U.S. Department of Transportation
National Highway Traffic Safety Administration

NHTSA

DOT HS 813 583 July 2024

A Framework for Assessing Pedestrian Exposure Using GPS and Accelerometer Walking Data

This page is intentionally left blank

DISCLAIMER

This publication is distributed by the U.S. Department of Transportation, National Highway Traffic Safety Administration, in the interest of information exchange. If trade or manufacturers' names or products are mentioned, it is because they are considered essential to the object of the publication and should not be construed as an endorsement. The United States Government does not endorse products or manufacturers.

Suggested APA Format Citation:

Douglas, G., Boyle, L. N., Saelens, B., Mooney, S., Moudon, A. V., Ebel, B. E., & Kim, H. (2024, July). *A framework for assessing pedestrian exposure using GPS and accelerometer data* (Report No. DOT HS 813 583). National Highway Traffic Safety Administration.

This page is intentionally left blank

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 813 583		

4. Title and Subtitle A Framework for Assessing Pedestrian Exposure Using GPS and Accelerometer Walking Data	**5. Report Date** July 2024 **6. Performing Organization Code**
7. Author(s) Grace Douglas, Linda Ng Boyle, Brian Saelens, Stephen Mooney, Anne Vernez Moudon, Beth E. Ebel, & Haena Kim	**8. Performing Organization Report No.**
9. Performing Organization Name and Address University of Washington PO Box 352650 Seattle, WA 98195	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.**
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration Office of Behavioral Safety Research 1200 New Jersey Avenue SE Washington, DC 20590	**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code**

15. Supplementary Notes
Kristie Johnson, Ph.D., was the contracting officer's representative on the project.

16. Abstract

Crashes involving pedestrians represent 19% of all motor-vehicle-related fatalities in the United States. Quantifying safe pedestrian travel requires a better understanding of the likelihood that a pedestrian will be involved in a potentially harmful pedestrian-vehicle interaction. This is achieved by obtaining a good pedestrian exposure metric, which will let traffic safety experts and policymakers differentiate between emerging risks and changing patterns of exposure. From this exposure metric, countermeasures can be tailored for different pedestrian walk patterns. The study team first examined pedestrian exposure metrics used in past studies. The team then used data collected from electronic devices from two previous longitudinal studies on walking in the Seattle, Washington, area. Estimates of pedestrian exposure were built on "walking bouts," which are time segments of physical activity during which walking was measured to have occurred. The unit of analysis was at the intersection level. Two models were developed: a zero inflated negative binomial model that examined the likelihood of walking at any intersection in Seattle, and a negative binomial model to examine the frequency of walking at intersections that included at least 10 walking bouts. Explanatory variables included geospatial locations, micro-environmental factors, and macro-environmental factors. This study provides a framework quantifying pedestrian exposure that can be used by other transportation municipalities to capture pedestrian exposure.

17. Key Words	18. Distribution Statement
Pedestrian safety, pedestrian exposure, GIS-linked data, built environment, walking bouts, ZINB, walking bouts	This document is available to the public from the DOT, BTS, National Transportation Library, Repository & Open Science Access Portal, rosap.ntl.bts.gov

19 Security Classif. (of this report)	20. Security Classif. (of this page)	21 No. of Pages	22. Price
Unclassified	Unclassified	133	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

This page is intentionally left blank

Table of Contents

This page is intentionally left blank

List of Figures

List of Tables

This page is intentionally left blank

List of Acronyms

ACS	American Community Survey
ACTION	Assessing Choices in Transportation in our Neighborhood
GBM	gradient boosted model
IRR	incidence risk ratio
LASSO	Least Absolute Shrinkage and Selection Operator
MAE	mean absolute error
ML	machine learning
MSE	mean squared error
NB	negative binomial
NBPDP	National Bicycle and Pedestrian Documentation Project
NHTS	National Household Travel Survey
OR	odds ratio
PUMS	Public Use Microdata Sample
RMSE	root mean squared error
SCRI	Seattle Children's Research Institute
TAZ	Transportation Analysis Zone
TRAC	Travel Assessment and Community
UFL	Urban Form Lab
UW	University of Washington
WB	walking bout
WSDOT	Washington State Department of Transportation
ZINB	zero-inflated negative binomial

This page is intentionally left blank

Executive Summary

Overview

Pedestrian exposure has been defined as the number of potential opportunities for a pedestrian to be involved in a crash or other harmful situation with a moving vehicle on or near the roadway environment (Greene-Roesel et al., 2007; Ryus et al., 2014). Pedestrian exposure is an important metric to transportation officials and city planners to ensure that policies and road improvements designed to enhance road safety accurately reflect the volume of pedestrians. However, pedestrian exposure is challenging to measure accurately and efficiently given obstacles such as high costs, limited sampling locations, individual variability, and limited generalizability. This project builds an estimate of pedestrian exposure based on individual walking activity captured using accelerometers and GPS devices.

Past research has shown that pedestrian exposure can be estimated at various scales: (1) area-based, (2) point/segment-based, and (3) trip-based. Based on an extensive program scan, the research team selected a point-based scale, with the point being the intersection level. *Walking bouts*, derived from the integration of accelerometry and GPS, were computed. Walking bouts are periods of time where GPS and accelerometer traces have identified segments of physical activity during which walking is estimated to have occurred. The total walking bouts within an intersection was computed within a 50 m buffer.

Zero-inflated negative binomial and the negative binomial models were created to examine environmental correlations between the number of walking bouts near or through intersections. The models examined the impact of roadway slopes as well as the proximal micro- and macro-environmental features. The ZINB predicts the likelihood of observing walking bouts at an intersection using a dataset of all intersections in Seattle, Washington. The NB model predicts the likelihood of higher frequency of walking bouts using a subset of intersections with at least 10 walking bouts.

Data

The data used in this project included data from de-identified longitudinal GPS and accelerometer data that were collected from hundreds of adults as part of two research projects: the Travel Assessment and Community project and the Assessing Choices in Transportation in our Neighborhood project. These projects were funded by the National Institutes of Health (NIH) for a collaboration between the Seattle Children's Research Institute and the University of Washington. Data collection was conducted from 2008 to 2018 and occurred in three waves for each project.

The ZINB and NB models used data from TRAC Waves 1 and 2 and ACTION Wave 1. The ZINB model used a dataset that included all intersections in Seattle (n = 14,073 intersections). The NB model used a subset of data that included intersections with at least 10 walking bouts in Seattle (n = 3,047 intersections). Because the data used for the NB model included only intersections with walking bouts, each intersection could be linked to people walking there. Accordingly, pedestrian characteristics were also considered in the NB model. ACTION Waves 2 and 3 were used to test and validate the models.

Statistical Models

Least Absolute Shrinkage and Selection Operator regression was used to select the most relevant environmental predictors and minimize multicollinearity. This process identified 26 predictor variables for consideration in the ZINB and NB model. These variables are discussed in detail in the main report.

The ZINB model (all intersections) revealed that 18 of the 26 predictor variables were significantly associated with the likelihood of observing walking bouts at an intersection. For example, the following nine variables were associated with a higher likelihood of presence and frequency of observed walking:

- longer bike lane lengths,
- presence of crosswalk warning sign,
- presence of bike and pedestrian signs,
- presence of traffic signal,
- higher bus ridership density,
- higher population density,
- park presence,
- longer trail length, and
- higher job density.

Additionally, these three variables were significantly associated with a decreased likelihood of walking bouts:

- higher maximum slope,
- presence of one-way signs, and
- presence of a park-and-ride facility.

The NB model (10+ walking bouts) showed 22 out of 26 predictors significantly affected the likelihood of higher number of walking bouts at intersections where walking activity was observed. Demographic predictors considered such as pedestrian age and income were positively associated with increased walking bout frequency with older pedestrians and pedestrians with incomes in the $40,000 to $69,000 range displaying more walking. In general, most conclusions confirmed the ZINB model, however the NB model showed the presence of stop signs and longer trail lengths were negatively associated with the number of walking bouts at intersections. Model validation was also conducted in terms of stability and accuracy.

Summary

This study provides a framework for using environmental predictors to estimate pedestrian exposure that can be used by other municipalities. The use of electronic device data to measure pedestrian exposure was shown to be useful, providing information on relative frequency of pedestrian activity at a highly disaggregated level. Walking bouts can also be aggregated up to the person-, trip-, and intersection-level to estimate exposure patterns and used with crash data to estimate risk of pedestrian-vehicle crashes or severe injuries accounting for variation in pedestrian exposure.

In terms of modeling, this study used ZINB and NB to identify factors associated with walking activities at the intersection level. These models were used to examine walking bouts as a binary outcome (walking: yes, no) and as a count. However, several other analytical methods can be considered given the research questions of interest. More specifically, walking bouts can also be used on a more continuous timeline to address research questions related to walk duration, travel time, and even impact of seasonal changes.

The ACTION and TRAC data contain information on pedestrian walk behavior across King County, Washington. Our model centered on Seattle, which is highly urbanized. However, this dataset can also provide insights on pedestrians from less urbanized areas outside Seattle but in King County.[1] This may be of future interest for understanding the impact of equity-related factors (e.g., accessibility to transit, availability of sidewalks, and median household income) on pedestrian safety.

Our results also identified environmental predictors that may be useful for pedestrian exposure analyses. These variables were identified as significant either during variable selection or modeling. Significant micro-environment (location-based) variables of interest include average roadway width, maximum speed limit, total sidewalk length (ft), and the presence of various pedestrian- and vehicle-related traffic control signs and devices (i.e., one-way signs, stop signs, crosswalk signs, and traffic signals). Significant macro-environment (population-based) variables of interest include total trail length (ft), public school enrollment count, bus ridership density, job density, and various land uses (manufacturing, service, transportation). Many of these variables are available at the local and state level. For example, one of the most significant variables from our study was maximum slope percentage, a variable that should be available in most transportation municipalities.

There were some limitations to this study. Notably, our analyses were conducted at the intersection-level only and may not include pedestrian walking activities on or near midblock areas where pedestrian-vehicle interactions can occur. This limitation was partially addressed by attributing any walking bout that happened within 50 m of an intersection to that intersection. This study was a first step in using walking bouts that combined accelerometry and GPS as a measure of pedestrian exposure. A major contribution of this project is the framework that has been developed to collect, process, and analyze such data.

In summary, the presented framework can be used by transportation agencies interested in quantifying pedestrian exposure. The framework can be adapted to different levels of analysis. While not fully exhaustive, the list of predictor variables provided in this report serve as a starting point for other localities to examine pedestrian behavior. Additional environmental factors that exist within a specific locality can easily be incorporated.

[1] Editor's note: In the 2020 U.S. Census King County's population was 2,269,675, and Seattle's was 738,172, so only about a third of King County's population lies within the city.

This page is intentionally left blank

Introduction

There was a slight increase (<1%) in pedestrian fatalities along with a 12 percent increase in pedestrian injuries in the United States from 2021 to 2022 (NCSA, 2024). In 2022 the number of pedestrian fatalities increased by 52 from 2021, with an estimated 7,522 pedestrians killed, and an estimated 67,336 pedestrians injured in motor-vehicle-related crashes. Pedestrian exposure is defined as the number of potential opportunities for a pedestrian to be in contact with a harmful situation and may be a contributing factor to this increase (Greene-Roese et al., 2007; Ryus et al., 2014; Qu et al., 2022).

Efforts such as the National Roadway Safety Strategy (U.S. DOT, 2022) and programs such as Vision Zero (City of Seattle, 2017) are designed to make travel safer for all road users. Many cities have adopted innovative policies to reduce vehicle congestion and build infrastructure that supports active transportation modes, such as walking and biking. These policies aim to improve the health, safety, and livability of residents and visitors. However, creation of safer travel requires a better understanding of both exposure data (e.g., an increase in pedestrian traffic, unsafe driving speeds, pedestrian path behavior) and pedestrian injury counts. This data enables traffic safety experts and policymakers to differentiate between emerging risks (e.g., an increase in the rates of pedestrian/vehicle conflicts) and changing patterns of exposure (e.g., growth in pedestrian traffic). This, in turn, can help to better tailor countermeasure approaches.

Program Scan

A program scan was conducted to (1) identify variables that could be considered for pedestrian exposure as measured and used across various localities, (2) assess modeling approaches, and (3) identify datasets that contain relevant predictor variables.

The program scan consisted of programs and data sources from research, government, and professional associations. Peer-reviewed journal articles, publicly available software, smartphone and tablet applications, and devices were included in the scan. Existing data collection programs, apps, or devices that included relevant pedestrian location and exposure information were identified. Documents and projects that provided data sources and methods to measure pedestrian exposure were extensively reviewed for validity and availability.

As part of their Vision Zero plans (Vision Zero Network, 2021), many U.S. cities have implemented approaches to measure pedestrian exposure. The project team leveraged its working relationships with transportation experts from several Vision Zero cities to gather information on methods used to capture pedestrian exposure. Many phone interviews were conducted with city planners and researchers involved in pedestrian studies across the United States.

The team also reviewed several datasets available to the project team in Washington State. In addition to WSDOT crash data and Puget Sound geocoded roadway data, the team reviewed pedestrian data collected as part of two previous NIH-funded projects, TRAC and ACTION. The program scan revealed several measures used to capture pedestrian exposure described in the next section. As part of the review, the team also identified the advantages and disadvantages of each pedestrian exposure measure (see Appendix A for details).

Pedestrian exposure measures used in past studies

Pedestrian exposure to potential pedestrian-vehicle conflicts can be examined at various scales and units. They primarily fall into three categories: area-based (Table A 1), point/segment-based (Table A 4), and trip-based (Table A 6). A synthesis of past studies in each of these categories is included in Appendix A, which provides the following information for each study.

- Exposure type: The type and unit of measure used for exposure
- Reference: The paper that used the noted exposure type
- Outcome of interest: the pedestrian exposure measure used in this study as a dependent variable (e.g., pedestrian crash frequency and injury severity)
- Study location: Location of pedestrian exposure measurement
- Spatial unit: Spatial unit of analysis used in the study (e.g., transportation analysis zone , census tract, individual trips, etc.)
- Area extent: Area of pedestrian exposure (e.g., neighborhood, an entire city, county, region, State, etc.)
- Data sources: The data used for exposure analysis
- Pedestrian exposure model type: Type of model used for exposure estimation for studies where observed data were used to estimate exposure in unobserved locations. (This does not include any other model types with a dependent variable being pedestrian crash frequency or injury severity.)
- Significant variables: Indirect variables that are statistically significant (a = .5) with pedestrian exposure estimation

The most useful metric of pedestrian exposure in transportation safety is pedestrian volume counts collected at point locations. However, given costs and sampling issues, proxy variables are often used.

Area-based pedestrian exposure

Area-based exposure measures are proxies for pedestrian volumes because they capture the densities of people and activities. Most area-based measures can be obtained from secondary data (e.g., U.S. Census Bureau) and are often available over time. Area-based pedestrian exposure measures can be categorized as:

- Area density (e.g., population, employment, residents, etc.), or
- Self-reported walking activity, including:
    - Walking distance,
    - Walking duration, and/or
    - Number of individual trips.

Area density

Area density such as population, employment, and resident densities is often used to approximate area-based pedestrian exposure, especially in pedestrian and bicyclist safety analysis (estimating crash or injury severities). Densities can be estimated within certain buffer sizes (e.g., distances from crash locations) or in predefined spatial units such as a TAZ or census tract. The main density data (population, residential, school) came from publicly available government websites, such as the U.S. Census Bureau, regional or local departments of transportation, and the National Center for Education Statistics.

Past aggregate demand models have found that population density is positively correlated with pedestrian volumes (Hankey & Lindsey, 2016; Jamali & Wang, 2017; Wang & Kockelman, 2013). This means that as population density increases, pedestrian volume increases as well. This also means that as population density decreases, pedestrian volume decreases as well. However, density data does not account for the variability of individual pedestrian activities such as pedestrians' walking distance and time (Mooney et al., 2016).

A summary of the advantages and disadvantages of density data for pedestrian exposure is provided in Table A 1 and Table A 2.

Self-reported walking activity

Individual walking activities can be used as measures of pedestrian exposure in terms of walking distance and duration, and number of individual trips. The main data source for area-based individual information is survey data. As self-reporting survey data obtains detailed information on individual characteristics such as age and gender, exposure measures can also be studied by different demographics.

In the United States, two national-level surveys that include pedestrian information are the American Community Survey and the National Household Travel Survey. Local and regional government agencies also conduct surveys either regularly or based on the needs of transportation-related projects. A summary of the advantages and disadvantages of self-reported walking activity data as a pedestrian exposure measure is provided in Table A 3.

Point/Segment-based pedestrian exposure

Government agencies and public organizations regularly collect pedestrian counts at points (e.g., intersection or midblock) or segments (e.g., a single face of a city block) using an agency-approved, standardized protocol. Direct measures of pedestrian volume include counts of individual pedestrians passing through specific points or segments. These collected count data can be used to monitor changes in pedestrian behavior and volumes over time. They are often used to assess whether traffic and road improvements are needed to enhance pedestrian safety. A summary of past studies that use point/segment-based pedestrian exposure is provided in Table A 4.

Typically, pedestrian and bicycle counts are directly measured by data collectors or automatic counting devices at studied locations. Because the counts are collected at a limited number of locations, they provide only a snapshot of pedestrian walking and is limited for use in area-wide pedestrian exposure estimations (e.g., network, entire cities). They are often highly accurate for a limited spatial scale. The count data in regression models are estimates of pedestrian volumes at points within areas, which are then aggregated to create area-based measures. These pedestrian counts can be used as

- a direct measure of exposure at a specific location, and
- an indirect measure of exposure at the areawide level by estimation and modeling.

Transportation-facility-specific exposure measures are most often used in studies that focus on high-crash locations or specific study locations for pedestrian safety projects. Indirect measures use point-based data to estimate pedestrian exposure in broader areas.

Turner et al. (2017) summarized different model types and corresponding pedestrian exposure estimates into four methodologies: direct demand, regional travel demand, simulation-based traffic, and special-focused.

Direct demand models have been used widely in many pedestrian safety studies. Schneider et al. (2012) summarized the studies conducted on pedestrian intersection volume models (direct-demand models) and the methods and statistically significant predictive variables from each model. The most common predictive variables in the intersection-based pedestrian models were population density, employment density, and transit accessibility. As expected, the coefficients for each variable were different and greatly depended on the community that the model was targeted. As noted by Schneider et al., the characteristics in individual communities greatly influence the estimation of pedestrian volumes.

Regional travel demand models are historically most popular for estimating non-motorized travel, which are based on traditional trip-based forecasting models. This modeling is done in four steps: trip generation, trip distribution, mode share, and traffic assignment.

Simulation-based traffic models rely on higher computational power and are often used for large-scale pedestrian networks. Hong et al. (2016) used data composed of pedestrian dynamics, area dynamics, and network topology measures.

Table A 4 also highlights special focused models used to address issues related to specific corridors and for sub-area planning. Regional models within this domain include trip generation and flow models, while a more specific corridor focused model (discrete choice) has been proposed to analyze crossing behavior. Additionally, GIS-based models are used in this context to model alternative land use or transportation investment strategies.

Trip-based pedestrian exposure

Trip-based pedestrian exposure measures can be categorized as

- space-time walking path estimation,
- crossing behavior, or
- physical activity/walking bouts.

As shown in Table A 6, studies showed various combinations in terms of their units of exposure, spatial resolution, measurement type, area extent, and model types.

Space-time walking path estimation

The space-time approach was developed by Lam et al. (2014). This approach overlays pedestrian activities with crash frequencies to estimate pedestrian exposure levels at the time of a crash. The pedestrian path is estimated based on the shortest path algorithm using origin and destination data. Pedestrian exposure is defined for each space-time walking path as the product of the walked distance and crash frequency. This approach is relatively new, and one major downside is that pedestrian exposure (as an outcome of the model) can potentially be set to zero for segments with no reported crashes (Jamali & Wang, 2017).

Crossing behaviors

Pedestrian exposure at crossing locations can be estimated with crossing behaviors. Crossing constitutes a different exposure weight than, for example, just walking on the sidewalk. Crossing involves intentionally walking in front of motorized vehicles whose engines are activated.

Researchers have studied this specific condition of pedestrian exposure using discrete choice models (e.g., sequential logit, nested logit model) to estimate the probability of crossing being chosen along a pedestrian trip. Papadimitriou et al. (2012) and Lassarre et al. (2007) controlled, trip-related variables (e.g., walk trip distance, crossing distance) in their choice models.

Physical activity/Walking bouts

Walking bouts were introduced in Kang et al. (2013). They represent pedestrian walking activity with high levels of detail. The data is recorded for each trip, but information can be examined at the individual level. Trip data is typically measured using electronic devices such as accelerometers and GPS data loggers. Periods in which accelerometer and GPS readings are consistent with walking can be classified as "walking bouts," such that the locations and times where walking bouts occur represent a highly spatiotemporally accurate recording of one pedestrian's exposure. These data can be used in concert with individual characteristics and sampling assumptions to estimate exposures for other pedestrians.

The challenges that exist with this data include the loss of GPS signals in urban areas, inside buildings, under high- or low-temperature conditions, and when device batteries are low. Because missing data in accelerometer and GPS data collections may not be randomly distributed in time and space, additional self-reporting travel diaries can be collected to identify possible walking behaviors when GPS data is missing.

Pedestrian exposure data sources

There are various data sources for pedestrian exposure; this includes counts of pedestrian volumes, surveys and travel diaries, and continuous data from GPS and accelerometers. This section reports on existing data sources that have been used with pedestrian volumes, surveys, and travel diaries to examine and model pedestrian location and exposure. Continuous data from GPS and accelerometers was used for this study and is discussed in the next section.

Many urban cities in the United States have implemented pedestrian count programs and the data from these programs can be useful for models that account for pedestrian exposure. However, for counts to be used as reliable exposure measures, the sampling of count locations needs to be carefully determined and defined. Without a sampling scheme, counts will only be useful for or applicable to the location where they are carried out, but they would not be as useful for corridor-or jurisdiction-level studies. For pedestrian counts to be generalizable, count data needs to be collected at various times (time of day, day of week, monthly, yearly, etc.) and locations, and for various environmental changes.

Sources of pedestrian volumes

Pedestrian counts directly measure non-motorized pedestrian travel exposure. There are national reports that have summarized the data collection methods used for counting pedestrians and bicyclists. These reports include

- Alta/ITE’s *National Bicycle and Pedestrian Documentation Project* (Jones et al., 2009),
- FHWA's *Traffic Monitoring Guide* (FHWA, 2016),
- National Cooperative Highway Research Program 797 report, *Guidebook on Pedestrian and Bicycle Volume Data Collection* (Ryus et al., 2014), and
- FHWA’s *Exploring Pedestrian Counting Procedures* report (Norback et al., 2016).

Appendix G summarizes the pedestrian and cyclist counting procedures from these reports, highlighting the strengths and weaknesses of each method. They report that budget constraints and commercially available technology affect the quality and quantity of data collection. (Tool availability continues to grow and those presented here are not exhaustive as newer tools have been released since this review was conducted.)

Some widely used data collection tools include automated count technologies and location-based counting methods. Automated count technologies are used for longer-term counts and require fewer person-hours. However, the accuracy of the collected data is dependent on the counter configuration, installation, calibration, reliability, and level of use. Some agencies have developed a standard error correction method and an equipment adjustment factor to account for such errors (FHWA, 2016). Newer technologies include cellphone-based app counters and route trackers (Louch et al., 2016).

Location-based counting methods include screen line counts or intersection crossing counts. A screen line count is a method that counts the number of pedestrians each time they pass a specific point. For high-crash locations, an intersection crossing counting method is recommended to record information on cross streets and pedestrian/bicycle/car turning movements (Norback et al., 2016).

The National Bicycle and Pedestrian Documentation Project is an effort to provide a national database of annual bicycle and pedestrian count information (Jones et al., 2009). The data can be used to examine associations between various factors and bicycle and pedestrian activity. The national count date for this project is mid-September to represent a peak period for walking and bicycling. It includes at least one weekday and one weekend day. Additional surveys and counts can be collected in January (winter), May (spring), and July (summer) to account for seasonal data (Vision Zero Miami, 2018).

Although a minimum number of count locations is not determined for the NBPDP, counting at more than one location is recommended. Criteria for selecting count locations include urban, suburban, and rural locations where the volumes of bicyclists and pedestrians are expected to be high based on historical data (NBPDP, 2010). NBPDP provides an option to automatically count volumes using EcoCounter[2] technology. Participating agencies and organizations provide data to the NBPDP national database. For public viewing, an annual summary report of trends is then published with information regarding volumes by user group, volume comparison to different locations and trip purposes and trip origin and destination.

Vision Zero measurement efforts

As part of the Vision Zero initiative, mayors of U.S. cities committed to eliminating traffic fatalities and severe injuries. Vision Zero cities work in partnership with city departments that include law enforcement, transportation, and public health sectors (Vision Zero Network, 2021). The cities work collaboratively to develop efficient approaches to measure pedestrian exposure.

The team gathered information from research, phone interviews with state and local government officials, professional associations, and used publicly available documentation on how to collect pedestrian exposure measures. Vision Zero cities were also identified as part of the program scan. In addition to Seattle, there were nine other Vision Zero Focus cities examined in more depth.

1. Los Angeles, California
2. San Francisco, California
3. Washington, DC
4. Chicago, Illinois
5. Fort Lauderdale, Florida
6. Boston, Massachusetts
7. New York, New York
8. Portland, Oregon
9. Austin, Texas

Table A 8 in the Appendix provides details on the 23 Vision Zero cities examined and their corresponding measures for pedestrian exposure. For example, in Seattle manual pedestrian and cyclist counts, in addition to automatic counters and GPS and accelerometer data, have been used as measures for pedestrian exposure. However, a pedestrian exposure model has not yet been developed. The model developed in this project in turn serves to address this need.

Summary of literature review

The program scan showed that there are several units of measurement for pedestrian exposure: (1) area, (2) point/segment, and (3) trip. There were several limitations identified with existing pedestrian data and models (see Appendix A). Area-based metrics (e.g., population density, residential density, job density) are widely used because they are often the most readily available. They have been used as proxies for pedestrian volumes because they capture the densities (or concentration) of people and activities. However, for these metrics to be useful, they should also be collected and examined over many time periods (e.g., hours, days, weekly,

---

[2] Eco-Counter Canada/USA, Montreal, Canada

monthly). Another limitation of these high-level estimates is that they do not provide the granularity that is needed for inter-urban-specific safety solutions.

Self-reported walking activity is also widely used but highly subjective and limited to the participant's ability to provide accurate and reliable information. The level of participation for self-reported information can also vary greatly, making it difficult to develop a robust model.

Based on the program scan, the most consistent and reliable exposure measure appears to be pedestrian volumes at point locations. The major challenges with this type of exposure data are cost and limited sampling locations. The high costs of collecting pedestrian volume data have led researchers to use proxy variables that are associated with pedestrian volumes such as population density and traffic volumes. Limited sampling locations may be mitigated with newer cellphone-based apps for counts and route trackers (Louch et al., 2016). App functionality evolves as technology improves.

Based on the program scan, there are three key indicators of a good metric and model of pedestrian exposure:

- accurately reflects the pedestrian density for the study location,
- scales to different spatial resolutions (e.g., intersections, transportation analysis zones), and
- generalizable to different spatial contexts (e.g., corridors, cities, counties).

In practice, there are tradeoffs between accuracy, scalability, and generalizability. The more flexible a model, the more accurate it is likely to be in the spatial context in which it is generated but the less likely it is to generalize to all pedestrian situations.

Given these key indicators, the research team used walking activities from two previous projects (TRAC and ACTION) to capture pedestrian exposure in Seattle. The walking activities are defined in terms of walking bouts, or physical activities recorded from personal accelerometers and include data on geospatial locations from personal GPS devices and self-reported travel diaries that provide supplemental information on trip date, duration, and purpose.

Walking bouts provide the foundation for measuring pedestrian exposure and can be aggregated to the individual-, trip-, and intersection-level. Walking bouts are naturalistic observations of pedestrians' walk patterns over various locations and times. For that reason, they can also be examined across different time scales: hourly, daily, weekly, and monthly. As described by Schneider et al. (2012), a GIS-based model requires land use measures such as population density, nearby employment, nearby commercial space, transportation system characteristics (e.g., bus ridership) and other social measures (e.g., number of car-free households nearby). Several micro- and macro-environmental factors are considered in the modeling approach. The details of the data processing approach and merging of databases for modeling is described in the next section.

Method

The primary objectives of this project were to develop an operational definition of pedestrian exposure and create a representative pedestrian exposure measure that can be modeled analytically. For this study, walking bouts (described below) were used to capture pedestrian volumes at intersections (the point location) in Seattle.

The first step in developing the pedestrian exposure model was to combine all data into a format that is defined at the intersection level. The project team used TIGER/Line spatial databases (data.census.gov) to identify over 14,000 intersections in King County.

A framework was developed to combine all environmental variables and walking bouts at the intersection level (see Figure 1). As part of the framework, walking behavior was first separated from non-walking behavior (Kang et al., 2013). The walking activities were then mapped to intersections of interest. Any GPS errors were reviewed and cleaned or removed. GPS error cleaning was the highest source of data elimination (15.4%). From there, buffers were placed around each intersection with walking bout counts and micro-environmental factors identified within 50 m buffers and macro-environmental variables identified within 400 m buffers. There was one spatial variable, slope, and it was recorded at the 400 m buffer level. The final dataset included each intersection as a row of observation and each predictor variable (micro, macro, spatial) and walking bout count as a column.

The data sources used for this study can be described in four categories.

- Pedestrian data: Project TRAC and Project ACTION were used to operationalize pedestrian exposure through physical activity. This data was collected in previously funded NIH studies.
- Intersection: TIGER/Line spatial databases are collected by the Census Bureau and provide shapefiles and geodatabases detailing topological information for spatial computations.
- Micro-environment: Seattle Department of Transportation collects city-specific data about the road environment, for example speed limits and pedestrian facilities.
- Macro-environment: Census block level data from the 5-year American Community Survey and the Public Use Microdata Sample were collected for density information (population, residential, land use, etc.).

The remainder of the methods section describes the datasets, data processing, descriptive statistics, correlation analysis, and variable selection using LASSO.

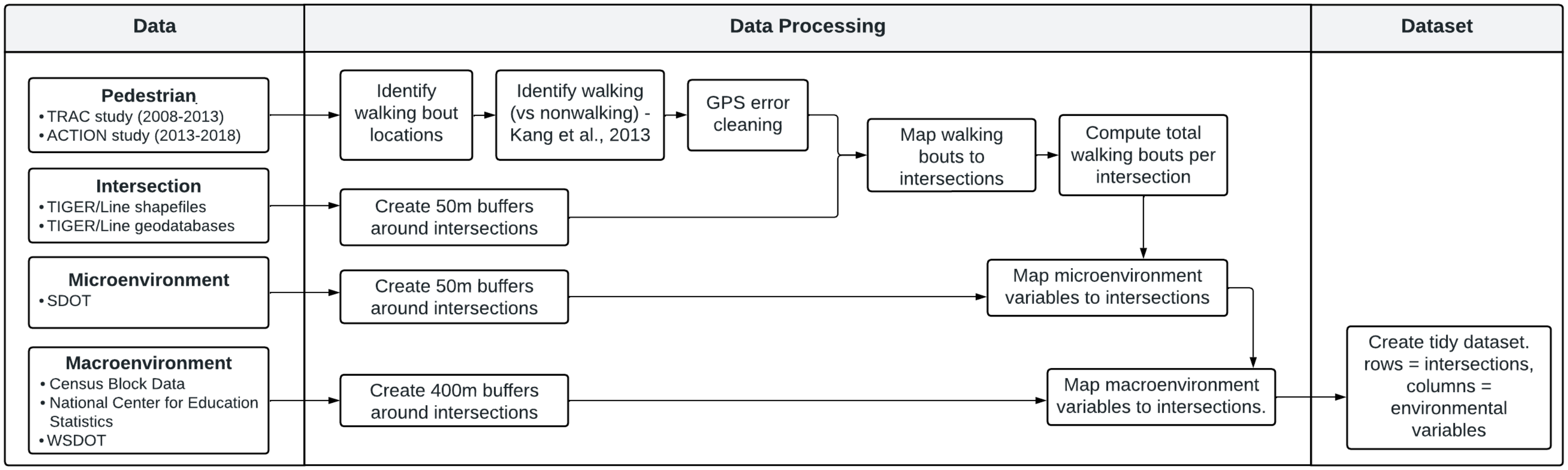


Figure 1. Framework to process pedestrian, intersection, micro-environment, and macro-environment data sources

Walking bouts

For this study, the basic building block on which estimates of pedestrian exposure were built was the "walking bout," based on a physical activity during which walking is estimated to have occurred. Physical activity data were collected from study participants who wore electronic accelerometer and GPS devices. Recorded physical activity was supplemented with information from participant travel diaries. We developed an algorithm to estimate walking bouts based on these inputs. This algorithm is described below and in greater detail in Kang et al. (2013).

The data used in this project includes de-identified longitudinal GPS and accelerometer data from the TRAC and ACTION projects. Both were funded by the NIH in collaboration between the Seattle Children's Research Institute and the University of Washington.

The TRAC data included three waves of data collected to coincide with the opening of light rail in King County (before light rail, 1 to 2 years after light rail opened, and then 3 to 4 years after light rail opened). The ACTION data also included three waves to coincide with opening of two new bus rapid transit lines in King County (before bus rapid transit, 1 to 2 years after bus rapid transit lines opened, and 3 to 4 years after bus rapid transit lines opened). Both sets of data include GIS data layers (roadway borders, shoulders, centerlines, sidewalks, etc.) and individual-level GPS points, accelerometer data points (sampled every 30 seconds), and corroborating travel diaries about the trips (travel times, mode, and activity purpose). Walking data was extracted from the integration of personal monitoring devices (GPS and accelerometers) worn by study participants. To summarize, the total participants per wave, per study for which this data was collected is shown in Table 1.

Table 1. Number of participants in TRAC and ACTION study

Study	Waves	Duration	Total participants
TRAC	1: Baseline	2008 and 2009	707
	2: 1- to 2 years after rail	2010 and 2011	581
	3: 3- to 4 years after rail	2012 and 2013	525
ACTION	1: Baseline	2013 and 2014	590
	2: 1- to 2 years after BRT*	2015 and 2016	398
	3: 3- to 4 years after BRT*	2017 and 2018	382

**BRT = Bus Rapid Transit*

Sampling biases

There were two major sources of biases. First, study participant behavior most likely did not represent walking behavior among all adults in the region. This was partially addressed by weighting the subjects demographically, to be more representative of the population in the area. While it was not possible to capture the entire population (e.g., some county residents are children and the participants were all adults), it does minimize some sampling biases.

There is also a bias given the spatial selection in the study – participants were invited based on their residential proximity and access to public transit services. The team attempted to minimize this bias using more advanced modeling strategies as well as considered other distance measures as a predictor in the exposure model (i.e., minimum, mean distance from home address).

Procedure for collecting walking data using electronic devices

For seven consecutive days, participants wore accelerometers and GPS units. Each participant was asked to record all trips in a standardized travel diary. The accelerometer recorded data every 30 seconds. Physical activity bouts were identified for time intervals in which accelerometer activity counts were higher than 500 counts per 30-second epoch (cpe) for at least 7 minutes (5 minutes in duration with 2-minutes tolerance of lower physical activity intensity). The start and end location could be the same location, but an accelerometer indication of physical activity was not considered to be a walking bout if it occurred at the same location. For example, walking in place on a treadmill would not be considered a walking bout, but walking around a shopping mall would be a walking bout. Travel diaries noted walking purpose, which was classified as utilitarian, recreational, or both.

The choice of device was intentional. The accelerometer was a GT1M[3] for TRAC wave 1 and a GT3X for TRAC wave 2 and thereafter. A study conducted by Carr and Mahar (2012) compared several devices and showed that the GT3X was able to identify over 80% of sedentary behaviors and 60% of light-intensity walking time based on intensity output. The GPS unit was a DG-100[4] GPS data logger for TRAC Wave 1 and a BT1000XT[5] for TRAC wave 2 and thereafter.

A decision tree algorithm (Figure 2) classified the physical activity bouts into walking and non-walking bouts. One hundred physical activity bouts were randomly sampled to test the algorithm. The algorithm's accelerometer and speed conditions are slightly modified from Kang et al. (2013). The rules associated with the decisions are described in Section 2.1.3 Classification of walking or non-walking. The criteria reviewed for walking bouts included GPS-derived walking speed, valid GPS temporal coverage, and minimum spatial extent of walking bouts.

---

[3] Both manufactured by ActiGraph LLC, Pensacola, Florida.

[4] Manufactured by GlobalSat WorldCom Co., Taipei, Taiwan.

[5] Manufactured by Qstarz International Co., Taipei, Taiwan.

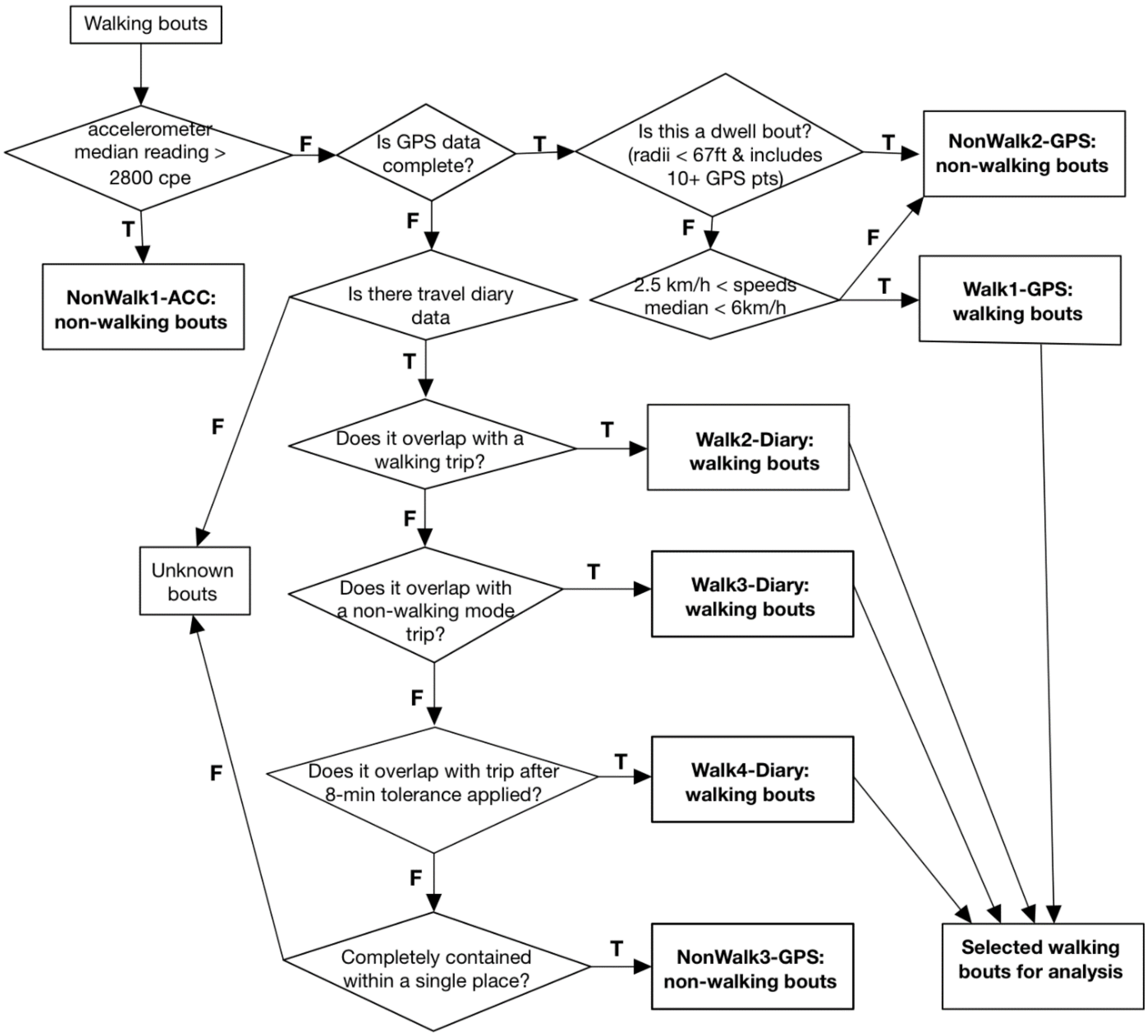


Figure 2. Summary of decision tree algorithm, modified from Kang et al. (2013)

GPS-derived walking measures

Walking bouts were defined as continuous walking with breaks that were less than or equal to 2 minutes within a 7-minute rolling window. Walking bout speeds are defined as the median of available GPS speeds within that bout. The definition accounted for noncontinuous movement often associated with walking (e.g., walking in an urban area where one might stop at an intersection) and for distinguishing walking from running or very slow movement (Kang et al., 2013). Median speed was selected rather than mean speed, which could be biased by a few GPS records from poor signals (Mooney et al., 2016).

GPS and travel diary data

Although GPS data is useful to capture location and speed, it has known limitations in terms of data completeness. Krenn et al. (2011) conducted a review that examined estimation of physical activity using GPS units with accelerometer or travel diaries, finding 17 of 24 studies had missing or unusable GPS data, ranging from 2.5% to 92% of the observed time. Missing GPS data can result from signal loss (in dense urban areas and inside buildings), signal dropout, temperature drops or increases, and low unit power. When used in real-world conditions, most GPS units will experience some rate of missingness and uncertainty relative to participant location. Analyses that rely only on accelerometer and GPS may bias estimated outcomes, since GPS data loss is not randomly distributed in time and space. The team found that combining travel diary data with accelerometer and GPS data helped to identify walking behaviors and improved the estimation of walking bouts where GPS data was missing or inaccurate. A similar procedure was completed for missing accelerometer data, although this did not happen with as much frequency. If there were no accelerometer readings or GPS data, this traversal was removed. If the absence of an accelerometer reading could be supplemented by both a GPS reading and travel diary for the duration of the trip, the traversal was classified and kept for analysis as a walking bout.

Dwell bouts or non-walking activities

Non-walking activity is defined as physical activity bouts that occurred at a single location; these are labeled “dwell bouts.” Dwell bouts were defined based on the following steps.

1. A distance marker was added between each GPS point and all other points within the bouts.
2. Points were then selected if they had a sum distance below the 95th percentile of the sum distances of all points in the bout.
3. A minimum bounding circle was then created to fully contain the selected point.
4. The circle’s radius was then obtained.
5. Bouts with radii less than or equal to 66 ft were considered as dwell bouts.

For step 5 the threshold distance of 66 ft (20 m) was selected given that Wu et al. (2010) showed that 95% of GPS points fall within this distance using a similar GPS model. Some non-dwell bouts with fewer GPS points had less than 66-ft radii. To prevent misclassification, bouts with radii less than or equal to 66 ft and more than or equal to 10 GPS points were defined as dwell bouts.

The accelerometer data was the most complete and accurate when compared to the GPS data, and GPS data was more reliable and accurate than travel diary data. The walking bout algorithm was structured to prioritize data obtained from the more reliable and accurate instruments first among available data. For example, a bout that includes complete GPS data and a median speed of 0.25 km/h is less than the threshold for walking. (As indicated by the tree algorithm, walking speed is a median speed from 2.5 km/h to 6 km/h.) Hence, this would be considered non-walking even though the travel diary may have recorded that walking had occurred.

Classification of walking or non-walking

There were seven scenarios observed for which walking and non-walking could be identified. All scenarios were based on accelerometer data, but not all had complete accelerometer or GPS data. The duration of physical activity bouts was derived from the accelerometer. Depending on the completeness of the GPS data and the diary entries, the seven scenarios were defined as follows.

1. Walk1-GPS: GPS and accelerometer data.
    a. GPS-derived case with non-dwell points and satisfactory walking speed: bouts with complete GPS data, with non-dwell GPS points, and with median GPS speeds ranging from 2.5 km/h to 6 km/h.
    b. While walking speed was not calculated for each person, a very low threshold for both accelerometer activity counts and walking speeds was used to be as inclusive as possible of very slow walkers.
2. Walk2-Diary: Accelerometer data overlapped with travel diary.
    a. Walking bouts with no or incomplete GPS data but that have any overlap in time with a walking trip recorded in the travel diary.
3. Walk3-Diary: Accelerometer data overlapped with the diary-based, non-walking trip.
    a. Bouts with no GPS data have some time overlap with a non-walking trip in the travel diary.
    b. If walking is involved for otherwise non-walking trips (e.g., walking to and from transit stops), it is assumed that such walking trips are not recorded in the diary because they were not the primary travel mode.
    c. When the travel diary recorded non-walking trips (e.g., car, transit, and bike trips) and accelerometer data shows peaks, adjacent bouts were considered as unreported walking trips.
4. Walk4-Diary: Overlap within 8-minute tolerance of diary-based trip.
    a. Bouts overlapping within an 8-minute time buffer around walking trips recorded in travel diaries.
    b. Recorded times on travel diaries and accelerometer did not match perfectly. This could be due to inaccurate time recall and rounding times on travel diaries.
    c. The 10-minute window of a diary-based trip is used to minimize the false-positive errors.
5. NonWalk1 – ACC: Upper bound of accelerometer counts.
    a. Bouts of vigorous physical activity with a mean count of more than 2,800 cpe.
    b. Note: This was validated with three physical activity bouts that were self-reported to be indoor exercising, which showed mean counts from 2,874 to 3,360 cpe.
6. NonWalk2 – GPS: GPS-derived dwell and speed.
    a. Bouts that fall within the definition of dwell bouts and have out-of-range GPS median speed for walking.
7. NonWalk3-Diary: Occurring within a diary-based place.
    a. Bouts with no or incomplete GPS data but bout durations fell completely within a reported single place (e.g., home) and not within an 8-minute tolerance of a reported trip.

These seven scenarios were consolidated into the decision-tree algorithm shown in Figure 2. The number of walking bouts per person provides insight into the frequency of walking, independent of the individual's walking distance or walking time. Given that trip diaries are used in conjunction with electronic devices, the walking bout algorithm captures the number and frequency of walking for activities such as commuting to school or work, visiting friends, shopping, or sightseeing. As noted earlier, walking bouts can be assessed at an individual level, as a function of a person, household, or location. For larger areas, data on walking bouts can provide insights on changing spatial and temporal patterns of pedestrian exposure. Walking bouts can also provide insights on characteristics that generate similar trip purposes or routes.

Once the data was extracted into walking bouts (Figure 2), additional processing was applied to count the walking bouts by intersection. Figure 3 and Figure 6 show the process for Project ACTION and Project TRAC data, respectively.

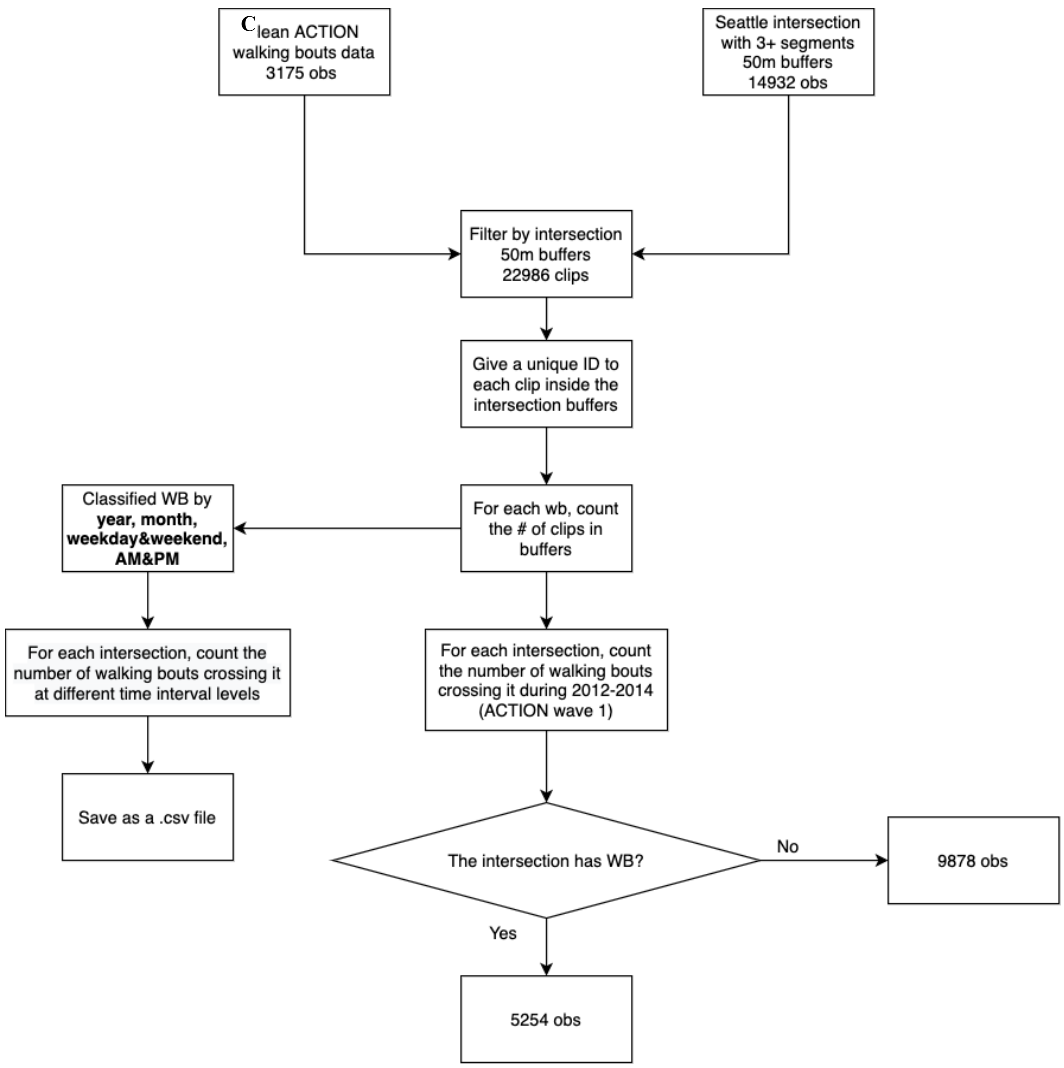


Figure 3. Project ACTION results from the Kang et al. (2013) walking bout processing framework

The “cleaned” walking bout data (Figure 2) was mapped to the Seattle intersections. These intersections were sourced from Census TIGER/Line geographic system, which is government-collected spatial data that includes geographic identification and topological information. These intersections were filtered to 3+ segments to focus the scope to intersections that have more interaction with road users. These intersections were also buffered by 50 m to increase coverage and account for imperfections in the GPS data. For example, if GPS data was slightly inaccurate, or the intersection had large sidewalks or pedestrian gathering areas around them, this walking data may not be captured as occurring at the intersection that it did. These buffers also help assume a larger range or more generalized impact of the walking activity observed. These buffers assume that exposure experienced by a pedestrian is the same within this 50 m buffer range, for example on either side of the road, or opposite sides of the intersection.

After filtering the walking bouts by these intersections, the walking bouts were “clipped” using the geospatial software, QGIS. The number of clips within each buffer represents the number of walking bouts that passed through this buffer, and therefore were ultimately counted and used as the outcome variable, walking bout count. Figure 4 and Figure 5 show how the walking bout counts were computed by 50 m intersection buffer.

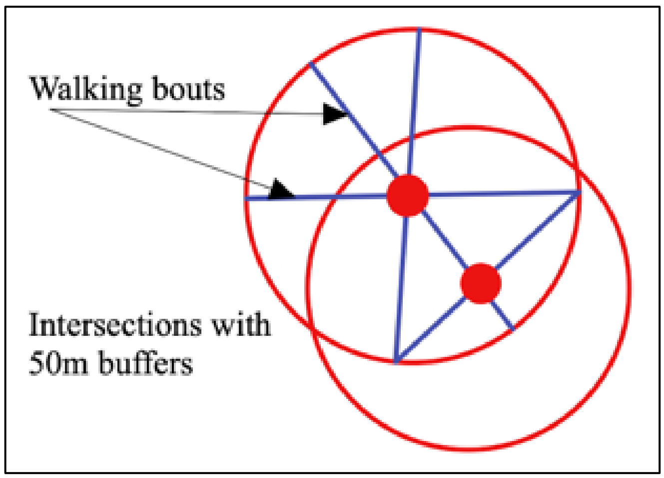


Figure 4. Demonstration of single buffer containing more than 1 intersection. All walking bouts counted for a single buffer are attributed to all intersections falling within a single buffer

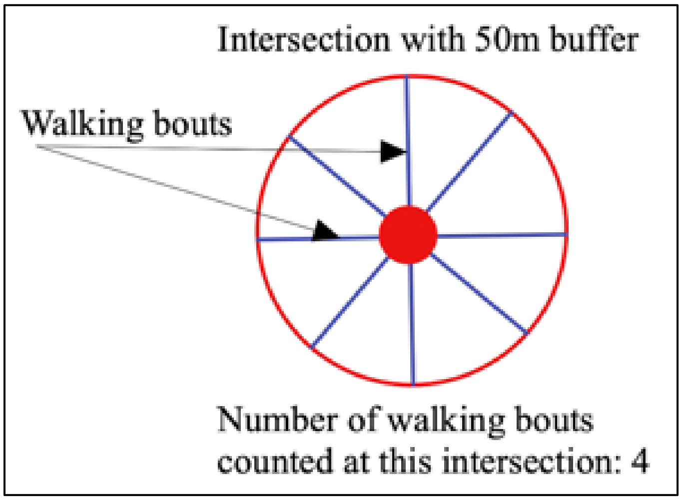


Figure 5. Demonstration of walking bout "count" by intersection 50 m buffers

Additional processing was computed to aggregate walking bouts by datetime, which was used to visualize differences in walking activity over the course of Project TRAC and Project ACTION. This data could also be used longitudinally for a more focused analysis of each projects' participants.

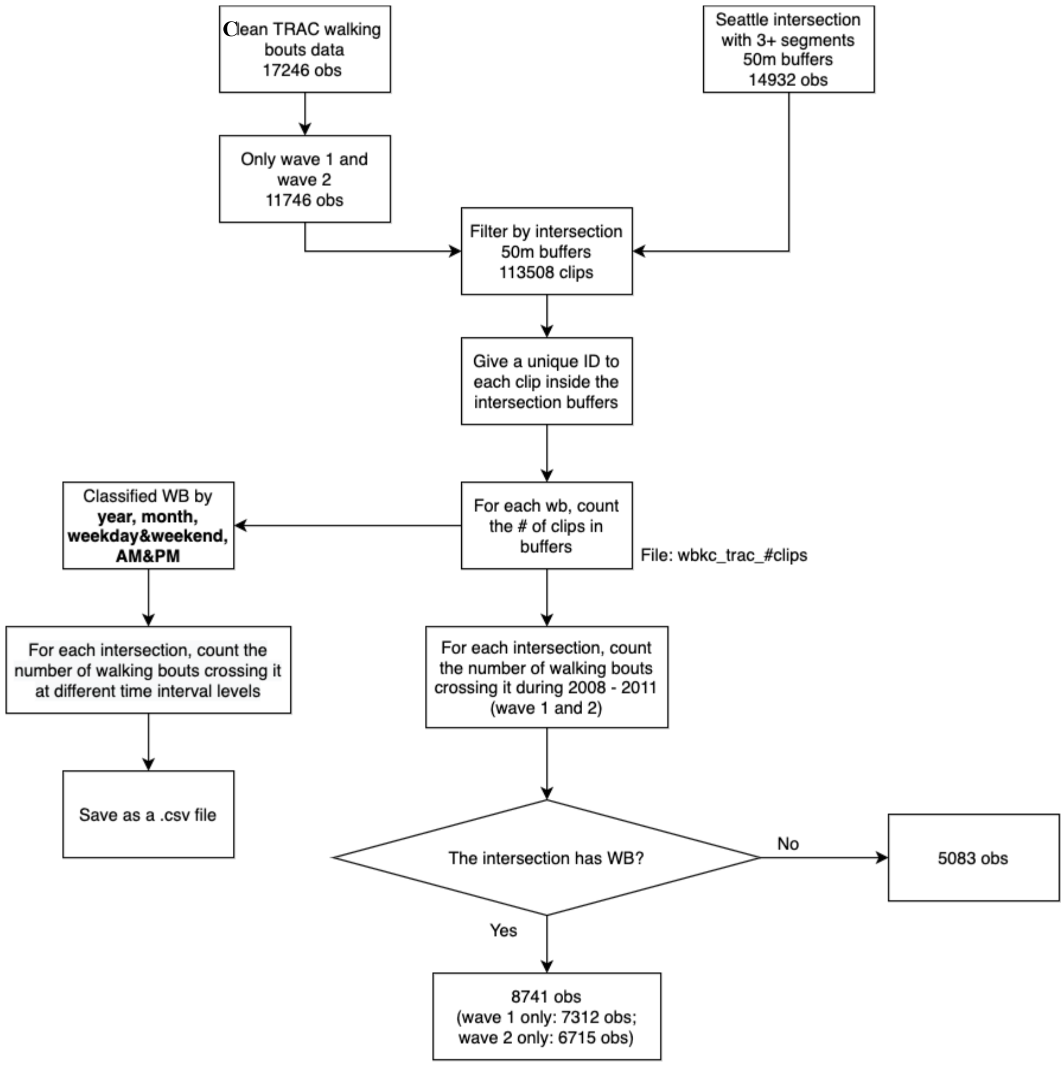


Figure 6. Project TRAC results from the Kang et al. (2013) walking bout processing framework

Using walking bouts to estimate pedestrian exposure

While walking bouts have been used to study health, they have not been used to estimate pedestrian exposure. In the processing described previously, physical activity data were developed into walking bouts, and walking bout counts were processed at the intersection level. The following steps were used to estimate pedestrian exposure using the walking bout counts:

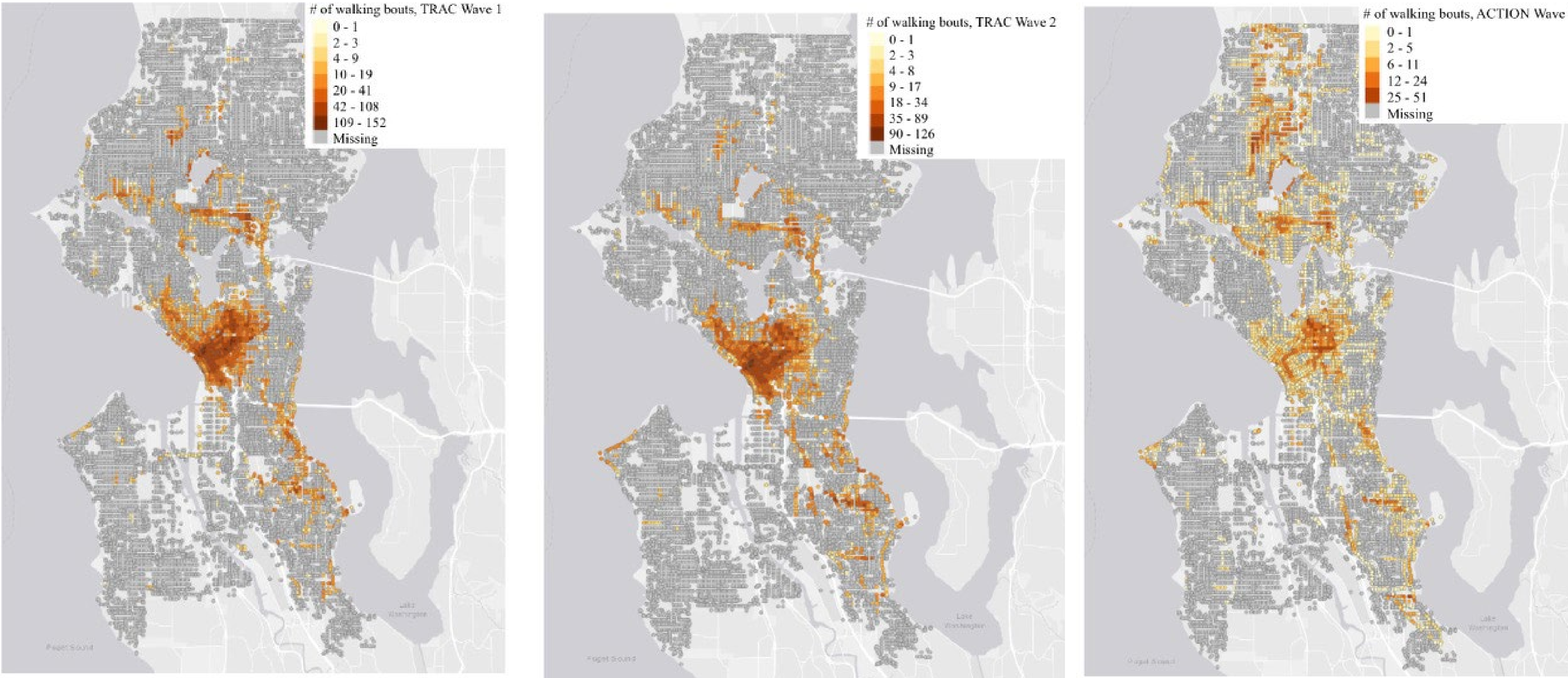
1. Compute spatial and temporal metrics for each validated Seattle pedestrian count location and at walking bout locations. Locations where the Seattle collected pedestrian counts can be found at: https://data.seattle.gov/browse/select_dataset?limitTo=datasets&sortBy=relevance&utf8=%E2%9C%93&q=pedestrian
2. Compute pedestrian counts at each intersection using walking bouts. The spatial metric of interest included area density of development, land use, and transportation infrastructure (bus stops, light rail, etc.). The temporal metric of interest included pedestrians per day, per time of day (morning and afternoon commute, lunchtime, etc.), per month, and per season.
3. Create models to predict pedestrian exposure using Seattle pedestrian count data. The models were conducted given the spatial and temporal segmentations from Step 1. The explanatory variables were based on the list of variables identified from the program scan. Example measures included population density, employment density, and retail proportion.
4. Validate the use of walking bouts. Models were developed to predict pedestrian exposure using walking bouts. The model outcomes (magnitude and direction of the parameter coefficients) were compared to data from the Seattle (Step 2). A goal was to examine the impact and value of each spatial metric. Cross-validation was done to ensure that the model was the most accurate for modeling pedestrian exposure.
5. Use the predictive model with the number of walking bouts for various spatial metrics to estimate pedestrian counts at unmeasured locations. This included the additional walking bouts for locations where pedestrian counts were not available. The data were aggregated to the various spatial areas of interest (defined in Step 1). The team validated the ability of the model to provide consistent, reliable estimates using a subset of the dataset, as well as the ability to provide generalizable outcomes for other datasets.

Descriptive statistics

This section familiarizes the reader with the data used in pedestrian exposure modeling with summary statistics and visual data. The walking bouts were split into two datasets: Dataset 1 included all intersections in Seattle (with 3+ segments), and Dataset 2 included only intersections where 10+ walking bouts were observed. To examine walking prevalence (where walking is known to have occurred), Dataset 2 also included demographic variables of the participants. All variables and corresponding data sources considered in this section and the forthcoming models are provided in Appendix B.

Post-processed walking bouts

Figure 7 shows the walking bouts counts for TRAC (waves 1 and 2) and ACTION (wave 1) identified after the processing framework was applied. As a reminder, there were three waves each for Project TRAC and Project ACTION. TRAC wave 3 and ACTION wave 2 were used in the validation of the pedestrian exposure model. The figure shows that Project TRAC participants were mostly concentrated in the downtown area, with limited dispersion to the north and south regions of Seattle, while participants in Project ACTION have higher pedestrian dispersion.



(a) TRAC Wave 1 (b) TRAC Wave 2 (c) ACTION Wave 1

Figure 7. Post-processing and counting of walking bouts within intersection buffers for (a) TRAC Waves 1, (b) 2, and (c) ACTION Wave 1

Environmental predictors

Correlation analysis between explanatory variables was conducted to identify any potential multicollinearity issues. Figure 8 shows the distribution of explanatory variables for the following top 8 variables that showed the highest correlation. Please refer to the numerated list to identify variables in Figure 8. The correlations for all environmental variables considered can be found in Appendix D.

- Residential density from the U.S. Census (RES_CENSUS_C_ACRE)
- Population density from the U.S. Census (POP_CENSUS_C_ACRE)
- Residential density from Urban Form Lab (UFL) Smart map (RES_UNIT_C_ACRE)
- Job density from the U.S. Census (JOBS_C_ACRE)
- Total crosswalk count (TOTAL_CROSSW_COUNT)
- Residential land use (RES_LU_PER)
- Total bike lane length (ft) (TOTAL_BIKEL_FT)
- Traffic sign presence (TRAFF_SIGN_Y)

These variables were explored in greater detail because they showed the highest correlations amongst each other. Additional visuals depicting the dispersion of the top eight predictors throughout Seattle are provided below and in Appendix C.

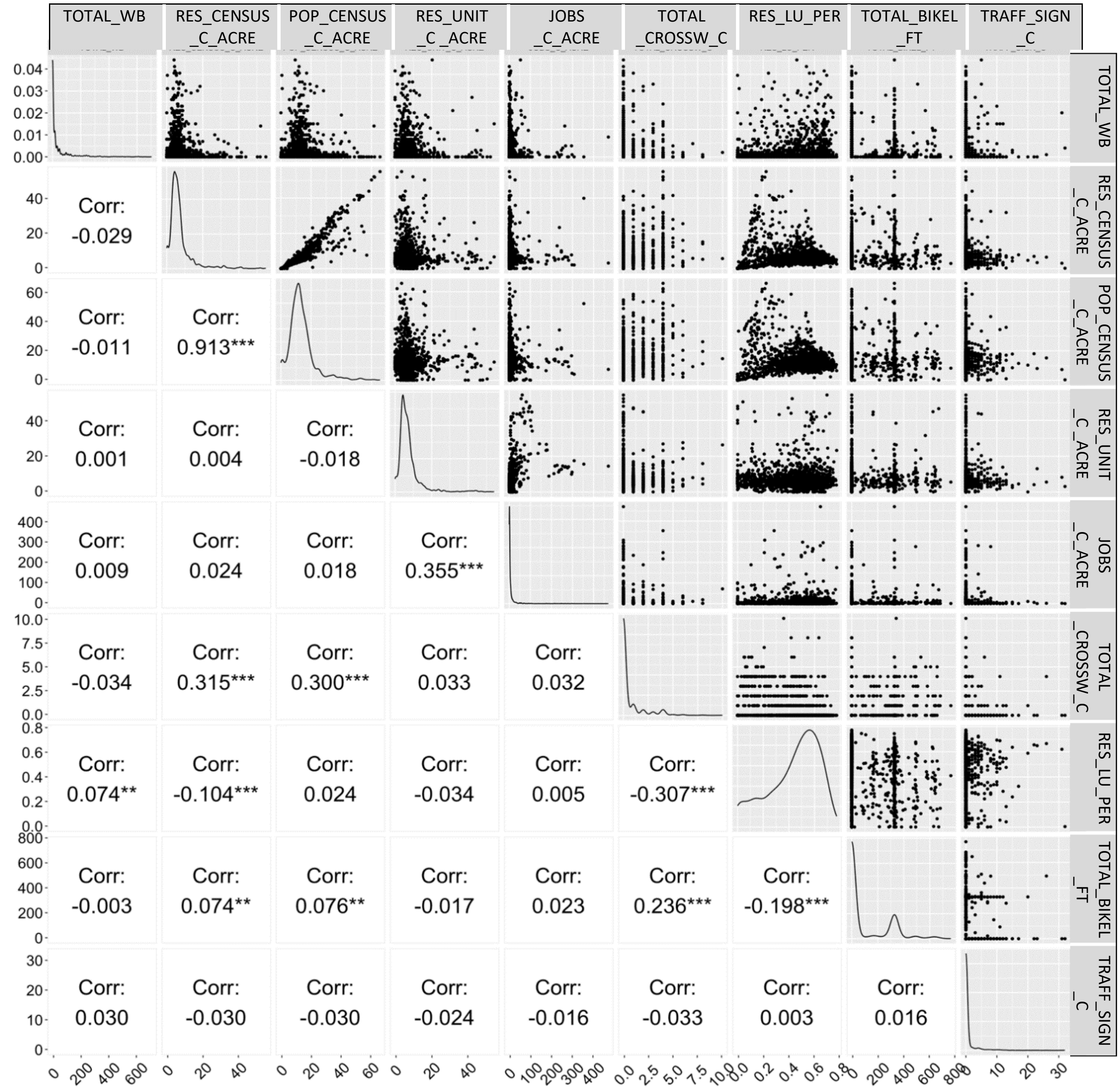
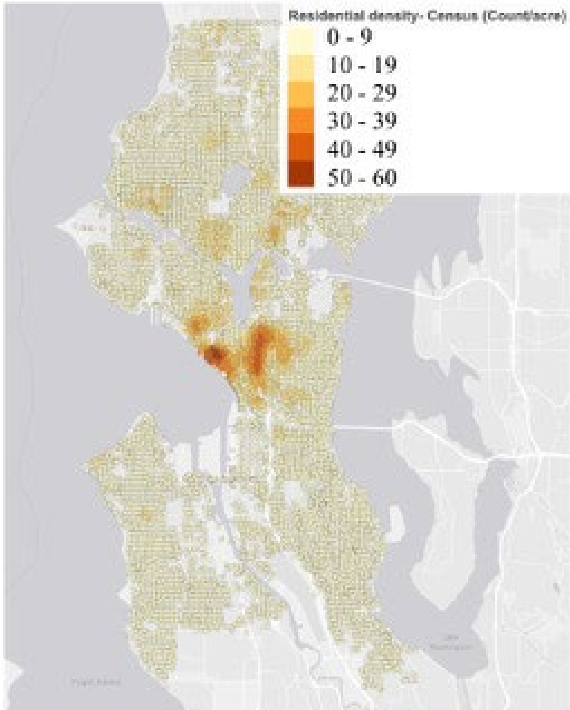
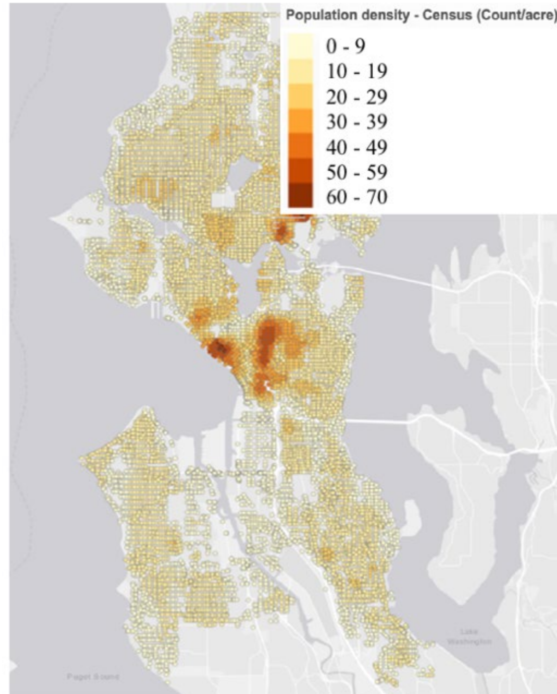


Figure 8. Correlation analysis and data distribution of highly correlated variables (α = .05)

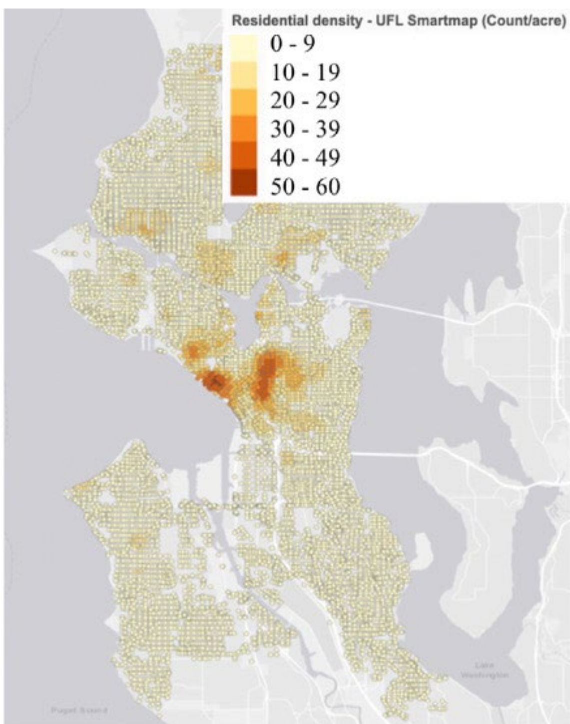
Figure 8 shows the distribution of the variables that are highly correlated. The proportion of residential land use (RES_LU_PER) at each intersection was further examined in Figure 9 (d). This variable represents the percent of residential land use at the intersection. Low percentage of residential land use in dense, urbanized areas (e.g., downtown) may indicate that there are many people residing in high rise buildings. This is confirmed given the concentration of people in the downtown area (see other variables [a to c] in Figure 9). The focus was to explore the relationship between total walking bouts and actual number of people who reside in the area. After extensive review, the population density from the Census appeared to be the most representative; we therefore excluded other similar variables (residential density from the Census, residential density from the Urban Form Lab Smartmap).



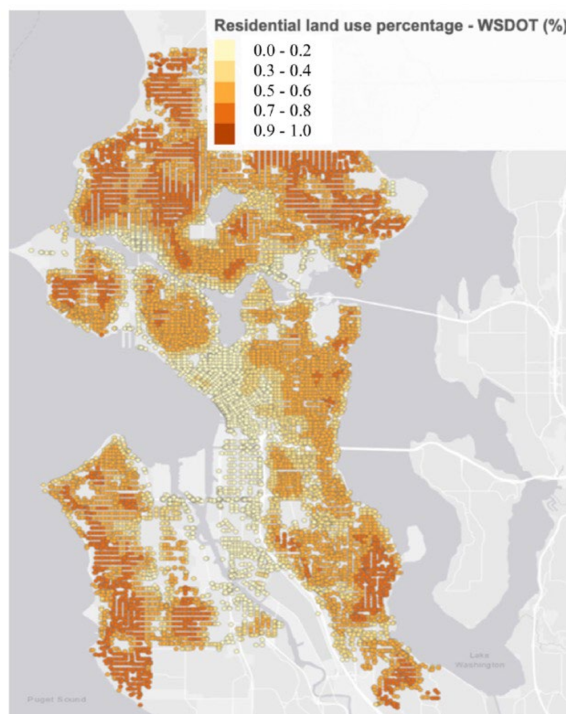
(a) *Residential census density*



(b) *Population census density*



(c) *Residential density*



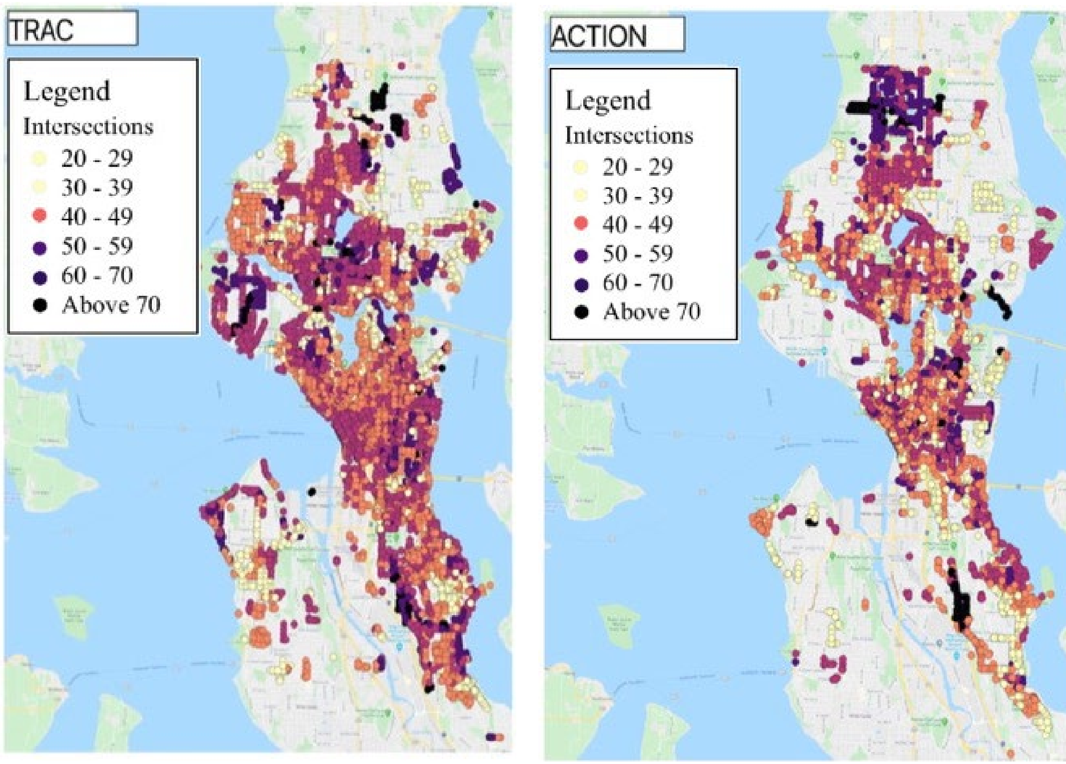
(d) *Residential land use percentage*

Figure 9. Seattle density and land use by 50 m intersection buffers

Demographic predictors

Age

The mean age by location for the TRAC and ACTION data is shown in Figure 10. The ACTION data was obtained from residences near bus rapid transit stations. For that reason, we noticed that there were many older adults near transit stations given the proximity to several housing developments for older adults.



(a) TRAC (b) ACTION

Figure 10. Average age of pedestrians at intersections in the (a) TRAC and (b) ACTION data

Sex

The number of female and male participants are shown in Figure 11. In general, the number of males and females was about the same in both the TRAC and ACTION data. The TRAC data are binned into quintiles, representing an even count dispersal within each sex, study combination. The ACTION data contains a smaller subset of pedestrians (only one wave), and thus was discretized even further into seven (ACTION, Female) and six (ACTION, Male) bins.

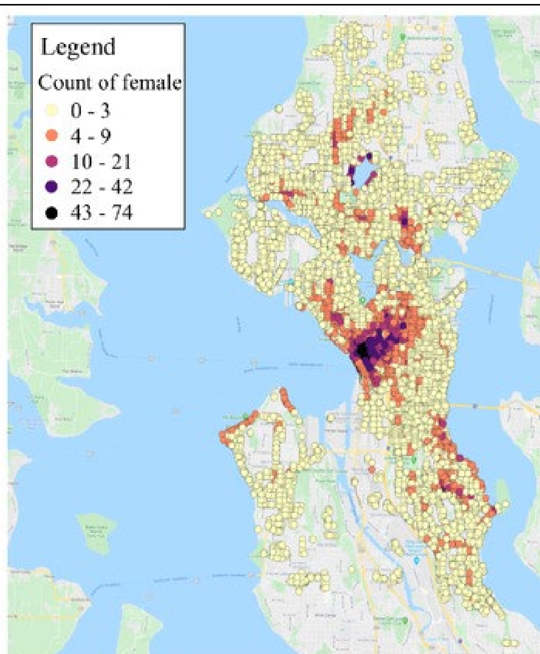
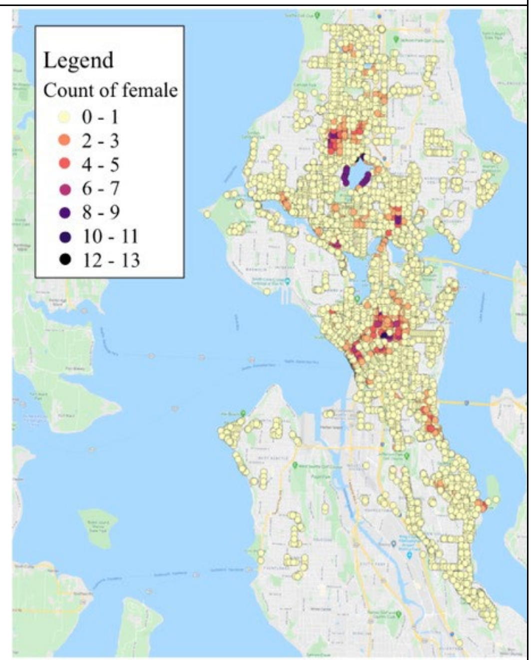
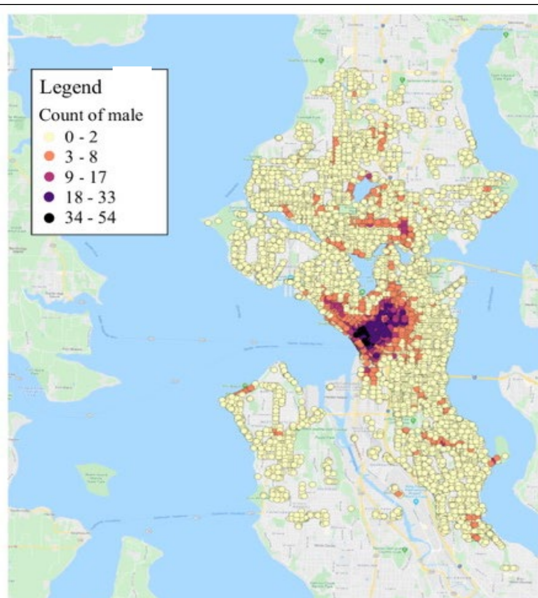
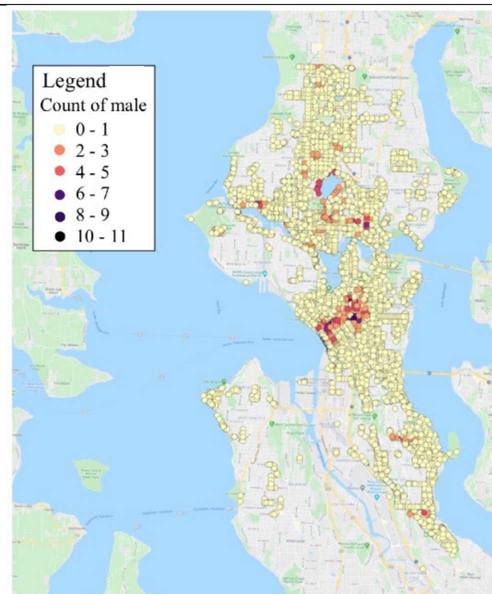
<table>
<tr><th>Variable</th><th>TRAC data</th><th>ACTION data</th></tr>
<tr><td>Females</td><td>
</td><td>
</td></tr>
<tr><td>Males</td><td>
</td><td>
</td></tr>
</table>

Figure 11. Sex (females, males) of pedestrians in TRAC and ACTION project by intersection.

Race

In some intersections, the number of pedestrian participants that self-reported as white were several times larger than the count of those that self-reported as non-white (Figure 12), especially in areas like downtown Seattle. This could also be a possible source of bias in the developed model.

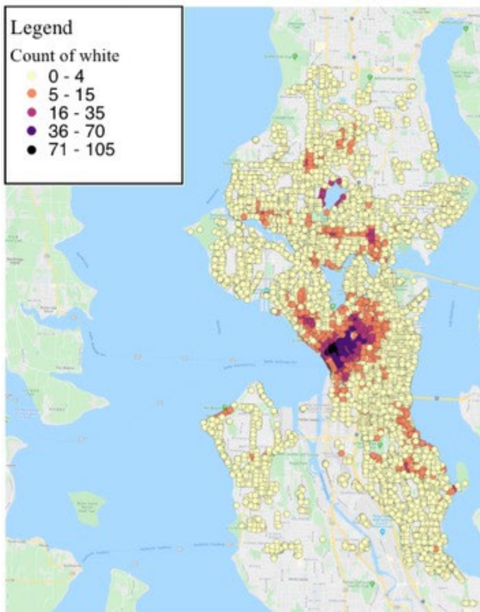
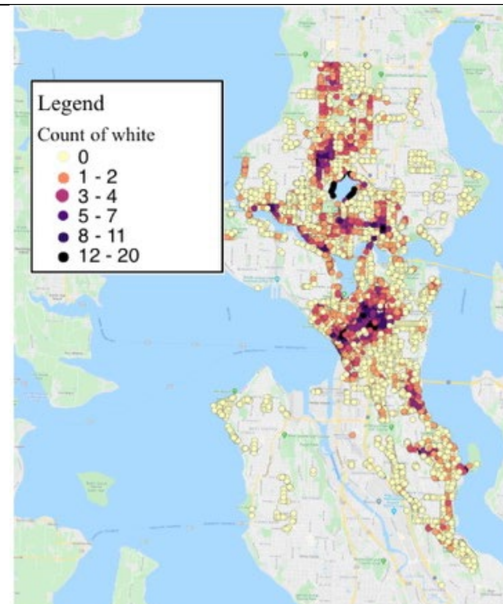
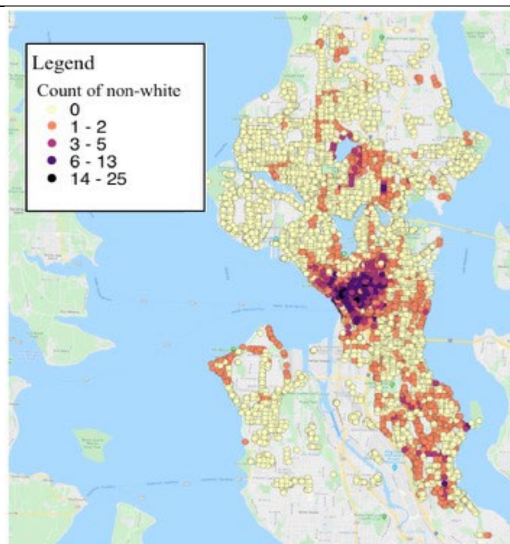
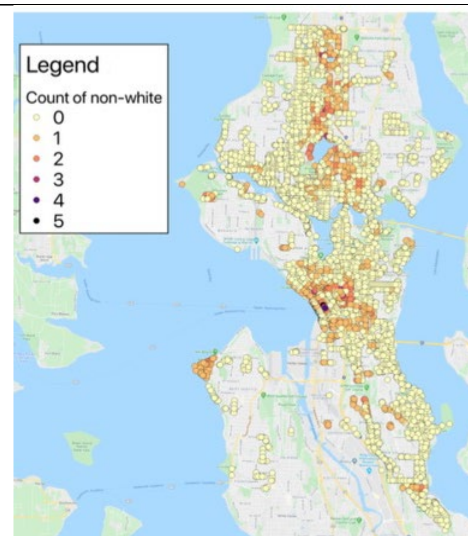
Variable	TRAC data	ACTION data
Count of white pedestrian participants		
Count of non-white pedestrian participants		

Figure 12. Self-reported race (white, non-white) of pedestrians in the TRAC and ACTION projects by intersections

Employment

In ACTION data, counts of unemployed participants in some intersections, like those in downtown Seattle and Northgate, are much larger than employed subjects. This is partly because of the definition of "employed" in the survey. "Employed" is defined as "working outside your home" so retired participants, students without any internships, and participants working from home are not considered "employed." The term "employed" is used because walking bout data includes activities of the observed pedestrians walking for commuting and non-commuting. This way certain walking bouts can be weighted with purpose; if a person makes the same walk to the same destination twice a day and is employed, this may be attributed to commuting.

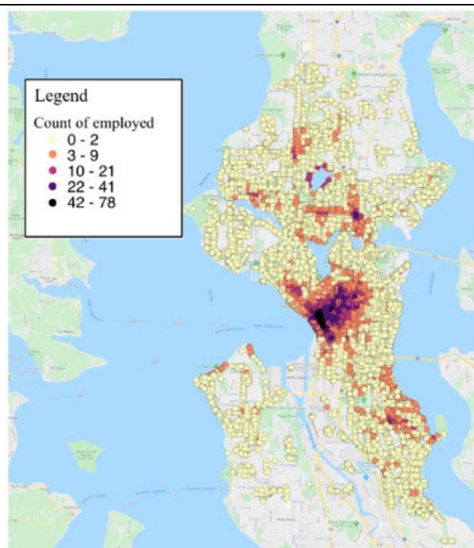
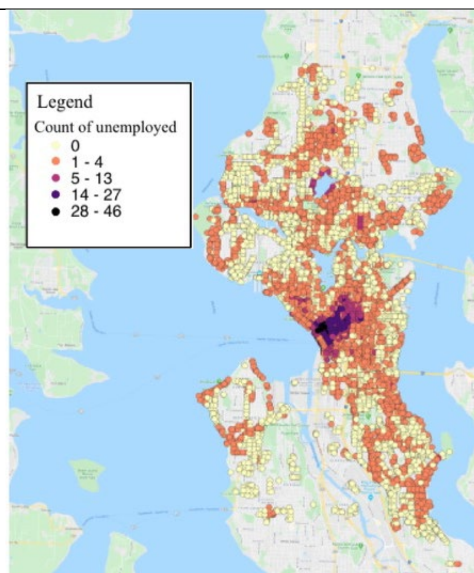
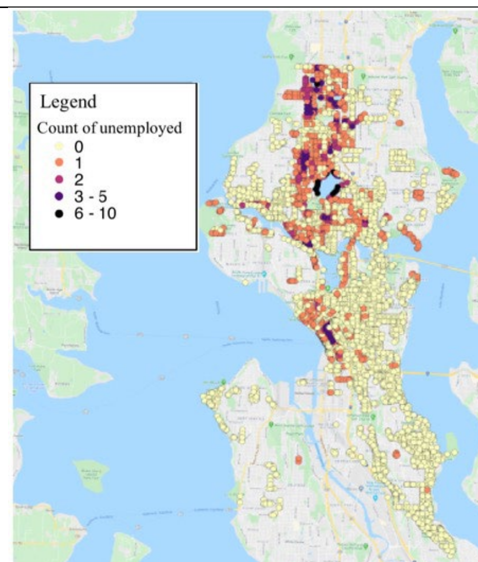
<table>
<tr><th>Variable</th><th>TRAC data</th><th>ACTION data</th></tr>
<tr><td>Employed</td><td>
</td><td>
</td></tr>
<tr><td>Unemployed</td><td>
</td><td>
</td></tr>
</table>

Figure 13. Employment status (employed, unemployed) of pedestrians in TRAC and ACTION project by intersections

After walking bouts, environmental variables, and demographic variables were processed at the intersection level, the two datasets for modeling were developed.

Correlation analysis

Dataset 1

Dataset 1 (n = 14,073) included all intersections in Seattle. Figure 14 shows that several variables were highly correlated with each other.

- "Crosswalk count" was highly correlated with "Traffic signal presence" ($\rho = .7$). "Crosswalk count" was removed but "Crosswalk warning sign presence" was retained to represent the relationship of crosswalk with walking bouts.
- "Stop sign" was highly correlated with "Max speed limit" ($\rho = .6$), therefore it was removed from the model for Dataset 1.
- "Service land use," "Job density", "Median household income", and "White population" exhibited high correlations ($\rho > .5$).
- "Culture land use" includes all publicly accessible social and entertainment areas such as public theaters, shopping malls, and parks. To avoid multicollinearity with other land use variables, this variable, "Culture land use', was processed as "Park presence," which is the variable used in the model.

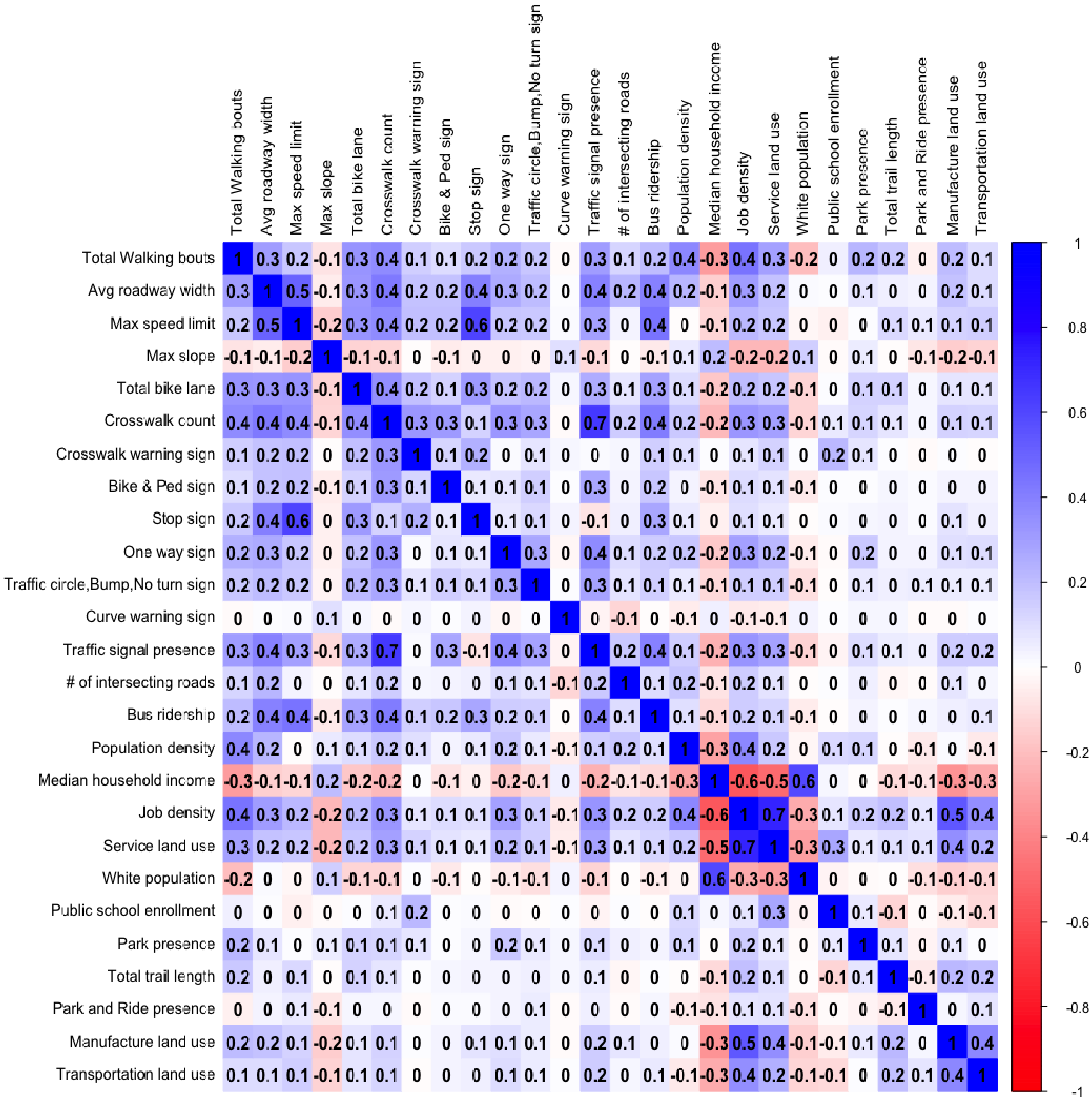


Figure 14. Correlation plot for Dataset 1 (n = 14,073)

Dataset 2

There were 8,547 intersections in Seattle with at least one walking bout. To account for individual pedestrian characteristics, Dataset 2 was further filtered to include only those intersections with at least 10 walking bouts in Seattle (n = 3,047). Figure 15 shows the correlation plot for Dataset 2. “Stop sign presence” no longer induced any multicollinearity (as it did for Dataset 1, therefore it was included in the model for Dataset 2.

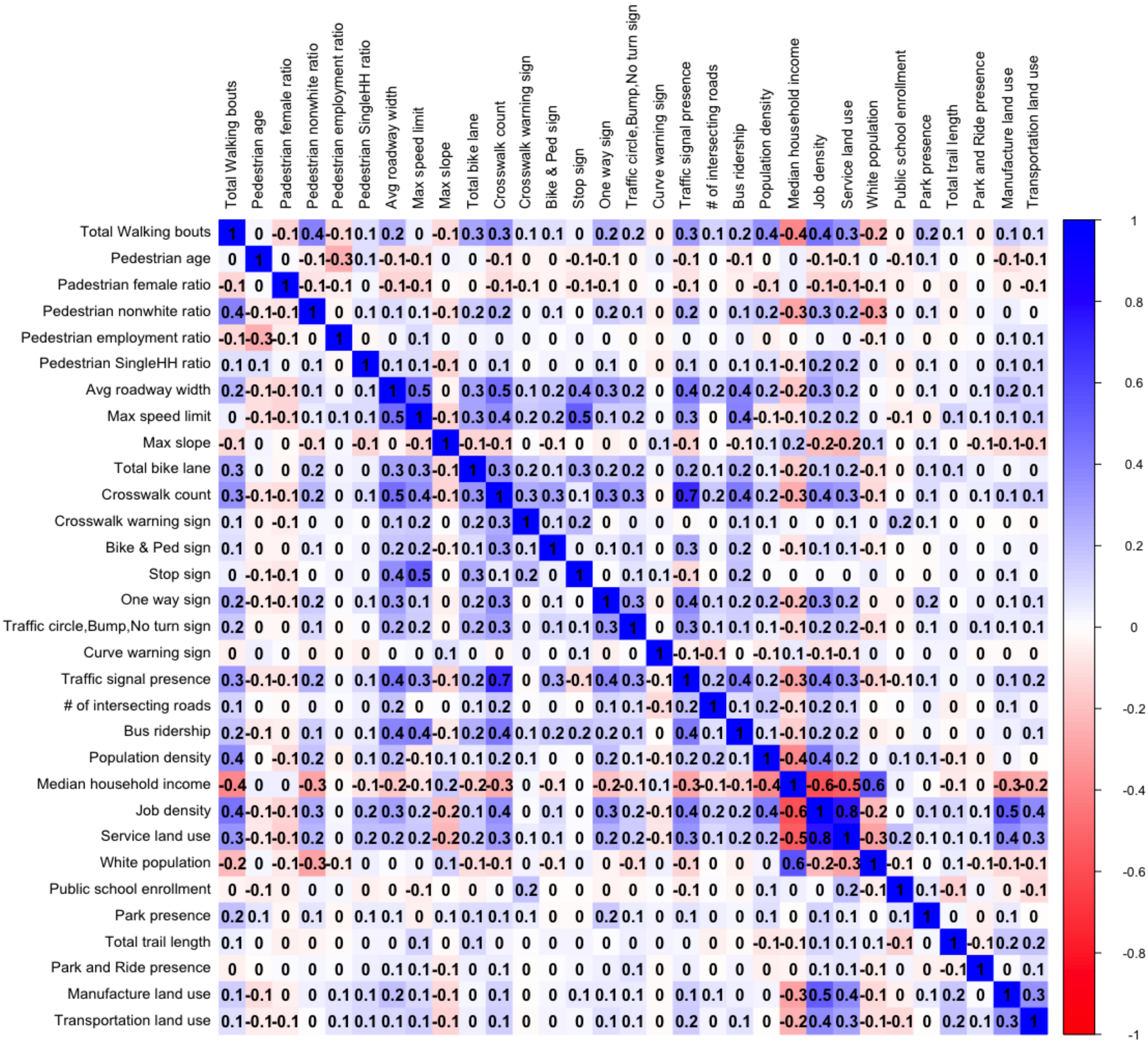


Figure 15. Dataset 2, intersections with 10+ walking bouts (n = 3,047)

Variable selection

The team used LASSO regression, a penalized regression technique, to select the most relevant environmental predictors. For example, if the outcome y_i is linearly associated with estimators x_{i1} to x_{ij}, then the coefficient β_j of an elastic net regularization method can be estimated by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i j\right)^2 + \lambda_1\sum_{j=1}^{p}\left|\beta_j\right| + \lambda_2\sum_{j=1}^{p}\beta_j^2 = RSS + LASSOpenalty$$

where λ_1 and λ_2 are tuning parameters based on the data, and β_0 is the intercept. By incorporating an elastic net regularization (the combination of LASSO penalty terms), the equation appropriately penalizes the situation where too many predictors are added to the model. This helps to prevent model overfitting while ensuring the successful selection of variables. A 10-fold cross-validation strategy was useful to assess model performance without overfitting to the data.

In this study, y=number of walking bouts within a 50 m buffered intersection. The estimators, x_{i1} to x_{ij}, represent explanatory variables. The list of variables considered are provided in Appendix E: LASSO for variable selection.

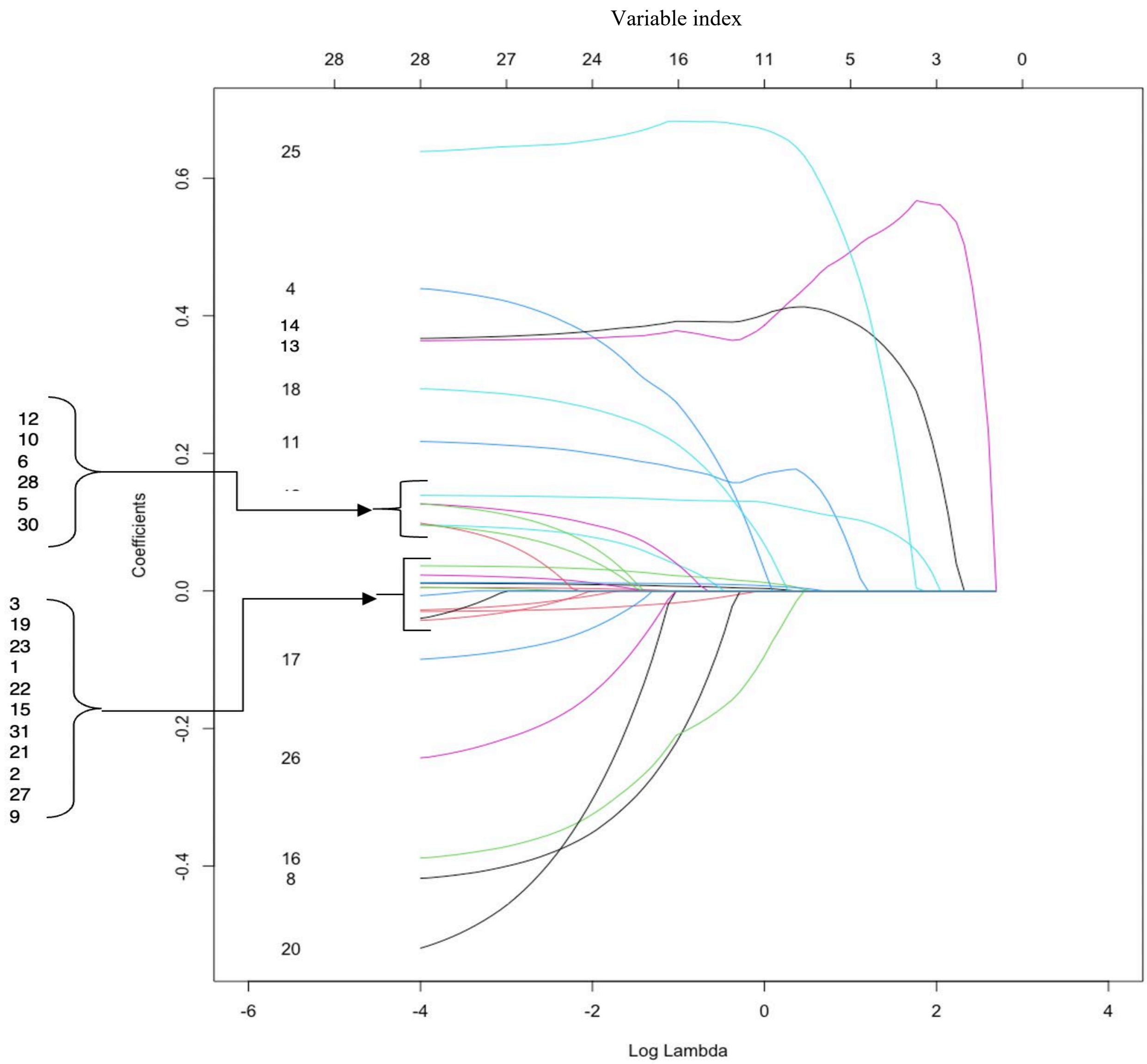
Dataset 1

A LASSO Poisson regression was built to estimate the number of walking bouts with the 31 explanatory variables (23 numerical variables; 2 categorical variables that were represented with 6 dummy variables (3 levels/factor) to represent maximum posted speed with the reference level of 20 mph (dummies: 25, 30, 35, 40 mph) and roadway segment count with a reference level of 4 segments (dummies: 3, 5+ roadway segments).

1. Average roadway width
2. Max slope
3. Total bike lane length
4. Sidewalk length
5. Presence of crosswalk warning sign
6. Presence of bike and pedestrian sign
7. Presence of stop sign (REMOVED)
8. Presence of one-way sign
9. Presence of traffic circle, Bump, No turn sign
10. Presence of curve warning sign
11. Traffic signal presence
12. Bus ridership
13. Job density
14. Population density

15. White population
16. Median household income
17. Public school enrollment
18. Park presence
19. Total trail length
20. Park-and-Ride presence
21. Manufacture land use
22. Transportation land use
23. Service land use
24. Max speed limit (20 mph) (REMOVED)
25. Max speed limit (25 mph)
26. Max speed limit (30 mph)
27. Max speed limit (35 mph)
28. Max speed limit (40 mph)
29. Number of intersecting segment (4) (REMOVED)
30. Number of intersecting segment (3)
31. Number of intersecting segment (>=5)

Figure 16 summarizes the path trajectory of the fitted sparse regression parameters. Each curve shows how the regression coefficient of a variable changes according to the value of lambda. The figure should be read from right to left - lambda from small to large. Variables with coefficient values (y-axis) that quickly become zero are considered weaker and are more likely to be insignificant in the model. The numbers in the figure refer to the variable list above.



Note: Each variable is numbered from the list found previously in this section.

Figure 16. Variable convergence of LASSO for Dataset 1

Using a 10-fold cross validation, the best model and the corresponding coefficients were identified. The selected best model was selected based on the best lambda value (0.018) with the smallest model deviance. The best model excluded stop sign presence (STOP_SIGN). This supported the reasoning to exclude stop sign presence as noted in the previous section.

Figure 17 identifies the lowest point that corresponds to the best model with the 28 variables (excluding stop sign presence and two referenced variables from maximum posted speed and roadway segment count).

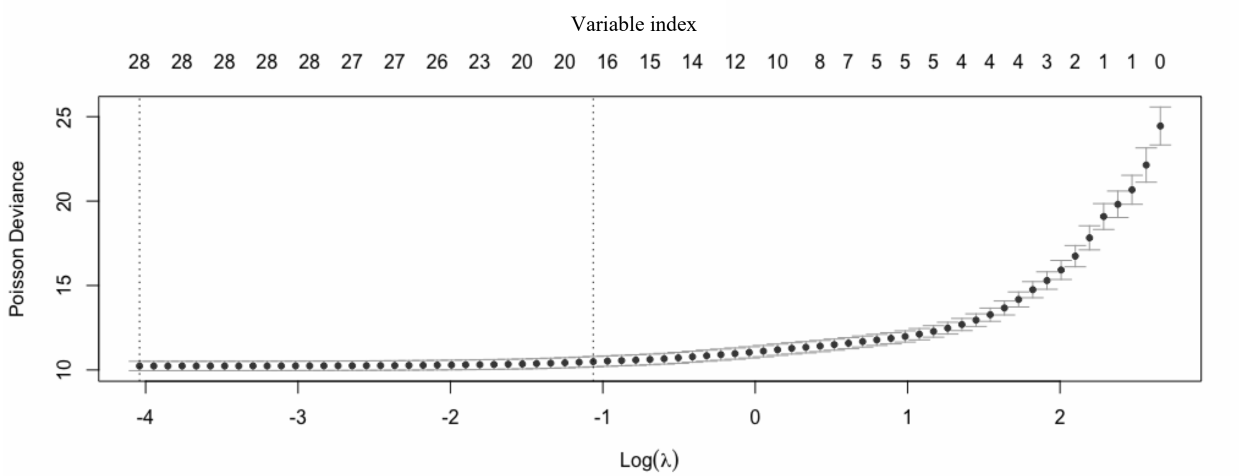


Figure 17. 10-fold cross validation of LASSO for Dataset 1

Figure 16 provides insights on the specific variables that hold weight (significance) in the model while Figure 17 provides insight on the total number of variables that minimize deviation in the model. Figure 17 shows that the model could include 28 variables (out of 31) with stable behavior, while Figure 16 helps to identify those variables for inclusion. The top x-axis represents the independent variable index in both Figures.

Dataset 2

LASSO Poisson regression was built to estimate number of walking bouts with 40 explanatory variables (23 numerical variables; 2 categorical variables with 6 dummy variables to represent maximum posted speed with reference level of 20 mph (dummies: 25, 30, 35, 40 mph) and roadway segment count with the reference level of 4 segments (dummies: 3, 5+ roadway segments); and 6 individual-level variables with 3 dummy variables for median household income of pedestrians (<$40K, $71K-$99K, >$100K). Additionally, Dataset 2 considered the following demographic variables:

- age,
- sex,
- race, and
- employment status.

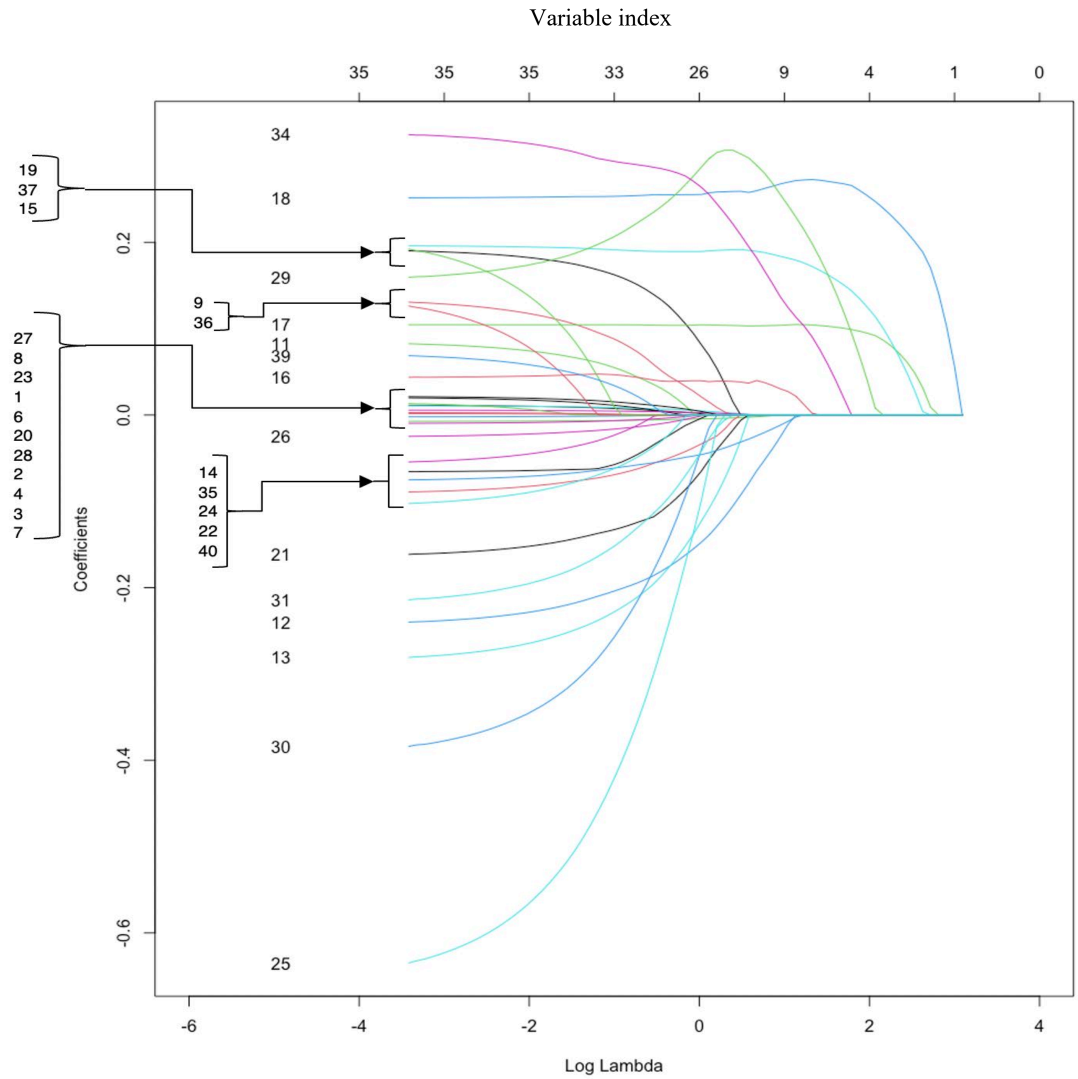
As noted earlier, Dataset 2 includes demographic variables associated with pedestrians. Because this dataset only includes intersections with 10+ walking bouts, participant information can be aggregated to the intersection-level, which protects the privacy of the participant while allowing the model to consider demographic attributes. The demographic predictors were processed within a 50 m buffer. The full list included 40 variables.

1. Pedestrian age
2. Pedestrian female ratio
3. Pedestrian participant non-white ratio
4. Pedestrian employment ratio

5. Pedestrian single household (HH) ratio (REMOVED)
6. Average roadway width
7. Max slope
8. Total bike lane length
9. Sidewalk length
10. Presence of crosswalk warning sign
11. Presence of bike and pedestrian sign
12. Presence of stop sign
13. Presence of one-way sign
14. Presence of traffic circle, Bump, No turn sign
15. Presence of curve warning sign (REMOVED)
16. Traffic signal presence
17. Bus ridership
18. Job density
19. Population density
20. White population
21. Median household income
22. Public school enrollment
23. Park presence
24. Total trail length
25. Park and Ride presence
26. Manufacture land use
27. Transportation land use
28. Service land use
29. Pedestrian median HH income (40K – 69K)
30. Pedestrian median HH income (<40K)
31. Pedestrian median HH income (70K – 99K)
32. Pedestrian median HH income (> 100K) (REMOVED)
33. Max speed limit (20 mph) (REMOVED)
34. Max speed limit (25 mph)
35. Max speed limit (30 mph)
36. Max speed limit (35 mph)
37. Max speed limit (40 mph)

38. Number of intersecting segment (4) (REMOVED)
39. Number of intersecting segment (3)
40. Number of intersecting segment (>=5)

Figure 18 summarizes the path trajectory of the fitted sparse regression parameters. Each curve shows how the regression coefficient of a variable changes according to the value of lambda. The figure should be read from right to left – lambda from small to large. Variables where their coefficient value (y-axis) quickly become zero are considered weaker and are more likely to be insignificant in the model. The numbers refer to the variable list above.



Note: Each variable is numbered from the list found previously in this section.

Figure 18. Variable convergence of LASSO for Dataset 2

Using a 10-fold cross validation, the best model and the corresponding coefficients were identified. The best model was selected based on the best lambda value (0.04) with the smallest model deviance. The best model excluded the presence of a crosswalk warning sign (CW_SIGN_Y).

Figure 19 identifies the lowest point that corresponds to the best model with the 36 variables (excluding presence of crosswalk warning signs and three referenced variables from median household income of pedestrians, maximum posted speed, roadway segment count).

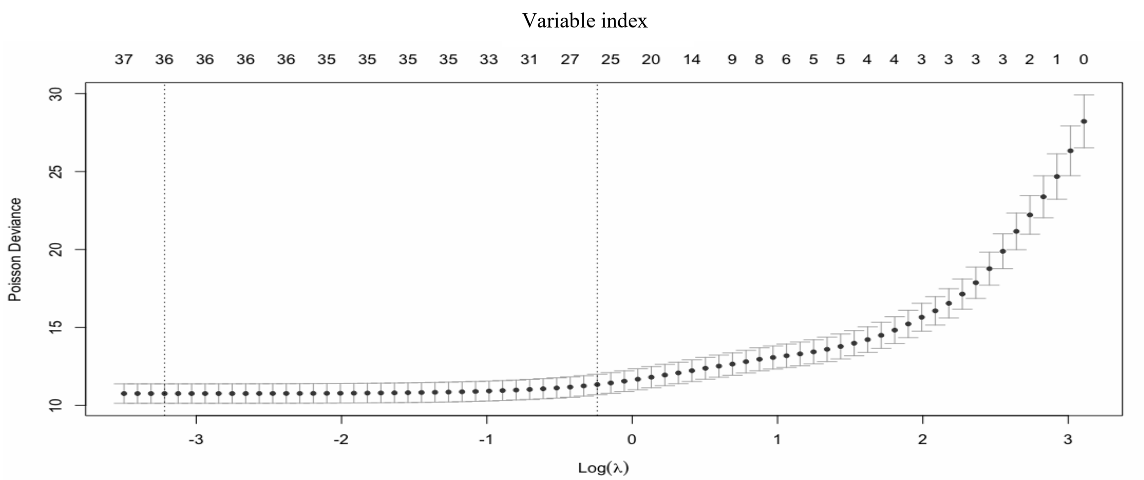


Figure 19. 10-fold cross validation of LASSO for Dataset 2

LASSO Poisson regression models provided insights on excluding the presence of stop sign for building models for Dataset 1 (all intersections) and excluding the presence of crosswalk warning signs for building models for Dataset 2 (intersections with 10+ walking bouts).

Figure 18 provides insights on the specific variables that hold weight (significance) in the model while Figure 19 provides insight on the total number of variables that minimize deviation in the model. Figure 19 shows that the model could include 36 variables with stable behavior, while Figure 18 helps to identify those variables for inclusion. The top x-axis represents independent variable index in both Figures.

Summary of variable selection

Two LASSO regression models were used for variable selection, one for Dataset 1 and one for Dataset 2. Based on these models, the presence of a stop sign variable for Dataset 1 and crosswalk warning sign variable for Dataset 2 were excluded. A correlation analysis was also conducted to confirm the findings from the variable selection of the LASSO regression described previously.

This page is intentionally left blank

Results

Analytical models

There were two statistical models computed to examine the likelihood of walking and likelihood of higher frequency of walking. A zero-inflated negative binomial model was used for Dataset 1 (all intersections). A negative binomial model was used for Dataset 2 (intersections with 10+ walking bouts).

Finalizing the zero-inflated negative binomial model

ZINB regression was used to model count variables with excessive zeros. A zero-inflated model assumes that a zero outcome is due to two different processes (Minami et al., 2007). In the presence of walking bouts, the two processes include (1) an intersection with walking bouts compared to (2) an intersection without walking bouts. If the intersection is without walking bouts, the only outcome possible is zero. If the intersection has walking bouts, it is then a count process.

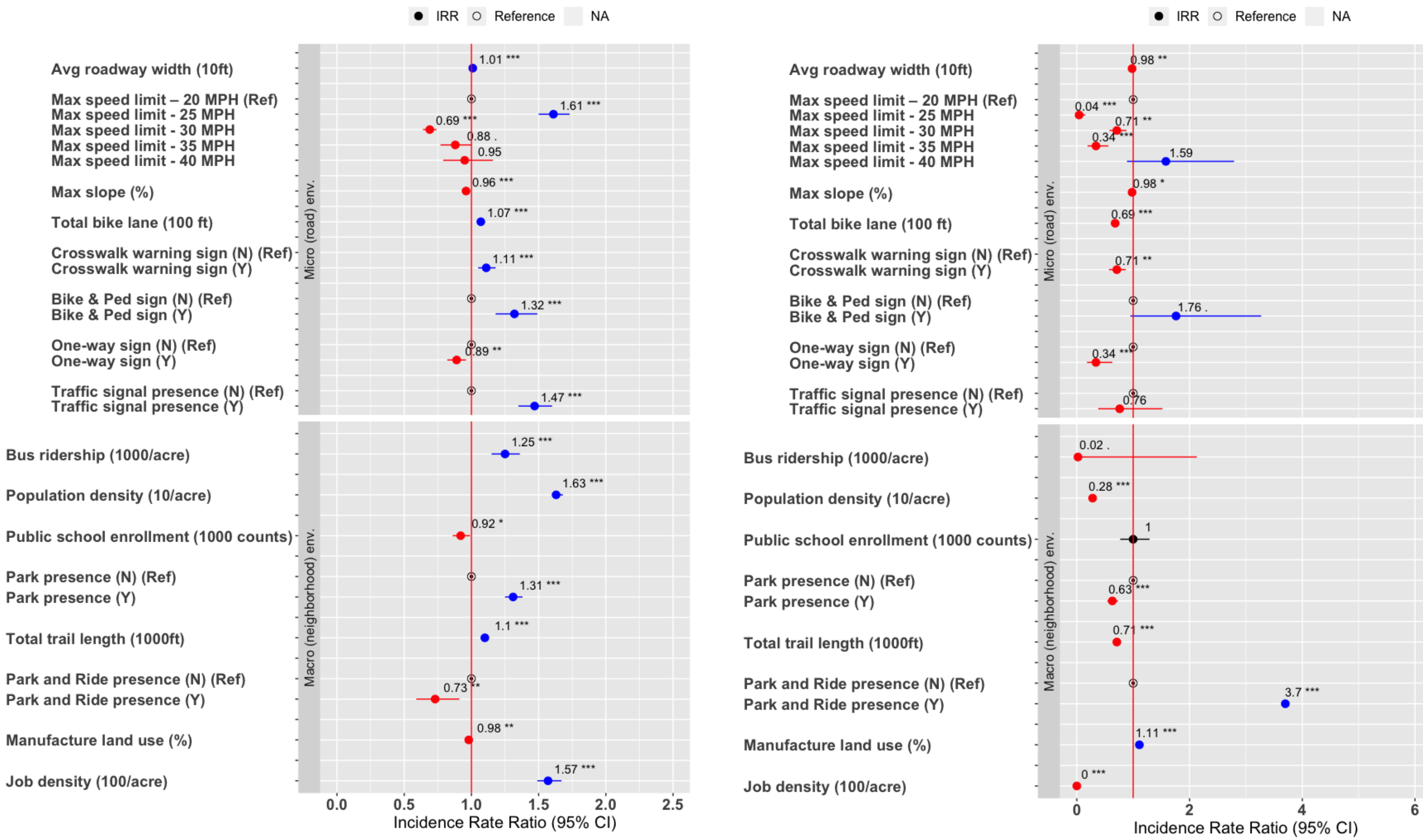
For the forthcoming model, the two parts of the zero-inflated model are a binary logit model (likelihood of zero walking bouts) and a negative binomial count model, to predict the likelihood of increasing number of walking bouts. Interpretation of the ZINB model results require understanding the relationship between the two sub-model. Reviewing only one portion of the ZINB model is insufficient to capture the impact of a predictor, both portions (zero-inflated, negative binomial) should be interpreted jointly. It is also important to remember that the binary logit distribution (ZI) only provides insight on the variation in the *presence* of observed walking (walking bout count does not equal 0). The count distribution (NB) then assesses variation in the *frequency* of observed walking (given walking bout count does not equal 0). The correlation analysis (and a variance inflation factor check) showed that four variables (Median household income, Job density, Service land use, and White population) were highly correlated.

To minimize multicollinearity, four separate ZINBs were created and compared. The dependent variable for all four models was walking bout counts. The model with "Job density" had the best model fit with the lowest AIC (Akaike information criterion – a measure of model fit with smaller values indicating better model performance) value (see Appendix F for detailed results). Using the model with job density as a predictor, estimates of the incident risk ratio (IRR) for the NB portion and odds ratio (OR) for the logistic (zero inflation) portion were obtained.

Dataset 1, all intersections

As shown in Figure 20a (the negative binomial part), the baseline number (model intercept) of walking bouts observed within a 50 m intersection buffer is 2.52 (see Appendix F, Table F 1 – Predictor Job Density) among the intersections which have walking bouts. A unit increase in bike lane length increases the baseline number of walking bouts by 1.07. When compared to a maximum posted speed of 20 mph, a maximum posted speed of 25 mph increased the number of walking bouts by 1.61 folds, whereas a maximum posted speed of 30 mph decreased the number of walking bouts by 0.69 times.

The zero inflated portion (Figure 20b) shows the likelihood that the predictor variable impacts the likelihood of having zero walking bouts at an intersection (dependent variable). The baseline odds of the intersection having no walking bouts is 8.27 (see Appendix F). The baseline odds of the intersection having no walking bouts decreased by 0.71 with the presence of a crosswalk warning sign. Intersections with maximum posted speeds of 25 mph and 35 mph were less likely to have no walking bouts (i.e., the intersections were more likely to have walking bouts) than intersections with a maximum posted speed of 20 mph (by 0.04 and 0.34 respectively). This aligns with the negative binomial portion (Figure 20a) which showed that 25 mph maximum posted speeds were associated with an increase in the number of walking bouts. When a park-and-ride facility is present at an intersection, the odds of the intersection having no walking bout increased by 3.7.



(a) Negative binomial portion of ZINB

(b) Zero-inflated portion of ZINB

Figure 20. ZINB Model result for all intersections (Dataset 1)

Dataset 2, intersections with 10+ walking bouts

A negative binomial model was created using Dataset 2 (Figure 21). The mean (baseline) number of walking bouts (dependent variable) within the 50 m buffer around an intersection is 13.18 (intercept). Presence of stop signs and one-way signs decreased the baseline number of walking bouts by 0.87 and 0.78 times, respectively.

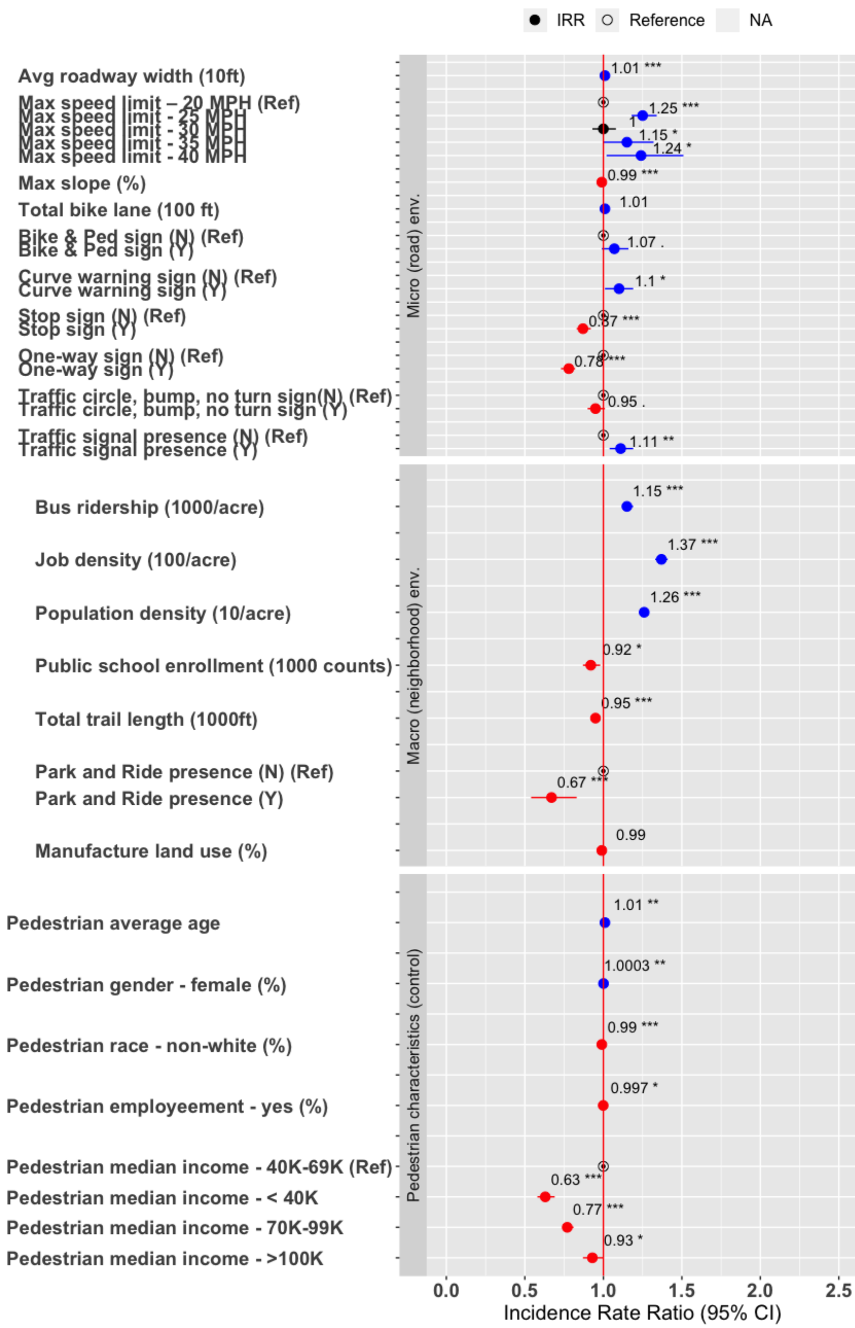


Figure 21. NB model result for intersections with 10+ walking bouts (Dataset 2)

Model validation

Validation entailed examining model stability and model accuracy.

Model stability

For model stability, ACTION Waves 2 and 3 were used. The data used in model development (TRAC Waves 1 and 2, and ACTION Wave 1) was considered the **training** set while the data used for validation (ACTION Waves 2 and 3) was considered the **testing** set. In general, the models performed well during the testing phase in predicting the number of walking bouts at an intersection.

ZINB using Dataset 1 (all intersections). In the training dataset, 26 variables were selected using LASSO regression. Of these, 18 were statistically significant predictors. In the testing dataset, 29 variables were selected for ZINB model inclusion, with 17 variables showing statistical significance. Thirteen of the 18 significant variables in the training set were also significant in the testing set. Four additional variables were selected by the testing model: percentage of non-white pedestrian participants, service land use percentage, median household income, and roadway segment count.

NB using Dataset 2 (intersections with 10+ walking bouts). In the training dataset, 26 variables were selected using LASSO regression. Of these, 22 predictors were statistically significant. In the testing dataset, 21 variables were selected for NB model inclusion, with 15 showing statistical significance. Fifteen of the 22 significant variables in the training set were also significant in the testing set. That is, every significant variable in the testing set was also significant in the training set. There were 3 variables that were *not* selected by the testing model but were selected by the training model: percentage female, job percentage, and presence of a curve sign.

Examples of similarities (Table 2) and differences (Table 3) between the training dataset and the testing dataset are shown below.

Table 2. Similarities in training and testing datasets

Predictor	**Notes**	
	Zero-inflated part	**Negative binomial part**
Population density	Higher population density → Lower probability of zero walking bout counts	Higher population density → Higher number of walking bout counts
Traffic signals	With signals → Lower probability of zero walking bout counts	With signals → Higher number of walking bout counts
Max roadway slope	Steeper slope → Higher probability of zero walking bout counts	Steeper slope → Lower number of walking bout counts

Table 3. Differences in training and testing datasets

Predictor	Notes
Bus ridership per acre	Significant in testing model only. This can be due to the following: • The training data (mostly TRAC) was from a light rail study • The testing data (ACTION) was from a rapid bus study
Average roadway width	Significant in testing model only, but coefficients are comparable
Maximum speed limit – 35 mph	Only significant in the training model

Additional model stability – Expected model variation

Data visualization showed that TRAC and ACTION data differed geographically. However, this variation offered a unique opportunity to examine model generalizability, which has been showcased in the validation effort. To better understand if the testing model can be used to validate the training model's variable selection, a subset of TRAC only data (a benchmark dataset) was selected to further estimate the expected model variation.

For the ZINB model, 26 variables from the training dataset were selected by LASSO for ZINB (Dataset 1) model inclusion. Of these variables, 18 showed statistical significance. In the benchmark dataset (TRAC only), 25 variables were selected for ZINB model inclusion, with 17 showing statistical significance. The benchmark model showed significant similarities to the testing set. This suggests that the pedestrian exposure model (using the TRAC data) is sensitive to the micro-environment and does reflect the high prevalence of walking bouts around downtown. There was one variable, total sidewalk length, that the benchmark model did not show as significant. However, this variable was significant ($p = 0.047$) in the training model, showcasing the natural variation that can appear due to variations in data collection.

For the NB model, 26 variables from the training dataset were selected by LASSO, with 22 of them showing statistical significance. In the benchmark dataset (TRAC only), 27 variables were selected by LASSO, 23 of these were statistically significant. The significant predictors in the NB model were identical between the training and benchmark datasets, with one exception: roadway segment count. This variable was significant in the benchmark model but was *not* selected for inclusion in the training NB model.

Model accuracy

The testing data was used in the trained pedestrian exposure model to assess the model's predictive performance. Mean squared error and mean absolute error were examined for the negative binomial model (Dataset 2). These metrics were not computed for the ZINB model (Dataset 1) because of the high prevalence of zeros. The MSE and MAE from the NB training model were compared with those from the NB testing model. The training dataset had 108,658 walking bouts while the testing dataset had 52,360 walking bouts.

MSE is a measure of the average squared difference between the estimated and actual values of the response variable. It is calculated by taking the sum of the squared differences between the predicted and actual values, divided by the number of observations. The lower the MSE, the better the model performance, as it indicates a smaller average difference between the predicted and actual values. The training dataset MSE is 2532.32, which means that on average, the squared difference between the predicted and actual values of the response variable is around 2532.32. The adjusted testing dataset MSE is 2295.29, which suggests that the model performance is slightly better on the testing dataset than on the training dataset.

MAE is another measure of the error between paired observations, and it is calculated by taking the absolute difference between the predicted and actual values, and then averaging across all observations. The lower the MAE, the better the model performance, as it indicates a smaller average difference between the predicted and actual values. The training dataset MAE is 32.32, which means that on average, the absolute difference between the predicted and actual values of the response variable is around 32.32. The adjusted testing dataset MAE is 28.95, which suggests that the model performance is slightly better on the testing dataset than on the training dataset, and the model can generalize well to new data. Overall, these metrics suggest that the negative binomial model was a reasonable fit to the data and can predict the count of walking bouts at an intersection.

Discussion

A ZINB model was developed to examine the likelihood of walking bouts at the intersection level. An NB model was also created that included intersections where 10+ walking bouts were recorded.

The number of walking bouts was associated with several variables. The ZINB showed that several variables increased the number of walking bouts, including bike lane length, presence of crosswalk warning sign, presence of bike and pedestrian sign, presence of traffic signal, bus ridership density, population density, park presence, trail length, and job density. Variables that were associated with a decrease in walking bouts included maximum roadway slope, one-way sign, and presence of park-and-ride facility. The model results align with expectations on the impact of variables on the number of walking bouts. For example, it is reasonable to see why pedestrians may opt for a different or longer route to avoid roads with steep inclines (high maximum slope percentage). This inverse relationship has been supported in past studies (Kang, 2017; Meeder, 2017).

The NB model highlighted the association of variables on pedestrian exposure at intersections with walking bouts, key components among them being presence of stop sign and trail length. (As a reminder, stop sign presence was not included in Dataset 1 (all intersections) because it was not chosen during variable selection.) Stop sign presence and trail length were negatively associated with more than 10 walking bouts. This can be translated into the finding that a large number (10+) of walking bouts is less likely to occur at an intersection with stop signs. From a practical perspective, these findings indicate that the stop signs are placed at intersections that require more road user interaction (e.g., residential areas are highly controlled). As trails are more likely located outside of crowded urban areas, they were a good indicator of presence of walking bouts but not for a larger number (10+) of walking bouts.

Exploratory machine learning approaches, pedestrian density estimation

To explore the potential offered by using machine learning (ML) models rather than conventional statistical models, three additional model types were fit to predict walking bouts.

1. **Negative binomial model**: used as a reference, representing the conventional statistical modeling, as discussed previously.
2. **Random forest model**: used to represent the flexible and simple modeling approach, not affected by linearity assumptions.
3. **Gradient-boosted tree model**: retains some of random forest's flexibility with less risk of overfitting by training models on subsets of variables rather than all variables and combining results from these models in principled ways.
4. **Random forest model** with latitude and longitude withheld: Initial results from the naive random forest model (model #2) indicated that, because the dataset included the latitude and longitude of each intersection and because spatial autocorrelation of walking bouts is high, the model was primarily picking up on spatial autocorrelation. The random forest model corrected for the risk of autocorrelation artifacts by withholding intersection coordinates from the random forest model.

For each model, a 10-fold cross-validation was used to fit and tune each model on a training set consisting of 50% of the intersections. The resulting model was then applied to a test set of the remaining 50% of the intersections to estimate model fit and error.

The primary concern in this exploration was to explore the tradeoff between minimizing random error at the cost of potentially introducing systematic error due to overfitting bias. The models were measured using conventional fit metrics: root mean squared error (RMSE), mean absolute error (MAE), and R^2 (Table 4). Plotted maps of predicted walking bouts and spatial error patterns for each model were useful to look for patterns that might indicate systematic error (Figure 22 and Figure 23).

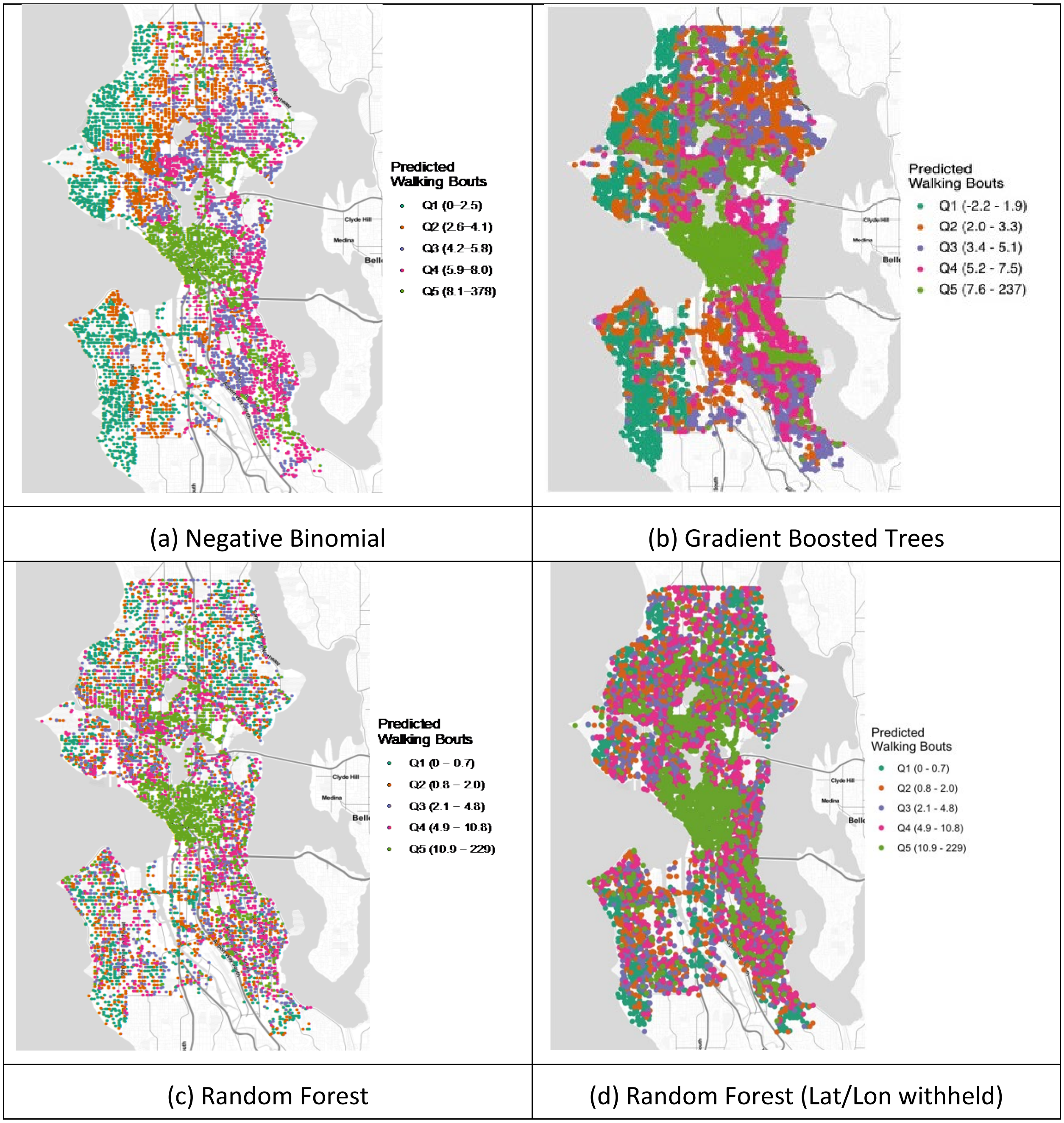
From this investigation, it was concluded that a naive use of machine learning is at some risk of overfitting bias. In particular, the finer scale of differences in predictions of the random forest model (Figure 22) is both consistent with random forest making better use of the available covariate set and consistent with random forest making use of spatial autocorrelation intrinsic in walking bouts (i.e., because one walking bout typically visits two or more adjacent intersections, intersections adjacent to one with a walking bout are more likely to have one or more walking bouts, on average). The gradient boosted model is flexible to counteract the "random forest" issue with overfitting through regularization, early stopping, and cross-validation techniques. However, the RMSE for GBM is higher and the R^2 is lower than the random forest counterparts. To maintain the low error from random forest and avoid overfitting, a random forest without the latitude and longitude of the data was also investigated. The research team intended this would scale model generalizability, meaning these results would be applicable across space. Developing this model requires highly educated tuning and validation to ensure that the model is not overfitting to the training data. The ML approach is not recommended when considering demographic attributes as predictors – as results of this overfitted model can lead to harmful and inaccurate interpretations.

In summary, this approach was promising but would require a deeper dive into ML details and potentially a dataset collected from another cohort to fully validate the approach. Because of its greater interpretability and lower potential for systematic bias due to autocorrelation, the research team moved forward with the negative binomial model.

Table 4. Model fit metrics for several modeling choices

Model	RMSE	MAE	R^2
Negative Binomial (Reference)	15.7	7.3	0.56
Gradient Boosted Model	11.0	5.8	0.77
Random Forest	8.5	3.9	0.87
Random Forest (without: LAT/LONG)	9.9	4.9	0.81

*The gradient boosted model is a composite model that ultimately combines many less effective models to reduce MSE of the overall model.



(a) Negative Binomial

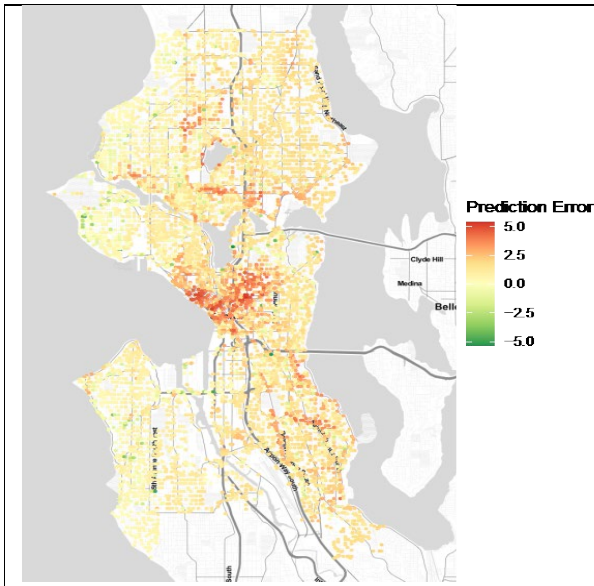
(b) Gradient Boosted Trees

(c) Random Forest

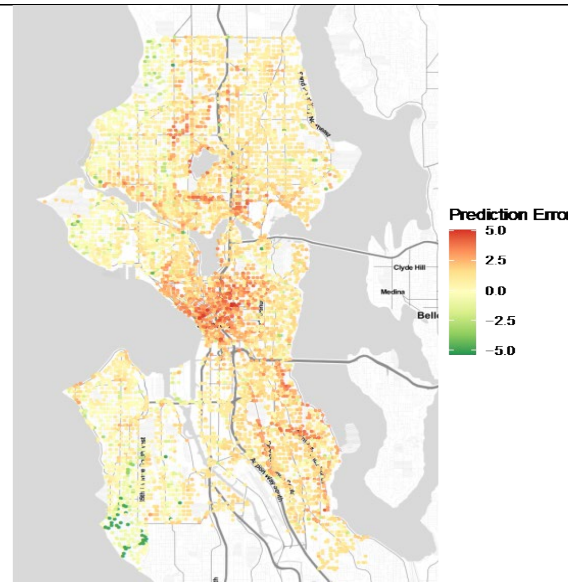
(d) Random Forest (Lat/Lon withheld)

Figure 22. Predicted number of walking bouts using Dataset 1 for (a) an initial negative binomial, (b) gradient boost model, (c) random forest, and (d) random forest without latitude and longitude.

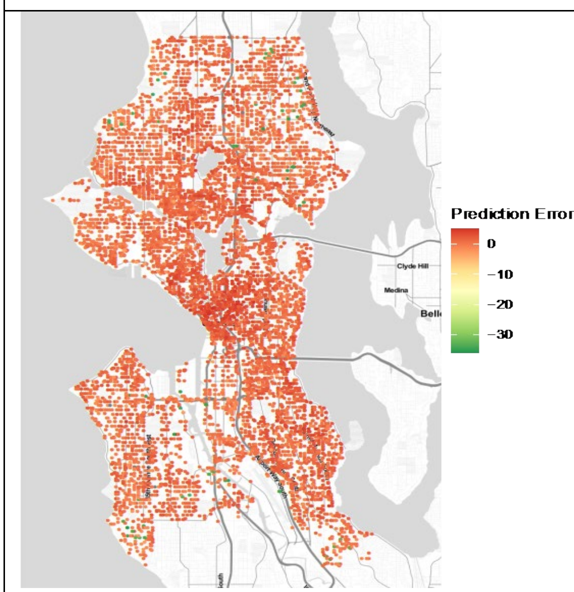
The number of walking bouts are color coded based on interquartile range



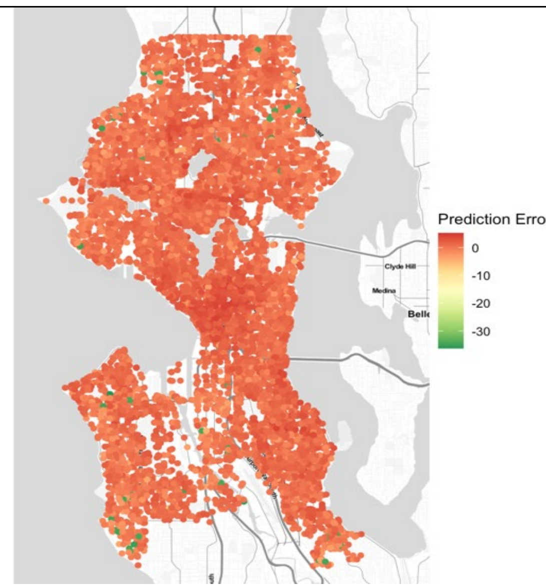
(a) Negative Binomial



(b) Gradient Boosted Trees



(c) Random Forest



(d) Random Forest (Lat/Lon withheld)

Figure 23. Log Prediction Error by modeling type using Dataset 1 for (a) an initial negative binomial, (b) gradient boost model, (c) random forest, and (d) random forest without latitude and longitude

This page is intentionally left blank

Conclusions

Pedestrian exposure was defined in this study as the period when a pedestrian is at risk of being struck and injured by a vehicle. The study developed a framework that defines pedestrian exposure at the intersection-level using macro- and micro-environmental factors to predict pedestrian exposure. Intersection-level findings can be used to assess the potential effect of intersection-level countermeasures such as pedestrian crosswalk designs, sidewalk length, and curb extensions. Additionally, measures with this specificity can be aggregated to larger scales to address area-based research questions.

The study used a pedestrian trip-based measure, walking bouts, to help operationally define and model pedestrian exposure. GPS and accelerometer data were used to transform people's walking bouts into intersection-level data. Traffic-related environmental attributes were also assessed at the intersection-level. Data were captured within buffer regions around each intersection, sized at 50 m and 400 m to define the micro- and macro-environment around each intersection. Buffering the intersections offered flexibility to associate walking bouts with attributes of the surrounding built environment to model pedestrian exposure.

Separate analytic models were developed for two different datasets: the first comprised all intersections in Seattle regardless of whether a walking bout was observed, and the second analyzed intersections where walking bout counts greater than 10 were observed. The first dataset examined the likelihood of observing any amount of walking, while the second analyzed the environmental factors that impact the frequency of observed walking.

The study identified existing sources of pedestrian data, developed an operational definition of pedestrian exposure, measured, and examined pedestrian exposure, and developed analytical models to assess exposure in King County. As highlighted in this report, the number of walking bouts was positively associated with bike lane length, presence of crosswalk warning sign, presence of bike and pedestrian signs, presence of traffic signal, bus ridership density, population density, park presence, trail length, and job density in a ZINB model. Variables that were negatively associated with the number of walking bouts included maximum roadway slope, presence of a one-way road sign, and presence of park-and-ride facility. These model results align with the research team's hypotheses and previous literature. For example, areas around park-and-rides may be expected to have less walking activity because the function of the facility is to drive there, and then transition to another mode of (moving) transit. It was observed that presence of park-and-rides had a positive association with the presence of walking bouts, however a negative association with the frequency of walking bouts. This is expected, as walking should occur (people moving from vehicle to other transit mode), however people may not walk enough to hit the minimum walking bout threshold limits (continuous walking for ~7 minutes). This pedestrian exposure model is specific enough to capture these nuances: walking is present around park-and-rides, but most likely not at a high enough frequency to accumulate significant exposure.

Future research should seek to validate the models in other cities, where conditions (topography, weather, development densities, policies, transit systems, roadway design, etc.) are different from those of Seattle/King County. For this framework to be successfully used in other cities, a sampling scheme will need to be identified and established before data collection begins. The micro- and macro-environmental predictors introduced in this project can serve to quantify pedestrian exposure in other cities or regions. It will be important to test the extent to which predictors remain significant and retain strength across cities and regions. Extending the analyses to other localities will also help identify whether the data are broadly available or whether other proxy variables will need to be used for nationwide applications.

The relationship between pedestrian exposure and safety is another important avenue for future research. Increases in pedestrian exposure are only desirable if they are not correlated with increases in the number of crashes or the number of injuries and fatalities. An improved understanding of whether the predictors of pedestrian exposure may be correlated or may interact with crash occurrence and outcomes is warranted. In particular, DOTs and State policy officials should work with researchers to ensure exposure does not lead to additional individual risk.

Overall, the framework developed during this project can be generalized and adaptable to variables on different spatial scales, offering flexibility for carrying out additional exposure analyses and for including alternative environmental factors of interest.

References

Amoh-Gyimah, R., Saberi, M., & Sarvi, M. (2016). Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. *Accident Analysis & Prevention, 93*, 147–159. https://doi.org/10.1016/j.aap.2016.05.023

Cai, Q., Lee, J., Eluru, N., & Abdel-Aty, M. (2016). Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident Analysis & Prevention, 93*, 14–22. https://doi.org/10.1016/j.aap.2016.04.015

Carr, L. J., & Mahar, M. T. (2012). Accuracy of intensity and inclinometer output of three activity monitors for identification of sedentary behavior and light-intensity activity. *Journal of Obesity, 2012*, 1–9. https://doi.org/10.1155/2012/460271

Chu, X. (2003, May 1-3). *The fatality risk of walking in America: A time-based comparative approach.* Walk21 Conference: Health Equity and the Environment, Portland, Oregon.

City of Seattle. (2017). *VisionZero 2017 progress report*. Seattle Department of Transportation. www.seattle.gov/Documents/Departments/beSuperSafe/VZ_2017_Progress_Report.pdf

Duncan, D. T., Aldstadt, J., Whalen, J., Melly, S. J., & Gortmaker, S. L. (2011). Validation of Walk Score for estimating neighborhood walkability: An analysis of four US metropolitan areas. *International Journal of Environmental Research and Public Health, 8*, 4160–4179. https://doi.org/10.3390/ijerph8114160

Federal Highway Administration. (2017). *Traffic monitoring guide (TMG): Chapter 4: Traffic monitoring for non-motorized traffic* (Report No. FHWA-PL-17-003). www.fhwa.dot.gov/policyinformation/tmguide/tmg_fhwa_pl_17_003.pdf

Frank, L. D., Kuntz, J. L., Chapman, J. E., Fox, E. H., Dickerson, J. F., Meenan, R. T., Berrigan, D., & Fortmann, S. P. (2019). The health and economic effects of light rail lines: Design, methods, and protocol for a natural experiment. *BMC Public Health, 19*, 200. https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-6518-6

Greene-Roesel, R., Diogenes, M. C., & Ragland, D. R. (2007, March 1). *Estimating pedestrian accident exposure*. University of California Traffic Safety Center. https://escholarship.org/content/qt8j8685jt/qt8j8685jt.pdf?t=krnbrs

Griswold, J., Medury, A., Huang, L., Amos, D., Lu, J., Schneider, R., & Grembeck, O. (2018, January). *Pedestrian safety improvement program: Phase 2* (Report No. CA18-2452). California Department of Transportation. https://rosap.ntl.bts.gov/view/dot/43656

Haddak, M. M. (2016). Exposure-based road traffic fatality rates by mode of travel in France. *Transportation Research Procedia, 14*, 2025–2034. https://doi.org/10.1016/j.trpro.2016.05.170

Hankey, S., & Lindsey, G. (2016). Facility-demand models of peak period pedestrian and bicycle traffic: Comparison of fully specified and reduced-form models. *Transportation Research Record, 2586*(1), 48–58.

Hong, J., Shankar, V. N., & Venkataraman, N. (2016). A spatially autoregressive and heteroskedastic space-time pedestrian exposure modeling framework with spatial lags and endogenous network topologies. *Analytic Methods in Accident Research, 10*, 26–46.

Jamali, A., & Wang, Y. (2017). Estimating pedestrian exposure for small urban and rural areas. *Transportation Research Record, 2661*(1), 84–94.

Jerrett, M., Su, J. G., MacLeod, K. E., Hanning, C., Houston, D., & Wolch, J. (2016). Safe routes to play? Pedestrian and bicyclist crashes near parks in Los Angeles. *Environmental Research, 151*, 742–755. https://doi.org/10.1016/j.envres.2016.07.029

Jones, M., Goldsmith, S., Litman, T., & Institute of Transportation Engineers Pedestrian and Bicycle Council. (2009). *National bicycle and pedestrian documentation project description*. National Bicycle and Pedestrian Documentation Project. http://www.bikepeddocumentation.org/application/files/2014/6671/8017/NBPD_Description_2009.pdf

Kang, B., Moudon, A. V., Hurvitz, P. M., Reichley, L., & Sealens, B. E. (2013). Walking objectively measured: Classifying accelerometer data with GPS and travel diaries. *Medicine and Science in Sports and Exercise, 45*(7), 1419–1428. https://pubmed.ncbi.nlm.nih.gov/23439414/

Kang, B., Moudon, A. V., Hurvitz, P. M., & Saelens, B. E. (2017). Differences in behavior, time, location, and built environment between objectively measured utilitarian and recreational walking. *Transportation Research Part D: Transport and Environment*, *57*, 185–194. https://doi.org/10.1016/j.trd.2017.09.026

Kang, M., Moudon, A. V., Kim, H., & Boyle, L. N. (2019). Intersections and non-intersections: A protocol for identifying pedestrian crash risk locations in GIS. *International Journal of Environmental Research and Public Health, 16*(19), 3565. https://pubmed.ncbi.nlm.nih.gov/23439414/

Krenn, P. J., Titze, S., Oja, P., Jones, A., & Ogilvie, D. (2011). Use of global positioning systems to study physical activity and the environment. *American Journal of Preventive Medicine, 41*, 508–515. https://doi.org/10.1016/j.amepre.2011.06.046

Lam, W. W. Y., Loo, B. P. Y., & Yao, S. (2013). Towards exposure-based time-space pedestrian crash analysis in facing the challenges of aging societies in Asia. *Asian Geographer, 30*(2), 105–125.

Lam, W. W. Y., Yao, S., & Loo, B. P. Y. (2014). Pedestrian exposure measures: A time-space framework. *Travel Behaviour and Society, 1*(1), 22–30.

Lassarre, S., Papadimitriou, E., Yannis, G., & Golias, J. (2007). Measuring accident risk exposure for pedestrians in different micro-environments. *Accident Analysis & Prevention, 39*(6), 1226–1238.

Louch, H., Davis, B., Voros, K., O'Toole, K., Piper, S. (2016). *Innovation in bicycle and pedestrian counts: A review of emerging technology.* Alta Planning + Design. https://altago.com/wp-content/uploads/Innovative-Ped-and-Bike-Counts-White-Paper-Alta.pdf

Loukaitou-Sideris, A., Liggett, R., & Sung, H.-G. (2007). Death on the crosswalk. *Journal of Planning Education and Research, 26*, 338–351. https://doi.org/10.1177/0739456x06297008

McAndrews, C., Beyer, K., Guse, C. E., & Layde, P. (2013). Revisiting exposure: Fatal and non-fatal traffic injury risk across different populations of travelers in Wisconsin, 2001–2009. *Accident Analysis & Prevention, 60*, 103–112.

Meeder, M., Aebi, T., & Weidmann, U. (2017). The influence of slope on walking activity and the pedestrian modal share. *Transportation Research Procedia*, *27*, 141–147. https://doi.org/10.1016/j.trpro.2017.12.095

Miller, H. J., Tribby, C. P., Brown, B. B., Smith, K. R., & Werner, C. M. (2015). Public transit generates new physical activity: Evidence from individual GPS and accelerometer data before and after light rail construction in a neighborhood. *Health & Place, 36*, 151–158. www.ncbi.nlm.nih.gov/pmc/articles/PMC4679466/

Minami, M., Lennert-Cody, C. E., Gao, W., & Román-Verdesoto, M. (2007). Modeling shark bycatch: The zero-inflated negative binomial regression model with smoothing. *Fisheries Research, 84*(2), 210–221. https://doi.org/10.1016/j.fishres.2006.10.019

Mooney, S. J., Sheehan, D. M., Zulaika, G., Rundle, A. G., McGill, K., Behrooz, M. R., & Lovasi, G. S. (2016). Quantifying distance overestimation from global positioning system in urban spaces. *American Journal of Public Health, 106*(4), 651–653. www.ncbi.nlm.nih.gov/pmc/articles/PMC4815998/

Moudon, A. V., Lin, L., Jiao, J., Hurvitz, P., & Reeves, P. (2011). The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. *Accident Analysis & Prevention, 43*(1), 11-24. https://doi.org/10.1016/j.aap.2009.12.008

National Center for Statistics and Analysis. (2024). *Quick facts 2022* (Report No. DOT HS 813 563). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813563

Nordback, K., Kothuri, S., Petritsch, T., McLeod, P., Rose, E., Twaddell, H., & Sprinkle Consulting, Inc. (2016). *Exploring pedestrian counting procedures* (Report No. FHWA-HPL-16-026). Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/64905

Osama, A., & Sayed, T. (2017). Evaluating the impact of connectivity, continuity, and topography of sidewalk networks on pedestrian safety. *Accident Analysis & Prevention, 107*, 117-125. https://doi.org/10.1016/j.aap.2017.08.001

Papadimitriou, E., Yannis, G., & Golias, J. (2012). Analysis of pedestrian exposure to risk in relation to crossing behavior. *Transportation Research Record, 2299*(1), 79-90. https://doi.org/10.3141/2299-09

Qin, X., & Ivan, J. N. (2001). Estimating pedestrian exposure prediction model in rural areas. *Transportation Research Record, 1773*(1), 89-96. https://doi.org/10.3141/1773-11

Qu, T., Gates, T. J., Xu, C., Seguin, D., & Kay, J. (2022). The disparate impact of COVID-19 pandemic on walking and biking behaviors. *Transportation Research Part D: Transport and Environment, 112*, 103494. www.ncbi.nlm.nih.gov/pmc/articles/PMC9574946/

Quistberg, D. A., Quistberg, A., Howard, E. J., Hurvitz, P. M., Moudon, A. V., Ebel, B. E., & Saelens, B. E. (2017). The relationship between objectively measured walking and risk of pedestrian-motor vehicle collision. *American Journal of Epidemiology, 185*(9), 810-821. www.ncbi.nlm.nih.gov/pmc/articles/PMC5411678/

Quistberg, D. A., Howard, E. J., Ebel, B. E., Moudon, A. V., Saelens, B. E., Hurvitz, P. M., & Rivara, F. P. (2015). Multilevel models for evaluating the risk of pedestrian–motor vehicle collisions at intersections and mid-blocks. *Accident Analysis & Prevention, 84*, 99–111. https://pubmed.ncbi.nlm.nih.gov/26339944/

Raford, N., & Ragland, D. (2004). Space syntax: Innovative pedestrian volume modeling tool for pedestrian safety. *Transportation Research Record: Journal of the Transportation Research Board, 1878*, 66–74. https://doi.org/10.3141/1878-09

Ryus, P., Ferguson, E., Laustsen, K. M., Schneider, R. J., Proulx, F. R., Hull, T., & Miranda-Moreno, L. (2014). *Guidebook on pedestrian and bicycle volume data collection*. (NCHRP Report 797). Transportation Research Board. https://escholarship.org/content/qt11q5p33w/qt11q5p33w.pdf

Saelens, B. E., Moudon, A. V., Kang, B., Hurvitz, P. M., & Zhou, C. (2014). Relation between higher physical activity and public transit use. *American Journal of Public Health, 104*(5), 854-859. https://doi.org/10.2105/AJPH.2013.301696

Schneider, R. J., Henry, T., Mitman, M. F., Stonehill, L., & Koehler, J. (2012). Development and application of volume model for pedestrian intersections in San Francisco, California. *Transportation Research Record, 2299*(1), 65–78. https://doi.org/10.3141/2299-08

Schneider, R. J., Henry, T., Mitman, M. F., Stonehill, L., & Koehler, J. (2013, March 1). *Development and application of the San Francisco pedestrian intersection volume model* (Report No. WP-2013-4). Safe Transportation Research and Education Center [University of California, Berkeley]. https://escholarship.org/uc/item/8cs2g40c

Tian, G., & Ewing, R. (2017). A walk trip generation model for Portland, OR. *Transportation Research Part D: Transport and Environment, 52*, 340–353. https://doi.org/10.1016/j.trd.2017.03.017

Tobey, H. N., Shunamen, E. M., & Knoblauch, R. L. (1983). *Pedestrian trip making characteristics and exposure measures* (Report No. FHWA/RD-83/062). Federal Highway Administration.

Turner, S., Sener, I. N., Martin, M. E., Das, S., Hampshire, R. C., Fitzpatrick, K., Molnar, L., Wijesundera, R. K., Colety, M., & Robinson, S. (2017). *Synthesis of methods for estimating pedestrian and bicyclist exposure to risk at areawide levels and on specific transportation facilities* (Report No. FHWA-SA-17-041). Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/36098

U.S. Department of Transportation. (2022, January). *National roadway safety strategy.* www.transportation.gov/NRSS

Vision Zero Miami. (2018). *What is Vision Zero?* [Web page]. www.miamidade.gov/global/initiatives/visionzero/home.page

Vision Zero Network. (2021, December 7). Vision Zero Network [Web page]. https://visionzeronetwork.org/

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention, 60*, 71–84. https://doi.org/10.1016/j.aap.2013.07.030

Washington State Department of Transportation. (2017). *2016 student travel survey state report.* www.seattle.gov/documents/Departments/SDOT/SRTS/2016%20Student%20Travel%20Survey%20Report.pdf

Wu, J., Jiang, C., Liu, Z., Houston, D., Jaimes, G., & McConnell, R. (2010). Performances of different global positioning system devices for time-location tracking in air pollution epidemiological studies. *Environmental Health Insights, 4*, EHI.S6246. https://doi.org/10.4137/ehi.s6246

Wyker, B., Bartley, K., Holder-Hayes, E., & Immerwahr, S. (2013, April). *Self-reported and accelerometer-measured physical activity: A comparison in New York City*. New York City Department of Health and Mental Hygiene. Epi Research Report, 2013(April), 1-12.

Yao, S., Loo, B. P. Y., & Lam, W. W. Y. (2015). Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: Space-time path vs. potential path tree methods. *Accident Analysis & Prevention,* 75, 320–332. www.sciencedirect.com/science/article/abs/pii/S0001457514003844

Zegeer, C. V., Richard Stewart, J., Huang, H., & Lagerwey, P. (2001). Safety effects of marked versus unmarked crosswalks at uncontrolled locations: Analysis of pedestrian crashes in 30 cities. *Transportation Research Record, 1773*(1), 56–68.

This page is intentionally left blank

Appendix A. Literature review on pedestrian exposure

Table A 1. Example studies with area-based pedestrian exposure

Exposure Type	Reference	Outcome	Location	Spatial unit	Area	Data source(s)	Statistical model	Significance at α= .05
Density (e.g., population, employment, residents)	Amoh-Gyimah et al. (2016)	Pedestrian and bicycle crash counts	Melbourne, AU	Census tract	City	AU 2011 population and housing census data	n/a	n/a
	Cai et al. (2016)	Pedestrian crash counts	FL	Transportation Analysis Zone	State	Florida DOT	n/a	n/a
	Quistberg et al. (2015)	Pedestrian–motor vehicle collisions	Seattle, WA	Crash location	City	● King County Dept of Assessment (KCAOS) Property Parcel 2007 and 2010 ● US Bureau of Labor Statistics 2007 and 2010	n/a	n/a
	Moudon et al. (2011)	Pedestrian injury	King County, WA	Crash location	County	● King County Assessor's Office ● UW Urban Form Lab (Moudon & Sohn, 2005)	n/a	n/a
	Loukaitou-Sideris et al. (2007)	Pedestrian-automobile collision density	Los Angeles, CA	Census tract	City	● US Census 2000 ● Census Transportation Planning Package (CTPP) 2000	n/a	n/a
Walking distance	Haddak (2016)	Fatality rates	France	Individual trips	Country	France National Travel Survey 2007-2008 (6 waves)	n/a	n/a
	Wang & Kockelman (2013)	Pedestrian exposure modeling	Austin, TX	Individual trips	City	Austin Travel Survey (2005/2006); 218 zones in Austin, TX	Weighted least squares regression	● Population counts ● Sidewalk lengths ● Zone size ● Lane-miles by road class

Exposure Type	Reference	Outcome	Location	Spatial unit	Area	Data source(s)	Statistical model	Significance at α= .05
	McAndrews et al. (2013)	Pedestrian injury	WI	Individual trips	State	● WI Add-On Sample to the 2001 NHTS ● WI Dept of Health Services (for population data)	n/a	n/a
Walking duration	Hong et al. (2016)	Pedestrian exposure modeling	Seattle, WA	Crosswalk	City	Microsimulations using VISSIM	n/a	n/a
	Haddak (2016)	Fatality rates	France	Individual trips	Nation	France National Travel Survey 2007-2008 in 6 waves	n/a	n/a
	McAndrews et al. (2013)	Pedestrian injury	WI	Individual trips	State	● WI Add-On Sample to the 2001 NHTS ● WI Dept of Health Services (for population data)	n/a	n/a
	Chu (2003)	Fatality risk of walking	US	Individual trips	Country	2001 NHTS	n/a	n/a
Number of trips	Jamali & Wang (2017)	Pedestrian exposure	US	Individual trips	Country	Household-level NHTS 2009 data in rural and small urban areas	Negative binomial regression	● Household size (+) ● Number of adults (+) ● Number of workers (-) ● Population density of the block group (+) ● Car ownership (-) ● Higher-income (+)
	Osama & Sayed (2017)	Number of pedestrian crashes	Vancouver, CA	Transportation Analysis Zone	City	● Census Canada 2011 ● TransLink Household travel survey 2011	n/a	n/a

Exposure Type	Reference	Outcome	Location	Spatial unit	Area	Data source(s)	Statistical model	Significance at α= .05
	Tian & Ewing (2017)	Home-based walk trip generation model	Portland, OR	Individual trips	City	Oregon Household Travel and Activity Survey of 2011	Negative binomial regression	● Land-use entropy within 0.5 mile ● Sidewalk quality within 0.25 mile ● Traffic calming prevalence within 0.5 mile ● Transit stop density within 0.5 mile ● Household size ● No. of children in the household
	Jerrett et al. (2016)	Pedestrian injuries	Southern CA	Census tract	Region	2001 Travel and Congestion Survey by the Southern California Association of Government	n/a	n/a
	Haddak (2016)	Fatality rates	France	Individual trips	Country	France National Travel Survey 2007-2008 in 6 waves	n/a	n/a
	McAndrews et al. (2013)	Pedestrian injury	WI	Individual trips	State	● WI Add-On Sample to the 2001 NHTS ● WI Dept of Health Services (for population data)	n/a	n/a

Table A 2. Area density for pedestrian exposure – Advantages, disadvantages, and possible improvements

Area density example	Data sources	Advantages	Disadvantages	Possible Improvements using GPS and accelerometer data
• Population density • Residential density • Employment density	• U.S. Census Bureau (publicly available) • Regional and local government	• Readily available for several geographies and time periods (Greene-Roesel et al., 2007) • Available to compare trends over time	• Proxy for pedestrian volume • High-level approximation of exposure • Predefined by spatial units (e.g., Transportation Analysis Zone [TAZ]) where the population may not distribute evenly, which may lead to erroneous results • Does not account for the variability of individual pedestrian activities such as pedestrians' walking distance and time (Mooney et al., 2016) • An actual number of pedestrians walking may not relate to the number of people that reside or work in the area (Greene-Roesel et al., 2007)	• Individual walking measure rather than area-based data • More representable and accurate of walking activity than proxy measure

Table A 3. Self-reported walking activity for pedestrian exposure – Advantages, disadvantages, and possible improvements

Self-reported walking activity	Data sources	Advantages	Disadvantages	Possible improvements using GPS and accelerometer data
• Frequency of walking • Duration of walking • Distance of walking	• National Household Travel Survey • American Community Survey • National Highway Traffic Safety Administration	• More detailed than pedestrian counts or volumes • Can record the purpose of the walking bout, from the perspective of the participant	• Subjective measure depending on a person's memory • Under-reported activity • Frequency of walking could be underestimated, walking activities such as a short walk to a vehicle or bus stop may not be reported • Duration of walking could be overestimated as people perceive their walking time longer than the actual duration (Chu, 2003)	• Objective measures (e.g., accelerometers and pedometers) • More accurate assessments of physical activity from which specific activities such as walking can be identified

Table A 4. Example studies with point/segment-based pedestrian exposure

Exposure type	Reference	Outcome of interest	Study location	Spatial unit	Area extent	Data sources	Pedestrian exposure model type	Significance at (α= .05) if any
Pedestrian volume	Hankey & Lindsey (2016)	Spatial estimates of pedestrian traffic	Minneapolis, MN	Intersections	City	• Minneapolis Department of Public Works (DPW) • Transit for Livable Communities (TLC)	Linear regression	• Retail area (+) • Open space area (+) • Population density (+) • Transit stop (+)
	Zegeer et al. (2001)	Pedestrian crashes	30 cities across the United States	Crosswalks	Region	• U.S. DOT – FHWA	Adjustment factors by the time of day and area type	n/a
	Raford & Ragland (2003)	Pedestrian volume	Oakland, CA	Intersections and Road segments	City	• CA Office of Traffic Safety through the Business, Transportation, and Housing Agency	Network Analysis Model (called Space Syntax)	n/a
	Qin & Ivan (2001)	Weekly pedestrian volumes	CT	Crosswalks	City	• U.S. DOT through the New England University Transportation Center and the Connecticut Transportation Institute	Generalized linear regression	• Campus area (+) • Tourist and downtown areas (+)
	Tobey et al. (1983)	Pedestrian exposure measure related to crash events	NY, MO, WA, FL, VA, Washington, DC	Crosswalks	Region	• U.S. DOT – FHWA	n/a	n/a

Table A 5. Pedestrian volume as pedestrian exposure – Advantages, disadvantages, and possible improvements

Examples data sources	Advantages	Disadvantages	Possible Improvements using GPS and accelerometer data
● Data collectors in the field ● Video recordings ● Automated counting devices	● More accurate than population density as individual pedestrians passing through designated points during an observed time interval are counted ● Useful when studying characteristics in a specific location	● Costly ● Time and location-dependent which reflects pedestrian activity in discrete areas and time periods only ● Could be misreporting and underreporting ● For individual pedestrians that cannot be identified, the estimated number of pedestrians may not be accurate when pedestrians cross multiple times (Greene-Roesel et al., 2007) ● Because the data are based on point locations, it may be difficult to assess pedestrian exposure over wide areas	● Able to measure pedestrian counts at any intersection traversed ● Can distinguish single vs. multiple crossings at same location

Table A 6. Examples of analysis on trip-based pedestrian exposure

Exposure type	Reference	Outcome of interest	Study location	Spatial unit	Area extent	Data sources	Pedestrian exposure model type	Significance at (α= .05) if any
Space-time walking path estimation	Lam et al. (2013); Lam et al. (2014); Yao et al. (2015)	Activity-based pedestrian exposure	Kwun Tong (KT) District in Hong Kong	Individual trips	County	• Transport Department of Hong Kong	• Space-time Path and prism • Potential path tree	n/a
Crossing behavior	Papadimitriou et al. (2012)	Pedestrian exposure related to crossing behavior	Athens, Greece	Road segments	Neighbor-hood	• Field survey using simple random sampling at the exits of Metro stations in Athens, Greece	• Sequential logit model	• Traffic signals (+) • Low traffic volume (+) • Presence of two lanes (-) • A change of trip direction (+) • crossing at the first road link (+) • increase in the percentage of the trip length (+) • increased walking speed (-)
	Lassarre et al. (2007)	Pedestrian risk exposure for pedestrians (micro-environment)	Florence, Italy and Athens, Greece	Crosswalks	City	• WHO-World Health Organization, 2006. Health effects and risks of transport systems: the HEARTS project. WHO Regional Office for Europe	Nested logit model	• Walking distance (-) • Crossing distance (-) • Vehicle volume (-) • Crosswalk (+)
Physical activity/ walking bouts	Kang et al. (2013)	Pedestrian walking activity	Seattle/ King County, WA	Individual trips	City	• Travel Assessment and Community project between July 2008 and July 2009 funded by National Institute of Health (NIH) and in collaboration with the Seattle	n/a	n/a

Exposure type	Reference	Outcome of interest	Study location	Spatial unit	Area extent	Data sources	Pedestrian exposure model type	Significance at (α= .05) if any
						Children's Research Institute (SCRI) and the University of Washington (UW)		
	Frank et al. (2019)	Physical activity, walkability	Portland, OR	Individual trips	City	• National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health-funded • Surveys were used from 2009 National Highway Transportation Survey	n/a	n/a
	Miller et al. (2015)	Physical activity, walkability	Salt Lake City, UT	Individual trips	City	• A custom web application and GIS-based Trip Identification and Analysis System (TIAS) (Westat, Inc.) supported the accelerometer and GPS data pre-processing.	n/a	n/a
	NYC Dept. of Health and Mental Hygiene (2013)	Physical activity, walkability	New York City borough (Bronx, Brooklyn, Manhattan, Queens, and Staten Island).	Individual trips	Region	• Physical Activity and Transit (PAT) survey conducted by the New York City Department of Health and Mental Hygiene	n/a	n/a

Table A 7. Examples of Previous Pedestrian Intersection Volume Models (Schneider et al., 2012)

General information			Pedestrian count information				Significant predictor variables				Model information		
Model Location	**Developed by**	**Intersec-tions Used for Model**	**Pedestrian Count Description**	**Type of Count Sites**	**Count Period(s) Used for Model**	**Weather During Counts**	**Land Use**	**Transport-ation System**	**Socio-economic Character-istic**	**Other**	**Model Output**	**Model Type**	**Validation Testing**
Charlotte, NC	UNC Charlotte (Pulugurtha & Repaka, 2008)	176	Pedestrians counted each time they arrived at the intersection from any direction	Signalized intersectio ns	7 am-7 pm	Clear weather conditions	● Population within 0.25 mi. ● Jobs within 0.25 mi. ● Mixed land use within 0.25 mi. ● The urban residential area within 0.25 mi.	Number of bus stops within 0.25 mi.	n/a	n/a	Total pedestrians approaching intersections from 7 am to 7 pm (shorter periods also modeled)	Linear	None reported
Alameda County, CA	UC Berkeley SafeTREC (Schneider et al., 2009)	50	Pedestrians counted every time they crossed a leg of the intersection (pedestrians within 50 feet of the crosswalk were counted)	Signalized and un-signalized intersectio ns	Tu, W, or Th: 12-2 pm or 3-5 pm, Sa: 9-11am, 12-2 pm, or 3-5 pm	All-weather conditions; weather adjustment factors were calculated from automated counters	● The population within 0.5 mi. ● Employment within 0.25 mi. ● Commercial properties within 0.25 mi.	BART (regional transit) station within 0.1 mi.	n/a	n/a	Total pedestrian crossings at intersections during a typical week	Linear	46 historic counts used for validation (30 additional intersections were counted for validation in 2009)
San Francisco, CA	San Francisco State (Liu & Griswold, 2009)	63	Pedestrians counted each time they crossed a leg of the intersection (no distance to crosswalk specified)	Signalized and unsignalize d intersectio ns	Weekdays 2:30-6:30 pm	Not reported	● Population density (net) within 0.5 mi. ● Employment density (net) within 0.25 mi. ● Patch richness density within 0.063 mi. ● Residential land use within 0.063 mi.	MUNI (light-rail transit) stop density within 0.38 mi. ● Presence of bike lane at intersection	n/a	Mean slope within 0.063 mi.	Total pedestrian crossings at intersections from 2:30-6:30 pm on typical weekday	Linear	None reported

General information			Pedestrian count information				Significant predictor variables				Model information		
Model Location	**Developed by**	**Intersec-tions Used for Model**	**Pedestrian Count Description**	**Type of Count Sites**	**Count Period(s) Used for Model**	**Weather During Counts**	**Land Use**	**Transport-ation System**	**Socio-economic Character-istic**	**Other**	**Model Output**	**Model Type**	**Validation Testing**
Santa Monica, CA	Fehr & Peers (Haynes et al., 2010)	92	Pedestrians counted each time they crossed a leg of the intersection (no distance to crosswalk specified)	Signalized and un-signalized intersectio ns	Weekdays 5-6 pm	Not reported	● Employment density within 0.33 mi. ● Within a commercially-zoned area	● Afternoon bus frequency ● Average speed limit on the intersection approaches	n/a	Distance from Ocean	Total pedestrian crossings at intersections from 5-6 pm on typical weekday	Linear	About 107 additional intersections were counted for validation
San Diego, CA	Alta Planning + Design (Jones et al., 2010)	80	Pedestrians counted each time they arrived at the intersection from any direction	Signalized and unsignalize d intersectio ns	Weekdays 7-9 am	Nice weather	● Population density within 0.25 mi. ● Employment density within 0.5 mi. ● Presence of retail within 0.5 mi.	● Greater than 6,000 transit ridership at bus stops within 0.25 mi. ● 4 or more Class I bike paths within 0.25 mi.	● More than 100 households without vehicles within ● 0.5 mi.	n/a	Total pedestrians approaching intersections from 7 am to 9 am	Log-linear	None reported
Montreal, Quebec	McGill University (Miranda-Moreno & Fernandes, 2011)	1018	Pedestrians counted each time they crossed a leg of the intersection (no distance to crosswalk specified)	Signalized intersectio ns	Weekdays 6-9 am, 11 am-1 pm, and 3:30-6:30 PM	Most counts during nice weather; weather variables were analyzed	● Population within 400 m ● Commercial space within 50 m ● Open space within 150 m ● Schools within 400 m	● Subway within 150 m ● Bus station within 150 m ● % Major arterials within 400 m ● Street segments within 400 m ● 4-way intersection	n/a	● Distance to downtown ● Daily high temperature >32C	Total pedestrian crossings at intersections over 8 count hours (shorter periods also modeled)	Log-linear (also used Negative binomial)	Counts at 20% of the intersections were compared to a model based on 80% of the intersections for validation

Table A 8. Summary of examples of pedestrian exposure measure data collection methods in U.S. cities

City	State	Pedestrian manual count	Bike manual count	Automatic counter (e.g., loop, counters)	Video technology	Survey	Pedestrian exposure model	Electronic devices (e.g., GPS, Accelerometer, Mobile application)
Seattle	WA	✓	✓	✓				✓
Bellevue	WA			✓	✓			
PSRC	WA					✓		
Portland	OR	✓	✓	✓				✓
Caltrans	CA	✓	✓			✓	✓	
San Francisco	CA	✓						
Santa Monica	CA				✓	✓	✓	
San Diego	CA	✓	✓	✓			✓	
Oakland	CA	✓					✓	
San Francisco	CA	✓						
Arizona	AZ			✓				
Tucson	AZ	✓	✓					
Salt Lake City	UT							✓
Colorado	CO			✓				
Austin	TX	✓		✓	✓			
Minneapolis	MN	✓	✓	✓				
Chicago	IL	✓						
FDOT	FL							✓
Jacksonville	FL			✓				

City	State	Pedestrian manual count	Bike manual count	Automatic counter (e.g., loop, counters)	Video technology	Survey	Pedestrian exposure model	Electronic devices (e.g., GPS, Accelerometer, Mobile application)
Orlando	FL	✓	✓	✓				
Miami	FL	✓	✓					
New York City	NY	✓	✓	✓		✓		✓
Boston	MA	✓	✓		✓			
Seattle	WA	✓	✓	✓				✓
Bellevue	WA			✓	✓			
PSRC	WA					✓		

Appendix B. All variables considered and data sources

Table B 1. All variables considered in analysis and data sources

Variable names	Abbreviation	Units	Type	Buffer size	Date	Removed	Data Sources
Total walking bouts - Outcome	TOTAL_WB	Count	Numeric	50 m	2008 - 2014		Seattle Children's Hospital and University of Washington (TRAC/ACTION)
Individual-level variables							
Pedestrian average age (DS2 only)	I_AVG_AGE	Age	Numeric	50 m	2008 - 2014		Seattle Children's Hospital and University of Washington
Pedestrian female ratio (DS2 only)	I_RATIO_GENDER_F	Ratio	Numeric	50 m	2008 - 2014		Seattle Children's Hospital and University of Washington
Pedestrian non-white race ratio (DS2 only)	I_RATIO_RACE_NW	Ratio	Numeric	50 m	2008 - 2014		Seattle Children's Hospital and University of Washington
Pedestrian employment ratio (DS2 only)	I_RATIO_EMP_Y	Ratio	Numeric	50 m	2008 - 2014		Seattle Children's Hospital and University of Washington
Pedestrian median household income (DS2 only)	I_MED_HHI	-	Categorical	50 m	2008 - 2014		Seattle Children's Hospital and University of Washington
Pedestrian single household ratio (DS2 only)	I_RATIO_SINGHH_Y	Ratio	Numeric	50 m	2008 - 2014		Seattle Children's Hospital and University of Washington

Variable names	Abbreviation	Units	Type	Buffer size	Date	Removed	Data Sources
Micro-environment variables							
Average roadway width	AVG_RW_WIDTH_FT	Feet	Numeric	50 m	2018		SDOT[1]
Median roadway class	MED_RW_CLASS	-	Categorical	50 m	2018	✓	SDOT[1]
Maximum posted speed limit	MAX_SPEEDL_MPH	MPH	Numeric	50 m	2018		SDOT[1]
Maximum slope percentage	MAX_SLOPE_PER	Ratio	Numeric	50 m	2018		SDOT[1]
Total bike lane length	TOTAL_BIKEL_FT	Feet	Numeric	50 m	2018		SDOT[2]
Total sidewalk length	TOTAL_SIDEW_FT	Feet	Numeric	50 m	2012		UFL
Total crosswalk count	TOTAL_CROSSW_C	Count	Numeric	50 m	2017	✓	SDOT[3]
Presence of Ped & Bike sign	TRAFF_SIGN_C	Count	Numeric	50 m	2018		SDOT[4]
Presence of stop sign	TRAFF_SIGN_C	Count	Numeric	50 m	2018		SDOT[4]
Presence of one-way sign	TRAFF_SIGN_C	Count	Numeric	50 m	2018		SDOT[4]
Presence of crosswalk warning sign	TRAFF_SIGN_C	Count	Numeric	50 m	2018		SDOT[4]
Presence of curve warning sign	TRAFF_SIGN_C	Count	Numeric	50 m	2018		SDOT[4]
Presence of traffic circle, bump, no turn sign	TRAFF_SIGN_C	Count	Numeric	50 m	2018		SDOT[4]
Traffic signal presence	TRAFF_SIGNAL_Y	Y/N	Binary	50 m	2018		SDOT[4]
Roadway segment count	RW_SEGM_C	-	Categorical	50 m	2018		SDOT[1]

Variable names	Abbreviation	Units	Type	Buffer size	Date	Removed	Data Sources
Bus ridership density	BUS_RIDERSHIP_ACRE	# /acre	Numeric	50 m	2012		UFL
Macro-environment variables							
Residential density	RES_UNIT_C_ACRE	# /acre	Numeric	400 m	2012	✓	UFL
Job density	JOBS_C_ACRE	# /acre	Numeric	400 m	2012		UFL
Residential census density	RES_CENSUS_C_ACRE	# /acre	Numeric	400 m	2010	✓	2010 Census Block data[1]
Population census density	POP_CENSUS_C_ACRE	# /acre	Numeric	400 m	2010		2010 Census Block data[1]
White population percentage	WHITE_CENSUS_PER	Ratio	Numeric	400 m	2010		2010 Census Block data[1]
Median household income	MED_MEDHHI_DOL	$	Numeric	400 m	2010		WA 2010 Census data Block Group level
Public school presence	PUB_SCHOOL_Y	Y/N	Binary	400 m	2015 - 2016	✓	National Center for Education Statistics[1]
Public school enrollment count	PUB_S_ENROLL_C	Count	Numeric	400 m	2015 - 2016		National Center for Education Statistics[1]
Park presence (land use code: 76)	PARK_Y	Y/N	Binary	400 m	2010		WSDOT[1]

Variable names	Abbreviation	Units	Type	Buffer size	Date	Removed	Data Sources
Total Trail length	TOTAL_TRAIL_FT	Feet	Numeric	400 m	2006		King County GIS Open Data
Park and Ride presence	PNR_Y	Y/N	Binary	400 m	2017		WSDOT[1]
Residential land use percentage (land use code: 11-19)	RES_LU_PER	Ratio	Numeric	400 m	2010	✓	WSDOT[2]
Manufacturing land use percentage (land use code: 21-39)	MANUFAC_LU_PER	Ratio	Numeric	400 m	2010		WSDOT[2]
Transportation land use percentage (land use code: 41-49)	TRANSPORT_LU_PER	Ratio	Numeric	400 m	2010		WSDOT[2]
Trade land use percentage (land use code: 50-59)	TRADE_LU_PER	Ratio	Numeric	400 m	2010	✓	WSDOT[2]
Service land use percentage (land use code: 61-69)	SERVICE_LU_PER	Ratio	Numeric	400 m	2010		WSDOT[2]

Notes:

2010 Census Block data[1] - www.ofm.wa.gov/washington-data-research/population-demographics/gis-data/census-geographic-files

King County GIS Open Data[1] - https://gis-kingcounty.opendata.arcgis.com/datasets/trails-in-king-county-trail-line

National Center for Education Statistics[1] - https://nces.ed.gov/ccd/elsi/tableGenerator.aspx

SDOT[1] - http://data-seattlecitygis.opendata.arcgis.com/datasets/seattle-streets

SDOT[2] - http://data-seattlecitygis.opendata.arcgis.com/datasets/existing-bike-facilities

SDOT[3] - http://data-seattlecitygis.opendata.arcgis.com/datasets/marked-crosswalks

SDOT[4] - https://data-seattlecitygis.opendata.arcgis.com/datasets/street-signs

WSDOT[1] - www.wsdot.wa.gov/mapsdata/geodatacatalog/Maps/noscale/DOT_PubTrans/ParkandRides.htm

WSDOT[2] - https://geo.wa.gov/datasets/washington-state-land-use-2010?selectedAttribute=LANDUSE_CD

This page is intentionally left blank

Appendix C. Visualization of significant predictors

Job density

Data source: UFL job density 400 m Smart map (internal data)

Processing: Extract the point values at intersections. The value represents the job density in the number of jobs per acre.

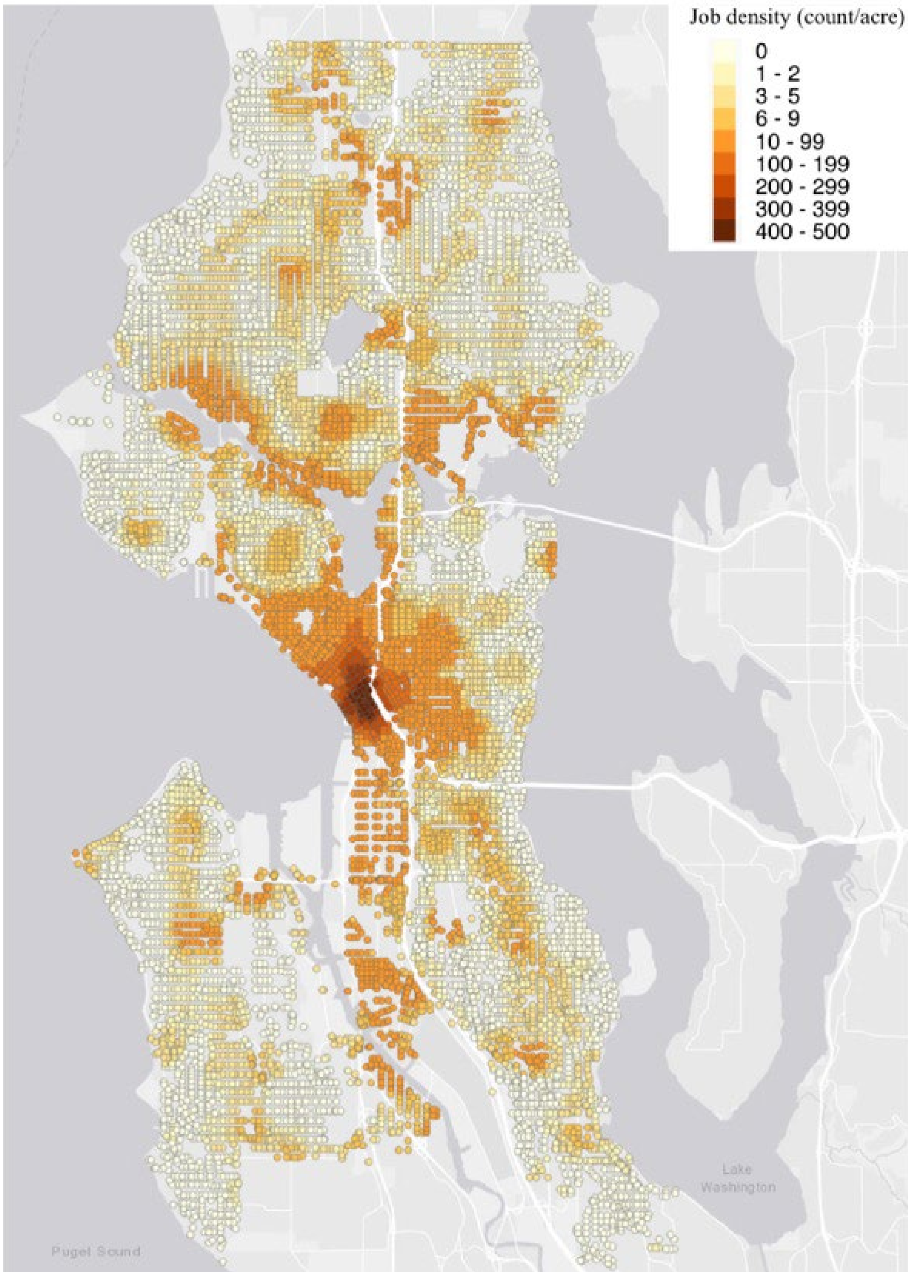


Figure C1. Map of intersections with job density (unit: per acre)

Crosswalk count

Data source: http://data-seattlecitygis.opendata.arcgis.com/datasets/marked-crosswalks

Processing: Crosswalk points were filtered by intersection 50 m buffers. The count of crosswalks in the buffer was calculated for each intersection. Figure C 2 is an overview of crosswalks in Seattle and Figure C 3 shows the crosswalks around a typical intersection.

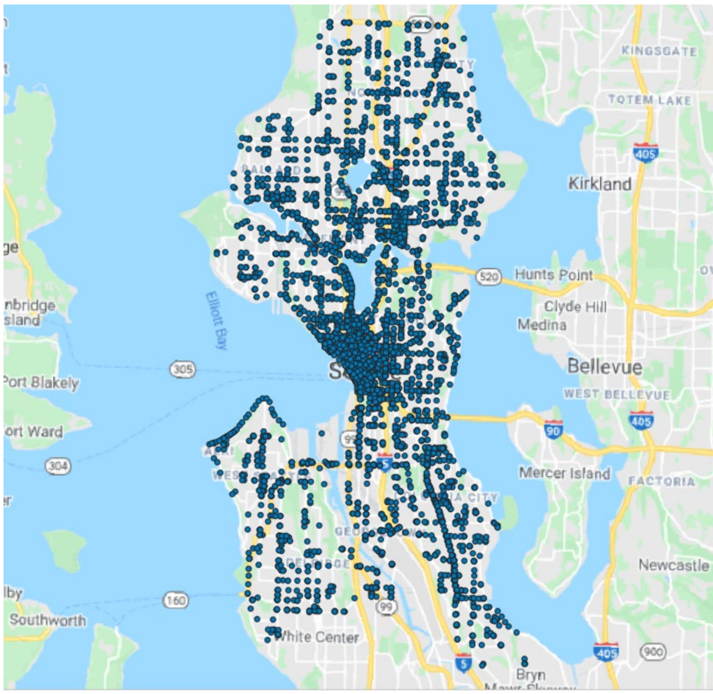


Figure C 2. Intersections with marked crosswalks

Figure C 3. Marked crosswalks around intersections within 50 m buffers

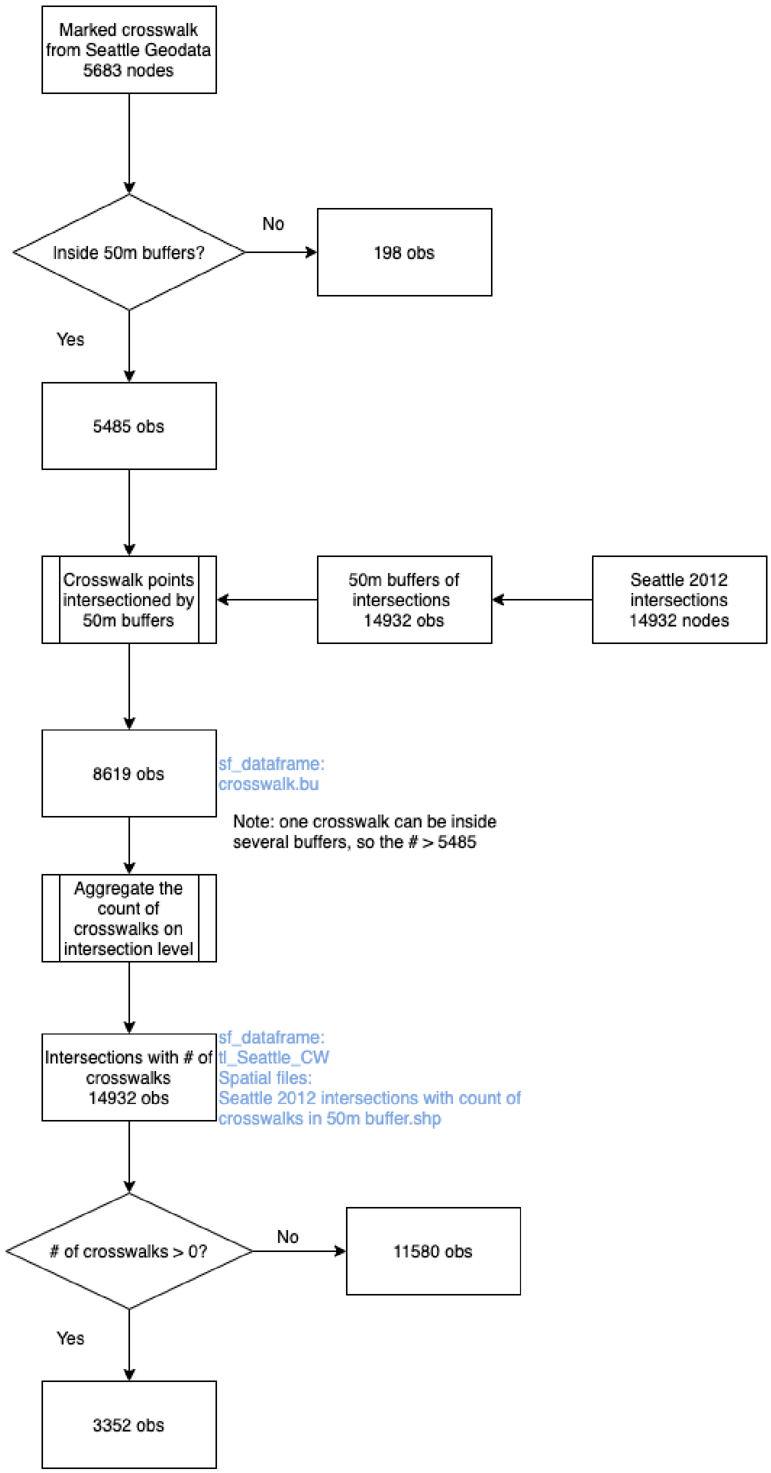


Figure C 4. Flow chart of data structure for crosswalk count

Figure C 4 shows the data structure of this variable, which was aggregated at the level of intersections. Among the 14,932 intersections in Seattle, only 3,352 have crosswalks in their 50 m buffers. From Figure C 5, most have less than 5 crosswalks, but there still exist outliers with 8 or 9 around them, especially in downtown (the deep blue nodes in Figure C 6).

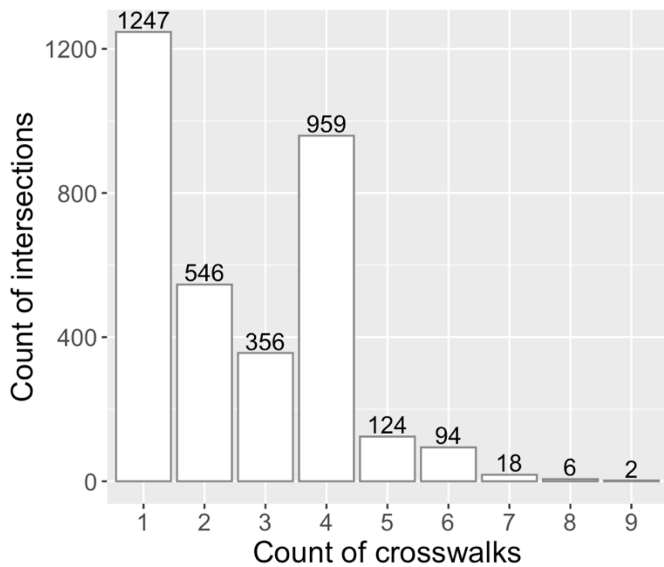


Figure C 5. Count of intersections by number of crosswalks in their 50 m buffers

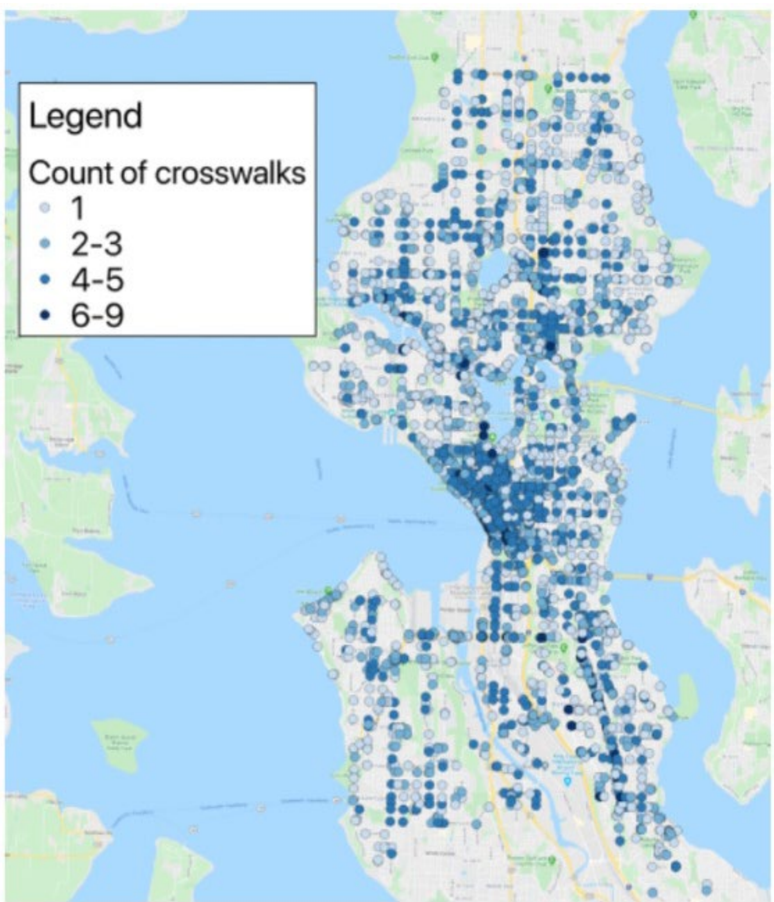


Figure C 6. Distribution of crosswalk counts in 50 m buffers in Seattle

Residential density

Data source: UFL residential density 400 m Smart map (internal data)

Processing: Extract the point values at intersections. The value represents the residential density in residents per acre.

Note: Residential density within a 400-m buffer was converted to the more commonly used density unit, number per acre. For example, if there are 1,000 residents in 502,700 m^2 (π * 400 m * 400 m), this means that there are 8 residents per acre (1,000 residents/502,700 m^2 * 4,046.86 m^2/acre).

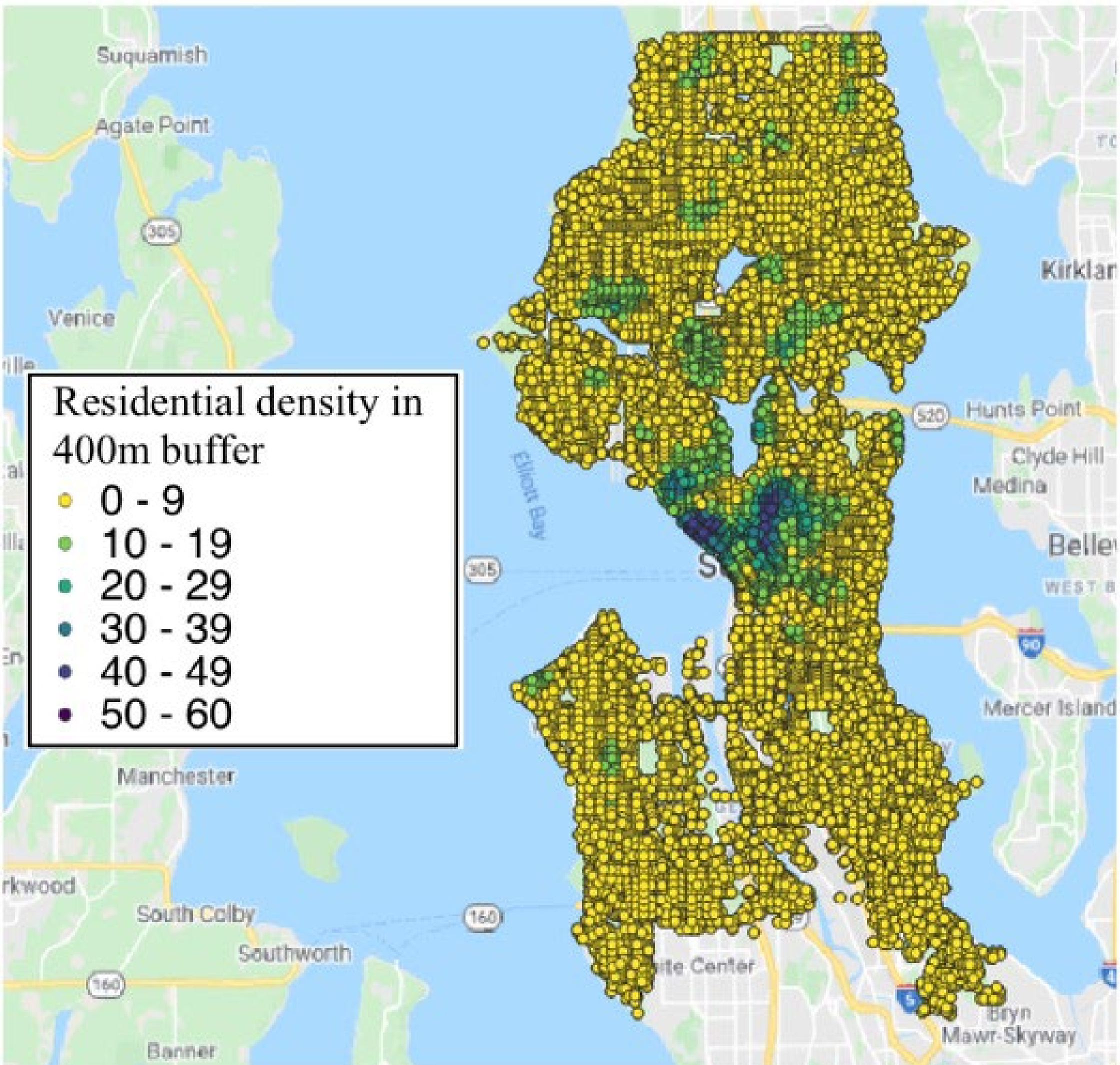


Figure C 7. Map of intersections with residential density (unit: per acre)

Total bike lane length

Data source: Seattle existing bike facilities (http://data-seattlecitygis.opendata.arcgis.com/datasets/existing-bike-facilities) including both bicycle-related streets and off-streets (trails).

Processing: Bike facilities were filtered by intersection 50 m buffers and summarized the length. Figure C 8 shows (left) the distribution of bike facilities and (right) gives the intersections based on the bike lane length.

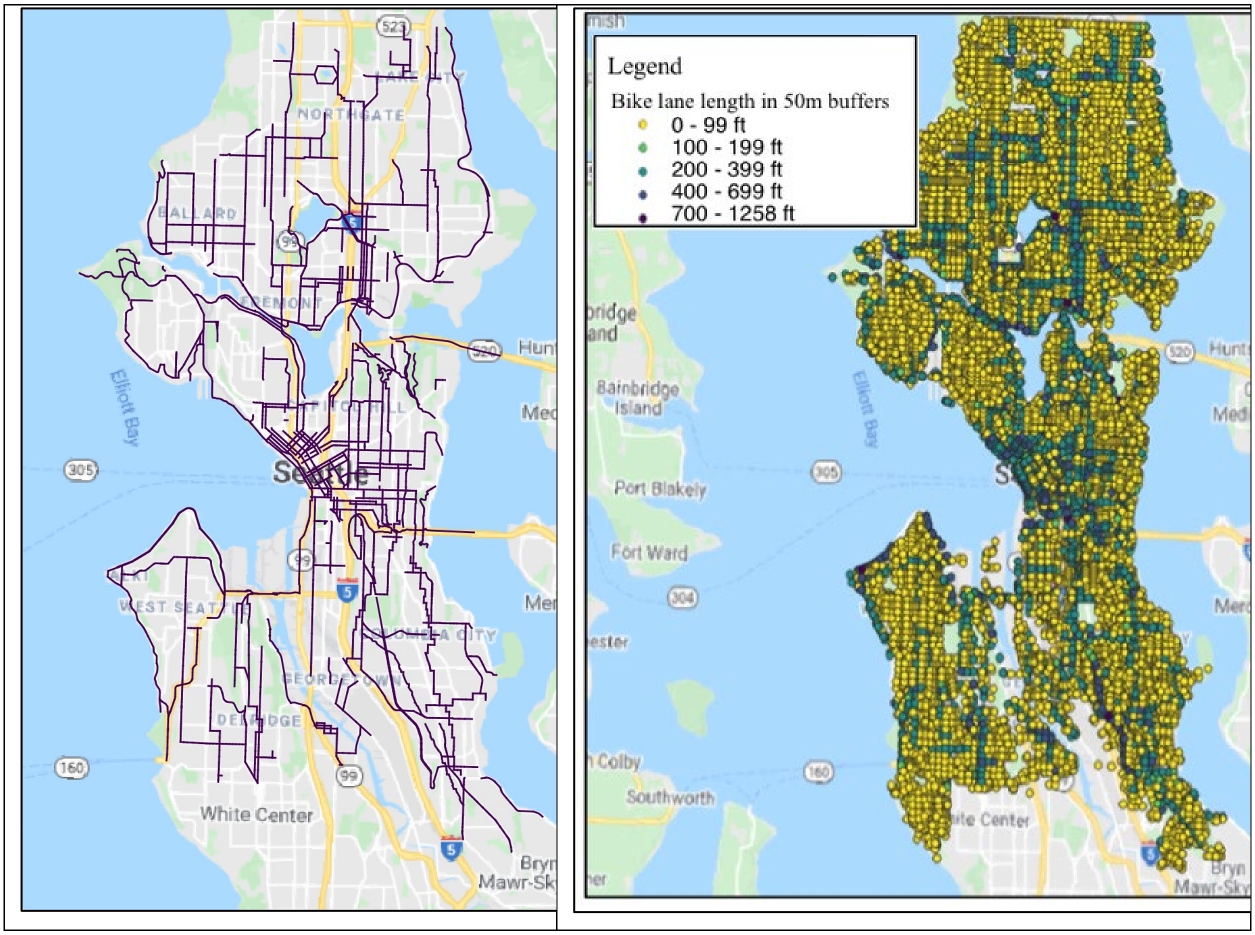


Figure C 8. Seattle existing bike facilities (left) and lengths of bike lanes by intersection (right)

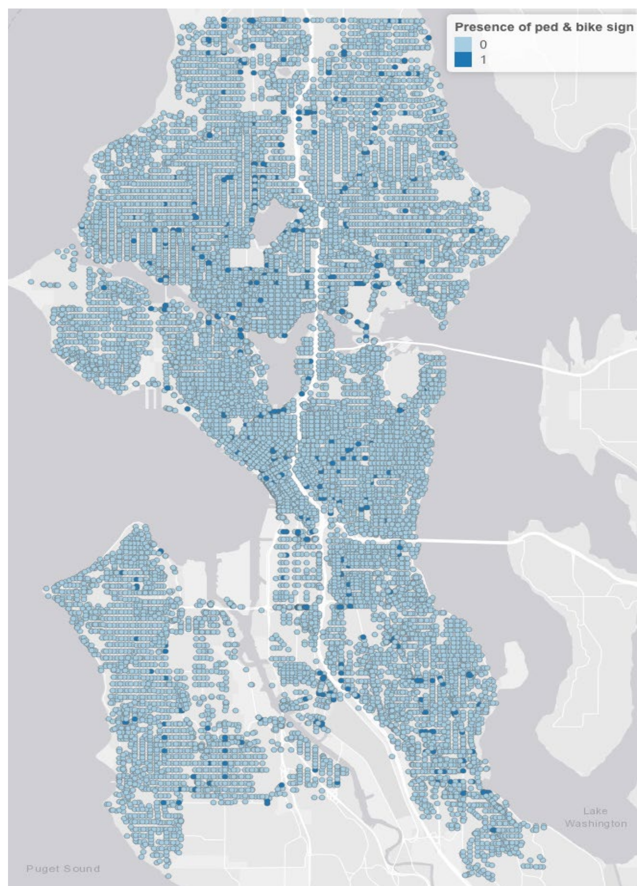
Traffic sign presence

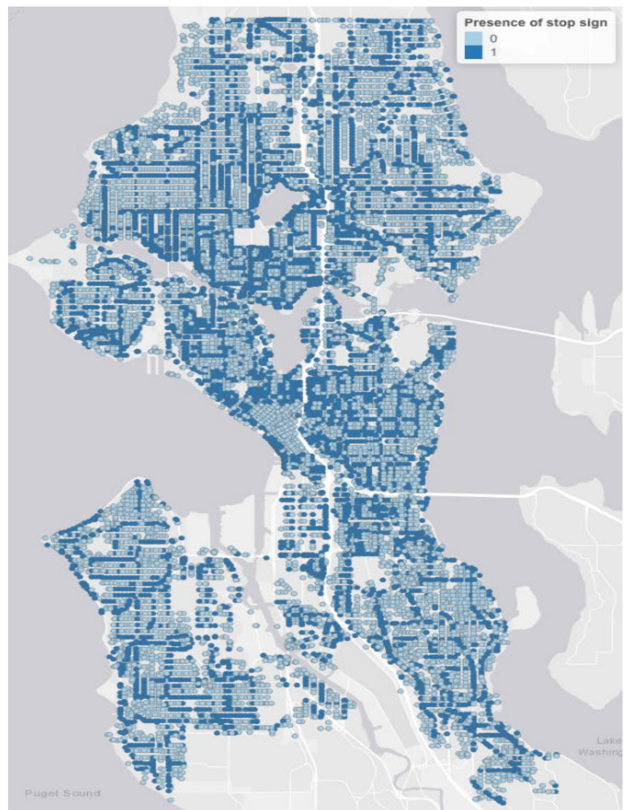
Data source: Seattle traffic signs from [http://data-seattlecitygis.opendata.arcgis.com/datasets/street-signs (open source)](http://data-seattlecitygis.opendata.arcgis.com/datasets/street-signs)

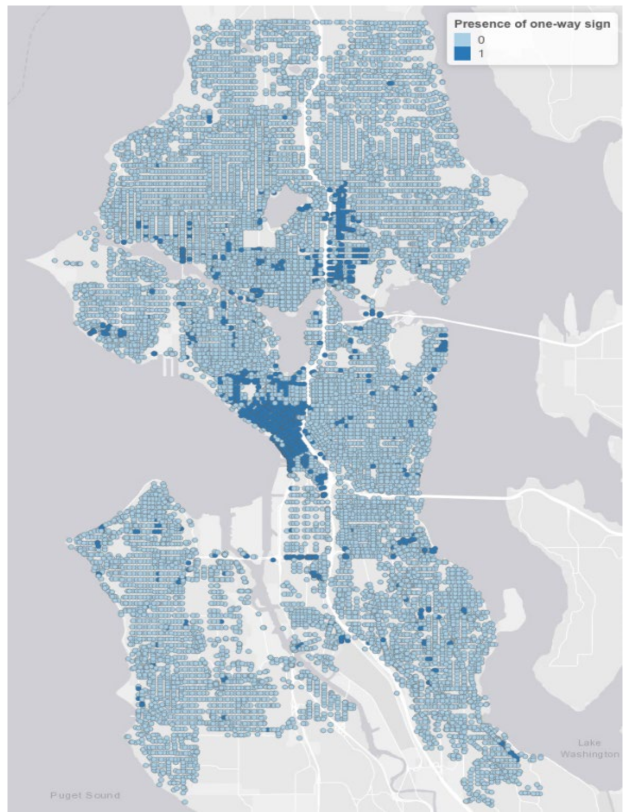
Processing: Conducted comprehensive review on Regulatory and Warning signs from the City of Seattle DOT - 2018**.** The included signs are summarized in Table C 1.

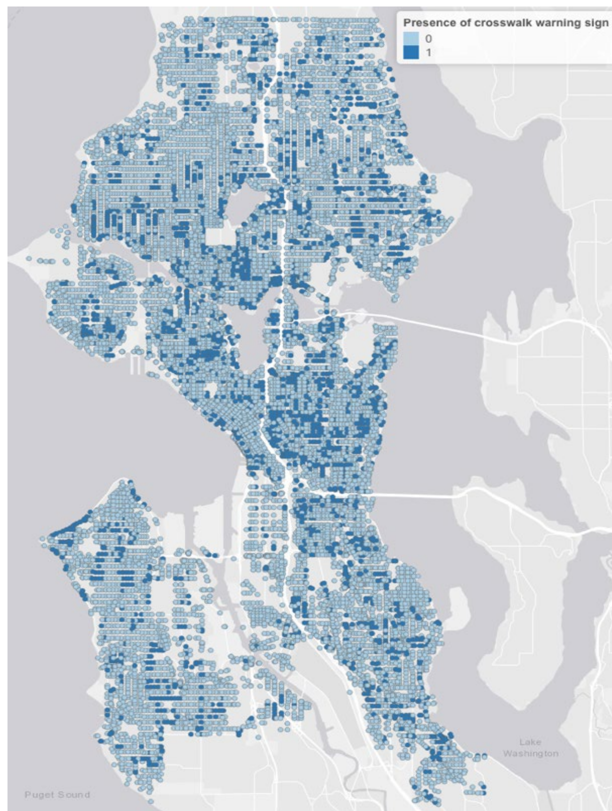
Light blue dots represent the intersections that did not have signs while dark blue dots show the intersection locations that did have signs. (Presence of signs are noted as 0: no, 1: yes). We used the presence of various signs, rather than number of signs at each intersection.

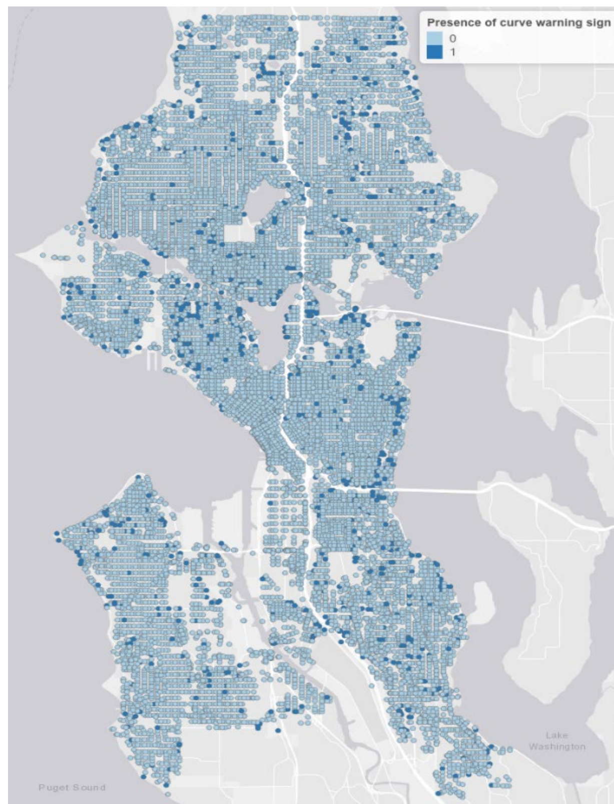
Table C 1. Summary of signs

Sign type	Description text	Frequency	Locations
Ped & Bike sign	[STOP] HERE [SWOOPING LEFT ARROW] FOR [PED]	213	
	PUSH BUTTON FOR [PED WALK] SIGNAL [LT ARROW]	210	
	PUSH BUTTON FOR [PED WALK] SIGNAL	194	
	PUSH BUTTON FOR [PED WALK] SIGNAL [RT-LT ARROW]	150	
	PUSH BUTTON FOR GREEN LIGHT	99	
	[BIKE] [PED]	64	
	[STOP SIGN] FOR [PED SYMBOL] [BIKE SYMBOL]	55	
	[BIKE] YIELD TO PEDS	54	

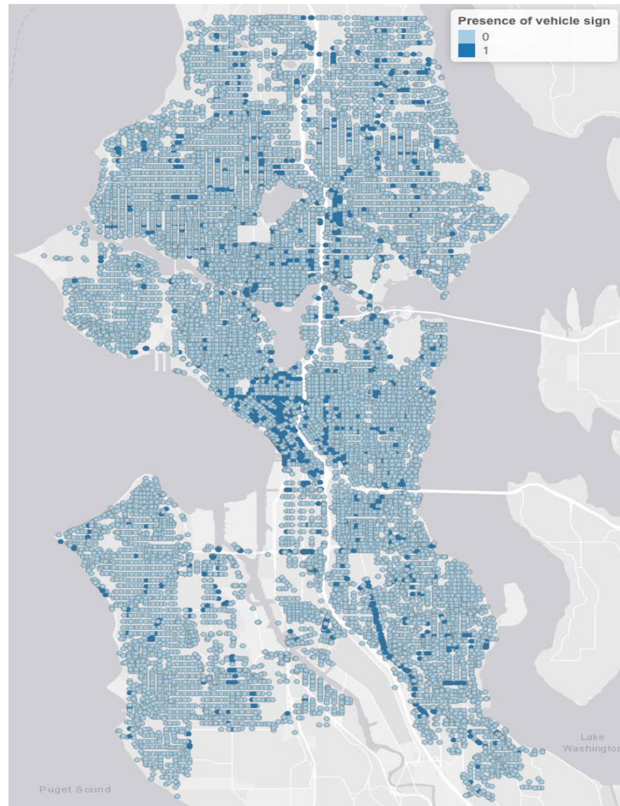
Sign type	Description text	Frequency	Locations
Stop sign	STOP	8,735	
	ALL WAY	1,305	

Sign type	Description text	Frequency	Locations
One-way sign	ONE WAY [RT ARROW?	1,562	
	ONE WAY [LT ARROW?	1,472	
	ONE WAY [RT ARROW]	193	
	ONE WAY [LT ARROW]	192	
	[LARGE RIGHT ARROW] ONE WAY	59	

Sign type	Description text	Frequency	Locations
Crosswalk warning sign	[SCHOOL PED]	3,063	
	[45L DW ARROW]	2,013	
	[PEDESTRIAN]	1,653	
	AHEAD	719	
	[45R DW ARROW]	357	
	[PEDESTRIAN IN X-WALK]	323	
	[PED] over [BIKE]	135	
	CROSSWALK	104	
	[PED] [BIKE]	86	
	STOP FOR ME [PEDESTRIAN SYMBOL] IT'S THE LAW	82	
	TRAIL X-ING	60	
	[BIKE] [PED]	50	

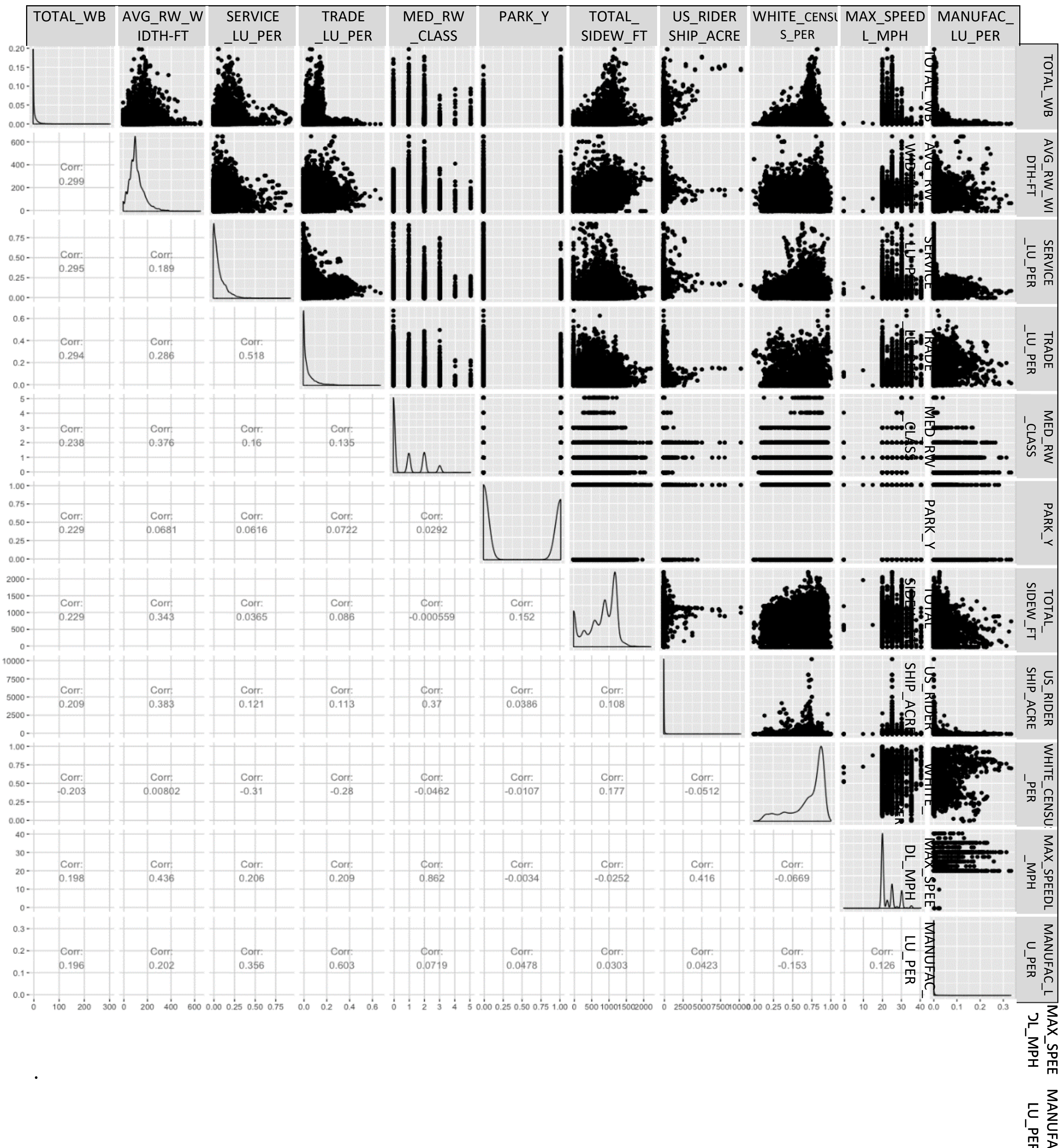
Sign type	Description text	Frequency	Locations
Curve warning sign	[RT LT ARROW]	316	
	[LT ARROW]	214	
	[RT ARROW]	213	
	[LT CHEVRON]	179	
	[45L CRV ARROW]	155	
	[45R CRV ARROW]	146	
	[90R CRV ARROW]	106	
	[90L CRV ARROW]	95	

Sign type	Description text	Frequency	Locations
Vehicle sign - No turn sign	[NO LEFT TURN]	560	
	[NO RIGHT TURN]	321	
	NO TURN ON RED	168	
	NO TURNS	31	
Vehicle sign - Traffic circle, bumps sign	BUMP	434	
	SPEED BUMPS AHEAD	267	
	[TRAFFIC CIRCLE 18 x 18 DIAMOND GRADE ORANGE]	43	
	SPEED BUMP AHEAD	26	
	BUMPS	22	



This page is intentionally left blank

Appendix D. Correlations for all environmental variables


TOTAL_WB
AVG_RW_WIDTH-FT
SERVICE_LU_PER
TRADE_LU_PER
MED_RW_CLASS
PARK_Y
TOTAL_SIDEW_FT
US_RIDERSHIP_ACRE
WHITE_CENSUS_PER
MAX_SPEEDL_MPH
MANUFAC_LU_PER
Corr: 0.299
Corr: 0.295
Corr: 0.189
Corr: 0.294
Corr: 0.286
Corr: 0.518
Corr: 0.238
Corr: 0.376
Corr: 0.16
Corr: 0.135
Corr: 0.229
Corr: 0.0681
Corr: 0.0616
Corr: 0.0722
Corr: 0.0292
Corr: 0.229
Corr: 0.343
Corr: 0.0365
Corr: 0.086
Corr: -0.000559
Corr: 0.152
Corr: 0.209
Corr: 0.383
Corr: 0.121
Corr: 0.113
Corr: 0.37
Corr: 0.0386
Corr: 0.108
Corr: -0.203
Corr: 0.00802
Corr: -0.31
Corr: -0.28
Corr: -0.0462
Corr: -0.0107
Corr: 0.177
Corr: -0.0512
Corr: 0.198
Corr: 0.436
Corr: 0.206
Corr: 0.209
Corr: 0.862
Corr: -0.0034
Corr: -0.0252
Corr: 0.416
Corr: -0.0669
Corr: 0.196
Corr: 0.202
Corr: 0.356
Corr: 0.603
Corr: 0.0719
Corr: 0.0478
Corr: 0.0303
Corr: 0.0423
Corr: -0.153
Corr: 0.126

.

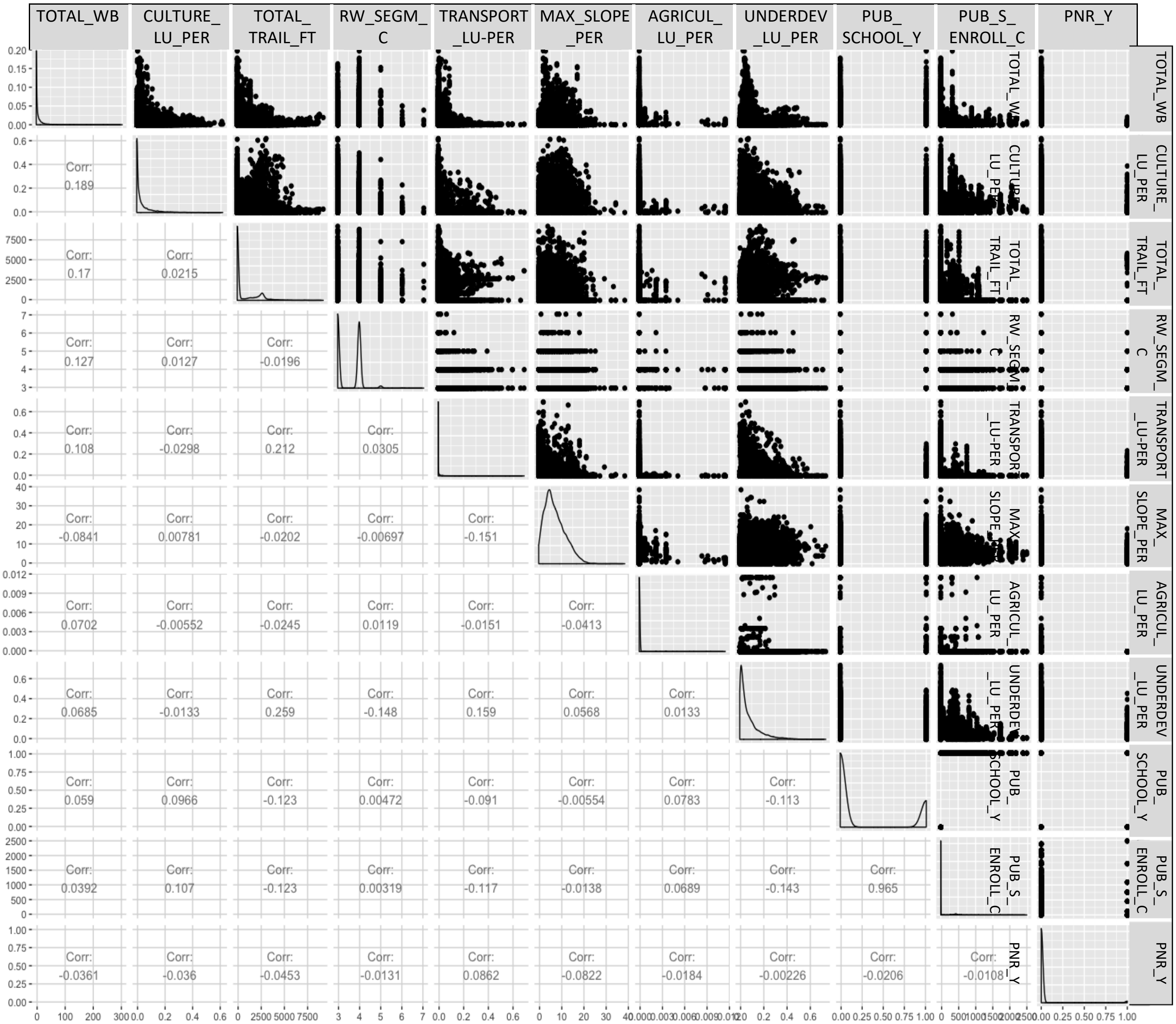


Figure D 1. Correlations and distributions with all variables in analysis

This page is intentionally left blank

Appendix E. LASSO for variable selection

The following 25 predictors were identified by the research team during environmental variable processing. This processing included LASSO regression; a statistical inference tool that helps select significant predictors by iteratively attempting to minimize model MSE.

1. Average roadway width
2. Maximum posted speed limit (categorical variable: 5 levels- 20, 25, 30, 35, 40 mph)
3. Maximum slope percentage
4. Total bike lane length
5. Sidewalk length
6. Presence of ped and bike sign
7. Presence of stop sign
8. Presence of one-way sign
9. Presence of crosswalk warning sign
10. Presence of curve warning sign
11. Presence of traffic circle, bump, and no turn sign
12. Traffic signal presence
13. Roadway segment count (categorical variable: 3 levels – 3, 4, 5+)
14. Bus ridership density
15. Job density
16. Population census density
17. White population percentage
18. Median household income
19. Public school enrollment count
20. Park presence
21. Total trail length
22. Park and ride presence
23. Manufacturing land use percentage
24. Transportation land use percentage
25. Service land use percentage

Appendix F. Model results

Table F 1. Dataset 1, all intersections (n= 14,073)

Predictors	**Median HH Income**			**Job Density**			**White Population**			**Service Land Use**		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
Negative Binomial Portion												
(Intercept)	3.86	3.42 – 4.35	**<0.001**	2.52	2.32 – 2.74	**<0.001**	2.78	2.46 – 3.15	**<0.001**	1.94	1.78 – 2.12	**<0.001**
Avg roadway width (10ft)	1	1.00 – 1.01	0.061	1.01	1.00 – 1.01	**0.001**	1	1.00 – 1.01	**0.037**	1.01	1.00 – 1.01	**<0.001**
Max speed limit – 20 mph *Reference*												
25 mph	1.75	1.62 – 1.88	**<0.001**	1.61	1.50 – 1.73	**<0.001**	1.77	1.64 – 1.91	**<0.001**	1.66	1.55 – 1.79	**<0.001**
30 mph	0.66	0.61 – 0.71	**<0.001**	0.69	0.64 – 0.74	**<0.001**	0.69	0.64 – 0.74	**<0.001**	0.61	0.57 – 0.66	**<0.001**
35 mph	0.79	0.69 – 0.90	**<0.001**	0.88	0.77 – 1.00	0.052	0.8	0.70 – 0.92	**0.002**	0.76	0.66 – 0.87	**<0.001**
40 mph	0.86	0.70 – 1.05	0.132	0.95	0.79 – 1.16	0.642	0.88	0.72 – 1.08	0.216	0.9	0.74 – 1.10	0.318
Max slope (%)	0.97	0.96 – 0.97	**<0.001**	0.96	0.96 – 0.97	**<0.001**	0.96	0.96 – 0.97	**<0.001**	0.97	0.96 – 0.98	**<0.001**
Total bike lane (100 ft)	1.07	1.05 – 1.09	**<0.001**	1.07	1.05 – 1.08	**<0.001**	1.07	1.05 – 1.08	**<0.001**	1.06	1.05 – 1.08	**<0.001**
Crosswalk warning sign (N) *Reference*												
Crosswalk warning sign (Y)	1.11	1.05 – 1.18	**<0.001**	1.11	1.05 – 1.18	**<0.001**	1.11	1.04 – 1.17	**0.001**	1.1	1.04 – 1.16	**0.002**
Bike and Ped sign (N) ref												
Bike and Ped sign (Y)	1.26	1.12 – 1.41	**<0.001**	1.32	1.18 – 1.49	**<0.001**	1.26	1.11 – 1.41	**<0.001**	1.23	1.10 – 1.38	**<0.001**
One-way sign (N) *Reference*												
One-way sign (Y)	1.11	1.02 – 1.20	**0.018**	0.89	0.82 – 0.96	**0.004**	1.16	1.06 – 1.26	**0.001**	1.06	0.98 – 1.15	0.174

Predictors	**Median HH Income**			**Job Density**			**White Population**			**Service Land Use**		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
Traffic signal presence (N) *Reference*												
Traffic signal presence (Y)	1.68	1.54 – 1.83	**<0.001**	1.47	1.35 – 1.60	**<0.001**	1.75	1.61 – 1.91	**<0.001**	1.54	1.41 – 1.68	**<0.001**
Bus ridership (1000/acre)	1.47	1.35 – 1.61	**<0.001**	1.25	1.15 – 1.36	**<0.001**	1.56	1.42 – 1.72	**<0.001**	1.34	1.24 – 1.45	**<0.001**
Population density (10/acre)	1.61	1.56 – 1.66	**<0.001**	1.63	1.59 – 1.68	**<0.001**	1.75	1.71 – 1.80	**<0.001**	1.67	1.63 – 1.72	**<0.001**
Public school enrollment (1000 counts)	0.87	0.81 – 0.94	**<0.001**	0.92	0.86 – 0.99	**0.027**	0.84	0.78 – 0.91	**<0.001**	0.83	0.77 – 0.89	**<0.001**
Park presence (N) *Reference*												
Park presence (Y)	1.42	1.35 – 1.50	**<0.001**	1.31	1.25 – 1.38	**<0.001**	1.4	1.33 – 1.48	**<0.001**	1.42	1.35 – 1.50	**<0.001**
Total trail length (1000ft)	1.09	1.07 – 1.11	**<0.001**	1.1	1.07 – 1.12	**<0.001**	1.11	1.08 – 1.13	**<0.001**	1.09	1.07 – 1.12	**<0.001**
Park and Ride presence (N) *Reference*												
Park and Ride presence (Y)	0.68	0.55 – 0.85	**0.001**	0.73	0.59 – 0.91	**0.005**	0.68	0.54 – 0.84	**<0.001**	0.68	0.54 – 0.85	**0.001**
Manufacture land use (%)	0.98	0.97 – 0.99	**0.003**	0.98	0.97 – 0.99	**0.005**				0.98	0.97 – 0.99	**0.001**
Transportation land use (%)	1	1.00 – 1.01	0.511				1	0.99 – 1.00	0.211	1	0.99 – 1.00	0.445
Median household income ($50,000)	0.74	0.71 – 0.78	**<0.001**									
Job density (100/acre)				1.57	1.48 – 1.67	**<0.001**						
White population (%)							1	1.00 – 1.00	**<0.001**			
Service land use (%)										1.03	1.02 – 1.03	**<0.001**

Predictors	**Median HH Income**			**Job Density**			**White Population**			**Service Land Use**		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
Zero-Inflated Portion												
(Intercept)	13.54	9.22 – 19.90	**<0.001**	8.27	6.32 – 10.80	**<0.001**	3.33	2.19 – 5.06	**<0.001**	9.59	7.07 – 13.03	**<0.001**
Avg roadway width (10ft)	0.97	0.95 – 0.99	**0.001**	0.98	0.96 – 0.99	**0.004**	0.96	0.94 – 0.98	**<0.001**	0.97	0.95 – 0.98	**<0.001**
Max speed limit – 20 mph *Reference*												
25 mph	0	0.00 – Inf	0.974	0.04	0.01 – 0.16	**<0.001**	0	0.00 – Inf	0.974	0	0.00 – Inf	0.97
30 mph	0.56	0.43 – 0.72	**<0.001**	0.71	0.58 – 0.88	**0.002**	0.55	0.42 – 0.72	**<0.001**	0.59	0.46 – 0.76	**<0.001**
35 mph	0.22	0.11 – 0.44	**<0.001**	0.34	0.20 – 0.58	**<0.001**	0.31	0.15 – 0.64	**0.001**	0.31	0.16 – 0.60	**<0.001**
40 mph	1.15	0.58 – 2.27	0.684	1.59	0.90 – 2.81	0.109	1.13	0.57 – 2.26	0.728	1.23	0.64 – 2.36	0.538
Max slope (%)	1.02	1.00 – 1.03	0.062	0.98	0.97 – 1.00	**0.024**	1.01	0.99 – 1.03	0.378	1	0.98 – 1.02	0.88
Total bike lane (100 ft)	0.69	0.62 – 0.77	**<0.001**	0.69	0.63 – 0.75	**<0.001**	0.62	0.53 – 0.73	**<0.001**	0.68	0.61 – 0.76	**<0.001**
Crosswalk warning sign (N) *Reference*												
Crosswalk warning sign (Y)	0.73	0.58 – 0.93	**0.009**	0.71	0.57 – 0.87	**0.001**	0.71	0.56 – 0.91	**0.007**	0.75	0.60 – 0.94	**0.013**
Bike and Ped sign (N) *Reference*												
Bike and Ped sign (Y)	2.1	0.96 – 4.58	0.062	1.76	0.95 – 3.27	0.073	2.38	1.04 – 5.48	**0.041**	1.92	0.88 – 4.19	0.1
One-way sign (N) *Reference*												
One-way sign (Y)	0.3	0.14 – 0.63	**0.001**	0.34	0.18 – 0.63	**0.001**	0.35	0.18 – 0.68	**0.002**	0.29	0.14 – 0.59	**0.001**

Predictors	Median HH Income			Job Density			White Population			Service Land Use		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
Traffic signal presence (N) *Reference*												
Traffic signal presence (Y)	0.11	0.02 – 0.76	**0.025**	0.76	0.38 – 1.52	0.444	0.15	0.03 – 0.89	**0.037**	0.23	0.06 – 0.93	**0.04**
Bus ridership (1000/acre)	0	0.00 – 0.04	**0.004**	0.02	0.00 – 2.13	0.098	0	0.00 – 0.07	**0.009**	0	0.00 – 0.12	**0.011**
Population density (10/acre)	0.2	0.16 – 0.24	**<0.001**	0.28	0.24 – 0.33	**<0.001**	0.21	0.17 – 0.25	**<0.001**	0.2	0.16 – 0.24	**<0.001**
Public school enrollment (1000 counts)	0.79	0.59 – 1.04	0.089	1	0.77 – 1.29	0.988	0.76	0.56 – 1.03	0.079	0.99	0.76 – 1.30	0.966
Park presence (N) *Reference*												
Park presence (Y)	0.58	0.49 – 0.68	**<0.001**	0.63	0.54 – 0.74	**<0.001**	0.57	0.48 – 0.67	**<0.001**	0.58	0.49 – 0.68	**<0.001**
Total trail length (1000ft)	0.64	0.58 – 0.71	**<0.001**	0.71	0.66 – 0.77	**<0.001**	0.66	0.59 – 0.73	**<0.001**	0.67	0.61 – 0.74	**<0.001**
Park and Ride presence (N) *Reference*												
Park and Ride presence (Y)	2.66	1.57 – 4.53	**<0.001**	3.7	2.16 – 6.35	**<0.001**	3.12	1.79 – 5.41	**<0.001**	2.9	1.71 – 4.93	**<0.001**
Manufacture land use (%)	0.97	0.93 – 1.01	0.102	1.11	1.07 – 1.15	**<0.001**				0.99	0.95 – 1.02	0.46
Transportation land use (%)	0.97	0.95 – 0.99	**0.003**				0.96	0.94 – 0.99	**0.002**	0.96	0.94 – 0.98	**<0.001**
Median household income ($50,000)	0.68	0.59 – 0.79	**<0.001**									
Job density (100/acre)				0	0.00 – 0.00	**<0.001**						
White population (%)							1.01	1.01 – 1.02	**<0.001**			
Service land use (%)										0.97	0.96 – 0.98	**<0.001**
Observations	14073	14073	14073	14073								
AIC	71091.46	70690.47	71166.32	70746.18								

Predictors	**Median HH Income**			**Job Density**			**White Population**			**Service Land Use**		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
Log-Likelihood	-35502.73	-35304.24	-35542.16	-35330.09								

Note – IRR: Incident Rate Ratios; CI: Confidence Intervals; p=p-value

Table F 2. Dataset 2, intersections with 10+ walking bouts (n=3,047)

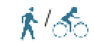
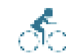
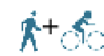
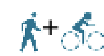
Predictors	**NB – Median HH Income**			**NB – Job Density**			**NB – White Population**			**NB – Service Land Use**		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
(Intercept)	11.21	8.79 – 14.29	**<0.001**	13.18	10.04 – 17.31	**<0.001**	6.39	4.87 – 8.39	**<0.001**	5.63	4.37 – 7.27	**<0.001**
Avg roadway width (10ft)	1.01	1.00 – 1.01	**0.001**	1.01	1.00 – 1.01	**<0.001**	1.01	1.00 – 1.01	**0.003**	1.01	1.00 – 1.01	**<0.001**
Max speed limit – 20 mph *Reference*												
Max speed limit - 25 mph	1.36	1.28 – 1.45	**<0.001**	1.25	1.18 – 1.34	**<0.001**	1.40	1.31 – 1.49	**<0.001**	1.34	1.25 – 1.43	**<0.001**
Max speed limit - 30 mph	0.95	0.88 – 1.02	0.169	1.00	0.93 – 1.08	0.956	0.94	0.87 – 1.01	0.094	0.88	0.82 – 0.95	**0.002**
Max speed limit - 35 mph	1.09	0.94 – 1.26	0.250	1.15	1.00 – 1.32	**0.043**	1.20	1.03 – 1.39	**0.014**	1.03	0.89 – 1.19	0.720
Max speed limit - 40 mph	1.11	0.91 – 1.36	0.333	1.24	1.02 – 1.51	**0.030**	1.06	0.87 – 1.31	0.550	1.12	0.92 – 1.38	0.270
Max slope (%)				0.99	0.99 – 1.00	**0.001**						
Total bike lane (100 ft)				1.01	1.00 – 1.02	0.147	1.01	1.00 – 1.02	**0.034**			
Bike and Ped sign (N) *Reference*												
Bike and Ped sign (Y)				1.07	0.99 – 1.16	0.077						
Curve warning sign (N) *Reference*												
Curve warning sign (Y)				1.10	1.01 – 1.19	**0.024**				1.07	0.98 – 1.16	0.120
Stop sign (N) *Reference*												
Stop sign (Y)	0.83	0.79 – 0.87	**<0.001**	0.87	0.83 – 0.92	**<0.001**	0.80	0.76 – 0.85	**<0.001**	0.84	0.80 – 0.89	**<0.001**

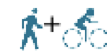
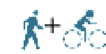
Predictors	NB – Median HH Income			NB – Job Density			NB – White Population			NB – Service Land Use		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
One-way sign (N) *Reference*												
One-way sign (Y)	0.95	0.90 – 1.01	0.079	0.78	0.73 – 0.82	**<0.001**				0.95	0.90 – 1.01	0.070
Traffic circle, bump, no turn signs (N) *Reference*												
Traffic circle, bump, no turn signs (Y)				0.95	0.90 – 1.01	0.082						
Traffic signal presence (N) *Reference*												
Traffic signal presence (Y)	1.19	1.12 – 1.28	**<0.001**	1.11	1.04 – 1.19	**0.001**	1.22	1.14 – 1.30	**<0.001**	1.18	1.10 – 1.26	**<0.001**
Bus ridership (1000/acre)	1.24	1.20 – 1.29	**<0.001**	1.15	1.12 – 1.19	**<0.001**	1.26	1.22 – 1.30	**<0.001**	1.23	1.19 – 1.28	**<0.001**
Population density (10/acre)	1.20	1.18 – 1.23	**<0.001**	1.26	1.23 – 1.28	**<0.001**	1.25	1.22 – 1.27	**<0.001**	1.25	1.22 – 1.27	**<0.001**
Public school enrollment (1000 counts)	0.86	0.81 – 0.92	**<0.001**	0.92	0.87 – 0.98	**0.013**	0.84	0.79 – 0.90	**<0.001**	0.84	0.79 – 0.90	**<0.001**
Park presence (N) *Reference*												
Park presence (Y)	1.08	1.03 – 1.13	**0.002**							1.07	1.03 – 1.12	**0.002**
Total trail length (1000ft)	0.94	0.93 – 0.96	**<0.001**	0.95	0.94 – 0.97	**<0.001**	0.94	0.93 – 0.96	**<0.001**	0.95	0.93 – 0.96	**<0.001**
Park and Ride presence (N) *Reference*												
Park and Ride presence (Y)	0.56	0.45 – 0.72	**<0.001**	0.67	0.54 – 0.83	**<0.001**	0.51	0.41 – 0.64	**<0.001**	0.58	0.46 – 0.73	**<0.001**
Manufacture land use (%)	0.98	0.97 – 1.00	**0.036**	0.99	0.98 – 1.00	0.140				0.99	0.97 – 1.00	0.109
Transportation land use (%)	1.01	1.00 – 1.01	0.147									
Pedestrian average age	1.02	1.01 – 1.02	**<0.001**	1.01	1.00 – 1.01	**0.001**	1.02	1.01 – 1.02	**<0.001**	1.02	1.01 – 1.02	**<0.001**
Pedestrian gender - female (%)				1.00	1.00 – 1.00	**0.008**	1.00	1.00 – 1.00	0.121	1.00	1.00 – 1.00	**0.007**

Predictors	NB – Median HH Income			NB – Job Density			NB – White Population			NB – Service Land Use		
	IRR	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***	***IRR***	***CI***	***p***
Pedestrian race - non-white (%)	0.99	0.99 – 0.99	**<0.001**	0.99	0.99 – 1.00	**<0.001**	0.99	0.99 – 1.00	**<0.001**	0.99	0.99 – 1.00	**<0.001**
Pedestrian employment - yes (%)				1.00	1.00 – 1.00	**0.025**						
Pedestrian median income - 40K-69K *Reference*												
Pedestrian median income - < 40K	0.56	0.51 – 0.61	**<0.001**	0.63	0.58 – 0.68	**<0.001**	0.61	0.56 – 0.67	**<0.001**	0.56	0.51 – 0.61	**<0.001**
Pedestrian median income - 70K-99K	0.77	0.74 – 0.81	**<0.001**	0.77	0.73 – 0.80	**<0.001**	0.79	0.75 – 0.83	**<0.001**	0.77	0.73 – 0.81	**<0.001**
Pedestrian median income - >100K	0.89	0.82 – 0.96	**0.002**	0.93	0.87 – 1.00	**0.034**	0.88	0.81 – 0.95	**0.001**	0.91	0.84 – 0.98	**0.010**
Pedestrian single household - yes (%)	1.00	1.00 – 1.00	**<0.001**				1.00	1.00 – 1.00	**0.001**	1.00	1.00 – 1.00	**<0.001**
Median household income ($50,000)	0.80	0.76 – 0.84	**<0.001**									
Job density (100/acre)				1.37	1.33 – 1.41	**<0.001**						
White population (%)							1.00	1.00 – 1.00	**<0.001**			
Service land use (%)										1.01	1.01 – 1.01	**<0.001**
Observations	3045			3045			3045			3045		
R^2 Nagelkerke	0.783			0.835			0.772			0.794		
AIC	25203.180			24878.977			25263.006			25134.978		
log-Likelihood	-12576.590			-12410.488			-12608.503			-12541.489		

Appendix G. Comparison of pedestrian and cyclist counting procedures

Table G 1. Comparison pedestrian and cyclist counting procedures

Technology	Strengths	Weaknesses	Typical Applications	User Type	Commercially Available Equipment	Cost
Inductance Loop	• Accurate when properly installed and configured • Uses traditional motor vehicle counting technology	• Capable of counting bicyclists only • Requires saw cuts in the existing pavement or pre-formed loops in new pavement construction • May have a higher error with groups	Permanent counts	 Or 	• Eco-Counter ZELT • RoadsysHI-TRAC CMU	$$
Magnetometer	May be possible to use existing motor vehicle sensors	• Commercially-available, off-the-shelf products for counting bicyclists are limited • May have a higher error with groups	Permanent counts		• Econolite AccuSense Mag	$-$$
Pressure sensor/pressure mats	Some equipment may be able to distinguish bicyclists and pedestrians	• Expensive/disruptive for installation under asphalt or concrete pavement	Permanent counts Typically unpaved trails or paths		• Eco-Counter SLAB • TRAFx Mountain Bike Counter	$$
Seismic sensor	Equipment is hidden from view	• Commercially-available, off-the-shelf products for counting are limited	Short-term counts on unpaved trails		• Diamond Traffic Traffic Tally 6 • Eco-Counter TUBE • Jamar TRAX Cycles Plus • MetroCount RidePod BT • TimeMark Delta NT, TimeMark Gamma NT	$$
Radar sensor	Capable of counting bicyclists in dedicated bike lanes or bikeways	• Commercially-available, off-the-shelf products for counting are limited	Short-term or permanent counts		• Econolite AccuSense MicroRadar • Roadsys SDR • Sensys Networks MicroRadar	$-$$

Technology	Strengths	Weaknesses	Typical Applications	User Type	Commercially Available Equipment	Cost
Video Imaging – Automated	Potential accuracy in dense, high-traffic areas	• Typically more expensive for exclusive installations • Algorithm development still maturing	Short-term or permanent counts		• Miovision Scout • Numina • Placemeter	$-$$
Infrared – Active	Relatively portable Low profile, unobtrusive appearance	• Cannot distinguish between bicyclists and pedestrians unless combined with another bicycle detection technology • Very difficult to use for bike lanes and shared lanes May have a higher error with groups	Short-term or permanent counts		• Diamond Traffic Trail Counter TTC-4420 • TrailMaster TM1550	$-$$
Infrared – Passive	Very portable with easy setup Low profile, unobtrusive appearance	• Cannot distinguish between bicyclists and pedestrians unless combined with another bicycle detector • Difficult to use for bike lanes and shared lanes requires careful site selection and configuration • May have a higher error when ambient air temperature approaches body temperature range • May have a higher error with groups • Direct sunlight on the sensor may create false counts	Short-term or permanent counts		• Eco-Counter PYRO • Roadsys HI-TRAC CMU • TRAFx Trail Counter	$-$$

Technology	Strengths	Weaknesses	Typical Applications	User Type	Commercially Available Equipment	Cost
Pneumatic Tube	Relatively portable, low-cost May be possible to use existing motor vehicle counting technology and equipment	• Capable of counting bicyclists only • Tubes may pose a hazard to trail users • Greater risk of vandalism	Short-term counts Bicyclists only		• Diamond Traffic Traffic Tally • Eco-Counter TUBE • Jamar TRAX Cycles Plus • MetroCount RidePod BT • TimeMark Delta NT, TimeMark Gamma NT	$-$$
Video Imaging – Manual Reduction	Can be a lower cost when existing video cameras are already installed	• Limited to short-term use • Manual video reduction is labor-intensive	Short-term counts Bicyclists and pedestrians separately		• CountingCars.com CountCloud • Miovision Scout • Various consumer video cameras with manual • reduction	$-$$$
Manual Observer	Very portable Can be used for automated equipment validation	Expensive and possibly inaccurate for longer duration counts	Short-term counts Bicyclists and pedestrians separately		n/a	$$-$$$

Bicyclists and pedestrians combined

Bicyclists and pedestrians separately

Bicyclists only

Pedestrian only

DOT HS 813 583
July 2024

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA

16282-071724-v4