

Correlates of the Built Environment and Active Travel: Evidence from 20 US Metropolitan Areas

Huyen T.K. Le,¹ Ralph Buehler,² and Steve Hankey¹

¹School of Public and International Affairs, Virginia Tech, Blacksburg, Virginia, USA

²School of Public and International Affairs, Virginia Tech, Alexandria, Virginia, USA

BACKGROUND: Walking and bicycling are health-promoting and environmentally friendly alternatives to the automobile. Previous studies that explore correlates of active travel and the built environment are for a single metropolitan statistical area (MSA) and results often vary among MSAs.

OBJECTIVES: Our goal was to model the relationship between the built environment and active travel for 20 MSAs spanning the continental United States.

METHODS: We sourced and processed pedestrian and bicycle traffic counts for 20 U.S. MSAs ($n = 4,593$ count locations), with 1–17 y of data available for each count location and the earliest and latest years of data collection being 1999 and 2016, respectively. Then, we tabulated land use, transport, and sociodemographic variables at 12 buffer sizes (100–3,000 m) for each count location. We employed stepwise linear regression to develop predictive models for morning and afternoon peak-period bicycle and pedestrian traffic volumes.

RESULTS: Built environment features were significant predictors of active travel across all models. Areas with easy access to water and green space, high concentration of jobs, and high rates of active commuting were associated with higher bicycle and pedestrian volumes. Bicycle facilities (e.g., bike lanes, shared lane markings, off-street trails) were correlated with higher bicycle volumes. All models demonstrated reasonable goodness-of-fit for both bicyclists (adj- R^2 : 0.46–0.61) and pedestrians (adj- R^2 : 0.42–0.72). Cross-validation results showed that the afternoon peak-period models were more reliable than morning models.

CONCLUSIONS: To our knowledge, this is the first study to model multi-city trends in bicycling and walking traffic volumes with the goal of developing generalized estimates of the impact of the built environment on active travel. Our models could be used for exposure assessment (e.g., crashes, air pollution) to inform design of health-promoting cities. <https://doi.org/10.1289/EHP3389>

Introduction

Many jurisdictions in the United States have stated goals to increase physical activity—such as walking and bicycling—to improve public health, reduce emissions, and increase livability (Hankey et al. 2017a; Jackson et al. 2013). The built environment influences rates of total physical activity (Frank et al. 2007; Giles-Corti et al. 2016; Sallis et al. 2006; Smith et al. 2017) and active travel (de Hartog et al. 2010; Götschi et al. 2016; Hankey and Marshall 2017; Oja et al. 2011; Smith et al. 2017). In addition to increasing physical activity (de Nazelle et al. 2011; Mueller et al. 2015, 2017), promotion of active travel may improve air quality (de Nazelle et al. 2011; Grabow et al. 2012; Macmillan et al. 2014; Rojas-Rueda et al. 2012), reduce crash rates (Buehler and Pucher 2017; Macmillan et al. 2014) and preventable deaths (Andersen et al. 2015; Mueller et al. 2017), and improve mood (Morris and Guerra 2015; Smith 2017).

Evidence of the association between features of the built environment and physical activity is well documented (Christian et al. 2017; Frank et al. 2016; Hankey and Marshall 2017; Ferdinand et al. 2012; Oja et al. 2011; Saelens et al. 2003; Winters et al. 2017). Population and employment density, mixed land use, street network design, and destination accessibility are among the

strongest predictors of active travel. Specifically, residents living in denser, highly connected neighborhoods with mixed land use have higher walking and bicycling rates for transport (Frank et al. 2007; Saelens et al. 2003; Saelens and Handy 2008; Sallis et al. 2013). The majority of studies on this topic focus on a single urban area. Providing generalizable information across many cities about the effect of the built environment on walking and bicycling would help cities assess what factors are most effective for planning health-promoting communities.

Direct-demand models are an alternative approach to traditional regional travel demand forecasting tools (e.g., travel demand models) for estimating traffic volumes on transportation networks (Kuzmyak et al. 2014). Developing travel demand models for bicycles and pedestrians has proven challenging due to limitations in availability of the required input data [although some progress is being made, e.g., route choice models (Hood et al. 2011) or specifying impedance factors (Iacono et al. 2010)]. Direct-demand modeling is a statistical–empirical approach (i.e., an analog to land-use regression models of air quality) that allows for the prediction of bicycle and pedestrian traffic volumes based on land-use and transportation network attributes. Therefore, these models are a potentially useful tool for generating high-resolution spatial estimates of pedestrian and bicycle volumes to inform exposure assessment and design of health-promoting cities. Direct-demand models are used widely to estimate demand for bicycle and pedestrian facilities during peak hours (Fagnant and Kockelman 2016; Griswold et al. 2011; Hankey and Lindsey 2016; Miranda-Moreno and Fernandes 2011; Pulgurtha and Repaka 2008; Schneider et al. 2009; Strauss and Miranda-Moreno 2013) and in a few cases for annual average volumes (Hankey et al. 2017b).

Most previous studies to develop direct-demand models have used traffic counts from a single city (Hankey and Lindsey 2016; Miranda-Moreno and Fernandes 2011; Tabeshian and Kattan 2014). The lack of databases that include pedestrian and bicycle traffic counts on a national scale (thus limiting spatial and temporal coverage) have precluded generalizability from these city-level studies. To the best of our knowledge, no study has combined pedestrian and bicycle count data across many metropolitan areas

Address correspondence to S. Hankey, School of Public and International Affairs, Virginia Tech, 140 Otey St., Blacksburg, VA 24061 USA. Telephone: (540) 231-7508. Email: hankey@vt.edu

Supplemental Material is available online (<https://doi.org/10.1289/EHP3389>).
The authors declare they have no actual or potential competing financial interests.

Received 20 January 2018; Revised 19 June 2018; Accepted 22 June 2018; Published 30 July 2018.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

(MSAs) to develop direct-demand models that are able to *a*) assess correlates of active travel and the built environment across many MSAs and climate regions in the United States, and *b*) more reliably predict bicycling and walking rates at locations without counts.

We aimed to address this research gap by developing a set of direct-demand models at the national level (i.e., with data from 20 U.S. MSAs) using only land-use and transportation network variables that are available at a national scale as predictors. Our work contributes to the growing literature on the impacts of the built environment on rates of active travel by assessing commonalities across many jurisdictions. An important potential application of our models is the ability to generate estimates of bicycling and walking rates in U.S. MSAs to assess exposure to crashes, air pollution, and other environmental hazards. Our models could also be used to estimate bicycling and walking rates in areas where traffic counts are few or unavailable.

Methods

We sourced, cleaned, and aggregated bicycle and pedestrian count data from 20 U.S. MSAs (Figure 1). The number of available years of data varied among MSAs (see Table S1). MSAs in our data set belong to eight of nine U.S. climate regions [based on categories by the National Oceanic and Atmospheric Administration (NOAA; Data Tools: Daily Weather Records; <https://www.ncdc.noaa.gov/cdo-web/datatools/records>)]. We compiled a corresponding database of land-use features to develop the direct-demand models;

given that a key goal of this study was to develop a tool that may allow for prediction in other jurisdictions in the continental United States, we focused on using comparable land-use data that is available at a national level. More detail on each component of this effort is described below.

Data Sources

We collected data from various publicly available sources (Table 1). First, we acquired all bicycle and pedestrian counts from the National Bicycle and Pedestrian Documentation Project (NBPDP) database (<http://bikepeddocumentation.org/>); then, we expanded this database by directly contacting additional jurisdictions to acquire available counts. We performed a Google search to identify jurisdictions that may have count data using the key words “bicycle count,” “bike count,” “pedestrian count,” and “ped count” among other count-related search terms. We downloaded count data when they were available online and sent additional requests to city staff to ensure we had complete data for that jurisdiction. The count data may not be exhaustive for all MSAs in our sample (i.e., additional count data might be available for each MSA); however, we were only able to obtain data that each jurisdiction was willing to share at the time of our request.

We performed a preliminary screening to eliminate jurisdictions that had a small number of count locations (i.e., fewer than 30 locations). Additionally, we grouped counts for jurisdictions in the same MSA (e.g., suburbs of central cities), resulting in 20 MSAs for inclusion in our analysis (Figure 1). The number of



Figure 1. Map of the MSAs with bicycle or pedestrian count data used in this study along with the nine climate regions based on NOAA’s designation (shown in gray). Created by the authors based on TIGER data (U.S. Census Bureau 2017).

Table 1. Description of data used to develop direct-demand models.

Type of data	Source	Unit of measurement	Areal unit of base data	Tabulation method	Year
Bicycle and pedestrian traffic counts	NBPDP; local agencies	AM/PM peak hour	Point	—	2000–2016
Land-use data					
Water	TIGER	Meter squared	Polyline	Buffer	2014
Park	Local agencies	Meter squared	Polyline	Buffer	Multiple
Housing units	ACS/SLD	Count	Block group	Buffer	2010
Total number of jobs	ACS/SLD	Count	Block group	Buffer	2010
Number of households	ACS/SLD	Count	Block group	Buffer	2010
University or college campus	TIGER	Meter squared	Polyline	Buffer	2014
Transportation-related data					
Number of zero-vehicle households	ACS/SLD	Count	Block group	Buffer	2010
Bicycle commute mode share	ACS	Percent	Block group	Buffer	2014
Walking commute mode share	ACS	Percent	Block group	Buffer	2014
Public transport commute mode share	ACS	Percent	Block group	Buffer	2014
Number of public transit stops	SLD	Count	Point	Buffer	2010
Local road network	TIGER	Meter	Polyline	Buffer	2010
Network density: facility miles of multi-modal links per mile squared ^a	SLD	Miles per mile squared	Block group	Buffer	2010
Street intersection density (weighted, auto-oriented intersections eliminated) ^a	SLD	Intersections per mile squared	Block group	Buffer	2010
Bicycle facility ^b	Google Earth	Type	Point estimate	Point	2000–2016
Socioeconomics					
Median household income	ACS	U.S. dollar	Block group	Buffer	2014
Population <18 y of age	ACS	Percent	Block group	Buffer	2014
Population 18–45 y of age	ACS	Percent	Block group	Buffer	2014
Population 46–65 y of age	ACS	Percent	Block group	Buffer	2014
Population >65 y of age	ACS	Percent	Block group	Buffer	2014
Weather data					
Precipitation	NOAA	Inch	—	—	2000–2016
Temperature	NOAA	Degrees Fahrenheit	—	—	—
Climate region	NOAA	—	Region	—	1984

Note: See Tables S1 and S2 for more information about types of counts and count methods. ACS, American Community Survey; NBPDP, National Bicycle and Pedestrian Documentation Project; NOAA, National Oceanic and Atmospheric Administration; SLD, Smart Location Database; —, not applicable.

^aFor a full description of these data, please see the Smart Location Database (SLD) User's Guide (Ramsey and Bell 2014, pp. 20–23).

^bBicycle facilities included off-street facilities (e.g., trails), on-street facilities (e.g., bike lanes, buffered bike lanes), and minor facilities (e.g., sharrows and bike boulevards).

years available for analysis varied among the MSAs, with 1–17 y of data available for each count location and with the earliest and latest years of data collection being 1999 and 2016, respectively. The inclusion of suburban count locations diversifies the built environment factors included in our database. However, many suburbs did not have bicycle and pedestrian counts; thus, our count database is skewed toward urban locations. Table S1 shows MSAs included in our analysis, types of jurisdictions (central city and/or suburbs) in each MSA, years, hour-of-day, and methods used for collecting counts.

We downloaded nationally available built environment variables (e.g., land-use and transportation-network characteristics) from the American Community Survey (ACS) 5-y averages (U.S. Census Bureau 2014), U.S. EPA's Smart Location Database (SLD) (U.S. EPA 2017), and TIGER (U.S. Census Bureau 2017). We also collected weather variables (i.e., temperature and precipitation) from the National Oceanic and Atmospheric Administration (NOAA Data Tools: Daily Weather Records database).

Because spatial data on bicycle facilities were not available for many MSAs, we used Google Earth (Pro version 7.3) imagery (which has a historical view option to track bicycle infrastructure changes over time) to assess whether bicycle facilities existed at each count location (see access dates in Table S1). If Google Earth views were blocked by cloud cover, or simply unavailable, we consulted other sources including Google Maps, Google Street View, and the MSA bicycle network shapefile (if available) to classify each count location.

Data Processing

A core task of our study was to assemble a national-scale database of bicycle and pedestrian traffic counts and to match those counts with corresponding land-use and transportation network variables. Below we describe our approach for data cleaning and aggregation and give descriptive statistics of the database.

Bicycle and pedestrian count data. As described above, we obtained traffic count data at 6,342 locations (12,231 bicycle count observations and 10,827 pedestrian count observations) by contacting individual jurisdictions and from the NBPDP database. Count methods varied by jurisdiction (see Table S1). For example, 13 MSAs employed manual counts (i.e., volunteer-based paper counts), two jurisdictions employed automated counters (e.g., loop detectors, pneumatic tubes, video counts, infrared, or radio beam counters), and five jurisdictions used a combination of both methods to collect traffic counts.

Because traffic counts are collected in different ways among jurisdictions, we developed a protocol for ensuring counts were comparable for all count locations. Our count database included both bidirectional (i.e., separate counts for each direction of traffic) and screenline (i.e., total volume including both directions) counts. For consistency, we converted all bidirectional counts to screenline counts that represent a total traffic volume (regardless of direction) for the street segment or intersection. Because our data set includes counts at street segments and intersections, we separated the two types of count locations for eventual use in

separate models (see below for details on model building) due to the difference in absolute volumes expected at these network locations. For example, multiple segments converge into each intersection; therefore, the traffic volume at each segment is smaller than at an intersection if all else is equal. For intersection counts with turning movements, we separated the counts into segment counts for each leg of the intersection for use in the segment models.

Depending on the jurisdiction, either raw (i.e., 15-min intervals) or aggregated (e.g., 2-h peak period) count data were available (see Table S1). We aggregated all 15-min counts into 2-h counts to allow for comparison across geographies. Most MSAs collect counts during morning (AM; 0700–0900 hours) and/or afternoon (PM; 1600–1800 hours or 1700–1900 hours) peak periods on weekdays. As such, we cleaned and aggregated the raw count data for weekday, peak-period counts only and excluded other hours and weekends from our analysis. Finally, we aggregated counts by season (see Table S1). Specifically, we averaged all counts at each individual count location, during the same time of day, in the same year, for each season. The fall season (August to November) accounted for the majority of the data set [i.e., 81% (9,901) of bicycle counts and 71% (7,664) of pedestrian counts]. Counts in other months accounted for a relatively smaller portion of the data set; thus, to increase the homogeneity among the MSAs and season for the purpose of developing the models, we removed nonfall counts from our database. We also excluded count locations where the location type (i.e., intersection or segment) was not clear [3.1% (309) of bicycle counts and 1.5% (117) of pedestrian counts].

The procedure described above resulted in our final data set for model-building, that is, separate counts of bicycles and pedestrians during two peak periods (morning and afternoon) and a single season (fall) for two types of network locations (intersections and segments). The resulting database allowed for developing four models (two network location types and two times of day) for each transportation mode at 4,593 count locations and including 9,592 counts of bicycle and 7,547 counts of pedestrian traffic.

Table 2 provides the number of traffic counts by peak period and network location type (see Table S2 for a summary of unique

count locations by time period and location type). The number of count locations is smaller than the number of counts given that a count location may be counted in one or multiple years. In total, 9,592 traffic counts were retained for bicycle traffic, of which the afternoon peak period accounted for 64% of the sample. Pedestrian traffic counts totaled 7,547 observations –60% for the afternoon peak periods. Most jurisdictions collected traffic counts during the afternoon; seven MSAs did not count bicyclists and pedestrians during the morning peak period. Traffic counts collected at intersections account for 49% and 46% of the total sample for bicycle and pedestrian traffic, respectively. Some MSAs did not have both bicycle and pedestrian counts, for example, Denver and Los Angeles (no pedestrian counts), and New York City (no bicycle counts). Thus, our models only reflect data that were available for each MSA, which might reduce the generalizability of our models.

Land-use and transportation network data. For each count location within an MSA, land-use and transportation network data were tabulated using the land use regression (LUR) tools for ArcGIS (Akita 2014). The tools allow for the measurement of areas of polygons, number of points, or distance of lines that fall inside a buffer. We tabulated land use, sociodemographic, and most transport-related variables (e.g., area of land-use types, weighted average income, percentage of population by age group, commute mode share) listed in Table 1 at 12 different buffer sizes (i.e., 100, 200, 300, 400, 500, 750, 1,000, 1,250, 1,500, 2,000, 2,500, and 3,000 m) following Hankey and Lindsey (2016). Presence and type of bicycle facility were obtained only at the count location itself using Google Earth (as described above) (see Table S1). We categorized bicycle facilities as follows: (a) on-street facilities including bike lanes, buffered bike lanes, and protected bike lanes/cycle tracks; (b) minor facilities including sharrows (shared lane markings) and bike boulevards; and (c) off-street facilities including trails and shared-use paths that are completely separated from vehicular traffic.

For the segment models, on- and off-street facility types were introduced as dummy variables, with the value 1 indicating that a facility is present in at least one travel direction. For the intersection models, we assigned a value of 1 for each facility present in

Table 2. Number of bicycle and pedestrian counts by peak period and location type.

MSAs	Bicycle				Pedestrian			
	AM Seg.	PM Seg.	AM Int.	PM Int.	AM Seg.	PM Seg.	AM Int.	PM Int.
Blacksburg, VA	101	101	—	—	72	72	—	—
Boston, MA	36	37	5	6	8	9	5	3
Champaign-Urbana, IL	255	255	66	66	—	—	121	121
Cleveland, OH	—	82	—	—	—	81	—	—
Columbus, OH	208	—	7	—	208	—	7	—
Denver, CO	—	—	47	73	—	—	—	—
Hartford, CT	1	11	3	60	1	11	3	60
Lawrence, KS	—	98	—	—	—	98	—	—
Los Angeles, CA	514	424	462	461	—	—	—	—
Madison, WI	73	73	91	144	73	73	—	—
Manhattan, KS	—	—	—	112	—	—	—	112
Minneapolis, MN	—	950	—	—	—	950	—	—
New York City, NY	—	—	—	—	1,022	1,022	—	—
Philadelphia, PA	192	192	—	—	162	165	—	—
Portland, OR	—	55	—	36	—	55	—	36
San Francisco, CA	—	1,084	—	305	—	2	—	78
Seattle, WA	16	5	254	249	16	5	256	249
St. Louis, MO	—	—	—	140	—	—	—	236
Tucson, AZ	4	4	1,054	1,052	3	3	1,064	1,065
Washington, DC	54	54	10	10	—	—	10	10
Total	1,454	3,425	1,999	2,714	1,565	2,546	1,466	1,970

Note: Data were derived from raw traffic count data obtained from each jurisdiction or from the National Bicycle and Pedestrian Documentation Project (NBPDP) during 1999–2016. Only traffic counts in fall (August to November) and weekday peak periods were included. AM Int., morning count at intersections; AM Seg., morning count at street segment; PM Int., afternoon count at street intersection; PM Seg., afternoon count at street segment; —, data not collected.

each travel direction on a leg of the intersection; the values were then summed for each facility type at each intersection. For example, if a four-leg intersection has eight travel directions, and bike lanes are present for two travel directions and sharrows are present for three travel directions, the intersection will receive a value of 2 for on-street facility and 3 for minor facility. As such, bicycle facilities at a four-leg intersection (e.g., on-street facility, off-street facility, minor facility) could receive values from 0 to 8 (no five- or six-leg intersections had facilities in all directions in our sample) and were treated as continuous variables.

Weather data. We obtained weather data from the NOAA Data Tools: Daily Weather Records database for each count date and each city. For each day a count was collected, we assigned the lowest temperature of the day for morning counts, the highest temperature for afternoon counts, and daily total precipitation for morning and afternoon counts in each MSA. We aggregated temperature and precipitation data along with the count data from August to November. We created dummy variables representing climate regions based on the NOAA designation (NOAA Data Tools: Daily Weather Records database; [Karl and Koss 1984](#)) ([Figure 1](#)) with the goal of capturing regional differences among MSAs in the data set. MSAs in our data set belong to eight of nine climate regions, with no count data from the West North Central region.

Direct-Demand Model Development

Once the data were cleaned and aggregated, we developed a set of direct-demand models for bicycle and pedestrian traffic during morning and afternoon peak periods. We modeled bicycle and pedestrian traffic (dependent variables) using land use, transportation network, weather, and sociodemographic variables as predictors (independent variables). The dependent variables (bicycle and pedestrian counts) followed a log-normal distribution and were log-transformed for model building. Zero-count values of the dependent variables were dropped from the data set with this procedure (2.9% of the total bicycle counts and 5.0% of the pedestrian counts used for modeling).

Core models. We applied a forward stepwise regression approach to select the variables most correlated with active travel among the potential predictor variables ([Table 1](#)) following the procedure used in a previous study ([Hankey and Lindsey 2016](#)). In this process, the independent variable with the highest correlation with the dependent variable (log of bicycle or pedestrian count) was selected first. Then, the independent variable with the highest correlation with the model residuals was added to the model. To avoid multicollinearity, our procedure did not select variables that were highly correlated with one of the previously chosen independent variables (based on variance inflation factor, i.e., $VIF \leq 5$). The process continued until the coefficient of the last variable included in the model was statistically insignificant (had a $p > 0.05$) or violated criteria for multicollinearity ($VIF > 5$). This variable was then removed and the model was considered complete. Each variable was measured at 12 buffer sizes and included for selection in the model building process. However, each variable was allowed to be selected only once among all buffer sizes (i.e., the buffer size that had the highest correlation with the residuals, among all buffers for each variable, was selected).

All coefficients were standardized to facilitate comparison across models and mode of transportation. Specifically, we calculated the 5th–95th percentile range for each variable. The standardized coefficients were then calculated by multiplying the regression coefficient by this factor: 5th–95th percentile range of the independent variable divided by the 5th–95th range of the dependent variable.

Sensitivity analysis. To test the sensitivity of our models to the input data, we developed two separate sensitivity analyses. First, we attempted to eliminate the year-to-year temporal variation in the bicycle and pedestrian count data set by averaging bicycle and pedestrian counts among all years at each count location. Count locations within MSAs with fewer than 3 y of counts were censored from the data set for this analysis and new models were developed to compare to the core models.

Second, we developed models using disaggregate employment (i.e., separate estimates of industry, retail, service, entertainment, office jobs) and open space data (i.e., separating park and water) rather than using aggregate predictor data (i.e., total employment, aggregate open space). All other independent variables from the core models were included in these disaggregate models. This analysis aimed to assess whether the disaggregate employment data can better approximate variation in land use in an urban area than total employment.

Cross validation. We performed cross validation using three methods: Monte Carlo–based random hold-out, systematic hold-out of individual MSAs, and a revised systematic hold-out that gradually introduces increments of data from the hold-out MSA to the model training data set.

Monte Carlo-based random hold-out. We randomly separated the data set into (a) a training data set (containing a random selection of 90% of the original sample used for the core models), and (b) a test data set (containing the remaining 10% of the original sample) to validate the results of the training models. The process was repeated 100 times for each model. This validation procedure assesses the performance of the models when used to predict bicycle and pedestrian volumes at locations without counts within the same jurisdictions in the model database.

Systematic hold-out. To further test the ability of our models to predict traffic volumes in MSAs outside of our sample, we performed a systematic hold-out of individual MSAs by sequentially holding out each MSA ($n = 20$) in our data set. In this procedure, the training model includes data from 19 MSAs and is tested using count data from the 20th MSA (this process was repeated iteratively for all MSAs). This validation procedure assesses model performance for MSAs that do not have any historical counts to include in model building.

Revised systematic hold-out. We also tested the improvement of model predictions when small amounts of count data are available for a jurisdiction. Specifically, to build on the systematic hold-out, we sequentially added increments of the count data from the 20th MSA to the training data set to assess whether a small amount of data from a given jurisdiction improves model predictions for that jurisdiction. Specifically, we incrementally included data from the 20th MSA (10%, 25%, 50%, 75%, 90%, and 100%) in the training data sets (in addition to all data from the other 19 MSAs), then tested model performance on the remaining data from that MSA. This analysis assesses (a) the prediction performance MSAs could expect by adding a small amount of data to our models, and (b) how each incremental increase in input data affects the prediction power of the models.

Predicting spatial patterns of bicycle and pedestrian traffic. An important application of our models is the ability to predict traffic volumes at spatial locations that do not have counts. The resulting spatial predictions could be used as an input to exposure assessment, for example, converting crash numbers to crash rates or assessing exposure to air pollution. We demonstrate this potential application by estimating spatial predictions of bicycle and pedestrian traffic volumes in Washington, DC, and Minneapolis, Minnesota, for every street segment. We used ArcGIS to create midpoints of all road segments for Washington, DC, and Minneapolis. We then applied models from the systematic hold-

out cross validation to predict traffic volumes at these MSAs. Using the models from the systematic hold-out simulates model performance for an MSA without any counts (i.e., out-of-sample prediction); therefore, the resulting maps illustrate the quality of results a jurisdiction could expect if they used our models without any local counts. In our example, we made predictions for afternoon peak-period counts in 2016 (the last year of count data) on a typical fall day [25°C (77°F) with no rain].

Results

We developed direct-demand models that included count data from 20 MSAs and predictor variables that are available at a national scale. All models showed relatively good model fit with adjusted R^2 ranging from 0.46 to 0.61 for bicycle traffic and 0.42 to 0.72 for pedestrian traffic. The full models are shown in Table 3 (unstandardized coefficients are in Table S3; correlations of all variables in the models are in Table S4). Because the dependent variables (bicycle and pedestrian counts) were log-transformed, the standardized coefficients are interpreted as percent change in the 5th–95th percentile range of pedestrian or bicycle volume. When referring to a percent increase in the discussion of model results below, we are referring to this relationship. Because the origins and destinations of bicycle and walk trips were unknown, we included socioeconomic characteristics of the neighborhoods around the count locations as control variables, but did not interpret them as main variables of interest.

Bicycle Models

Bicycle model fit (adj- R^2) ranged from 0.46 to 0.61 (Table 3). Transportation-network variables and land-use variables were selected in most bicycle models. Weather, climate region, and year were selected in the models as control variables.

Water and green space showed a positive correlation with bicycle traffic (Table 3). Each 5th–95th percentile change in water and green space (within a 200–500 m buffer) was associated with a 4–8% increase in the 5th–95th percentile range of bicycle traffic. Number of jobs was also correlated with higher bicycle traffic volumes in three of the models (5–29% increase). Count locations in close proximity to university or college campuses were positively associated with bicycle volume (12–14% increase); however, this variable was only selected in the intersection models.

Variables describing characteristics of the transportation network were selected in all models (Table 3). Off-street facilities (i.e., bike trails, shared-use paths) were correlated with higher bicycle volumes. For the segment models, count locations with off-street facilities had higher bicycle volumes (25–31%) as compared with count locations without such facilities. At intersections, each additional off-street facility (for each direction of traffic flow) was associated with a 13–18% increase in bicycle traffic. We found similar positive effects among models for on-street facilities, although the magnitude of effect was reduced (10% increase at segments, 12–15% increase at intersections). Minor facilities, such as bicycle boulevards or sharrows (shared lane markings), also had a small but positive correlation with bicycle traffic (3–6% increase at intersections).

Multimodal network density, measured as miles of streets that accommodate various modes of transport per mile squared, was a strong positive predictor for bicycle traffic, with a stronger relationship in the segment models as compared with the intersection models (7–43% increase on segments; 5–11% at intersections). Intersection density was selected in the two afternoon peak-period models (6–22% increase in bicycle volume). Fewer bicycles were found on local roads (5–25% decrease in bicycle volume at street segments).

The models also showed that neighborhoods with greater bicycle commute mode share (as reported by the ACS) were associated with higher bicycle traffic counts. Higher bicycling volumes were correlated with neighborhoods with lower car ownership rates, although this effect was only detected in two of four models.

Pedestrian Models

Model fit (adj- R^2) for the pedestrian models ranged from 0.42 to 0.72 (Table 3). Similar to the bicycle models, the pedestrian models selected various land-use variables as major predictors of pedestrian traffic. The effect of water and green space, however, was mixed and not as pronounced as was the case in the bicycle models (e.g., the effect size was between –15% and +9% change in pedestrian volume for the morning segment and afternoon intersection models, respectively). Household density was positively correlated with pedestrian volume at intersections during the afternoon peak period and segments during the morning peak period, with a wide range in magnitude among models (e.g., 42% increase in the morning segment model and 4% increase in the afternoon intersection model) partly owing to the different nature of these two types of count locations. Number of jobs (4–17% increase in all models except the afternoon segment model) and proximity to a university or college campus (4–20% increase in the afternoon models) were correlated with higher pedestrian volumes.

As with the bicycle models, multimodal network density was a strong positive predictor for pedestrian traffic in three models (8% for the afternoon intersection model, 3% increase for the segment models) (Table 3). Intersection density was positively associated with pedestrian volume during afternoon peak period (24% increase in pedestrian traffic). However, intersection density was not selected in the morning peak-period models. Unlike the bicycle models, local roads were positively associated with pedestrian traffic in three models (13–28% at intersections; 15% at segments during afternoon peak period).

Three pedestrian models also showed that neighborhoods with greater walking commute mode share were associated with higher pedestrian volumes (Table 3). Neighborhoods with a higher density of transit stops were positively correlated with pedestrian volumes (except for the afternoon segment model where it was inversely correlated). Conversely, areas with high public transit commute mode share were negatively correlated with pedestrian volumes in two models. This contradictory finding could potentially be a result of confounding effects among transit-related variables and walking commute mode share. For example, areas with higher levels of transit service could also be areas with high rates of walking. As expected, neighborhoods with a high number of zero-car households were correlated with higher pedestrian traffic (afternoon models only).

Sensitivity Analyses

We explored two sensitivity analyses: (a) temporally averaging pedestrian and bicycle counts across years as an alternate input to model-building (see Tables S5 and S6 and Figure S1), and (b) disaggregating employment data (from total employment to sector-based employment) as a proxy for land use types (see Tables S7 and S8 and Figure S2). Overall, results from our sensitivity analyses show that the signs of coefficients for all land-use and transportation-network variables are consistent across models (although the magnitude of coefficients varies). Bicycle facilities remained an important predictor of bicycle traffic in all bicycle models, suggesting that developing robust networks of facilities is important to attract bicyclists. Similarly, network variables such as intersection density and multimodal network density

Table 3. Direct-demand model results for bicycle and pedestrian volumes in 20 U.S. MSAs.

Independent variable	Bicycle				Pedestrian			
	AM Seg.	PM Seg.	AM Int.	PM Int.	AM Seg.	PM Seg.	AM Int.	PM Int.
Land use								
Water and green space	0.04 (300 m)	0.06 (200 m)	0.06 (400 m)	0.08 (500 m)	-0.15 (300 m)	—	—	0.09 (500 m)
Household density	—	—	—	—	0.42 (1,000 m)	—	—	0.04 (1,000 m)
Total number of jobs	—	0.29 (300 m)	0.05 (300 m)	0.14 (200 m)	0.17 (400 m)	—	0.08 (500 m)	0.04 (400 m)
University/college campus	—	—	0.12 (400 m)	0.14 (300 m)	—	0.04 (500 m)	—	0.20 (300 m)
Transport network								
Off-street bike facility	0.31	0.25	0.18	0.13	—	—	—	—
On-street bike facility	0.10	0.10	0.12	0.15	—	—	—	—
Minor bike facility	—	—	0.03	0.06	—	—	—	—
Intersection density	—	0.06 (400 m)	—	0.22 (2,500 m)	—	0.24 (750 m)	—	0.24 (1,500 m)
Multimodal network density	0.43 (3,000 m)	0.07 (100 m)	0.11 (3,000 m)	0.05 (500 m)	0.03 (100 m)	0.03 (100 m)	—	0.08 (100 m)
Local road	-0.25 (3,000 m)	-0.05 (200 m)	—	—	—	0.15 (1,250 m)	0.28 (2,000 m)	0.13 (750 m)
Other transport								
Bike commute share	0.30 (750 m)	0.17 (1,000 m)	0.26 (3,000 m)	0.15 (750 m)	—	—	—	—
Walking commute share	—	—	—	—	—	0.11 (100 m)	0.34 (100 m)	0.15 (400 m)
Transit stops	—	—	—	—	0.08 (400 m)	-0.24 (3,000 m)	0.07 (100 m)	0.08 (100 m)
Transit commute share	—	—	—	—	-0.26 (3,000 m)	—	—	-0.14 (3,000 m)
Zero-car households	0.11 (1,000 m)	—	—	0.04 (100 m)	—	0.46 (3,000 m)	—	0.08 (100 m)
Sociodemographics								
Income	0.20 (3,000 m)	-0.04 (1,000 m)	0.20 (2,500 m)	—	-0.10 (2,000 m)	—	—	-0.08 (100 m)
Population <18 y of age	—	—	-0.19 (3,000 m)	-0.08 (500 m)	—	—	—	—
Population 19–45 y of age	—	0.17 (500 m)	—	—	—	0.11 (100 m)	—	—
Population 45–65 y of age	-0.19 (100 m)	—	-0.12 (500 m)	-0.05 (100 m)	-0.05 (100 m)	—	—	—
Population >65 y of age	-0.09 (500 m)	0.04 (100 m)	-0.08 (100 m)	-0.07 (300 m)	—	0.09 (3,000 m)	—	—
Temperature	0.05	-0.04	—	—	0.08	—	—	—
Precipitation	-0.20	—	—	0.03	—	—	—	—
n	1,126	3,279	1,915	2,533	1,545	2,526	1,202	1,657
Adj-R ²	0.50	0.46	0.49	0.61	0.61	0.72	0.42	0.60
MSAs ^a	BBG, BOS, COL, LA, MAD, PHI, SEA, TUC, DC	BBG, BOS, CLE, HAR, LAW, LA, MAD, MIN, PHI, POR, SF, SEA, TUC, DC	BOS, CU, COL, HAR, LA, MAD, MAN, MAD, SEA, TUC, DC	BOS, DEN, HAR, LA, MAD, MAN, POR, SF, SEA, STL, TUC, DC	BBG, BOS, COL, HAR, MAD, NYC, PHI, SEA, TUC	BBG, BOS, CLE, HAR, LAW, MAD, MIN, NYC, PHI, POR, SF, SEA, TUC	BOS, CU, COL, HAR, SEA, TUC, DC	BOS, CU, HAR, LAW, MAN, POR, SF, SEA, STL, TUC, DC

Note: Buffer sizes are shown in parentheses. Results obtained using stepwise linear regression method. All dependent variables were log-transformed. The standardized coefficients are interpreted as percentage change in the 5th–95th percentile range of pedestrian or bicycle volume. All independent variables were significant at $p < 0.05$ level. All models included year and climate region as control variables (see Table S3 for model results with unstandardized coefficients and all control variables). Morning peak period (AM) is 0700–0900 hours. Afternoon peak period (PM) is 1600–1800 hours or 1700–1900 hours. AM Int., morning intersections model; AM Seg., morning segment model; PM Int., afternoon intersection model; PM Seg., afternoon segment model; —, data not collected.
^aChampaign-Urbana (CU) was excluded from the PM Intersection model because its count dates were unknown, thus we did not have weather variables for this MSA. City abbreviations: BBG, Blacksburg; BOS, Boston; CLE, Cleveland; COL, Columbus; CU, Champaign-Urbana; DC, Washington DC; DEN, Denver; HAR, Hartford; LA, Los Angeles; LAW, Lawrence; MAD, Madison; MAN, Manhattan; MIN, Minneapolis; NYC, New York City; PHI, Philadelphia; POR, Portland; SEA, Seattle; SF, San Francisco; STL, St. Louis; TUC, Tucson.

Table 4. Cross validation results.

Cross validation type	Test	Bicycle models				Pedestrian models			
		AM Seg.	PM Seg.	AM Int.	PM Int.	AM Seg.	PM Seg.	AM Int.	PM Int.
Random hold-out	Average test R^2	0.44	0.44	0.46	0.59	0.6	0.72	0.41	0.58
	Drop in R^2	0.06	0.02	0.03	0.02	0.01	0.00	0.01	0.02
	Average MSE	1.17	1.07	0.8	0.85	1.49	1.34	1.25	1.16
Systematic hold-out	Average test R^2	0.29	0.39	0.45	0.37	0.38	0.44	0.60	0.49
	Drop in R^2	0.21	0.07	0.05	0.23	0.23	0.28	-0.18	0.11
	Average MSE	2.95	2.77	1.70	1.28	5.24	3.23	2.43	2.27

Note: Results obtained using Monte-Carlo random hold-out and systematic hold-out cross validation method. AM Int., morning intersections model; AM Seg., morning segment model; MSE, mean square error; PM Int., afternoon intersection model; PM Seg., afternoon segment model.

were found to be strongly and positively associated with bicyclist and pedestrian activity.

The temporally averaged models were more parsimonious (i.e., fewer predictor variables were selected) as compared with the full models, potentially owing to the significant reduction in number of observations and variability by averaging counts in the temporally averaged models (see Tables S5 and S6). Model fit (adj- R^2) was similar to the core models (i.e., 0.39–0.61 for the bicycle models; 0.44–0.71 for the pedestrian models). When more consistent bicycle and pedestrian traffic count campaigns become available in the future (i.e., with repeated counts at locations over time), this method may be more appropriate for modeling the spatial patterns of bicycle and pedestrian activity by removing temporal effects.

Results from the models using disaggregate employment type (as a proxy for land-use type) were mixed (see Tables S7 and S8). Model fit (adj- R^2) was similar to the core models (0.48–0.61 for bicycle models; 0.46–0.73 for pedestrian models). Effects of the nonemployment variables on bicycle and pedestrian traffic were similar to the core models (Table 3). The disaggregate employment variables had mixed correlations with bicycle and pedestrian volumes, possibly owing to high correlations among the employment types (see Table S9). Because model results among employment types were sometimes counterintuitive, we suggest using the core models that use aggregate employment data for prediction (Table 3).

Model Validation Results

Based on model fit, the afternoon peak-period models generally performed better than the morning peak-period models (except for the bicycle morning peak-period segment model). To further test model performance, we conducted a series of random and systematic hold-out procedures to validate the models. We calculated adjusted R^2 and mean square error (MSE) as indicators of predictive power.

For the random hold-out validation, the average adjusted R^2 from the training models were similar to the values displayed in Table 3 for the full models. The training models were used to estimate traffic counts at locations in the test data set for comparison. In general, the validation results were robust with only modest changes in adjusted R^2 of 0.02–0.06 for the bicycle models and 0.00–0.02 for the pedestrian models. The average MSE ranges from 0.80 to 1.17 for the bicycle models and 1.16 to 1.49 for the pedestrian models (Table 4). Variables selected in the training models were mostly consistent with the variables in the full models indicating that our models have reasonable out-of-sample prediction. Cross-validation results for afternoon (PM) models are shown in Figure 2; results for morning (AM) models are shown in the Supplemental Material (see Figure S3).

As a more rigorous test of each model's predictive power, we used a systematic hold-out procedure. We sequentially held out each MSA (i.e., the training data set becomes the remaining 19 MSAs) and predicted bicyclist and pedestrian volumes at each

count location in the hold-out MSA (this process was repeated for each MSA). Drops in adjusted R^2 for the systematic hold-out (i.e., 0.05 to 0.23 for the bicycle models and -0.18 to 0.28 for the pedestrian models) are higher than those in the random hold-out procedure. The average mean square error of all models for individual MSAs is higher compared with that of the random hold-out approach, increasing from 1.28 to 2.95 for the bicycle models and from 2.43 to 5.24 for the pedestrian models (Table 4). The poor performance of the pedestrian models is likely attributable to outlier MSAs, for example, pedestrian counts in New York City. Figure S4 shows the cross validation results from the random hold-out and systematic hold-out approaches.

Last, we modified the systematic hold-out procedure to allow for incrementally adding data from the 20th MSA to the training data set to test whether a small amount of data from the 20th MSA has the ability to improve model predictions (similar to the systematic hold-out, this procedure was applied to all MSAs). We performed this validation for the afternoon peak-period models only given that the analyses reported above suggested that the morning peak-period models would likely require more count data to be reliable for prediction. Figure 3 shows an improvement in predictive power of the model, as suggested by the reductions in MSE, as more data from the 20th MSA is added to the training data set. This finding suggests that the predictive power of the models improves when adding a small portion of the count data from a given MSA rather than making predictions in that MSA with no count data. The practical implication of this finding is that a city or MSA with limited resources (and thus limited count data that is not sufficient to build a city-specific model) could potentially leverage our database to make reasonable predictions.

Example Application: Predicting Bicycle and Pedestrian Volumes in Two MSAs

Based on prediction results from the systematic hold-out cross validation, we generated spatial estimates of bicycle and pedestrian volumes for all roads and off-street trails in Washington, DC, and Minneapolis based on the afternoon peak-period segment models (Figure 4).

The maps show a pattern of high levels of active travel along main corridors and city centers with high density and mixed-use developments (Figure 4). The predictions are for afternoon peak-period bicycling and walking volumes that are mainly driven by local commuters and exercise-related trips. As such, tourist attractions in these MSAs may show lower than expected traffic in these maps.

Discussion

This study supports findings from previous research on the impact of the built environment on active travel (Fagnant and Kockelman 2016; Hankey and Lindsey 2016; Heesch et al. 2015; Saelens et al. 2003; Saelens and Handy 2008; Sallis et al. 2013; Smith et al. 2017; Winters et al. 2017). Our study expands on

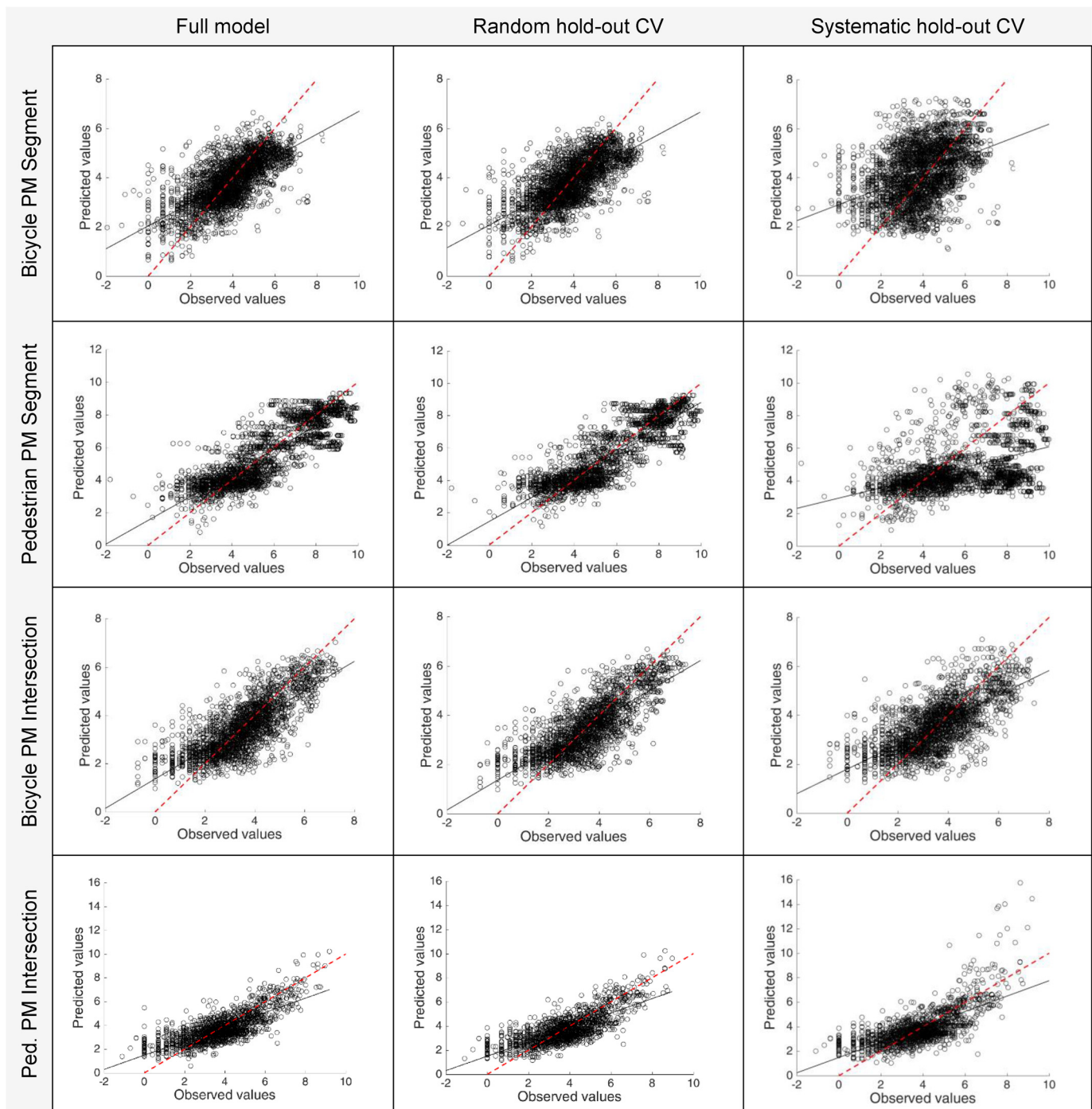


Figure 2. Full model and cross validation results. Plots of predicted vs. observed values of the afternoon (PM) peak-period models for the full model and each cross-validation (CV) approach. The dashed red line is the 1:1 line; the solid black line is the best fit line. [For cross-validation results for morning (AM) models, see Figure S3]. Note: Ped, pedestrian.

previous work by modeling this relationship across 20 MSAs rather than focusing on a single urban area or region. Consistent with findings from the single-city studies, we found that various land-use features and transportation network variables (e.g., job density, water and parks, network and intersection density) were positively correlated with pedestrian and bicycle traffic.

An important finding is that bicycle infrastructure was highly correlated with bicycle volumes and that this relationship held across many MSAs and climate zones. For example, off-street facilities (e.g., trails, shared-use paths) had the strongest association with bicycle traffic (~18 to 31% increase in bicycle traffic)

likely owing to the fact that users perceive these facilities as the safest type of infrastructure and they may therefore attract less-experienced bicyclists (Buehler and Dill 2016; Buehler and Pucher 2012). Major on-street (10–15% increase) and minor (3–6% increase) facilities were also correlated with bicycle traffic. Although many planning efforts are focused on high quality off-street and on-street facilities, minor facilities (e.g., sharrows) may play a useful role in completing the bicycle network. Our model results suggest that minor facilities have a positive correlation with bicycle traffic and may be a useful alternative when street width does not allow room for other bicycle facilities and

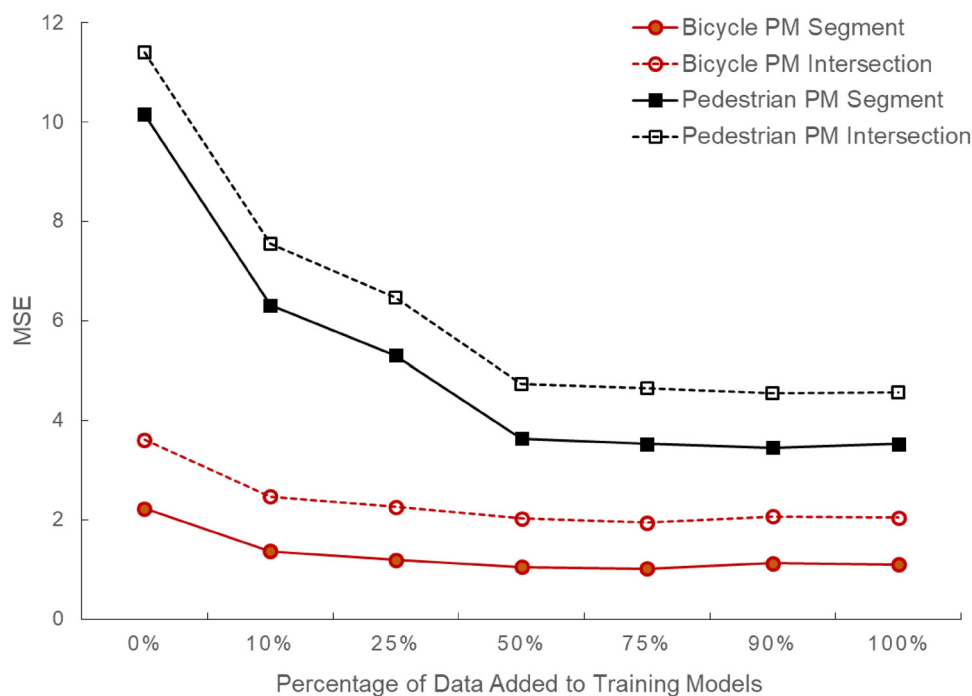


Figure 3. Results from the revised systematic hold-out procedure showing reduction of mean square error (MSE) when data is incrementally added from the 20th (hold-out) MSA to the training models. Values shown are averaged across the hold-out results for each individual MSA. PM, afternoon.

traffic volumes and speeds are low enough for cyclists to share the roadway (as recommended in [AASHTO \[2012\]](#) or [NACTO \[2014\]](#)). In total, the effects from off-street (18–31%), and on-street (10–15%) facilities were greater than the average land-use impact (mean land-use effect: 11%). We did not separate different types of on-street facilities (e.g., protected bike lanes vs. conventional bike lanes) because very few count locations (i.e., only 11 count locations had protected bike lanes) had higher level on-street facilities. Future work to address this limitation would be useful.

Our models have several limitations pertaining to data availability and modeling approach. For example, our goal was to develop a model that was capable of predicting bicycle and pedestrian volumes in all jurisdictions in the United States. However, many important predictors of bicycle and pedestrian traffic identified in single-city studies (e.g., road functional class, vehicular traffic, speed limit) were unavailable on a national scale and thus were not included in our models. We found few MSAs with bicycle and pedestrian count data in some regions of the United States (i.e., the Southeast and West North Central). This may reduce the generalizability of our study to MSAs in those regions. Additionally, most MSA-wide bicycle and pedestrian counts (as well as the counts used in this study) focus on the fall season (August–November) due to constraints in city- or MSA-level funding for count efforts. Methods for counting bicycles and pedestrians vary across the country. Site selection criteria differ across MSAs, and count sites are not randomly selected. Therefore, work to make bicycle and pedestrian traffic counts more consistent across jurisdictions would be useful for spatial modeling. In general, increased attention to developing robust traffic count campaigns for bicycles and pedestrians (similar to those for vehicular traffic) would greatly improve the performance of future modeling efforts.

A challenge when developing national-scale models is that land-use data vary across jurisdictions. Thus, models developed at the national scale use crude measures of land use (e.g., housing density, network density, total employment) rather than refined

land-use information (e.g., retail area, commercial area). We attempted to create a proxy for land use by splitting employment data by sector. However, this effort did not seem to fully capture the nuances of land-use patterns. Future efforts to develop more specific land-use patterns in a consistent format across the country would benefit the modeling approach described here.

Our modeling approach followed previous direct-demand models in an attempt to compare results among our multi-MSA models to previous single-city models. Our modeling approach does not account for spatial dependence (i.e., autocorrelation), which is an issue that should be addressed in future work. Furthermore, more work to expand our database and replicate this study would be useful. Specifically, adding additional MSAs, assessing how count sites were chosen, and adding counts during additional times-of-day and seasons would be useful. As cities grow their bicycle and pedestrian count campaigns, additional data (collected on a yearly basis) would allow for developing time-averaged models that focus exclusively on the spatial dimension of bicycle and pedestrian traffic.

Despite these limitations, our work has implications for designing health-promoting cities. For example, our models suggest that investing in compact development with supporting infrastructure (e.g., bicycle facilities) would help to promote active travel, which in turn could potentially increase total physical activity, reduce air and noise pollution, and gain public space for uses other than motorized traffic ([Donaire-Gonzalez et al. 2015](#); [Frank et al. 2005](#); [Sallis et al. 2009](#)). To our knowledge, our work provides the first set of direct-demand models that offer the potential for predicting bicycle and pedestrian volumes at a national scale (i.e., using count data from 20 MSAs). Previous direct-demand models, which were developed using data from a single city, are not able to make reliable predictions in other jurisdictions ([Fagnan and Kockelman 2016](#); [Griswold et al. 2011](#); [Hankey and Lindsey 2016](#); [Miranda-Moreno and Fernandes 2011](#)). This limitation of previous models does not allow for practitioners or researchers outside of the study area to apply results of the single-city models. In our study, the large spatial

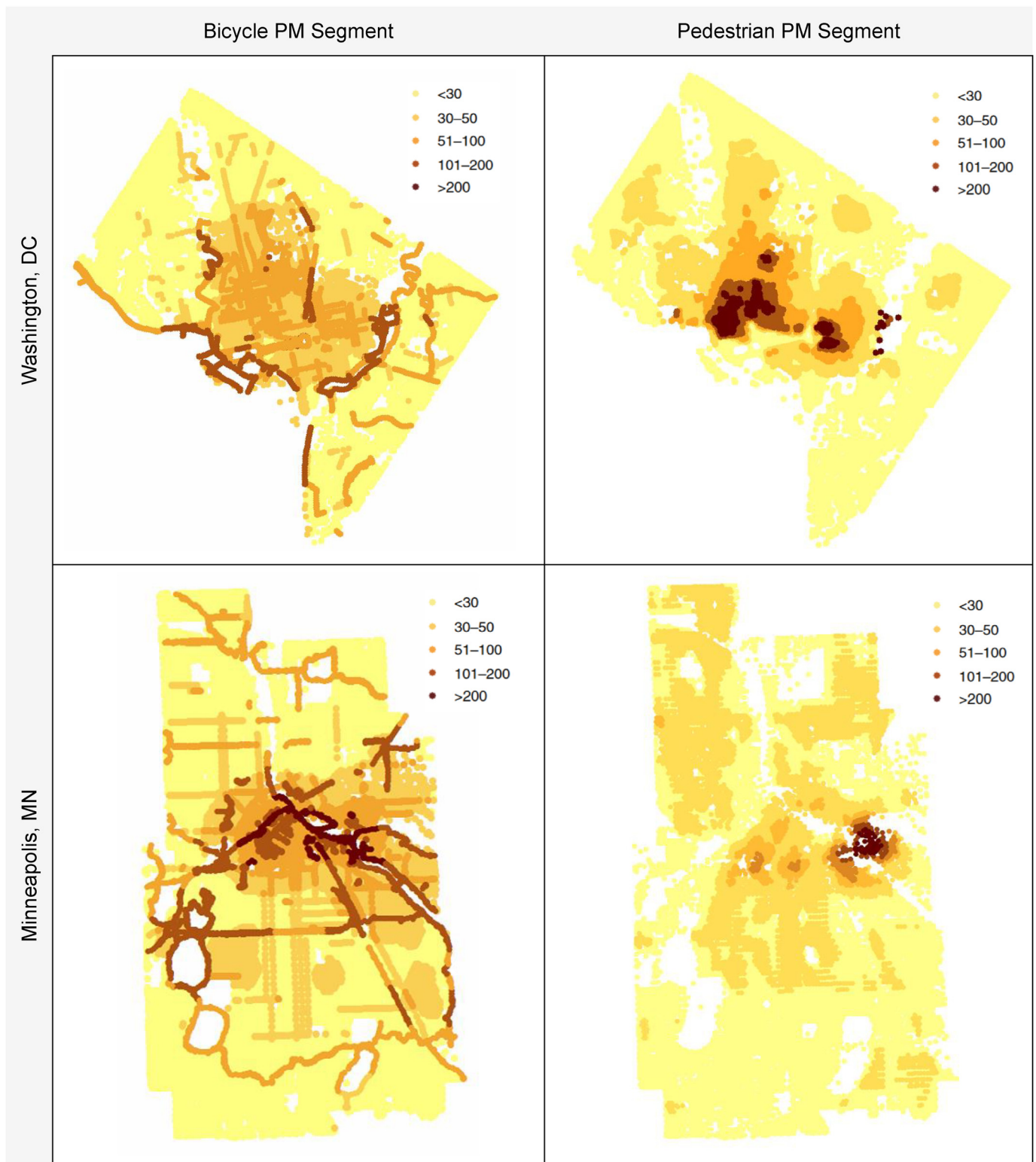


Figure 4. Spatial predictions of bicycle and pedestrian traffic volumes for all roads and off-street trails in Washington, DC, and Minneapolis, MN. Values represent total traffic volumes (i.e., number of bicyclists or pedestrians) during the 2-h afternoon peak period (PM segment).

and temporal coverage of the count data combined with the use of predictor variables that are available at a national scale allows for estimating active travel in areas where counts are inadequate or unavailable. Such models could aid the site selection process for bicycle and pedestrian facilities in order to increase active travel and physical activity levels.

We demonstrated how our models could be used to generate predictions of bicycle and pedestrian traffic volumes for MSAs outside of our sample more reliably than previous modeling efforts (i.e., spatial estimates of active travel in Washington, DC, and Minneapolis). By providing a model that can estimate bicycle and pedestrian traffic with high spatial resolution (i.e., street

segment) across many MSAs, future studies could explore how those spatial estimates could be used for exposure assessment to improve health. For example, single-city studies of the spatial patterns of cyclists' and pedestrians' exposure to air quality (Hankey et al. 2017a) could be expanded to a national scale; similarly, future work could explore patterns of crash rates using the same exposure surface across many cities or MSAs rather than relying on findings from a single city (Chen 2015; Toran Pour et al. 2017). In sum, creating generalizable guidance (via models that span many cities) may help planners and policy makers to assess strategies that simultaneously promote active travel (and thus physical activity) while reducing exposure to hazards (e.g., air pollution, crashes). Our work thus advances the development of models to estimate exposure to air pollution and crashes with larger spatial coverage (i.e., across MSAs) than previous studies.

Conclusions

We examined the relationship between the built environment and active travel in 20 U.S. MSAs (based on 4,593 bicycle and pedestrian count locations across all MSAs). Our models had reasonable goodness-of-fit for both bicycle traffic (adj- R^2 : 0.46–0.61) and pedestrian traffic (adj- R^2 : 0.42–0.72). Land-use and transportation-network variables are correlated with bicycle and pedestrian traffic; for example, water and green space, job density, proximity to university and college campuses, multimodal network and intersection density, as well as bicycle and walking commute mode shares are all consistently selected as predictor variables across models. Household density is also a strong predictor for pedestrian volume. One of the strongest predictors of bicycle volume is the presence of a bicycle facility (off-street: 13–31% increase; on-street: 10–15% increase compared with count locations without a facility) indicating the importance of providing supporting infrastructure to promote active travel.

Despite limitations regarding available input data, our models produced robust outcomes using a variety of cross-validation procedures; this finding is promising for efforts to develop more accurate estimates of bicycle and pedestrian traffic with high spatial precision across many cities and regions. We demonstrated an application of our models for out-of-sample prediction and for building city-wide spatial predictions of bicycle and pedestrian traffic. Outputs from our models could be used to assess exposure to air pollution, accidents with motor vehicles, or other environmental hazards. Furthermore, our models could be used to inform decisions on where to locate active travel facilities to promote bicycling and walking toward the goal of realizing health benefits from increased physical activity.

Acknowledgments

We thank K. Nordback, H. Hagedorn, M. Watkins, and B. Johnson and the staff from all participating metropolitan statistical areas (MSAs) for their support in preparing traffic count data. We thank the editor and reviewers for providing feedback that greatly improved the quality of our work. This study was funded by the Mid-Atlantic Transportation Sustainability University Transportation Center (MATS-UTC).

References

AASHTO (American Association of State Highway and Transportation Officials). 2012. *Guide for the Development of Bicycle Facilities*. 4th Edition. Washington, DC: AASHTO.

Akita Y. 2014. LUR Tools: ArcGIS Toolbox for Land Use Regression (LUR) Model. <http://www.unc.edu/~akita/lurtools/index.html> [accessed 10 July 2017].

Andersen ZJ, de Nazelle A, Mendez MA, Garcia-Aymerich J, Hertel O, Tjønneland A, et al. 2015. A study of the combined effects of physical activity and air pollution on mortality in elderly urban residents: the Danish Diet, Cancer, and Health

cohort. *Environ Health Perspect* 123(6):557–563, PMID: 25625237, <https://doi.org/10.1289/ehp.1408698>.

Buehler R, Dill J. 2016. Bikeway networks: a review of effects on cycling. *Transp Res* 36(1):9–27, <https://doi.org/10.1080/01441647.2015.1069908>.

Buehler R, Pucher J. 2012. Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation* 39(2):409–432, <https://doi.org/10.1007/s11116-011-9355-8>.

Buehler R, Pucher J. 2017. Trends in walking and cycling safety: recent evidence from high-income countries, with a focus on the United States and Germany. *Am J Public Health* 107(2):281–287, PMID: 27997241, <https://doi.org/10.2105/AJPH.2016.303546>.

Chen P. 2015. Built environment factors in explaining the automobile-involved bicycle crash frequencies: a spatial statistic approach. *Saf Sci* 79:336–343, <https://doi.org/10.1016/j.ssci.2015.06.016>.

Christian H, Knuiman M, Divitini M, Foster S, Hooper P, Boruff B, et al. 2017. A longitudinal analysis of the influence of the neighborhood environment on recreational walking within the neighborhood: results from RESIDE. *Environ Health Perspect* 125(7):077009, <https://doi.org/10.1289/EHP823>.

de Hartog JJ, Boogaard H, Nijland H, Hoek G. 2010. Do the health benefits of cycling outweigh the risks? *Environ Health Perspect* 118(8):1109–1116, PMID: 20587380, <https://doi.org/10.1289/ehp.0901747>.

de Nazelle A, Nieuwenhuijsen MJ, Antó JM, Brauer M, Briggs D, Braun-Fahrlander C, et al. 2011. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environ Int* 37(4):766–777, PMID: 21419493, <https://doi.org/10.1016/j.envint.2011.02.003>.

Donaire-Gonzalez D, de Nazelle A, Cole-Hunter T, Curto A, Rodriguez DA, Mendez MA, et al. 2015. The added benefit of bicycle commuting on the regular amount of physical activity performed. *Am J Prev Med* 49(6):842–849, PMID: 26228005, <https://doi.org/10.1016/j.amepre.2015.03.036>.

Fagnant DJ, Kockelman K. 2016. A direct-demand model for bicycle counts: the impacts of level of service and other factors. *Environ Plan B Urban City Sci* 43(1):93–107, <https://doi.org/10.1177/0265813515602568>.

Ferdinand AO, Sen B, Rahurkar S, Engler S, Menachemi N. 2012. The relationship between built environments and physical activity: a systematic review. *Am J Public Health* 102(10):e7–e13, PMID: 22897546, <https://doi.org/10.2105/AJPH.2012.300740>.

Frank L, Giles-Corti B, Ewing R. 2016. The influence of the built environment on transport and health. *J Transp Health* 3(4):423–425, <https://doi.org/10.1016/j.jth.2016.11.004>.

Frank LD, Saelens BE, Powell KE, Chapman JE. 2007. Stepping towards causation: do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Soc Sci Med* 65(9):1898–1914, PMID: 17644231, <https://doi.org/10.1016/j.socscimed.2007.05.053>.

Frank LD, Schmid TL, Sallis JF, Chapman J, Saelens BE. 2005. Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. *Am J Prev Med* 28(2 suppl 2):117–125, PMID: 15694519, <https://doi.org/10.1016/j.amepre.2004.11.001>.

Giles-Corti B, Vernez-Moudon A, Reis R, Turrell G, Dannenberg AL, Badland H, et al. 2016. City planning and population health: a global challenge. *Lancet* 388(10062):2912–2924, PMID: 27671668, [https://doi.org/10.1016/S0140-6736\(16\)30066-6](https://doi.org/10.1016/S0140-6736(16)30066-6).

Götschi T, Garrard J, Giles-Corti B. 2016. Cycling as a part of daily life: a review of health perspectives. *Transp Res* 36(1):45–71, <https://doi.org/10.1080/01441647.2015.1057877>.

Grabow ML, Spak SN, Holloway T, Stone B, Mednick AC, Patz JA. 2012. Air quality and exercise-related health benefits from reduced car travel in the midwestern United States. *Environ Health Perspect* 120(1):68–76, PMID: 22049372, <https://doi.org/10.1289/ehp.1103440>.

Griswold J, Medury A, Schneider R. 2011. Pilot models for estimating bicycle intersection volumes. *Transp Res Rec* 2247(1):1–7, <https://doi.org/10.3141/2247-01>.

Hankey S, Lindsey G. 2016. Facility-demand models of peak period pedestrian and bicycle traffic: comparison of fully specified and reduced-form models. *Transp Res Rec* 2586:48–58, <https://doi.org/10.3141/2586-06>.

Hankey S, Lindsey G, Marshall JD. 2017a. Population-level exposure to particulate air pollution during active travel: planning for low-exposure, health-promoting cities. *Environ Health Perspect* 125(4):527–534, PMID: 27713109, <https://doi.org/10.1289/EHP442>.

Hankey S, Lu T, Mondschein A, Buehler R. 2017b. Spatial models of active travel in small communities: merging the goals of traffic monitoring and direct-demand modeling. *J Transp Health* 7(pt B):149–159, <https://doi.org/10.1016/j.jth.2017.08.009>.

Hankey S, Marshall JD. 2017. Urban form, air pollution, and health. *Curr Environ Health Rep* 4(4):491–503, PMID: 29052114, <https://doi.org/10.1007/s40572-017-0167-7>.

Heesch KC, Giles-Corti B, Turrell G. 2015. Cycling for transport and recreation: associations with the socio-economic, natural and built environment. *Health Place* 36:152–161, PMID: 26598959, <https://doi.org/10.1016/j.healthplace.2015.10.004>.

- Hood J, Sall E, Charlton B. 2011. A GPS-based bicycle route choice model for San Francisco, California. *Transp Lett* 3(1):63–75, <https://doi.org/10.3328/TL.2011.03.01.63-75>.
- Iacono M, Krizek KJ, El-Geneidy A. 2010. Measuring non-motorized accessibility: issues, alternatives, and execution. *J Transp Geogr* 18(1):133–140, <https://doi.org/10.1016/j.jtrangeo.2009.02.002>.
- Jackson RJ, Dannenberg AL, Frumkin H. 2013. Health and the built environment: 10 years after. *Am J Public Health* 103(9):1542–1544, PMID: 23865699, <https://doi.org/10.2105/AJPH.2013.301482>.
- Karl TR, Koss WJ. 1984. *Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895–1983*. Historical Climatology Series 4-3. Asheville, NC:National Climatic Data Center.
- Kuzmyak JR, Walters J, Bradley M, Kockelman KM. 2014. “Estimating Bicycling and Walking for Planning and Project Development: A Guidebook.” NCHRP Rep. No. 770. Washington, DC:Transportation Research Board.
- Macmillan A, Connor J, Witten K, Kearns R, Rees D, Woodward A. 2014. The societal costs and benefits of commuter bicycling: simulating the effects of specific policies using system dynamics modeling. *Environ Health Perspect* 122(4):335–344, PMID: 24496244, <https://doi.org/10.1289/ehp.1307250>.
- Miranda-Moreno LF, Fernandes D. 2011. Modeling of pedestrian activity at signalized intersections: land use, urban form, weather, and spatiotemporal patterns. *Transp Res Rec* 2264(1):74–82, <https://doi.org/10.3141/2264-09>.
- Morris EA, Guerra E. 2015. Mood and mode: does how we travel affect how we feel? *Transportation* 42(1):25–43, <https://doi.org/10.1007/s11116-014-9521-x>.
- Mueller N, Rojas-Rueda D, Basagaña X, Cirach M, Cole-Hunter T, Davdand P, et al. 2017. Urban and transport planning related exposures and mortality: a health impact assessment for cities. *Environ Health Perspect* 125(1):89–96, PMID: 27346385, <https://doi.org/10.1289/EHP220>.
- Mueller N, Rojas-Rueda D, Cole-Hunter T, de Nazelle A, Dons E, Gerike R, et al. 2015. Health impact assessment of active transportation: a systematic review. *Prev Med* 76:103–114, PMID: 25900805, <https://doi.org/10.1016/j.ypmed.2015.04.010>.
- NACTO (National Association of City Transportation Officials). 2014. *Urban Bikeway Design Guide*, 2nd Edition. Washington, DC:Island Press.
- Oja P, Titz S, Bauman A, de Geus B, Krenn P, Reger-Nash B, et al. 2011. Health benefits of cycling: a systematic review. *Scand J Med Sci Sports* 21(4):496–509, PMID: 21496106, <https://doi.org/10.1111/j.1600-0838.2011.01299.x>.
- Pulgurtha SS, Repaka SR. 2008. Assessment of models to measure pedestrian activity at signalized intersections. *Transp Res Rec* 2073(1):39–48, <https://doi.org/10.3141/2073-05>.
- Ramsey K, Bell A. 2014. Smart Location Database. Version 2.0 User Guide. https://www.epa.gov/sites/production/files/2014-03/documents/sld_userguide.pdf [accessed 5 July 2018].
- Rojas-Rueda D, de Nazelle A, Teixidó O, Nieuwenhuijsen MJ. 2012. Replacing car trips by increasing bike and public transport in the greater Barcelona metropolitan area: a health impact assessment study. *Environ Int* 49:100–109, PMID: 23000780, <https://doi.org/10.1016/j.envint.2012.08.009>.
- Saelens BE, Handy SL. 2008. Built environment correlates of walking: a review. *Med Sci Sports Exerc* 40(7 suppl):S550–S566, PMID: 18562973, <https://doi.org/10.1249/MSS.0b013e31817c67a4>.
- Saelens BE, Sallis JF, Frank LD. 2003. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Ann Behav Med* 25(2):80–91, PMID: 12704009, https://doi.org/10.1207/S15324796ABM2502_03.
- Sallis JF, Certero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. 2006. An ecological approach to creating active living communities. *Annu Rev Public Health* 27:297–322, PMID: 16533119, <https://doi.org/10.1146/annurev.publhealth.27.021405.102100>.
- Sallis JF, Conway TL, Dillon LI, Frank LD, Adams MA, Cain KL, et al. 2013. Environmental and demographic correlates of bicycling. *Prev Med* 57(5):456–460, PMID: 23791865, <https://doi.org/10.1016/j.ypmed.2013.06.014>.
- Sallis JF, Saelens BE, Frank LD, Conway TL, Slymen DJ, Cain KL, et al. 2009. Neighborhood built environment and income: examining multiple health outcomes. *Soc Sci Med* 68(7):1285–1293, PMID: 19232809, <https://doi.org/10.1016/j.socscimed.2009.01.017>.
- Schneider RJ, Arnold LS, Ragland DR. 2009. Methodology for counting pedestrians at intersections: use of automated counters to extrapolate weekly volumes from short manual counts. *Transp Res Rec* 2140(1):1–12, <https://doi.org/10.3141/2140-01>.
- Smith O. 2017. Commute well-being differences by mode: evidence from Portland, Oregon, USA. *J Transp Health* 4:246–254, <https://doi.org/10.1016/j.jth.2016.08.005>.
- Smith M, Hosking J, Woodward A, Witten K, MacMillan A, Field A, et al. 2017. Systematic literature review of built environment effects on physical activity and active transport – an update and new findings on health equity. *Int J Behav Nutr Phys Act* 14(1):158, PMID: 29145884, <https://doi.org/10.1186/s12966-017-0613-9>.
- Strauss J, Miranda-Moreno LF. 2013. Spatial modeling of bicycle activity at signalized intersections. *J Transp Land Use* 6(2):47–58, <https://doi.org/10.5198/jtlu.v6i2.296>.
- Tabeshian M, Kattan L. 2014. Modeling nonmotorized travel demand at intersections in Calgary, Canada: use of traffic counts and Geographic Information System data. *Transp Res Rec* 2430(1):38–46, <https://doi.org/10.3141/2430-05>.
- Toran Pour A, Moridpour S, Tay R, Rajabifard A. 2017. Neighborhood influences on vehicle-pedestrian crash severity. *J Urban Health* 94(6):855–868, PMID: 28879440, <https://doi.org/10.1007/s11524-017-0200-z>.
- U.S. Census Bureau. 2014. “American Community Survey (ACS).” <https://www.census.gov/programs-surveys/acs> [accessed 2 February 2017].
- U.S. Census Bureau. 2017. “TIGER/Line® Shapefiles and TIGER/Line® Files.” <https://www.census.gov/geo/maps-data/data/tiger-line.html> [accessed 2 February 2017].
- U.S. EPA (U.S. Environmental Protection Agency). 2017. “Smart Location Mapping.” <https://www.epa.gov/smartgrowth/smart-location-mapping> [accessed 27 February 2017].
- Winters M, Buehler R, Götschi T. 2017. Policies to promote active travel: evidence from reviews of the literature. *Curr Environ Health Rep* 4(3):278–285, PMID: 28695486, <https://doi.org/10.1007/s40572-017-0148-x>.