to the literature on health benefits of urban parks (Fan, Das, and Chen 2011; UN-Habitat 2015). This finding is encouraging to urban planners who face increasing difficulties in acquiring large pieces of new parklands. A small park within walking distance works better for daily active travel.

Active travel is higher within four miles of downtown Minneapolis. This suggests that population densification with a few miles of city centers can promote active travel. Policy incentives should be used to direct future development toward the central city.

Transit-related variables have important effects on active travel. Transit-oriented development, not only promotes transit use, but also facilitates active travel. The effect is partly due to the active travel required to reach transit stops or egresses from these stops. With the intention of increasing transit use, local municipalities often fund station-area improvements around fixed transit infrastructure such as LRT to improve pedestrian and bicycle access to transit. Furthermore, because transit users become familiar with the surrounding environment while they walk to transit stops, they may engage in other types of active travel (such as walking to a local store) and become accustomed to an active living style. Local land use could reinforce the effect of transit infrastructure because density and land-use mix are also positively associated with active travel. Overall, a comprehensive package of strategies is needed to promote active travel effectively.

Furthermore, this study found that built environment characteristics tend to have nonlinear or threshold effects on daily active travel. More precisely, the influences of built environment variables are piecewise linear and they are

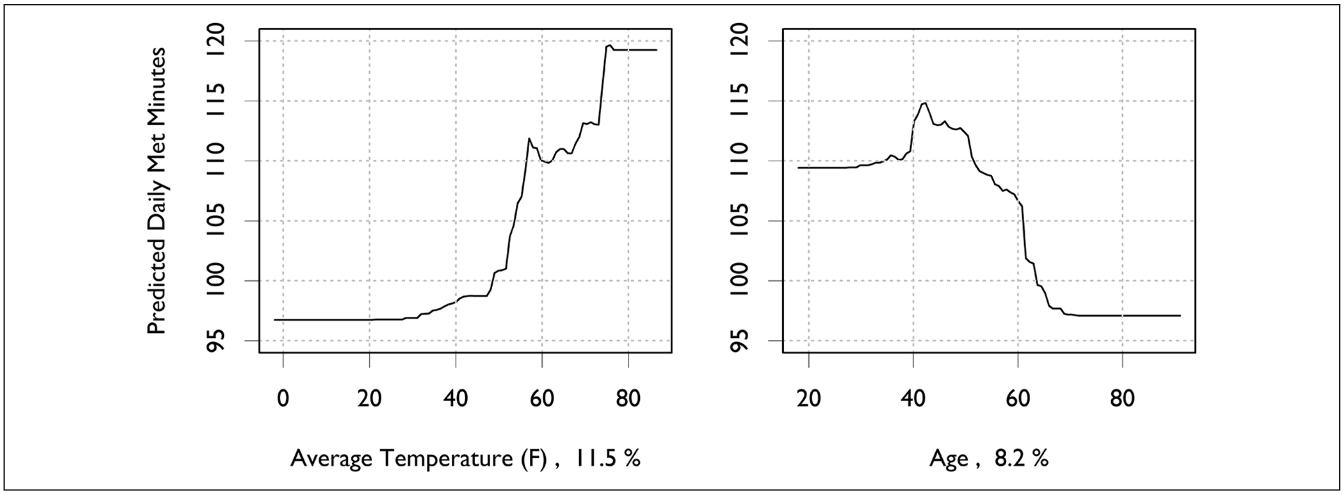

Figure 9. The relationships between active travel and other key variables.

effective only within narrow ranges. For example, a park beyond 0.2 miles of a residence has few impacts on active travel; when the entropy index is smaller than 0.5, land-use mix does not affect active travel. If we ignore these threshold effects, we will misestimate the effects of the built environment on active travel and offer false policy implications for planning practice.

It is worth noting that the thresholds obtained in this study may be context specific. Some scholars attempted to generalize the relationship between the built environment and travel behavior (Ewing and Cervero 2010; Stevens 2017). However, the relationship is likely to be context-dependent: the effect size of a built environment variable in one region may differ from another region (Næss 2019). Similarly, the thresholds of built environment variables based on six neighborhoods in the Twin Cities may not be transferable to another region, even the ones of a similar size. More studies are needed to test whether threshold effects are sensitive to local contexts.

Appendix

A. The Content of an Example Recruitment Postcard

Researchers at the University of Minnesota are conducting a study to understand the relationship between neighborhood environment and resident well-being. We are currently looking to recruit participants from Phillips, MN (current residents only). Participation is completely voluntary, and responses are confidential. The study offers a $50 Target gift card to compensate for your time and effort. If you are interested in participating, please contact us at [Phone number] or [Email address]. Don't forget to mention your neighborhood when you contact us!

B. User Guide of the Daynamica^{TM} *App*

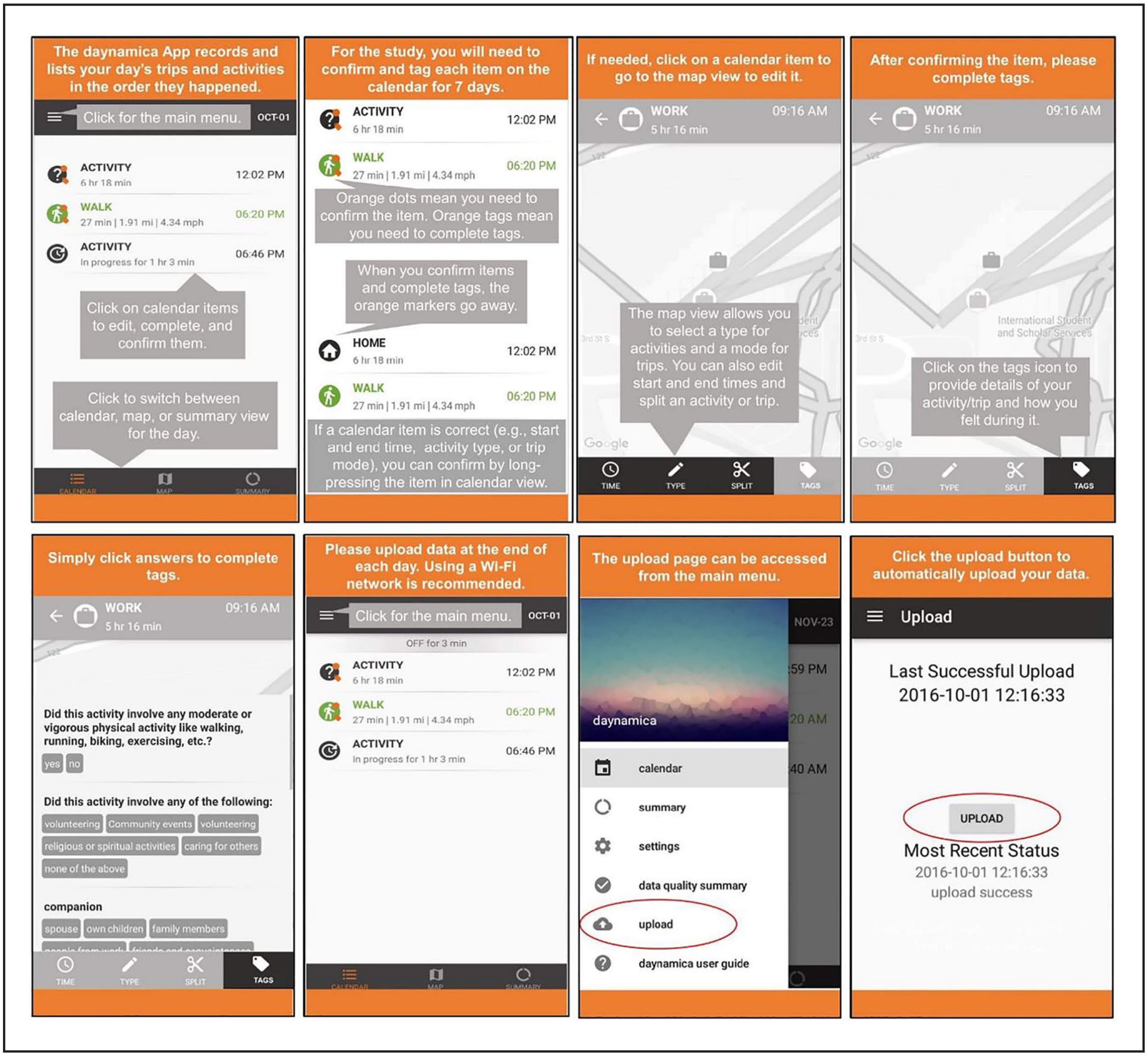

Figure 10. User guide of the Daynamica^{TM} APP.

C. Algorithm for gradient boosting decision tree (GBDT)

We borrowed this section directly from our previous work (Wu et al. 2019). Gradient boosting assembles M decision trees to produce a strong prediction model. For a sample of (y, x), the goal of GBDT is to fit a function of $\widehat{F(x)}$ to minimize the loss function $\psi\left[y, F(x)\right]$. We developed the model using the R-based "gbm" package (Ridgeway 2019). According to Friedman (2002), the output of a gradient boosting model in Step $m(0 < m \leq M)$ is

$$\widehat{F_m(x)} = \widehat{F_{m-1}(x)} + \xi \sum_{j=1}^{J} \gamma_{jm} I\left(x \in R_{jm}\right), \text{ where } 0 < \xi \leq 1, \quad (2)$$

where J is the number of regions partitioned by a decision tree, γ_{jm} is the value of optimal gradient for the region R_{jm}, which could make the current function $\widehat{F_m(x)}$ obtain the smallest loss; $I = 1$ if x falls into R_{jm} and $I = 0$ otherwise; ξ is the shrinkage parameter, also called learning rate. After M iteration of the steps above, we will have the final model.

Breiman et al. (2017) developed a measure to approximate the relative influence of an independent variable x_κ in predicting the dependent variable in a decision tree T:

$$I_\kappa^2(T) = \sum_{t=1}^{J-1} \hat{\tau}_t^2 I\left(v(t) = \kappa\right), \tag{3}$$

where $\hat{\tau}_t^2$ is the reduction in squared error when predictor x_κ is used as the splitting variable, which is $v(t)$ in the equation above, and J is the number of partitioned regions by the tree T. For a series of decision trees $\{T_m\}_1^M$, we can generalize equation (3) by averaging all trees obtained through the GBDT approach:

$$I_\kappa^2 = \frac{1}{M} \sum_{m=1}^{M} I_\kappa^2(T_m). \tag{4}$$

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Sustainable Research Network project of the National Science Foundation of USA: Integrated Urban Infrastructure Solutions for Environmentally Sustainable, Healthy and Livable Cities (Award No. 1444745).

ORCID iD

Jason Cao https://orcid.org/0000-0002-4403-9762

Note

1. The average population density in Minneapolis is about eleven person/acre.

References

Ainsworth, Barbara E., William L. Haskell, Stephen D. Herrmann, Nathanael Meckes, David R. Bassett Jr, Catrine Tudor-Locke, Jennifer L. Greer, Jesse Vezina, Melicia C. Whitt-Glover, and Arthur S. Leon. 2011. "2011 Compendium of Physical Activities: A Second Update of Codes and MET Values." *Medicine & Science in Sports & Exercise* 43 (8): 1575–81. doi:10.1249/MSS.0b013e31821ece12.

Baker, Dwayne Marshall, and Bumsoo Lee. 2019. "How Does Light Rail Transit (LRT) Impact Gentrification? Evidence from Fourteen US Urbanized Areas." *Journal of Planning Education and Research* 39 (1): 35–49. doi:10.1177/0739456X17713619.

Beenackers, Mariëlle A., Sarah Foster, Carlijn B. M. Kamphuis, Sylvia Titze, Mark Divitini, Matthew Knuiman, Frank J. Van Lenthe, and Billie Giles-Corti. 2012. "Taking Up Cycling after Residential Relocation: Built Environment Factors." *American Journal of Preventive Medicine* 42 (6): 610–15. doi:10.1016/j.amepre.2012.02.021.

Boarnet, Marlon G., Michael Greenwald, and Tracy E. McMillan. 2008. "Walking, Urban Design, and Health: Toward a Cost-Benefit Analysis Framework." *Journal of Planning Education and Research* 27 (3): 341–58. doi:10.1177/0739456X07311073.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 2017. *Classification and Regression Trees*. Routledge. doi:10.1201/9781315139470.

Brownson, Ross C., Christine M. Hoehner, Kristen Day, Ann Forsyth, and James F. Sallis. 2009. "Measuring the Built Environment for Physical Activity: State of the Science." *American Journal of Preventive Medicine* 36 (4 Suppl.): S99–S123.e12. doi:10.1016/j.amepre.2009.01.005

Cao, Xinyu J. 2015. "Examining the Impacts of Neighborhood Design and Residential Self-Selection on Active Travel: A Methodological Assessment." *Urban Geography* 36 (2): 236–55. doi:10.1080/02723638.2014.956420.

Cervero, Robert, and Michael Duncan. 2003. "Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area." *American Journal of Public Health* 93 (9): 1478–83. doi:10.2105/AJPH.93.9.1478

Cervero, Robert, S. Murphy, C. Ferrell, N. Goguts, Y -H. Tsai, G. B. Arrington, J. Boroski, et al. 2004. *Transit-oriented Development in the United States: Experiences, Challenges, and Prospects*. Vol. 102. Washington, DC: Transportation Research Board. doi:10.17226/23360

Cervero, Robert, Onésimo Sandoval, and John Landis. 2002. "Transportation as a Stimulus of Welfare-to-Work." *Journal of Planning Education and Research* 22 (1): 50–63. doi:10.1177/0739456X0202200105

Cohen, Deborah A., Terry Marsh, Stephanie Williamson, Kathryn Pitkin Derose, Homero Martinez, Claude Setodji, and Thomas L. McKenzie. 2010. "Parks and Physical Activity: Why Are Some Parks Used More Than Others?" *Preventive Medicine* 50 (Suppl.): S9–S12. doi:10.1016/j.ypmed.2009.08.020

Cohen, Deborah A., Thomas L. McKenzie, Amber Sehgal, Stephanie Williamson, Daniela Golinelli, and Nicole Lurie. 2007. "Contribution of Public Parks to Physical Activity." *American Journal of Public Health* 97 (3): 509–14. doi:10.2105/AJPH.2005.072447.

Ding, Chuan, Xinyu Cao, and Chao Liu. 2019. "How Does the Station-Area Built Environment Influence Metrorail Ridership? Using Gradient Boosting Decision Trees to Identify Non-Linear Thresholds." *Journal of Transport Geography* 77:70–78. doi:10.1016/j.jtrangeo.2019.04.011

Ding, Chuan, Xinyu Cao, and Yunpeng Wang. 2018. "Synergistic Effects of the Built Environment and Commuting Programs on Commute Mode Choice." *Transportation Research Part A* 118:104–18. doi:10.1016/j.tra.2018.08.041

Ding, Chuan, Xinyu (Jason) Cao, and Petter Næss. 2018. "Applying Gradient Boosting Decision Trees to Examine Non-Linear Effects of the Built Environment on Driving Distance in Oslo." *Transportation Research Part A: Policy and Practice* 110:107–17. doi:10.1016/j.tra.2018.02.009.

Duncan, Mitch J., Elisabeth Winkler, Takemi Sugiyama, Ester Cerin, Lorinne Dutoit, Eva Leslie, and Neville Owen. 2010. "Relationships of Land Use Mix with Walking for Transport: Do Land Uses and Geographical Scale Matter?" *Journal of Urban Health* 87 (5): 782–95. doi:10.1007/s11524-010-9488-7.

Elith, Jane., John R. Leathwick, and T. Hastie. 2008. "A Working Guide to Boosted Regression Trees." *Journal of Animal Ecology* 77 (4): 802–13. doi:10.1111/j.1365-2656.2008.01390.x.

Ewing, Reid, and Robert Cervero. 2010. "Travel and the Built Environment." *Journal of the American Planning Association* 76 (3): 265–94. doi:10.1080/01944361003766766.

Ewing, Reid, and Robert Cervero. 2017. "'Does Compact Development Make People Drive Less?' The Answer Is Yes." *Journal of the American Planning Association* 83 (1): 19–25. doi:10.1080/01944363.2016.1240044.

Fan, Yingling, Kirti V. Das, and Qian Chen. 2011. "Neighborhood Green, Social Support, Physical Activity, and Stress: Assessing the Cumulative Impact." *Health and Place* 17 (6): 1202–11. doi:10.1016/j.healthplace.2011.08.008.

Fan, Yingling, Julian Wolfson, Gediminas Adomavicius, Kirti Vardhan Das, Yash Khandelwal, and Jie Kang. 2015. "SmarTrAC: A Smartphone Solution for Context-Aware Travel and Activity Capturing." Center for Transportation Studies, University of Minnesota. http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=2424.

Frank, Lawrence, and G. Pivo. 1994. "Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single Occupant Vehicle, Transit, and Walking." *Transportation Research Record: Journal of the Transportation Research Board* 1466:44–52. doi:10.1073/pnas.1100480108

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29 (5): 1189–1232. doi:10.1214/aos/1013203451.

Friedman, Jerome H. 2002. "Stochastic Gradient Boosting." *Computational Statistics & Data Analysis* 38 (4): 367–78. doi:10.1016/S0167-9473(01)00065-2

Giles-Corti, Billie, Fiona Bull, Matthew Knuiman, Gavin McCormack, Kimberly Van Niel, Anna Timperio, Hayley Christian, et al. 2013. "The Influence of Urban Design on Neighbourhood Walking following Residential Relocation: Longitudinal Results from the RESIDE Study." *Social Science & Medicine* 77 (1): 20–30. doi:10.1016/j.socscimed.2012.10.016

Guan, Xiaodong, Donggen Wang, and Xinyu Jason Cao. 2020. "The Role of Residential Self-Selection in Land Use-Travel Research: A Review of Recent Findings." *Transport Reviews* 40:267–87. doi:10.1080/01441647.2019.1692965.

Guo, Zhan, and Peiyi Xu. 2013. "Duet of the Commons: The Impact of Street Cleaning on Car Usage in the New York City Area." *Journal of Planning Education and Research* 33 (1): 34–48. doi:10.1177/0739456X12459360.

Handy, Susan L., Marlon G. Boarnet, Reid Ewing, and Richard E. Killingsworth. 2002. "How the Built Environment Affects Physical Activity: Views from Urban Planning." *American Journal of Preventive Medicine* 23 (2 Suppl. 1): 64–73. doi:10.1016/S0749-3797(02)00475-0.

Heesch, Kristiann C., Billie Giles-Corti, and Gavin Turrell. 2015. "Cycling for Transport and Recreation: Associations with the Socio-Economic, Natural and Built Environment." *Health and Place* 36:152–61. doi:10.1016/j.healthplace.2015.10.004.

Isıkhan, Yılmaz, Erdem Karabulut Selen, and Celal Reha Alpar. 2016. "Determining Cutoff Point of Ensemble Trees Based on Sample Size in Predicting Clinical Dose with DNA Microarray Data." *Computational and Mathematical Methods in Medicine* 2016:1–9. doi:10.1155/2016/6794916.

Kitamura, Ryuichi, Patricia L. Mokhtarian, and Laura Laidet. 1997. "A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area." *Transportation* 24 (2): 125–58. doi:10.1023/A:1017959825565.

Litman, Todd. 2003. "Active Transportation Policy Issues." Victoria, Canada: National Roundtable on Active Transportation. https://www.vtpi.org/act_tran.pdf.

Mokhtarian, Patricia L., and Xinyu Cao. 2008. "Examining the Impacts of Residential Self-Selection on Travel Behavior: A Focus on Methodologies." *Transportation Research Part B: Methodological* 42:204–28. doi:10.1016/j.trb.2007.07.006.

Næss, Petter. 2019. "Meta-Analyses of Built Environment Effects on Travel: No New Platinum Standard." *Journal of Planning Education and Research*. Published electronically June 13. doi:10.1177/0739456X19856425.

Neves, Andre, and Christian Brand. 2019. "Assessing the Potential for Carbon Emissions Savings from Replacing Short Car Trips with Walking and Cycling Using a Mixed GPS-Travel Diary Approach." *Transportation Research Part A: Policy and Practice* 123:130–46. doi:10.1016/j.tra.2018.08.022.

Oliver, Lisa N., Nadine Schuurman, and Alexander W. Hall. 2007. "Comparing Circular and Network Buffers to Examine the Influence of Land Use on Walking for Leisure and Errands." *International Journal of Health Geographics* 6:1–11. doi:10.1186/1476-072X-6-41.

Physical Activity Guidelines Advisory Committee. 2008. "Physical Activity Guidelines Advisory Committee Report, 2008." Washington, DC: U.S. Department of Health and Human Services. https://www.scribd.com/document/64587293/Physical-Activity-Guidelines-Advisory-Committee-Report.

Pushkarev, Boris, J. M. Zupan, Boris Pushkarev, and Jeffrey M. Zupan. 1975. "Capacity of Walkways." *Transportation Research Record* 538:1–15. http://onlinepubs.trb.org/Onlinepubs/trr/1975/538/538-001.pdf.

Ridgeway, Greg. 2007. "Generalized Boosted Models: A Guide to the gbm Package." *Compute* 1 (4): 1–12. doi:10.1111/j.1467-9752.1996.tb00390.x

Ridgeway, Greg. 2019. "Package 'Gbm.'" The R Project for Statistical Computing. https://cran.r-project.org/web/packages/gbm/gbm.pdf.

Rodríguez, Daniel A., Asad J. Khattak, and Kelly R. Evenson. 2006. "Can New Urbanism Encourage Physical Activity? Comparing a New Urbanist Neighborhood with Conventional Suburbs." *Journal of the American Planning Association* 72 (1): 43–54. doi:10.1080/01944360608976723

Sallis, James F., Heather R. Bowles, Adrian Bauman, Barbara E. Ainsworth, Fiona C. Bull, Cora L. Craig, Michael Sjöström, et al. 2009. "Neighborhood Environments and Physical Activity among Adults in 11 Countries." *American Journal of Preventive Medicine* 36 (6): 484–90. doi:10.1016/j.amepre.2009.01.031

Schoner, Jessica, and Xinyu Cao. 2014. "Walking for Purpose and Pleasure." *Transportation Research Record: Journal of the Transportation Research Board* 2464:67–76. doi:10.3141/2464-09.

Singh, Abhilash C., Sebastian Astroza, Venu M. Garikapati, Ram M. Pendyala, Chandra R. Bhat, and Patricia L. Mokhtarian. 2018. "Quantifying the Relative Contribution of Factors to Household Vehicle Miles of Travel." *Transportation Research*

Part D: Transport and Environment 63:23–36. doi:10.1016/j.trd.2018.04.004.

Stead, D.. 2001. "Relationships between Land Use, Socioeconomic Factors, and Travel Patterns in Britain." *Environment and Planning B: Planning and Design* 28 (4): 499–528. doi:10.1068/b2677

Stevens, Mark R.. 2017. "Does Compact Development Make People Drive Less?" *Journal of the American Planning Association* 83 (1): 7–18. doi:10.1080/01944363.2016.1240044.

Strath, Scott J., Leonard A. Kaminsky, Barbara E. Ainsworth, Ulf Ekelund, Patty S. Freedson, Rebecca A. Gary, Caroline R. Richardson, Derek T. Smith, and Ann M. Swartz. 2013. "Guide to the Assessment of Physical Activity: Clinical and Research Applications: A Scientific Statement from the American Heart Association." *Circulation* 128 (20): 2259–79. doi:10.1161/01.cir.0000435708.67487.da.

Transportation Research Board and Institute of Medicine. 2005. "Does the Built Environment Influence Physical Activity? Examining the Evidence—Special Report 282." Vol. 282. Transportation Research Board. https://www.nap.edu/catalog/11203/does-the-built-environment-influence-physical-activity-examining-the-evidence.

UN-Habitat. 2015. "*Global Public Space Toolkit: From Global Principles to Local Policies and Practice. Nairobi: United Nations Human Settlements Programme (Un-habitat)*." https://unhabitat.org/node/143325.

Van Wee, Bert, and Susan Handy. 2016. "Key Research Themes on Urban Space, Scale, and Sustainable Urban Mobility." *International Journal of Sustainable Transportation* 10 (1): 18–24. doi:10.1080/15568318.2013.820998.

Wijk, Daniël C., Joost Oude Groeniger, Frank J. Lenthe, and Carlijn B. M. Kamphuis. 2017. "The Role of the Built Environment in Explaining Educational Inequalities in Walking and Cycling among Adults in the Netherlands." *International Journal of Health Geographics* 16 (1): 1–12. doi:10.1186/s12942-017-0083-y.

Wu, Xinyi, Tao Tao, Jason Cao, Yingling Fan, and Anu Ramaswami. 2019. "Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions." *Transportation Research Part D: Transport and Environment* 75: 1-12. https://doi.org/10.1016/j.trd.2019.08.018.

Yang, Pengyi, Yee Hwa Yang, Bing B. Zhou, and Albert Y. Zomaya. 2010. "A Review of Ensemble Methods in Bioinformatics." *Current Bioinformatics* 5 (4): 296–308. https://www.ingentaconnect.com/content/ben/cbio/2010/00000005/00000004/art00006.

Zhang, Lei, Jinhyun Hong, Arefeh Nasri, and Qing Shen. 2012. "How Built Environment Affects Travel Behavior: A Comparative Analysis of the Connections between Land Use and Vehicle Miles Traveled in US Cities." *Journal of Transport and Land Use* 5 (3): 40–52. doi:10.5198/jtlu.v5i3.266

Zhang, Ming. 2004. "The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong." *Journal of the American Planning Association* 70 (3): 344–60. doi:10.1080/01944360408976383.

Author Biographies

Tao Tao is a PhD student at the Humphrey School of Public Affairs, University of Minnesota, Twin Cities. He is interested in land-use and travel behavior.

Xinyi Wu is a PhD student at the Humphrey School of Public Affairs, University of Minnesota, Twin Cities. Her interests include transportation planning and planning for life satisfaction.

Jason Cao is a professor at the Humphrey School of Public Affairs, University of Minnesota, Twin Cities. He specializes in land-use and travel behavior.

Yingling Fan is a professor at the Humphrey School of Public Affairs, University of Minnesota, Twin Cities. Her research interests include transit and land-use planning, neighborhood design, and physical activity promotion.

Kirti Das is a PhD student at the Humphrey School of Public Affairs, University of Minnesota, Twin Cities. He is interested in satisfaction research.

Anu Ramaswami is the inaugural director of the M.S. Chadha Center for Global India at Princeton University. She is a leader of urban sustainable systems.