Development of an arterial-based pedestrian exposure and crash risk model for North Carolina

NCDOT Project 2025-08
FHWA/NC/2025-08
February 2026

Vikash V. Gayah, S. Ilgin Guler, Xiaohan Gu – Penn State University
Ian Hamilton, Lauren Blackburn – VHB, Inc.

PennState College of Engineering
LARSON TRANSPORTATION INSTITUTE

vhb

RESEARCH & DEVELOPMENT

Development of An Arterial-Based Pedestrian Exposure and Crash Risk Model for North Carolina

FINAL REPORT

Submitted to:
North Carolina Department of Transportation
Office of Research
(Research Project No. RP2025-08)

Submitted by

Vikash V. Gayah
The Pennsylvania State University
218 Engineering Collaborative Research and Education (ECoRE) Building
556 White Course Drive
University Park PA 16802
814-865-4014
gayah@engr.psu.edu

The Pennsylvania State University

February 2026

Technical Report Documentation Page

1. Report No. FHWA/NC/2025-08	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle *Development of an arterial-based pedestrian exposure and crash risk model for North Carolina*		5. Report Date February 2026
		6. Performing Organization Code
7. Author(s) Vikash V. Gayah, S. Ilgin Guler, Xiaohan Gu, Ian Hamilton, Lauren Blackburn		8. Performing Organization Report No.
9. Performing Organization Name and Address *The Pennsylvania State University (Larson Transportation Institute)* *Vanasse Hangen Brustlin, Inc*		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address North Carolina Department of Transportation Research and Development Unit 1549 Mail Service Center Raleigh, North Carolina 276699-1549		13. Type of Report and Period Covered *August 2024 – February 2026*
		14. Sponsoring Agency Code *RP2025-08*
15. Supplementary Notes		

16. Abstract

Pedestrian safety is an increasingly critical concern for transportation agencies in North Carolina and throughout the United States. Although pedestrians account for a relatively small share of overall travel and are involved in less than one percent of reported crashes in North Carolina, they represent a disproportionately high share of severe outcomes, comprising approximately 15 percent of traffic fatalities and serious injuries. Pedestrian fatalities in the state have continued to rise in recent years, consistent with national trends, highlighting the need for improved methods to identify high-risk locations and proactively address pedestrian safety. A key limitation of many systemic safety analyses is the lack of pedestrian exposure information, which is essential for distinguishing locations with elevated crash risk from those that simply experience higher levels of pedestrian activity.

The objective of this project was to enhance systemic pedestrian safety analysis for arterial roadway segments in North Carolina by explicitly incorporating pedestrian exposure into crash risk modeling. Specific objectives included identifying critical locations for pedestrian data collection, developing pedestrian exposure models for arterial roadways, and integrating exposure estimates into updated pedestrian crash risk factor models. The analysis focused on urban arterial segments with four or more lanes, higher traffic volumes, and limited access control. Both direct pedestrian count models and categorical exposure models were developed, with the categorical model—classifying pedestrian activity as low, medium, or high—recommended for practical application due to its strong performance and ease of implementation. The resulting exposure-adjusted crash risk models incorporate roadway, traffic, land use, transit, and socioeconomic factors and are generally consistent with prior findings, while representing a substantial methodological improvement through the inclusion of pedestrian activity. The final outcome of this research is a quantitative, exposure-based risk scoring framework that supports more accurate identification and prioritization of high-risk arterial segments. This framework provides a practical tool for the North Carolina Department of Transportation and local agencies to improve data-driven decision-making and target pedestrian safety investments more effectively.

17. Keywords *Pedestrian safety* *Pedestrian risk factors* *Pedestrian exposure*	18. Distribution Statement		
19. Security Classif. (of this report) unclassified	20. Security Classif. (of this page) unclassified	21. No. of Pages 46	22. Price

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

DISCLAIMER

The contents of this report reflect the views of the authors and are not necessarily the views of Pennsylvania State University. The authors are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the North Carolina Department of Transportation at the time of publication. This report does not constitute a standard, specification, or regulation.

ACKNOWLEDGEMENTS

This research was sponsored by the North Carolina Department of Transportation. The research team would like to thank the NCDOT technical panel that oversaw this work, including Daniel Carter, John (Hart) Evans, John Vine-Hodge, Brian Murphy and Celeste Semanick.

EXECUTIVE SUMMARY

Pedestrian safety remains a critical and growing concern for transportation agencies in North Carolina and nationwide. Although walking accounts for a relatively small share of total trips and pedestrians are involved in less than one percent of reported crashes in North Carolina, pedestrians represent a disproportionately high share (15 percent) of severe injury outcomes. Pedestrian fatalities in North Carolina have continued to rise, reaching a peak of 280 deaths in 2024. This trend mirrors national patterns, with more than 7,300 pedestrian fatalities reported across the United States in 2023. These statistics underscore the urgent need for improved methods to identify high-risk locations and better targeted pedestrian safety investments.

A fundamental challenge in pedestrian safety analysis is accounting for pedestrian exposure – the level of pedestrian activity at a given location – which is essential for distinguishing between locations that are inherently risky and those that simply experience higher pedestrian volumes. Prior research conducted under NCDOT Research Project 2022-11 identified several roadway and land use factors associated with severe pedestrian crashes on urban roadways. However, the effort was unable to explicitly incorporate pedestrian exposure due to limited and non-systematic pedestrian count data. Existing pedestrian counts were largely collected for traffic monitoring or project development purposes and were not designed to capture the underlying determinants of pedestrian activity.

The primary objective of this project was to address this limitation by developing updated pedestrian crash risk factors for arterial roadway segments in North Carolina that explicitly account for pedestrian exposure. Specific objectives included: (1) identifying critical arterial locations where pedestrian counts are most needed to support reliable exposure modeling; (2) developing improved pedestrian exposure models for arterial segments; and, (3) integrating pedestrian exposure into updated pedestrian crash risk models to better support systemic safety analysis.

The analysis focuses on target arterial roadway segments defined by the following characteristics: four or more travel lanes, average annual daily traffic (AADT) greater than 12,500 vehicles per day, posted speed limits of 35 mph or higher, and roadways that are not fully access controlled. Two pedestrian exposure modeling approaches were developed and evaluated: a direct exposure model that predicts daily pedestrian crossing volumes, and a categorical exposure model that classifies pedestrian activity into low (0-30 pedestrians per day), medium (31-200 pedestrians per day), or high (greater than 200 pedestrians per day) categories. While both approaches produced reasonable results, the categorical exposure model demonstrated strong performance and offers a simpler, more practical tool for implementation. As such, it is recommended for use by NCDOT in estimating pedestrian activity along target arterial segments.

The exposure modeling results indicate that pedestrian activity on arterial roadways is influenced by vehicular traffic volume, roadway functional classification, parcel density, high-intensity land use, proximity to colleges or universities, parks, or greenways, the presence of pedestrian infrastructure (sidewalks, crosswalks, and signals), non-motorized population characteristics, school presence, posted speed limits, alcohol sales locations, household vehicle ownership, educational attainment, age distribution, and median household income.

Updated pedestrian crash risk models were then developed incorporating the categorical pedestrian exposure measure alongside other established risk factors. These include vehicular traffic volume, number of lanes, high-intensity development near the roadway, densities of alcohol sales establishments and bus stops, population and school enrollment density, block length, and median household income. The results are generally consistent with findings from the earlier NCDOT research, but the explicit inclusion of pedestrian exposure represents a substantial methodological advancement. By accounting for pedestrian activity levels, the updated models provide a more accurate and defensible basis for identifying high-risk locations and prioritizing safety countermeasures. The final output of this project is a quantitative risk scoring framework that enables NCDOT to compute pedestrian crash risk for arterial segments using readily available roadway, land use, and demographic data. This framework supports more informed, exposure-adjusted systemic safety analyses and provides a practical tool to guide data collection, project prioritization, and investment decisions aimed at reducing pedestrian fatalities and serious injuries across North Carolina.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

Pedestrian safety is a significant concern for transportation planners and safety engineers both within North Carolina and across the country. While a relatively small number of trips are made via walking, and pedestrians were involved in less than 1 percent of crashes in North Carolina, pedestrians represented roughly 15 percent of traffic fatalities or serious injuries. As a result, pedestrian safety is an important emphasis area in North Carolina's Strategic Highway Safety Plan (SHSP; NCDOT, 2024a). Further, pedestrian fatalities have continued to increase in the state, with a peak total of 280 in 2024. More broadly across the United States, pedestrian fatalities are increasing annually; in fact, annual pedestrian fatalities within the U.S. claimed the lives of more than 7,300 people in 2023. To better identify factors that contribute to pedestrian fatalities/injuries and identify the highest risk locations for these crashes, it is critical to understand which locations have the most pedestrian activity so that the most risk-prone locations can be identified.

As part of the research report *2022-11: Quantification of Systemic Risk Factors for Pedestrian Safety on North Carolina*, the North Carolina Department of Transportation (NCDOT) discovered several risk factors that were most associated with serious injuries and fatalities among pedestrians involved in a crash on urban roadways. However, the final risk factor analysis did not include a critical aspect of pedestrian crash risk – pedestrian crossing activity or exposure – due to a lack of available data. The research collected traffic counts, including pedestrian counts, from NCDOT and local governments, but many of these counts reflected sites selected for annual traffic volume monitoring or broader project development; further, these sites were not selected to understand the dynamics that influence pedestrian activity, but rather as a secondary consideration to traffic monitoring and project development.

The overall objective of this project was to develop updated risk factors for arterial roadway segments in North Carolina, specifically considering pedestrian exposure at these locations. Specifical objectives to accomplish this goal included: 1) identifying critical locations where pedestrian counts are needed to develop a more reliable pedestrian exposure model for arterials and obtain pedestrian crossing (i.e., exposure) information from these locations; 2) developing an updated pedestrian exposure model for arterial segments in North Carolina; and, 3) incorporating this exposure estimate into updated pedestrian risk factor models for North Carolina arterials.

The remainder of this document is organized as follows:

- A literature review on pedestrian risk factors and exposure modeling.
- A summary of data sources and database development.

- A review of pedestrian exposure and risk factor estimation results.

The document concludes with a discussion of how these factors can be used to enhance NCDOT's safety planning and analysis process.

Chapter 2. Review of research literature

As a first step, the research team gathered information on existing practices and guidance related to modeling and the use of pedestrian risk factors from state transportation agencies and journal articles. This review was an update of the original *2022-11: Quantification of Systemic Risk Factors for Pedestrian Safety on North Carolina* research project. For simplicity, only the key takeaways that summarize how the literature review findings were used to inform the present research project are provided in the body of the report. The full literature review can be found in Appendix A.

The literature review revealed several key insights that informed the present project:

- A variety of statistical approaches have been used to estimate systemic risk factors for pedestrian safety performance on roadways. Of these, the most prevalent methods are binary logistic regression, which is used to estimate the risk of one or more pedestrian crashes occurring at a given location during a specified period of time. NB regression is used to estimate the number of pedestrian crashes that occur at a given location during certain periods of time. These methods are particularly useful because they provide an interpretable relationship between model coefficients and changes in crash risk or frequency. Both methods were considered in this project to develop pedestrian risk factors, with NB regression ultimately selected because the dependent variable was crash frequency (rather than the probability of a crash occurring on a segment).
- Common pedestrian risk factors identified in the literature (both academic literature and available state practices) and considered for inclusion in this project include the following:
    - Vehicular traffic volume and composition
    - Functional classification
    - Number of vehicle travel lanes
    - Presence of two-way left-turn lane
    - Speed limit
    - Presence of lighting
    - Presence of pedestrian features (e.g., sidewalk, protected median, crosswalk, leading pedestrian interval)
    - Presence of transit stops
    - Presence of on-street parking
    - Proximity to school
    - Number of alcohol sales
    - Land use mix

    - Street network connectivity
    - Area type (e.g., urban vs. rural; large city vs. small town)
- Pedestrian exposure also plays a large role in pedestrian safety risk. However, this exposure is generally difficult to represent directly across an entire roadway network. Instead, surrogates of exposure should be considered to control for the level of pedestrian activity at a given location. These include the following, which were considered for inclusion in this project:
    - Population density
    - Presence and/or proximity to schools
    - Presence of transit stops
    - Land use mix
    - Population characteristics (e.g., walking mode share, fraction of households without a vehicle, income)
- Pedestrian exposure models can be developed using NB regression or linear regression models to predict the level of pedestrian activity as a function of the above-listed surrogates. In this project, NB regression was selected due to its ability to handle count data.
- Failure to properly account for pedestrian exposure might lead to counterintuitive findings related to risk; for example, sidewalks might be associated with increased pedestrian crash risk because their presence typically indicates higher pedestrian activity. As a result, the research team considered various means to account for pedestrian exposure more accurately. Accordingly, this research included pedestrian exposure surrogates, as well as the development of pedestrian exposure models to predict pedestrian activity at locations where sufficient information was available.

Chapter 3. Analysis database development

The second task in this research was to gather publicly available data that could be used to estimate pedestrian risk factors as a part of this study. This section summarizes the data elements that were included as part of this process and were available for use in this project.

3.1 Data summary

Table 1 provides a summary of the data elements that were described in the proposal. Where data were available, these data elements were merged in a GIS format so individual elements could be spatially joined for analysis purposes. The remainder of this section provides additional details on these data.

Table 1. Summary of data collected

<table>
<tr><th>Data source</th><th>Data elements</th></tr>
<tr><td rowspan="5">NCDOT or other NC source</td><td>Pedestrian crashes</td></tr>
<tr><td>Pedestrian counts</td></tr>
<tr><td>Roadway segment data
o Functional classification
o Number of lanes
o Median presence
o AADT
o Posted speed limit</td></tr>
<tr><td>Pedestrian infrastructure data
o Sidewalk presence
o Shared use path presence
o Greenway presence
o Crosswalk presence</td></tr>
<tr><td>Pedestrian destination data
o Public parks
o Supplemental Nutrition Assistance Program (SNAP) benefits establishments
o K-12 schools
o University and college campuses
o Transit stops and routes</td></tr>
</table>

Data source	Data elements
	Intersection data o Geometry o Number of approaches o Control type (signalized or unsignalized) o Major / minor annual average daily traffic (AADT) o Posted speed limits o Number of lanes o Median presence
	Transportation Disadvantage Index
	Land parcels
U.S. Census Bureau	Socioeconomic/demographic data o Total population o Population by age o Population by race o Educational attainment o K-12 school enrollment by place of residence o Unemployment by place of residence o Median household income o Disabled population o Vehicle ownership by household o Poverty rate o Limited English proficiency households o Commute by mode o Total employment o Employment by industry; North American Industry Classification System (NAICS)
Other	Land cover information
	Alcohol sales establishments
	Building footprints (by building use)

3.1.1 Pedestrian crashes

Pedestrian crash information was obtained directly from NCDOT via ArcGIS Online. The research team identified all pedestrian crashes that occurred between 2007 and 2023, along with information on the exact crash location, injury severity level, action, and relative location of each unit involved in the crash. In 2020, the database also includes classifications for special pedestrian types (e.g., wheelchairs, skateboards, or scooters). The locations were geocoded so that crashes could be assigned to specific roadway segments or corridors, as needed.

3.1.2 Roadway characteristics

Roadway characteristic data were obtained directly from NCDOT. For this project, the research team identified the locations of all roadways within the NCDOT network, with roadways dynamically segmented, according to available attribute data. The project team re-segmented road centerlines to begin and end between intersections (i.e., block-length segments), with the dominant characteristic representing the segment. For example, if 60 percent of a segment had two lanes and the other 40 percent had four, the segment was classified as a two-lane segment.

Road characteristic data are primarily available for "On-System" roads (i.e., U.S. Routes, NC Routes, and Secondary Routes), although centerline and route information are available for all roads. Specific data elements associated with each roadway segment included the functional classification, number of lanes, AADT, median presence (and type), and posted speed limit. AADT data for 2015 to 2023 were obtained from NCDOT. Traffic volumes for On-System roadway segments were also obtained from NCDOT and merged with the individual roadway segments. For "Off-System" (i.e., Non-System) routes, a snapshot of AADT from 2022 was available from Streetlight via NCDOT.

3.1.3 Pedestrian infrastructure

Information on pedestrian infrastructure was obtained from NCDOT's PBIN via Connect NCDOT[1]. Specific data elements included the presence of sidewalks, shared use paths, greenways, and crosswalks. The comprehensiveness of these data varies by portion of the state and part of the road network, but sidewalks along "On-System" roads are generally reliable statewide.

3.1.4 Pedestrian surrogate information

The research team obtained locations of parks, stores that accept SNAP benefits, K-12 schools, and university campuses from NCDOT. These data were developed to support the state's Advanced Transportation through Linkages, Automation and Screening (ATLAS) program, as these features are potentially associated with increased pedestrian activity. In addition, the research team obtained information on transit stops and routes, since transit usage is typically associated with pedestrian activity. A comprehensive inventory of fixed-route transit stops statewide was available for the year 2022, developed

[1] https://connect.ncdot.gov/projects/BikePed/Pages/PBIN.aspx

as part of the Access in Appalachia Pilot Study. The research team attempted to obtain information on micromobility usage (e.g., shared bikes, e-bikes, and e-scooters), but these data were not available in a useful format. Therefore, micromobility data were not obtained for this project.

3.1.5 Intersection information

Intersection characteristics were obtained for every intersection within North Carolina from the NCDOT Mobility and Safety Unit's Intersection Inventory. Specific data elements that were obtained include intersection geometry, number of approach legs, traffic control type (primarily signalized or unsignalized), major and minor leg traffic volumes, number of lanes, presence of a median, and pedestrian signal presence.

3.1.6 Demographic, socioeconomic, and employment data

The research team obtained sociodemographic and socioeconomic data from the U.S. Census Bureau's five-year ACS[2] to serve as additional indicators and surrogates correlated with pedestrian exposure. These data were available at the Census block group level and contain information on population by various categories, such as race, age, sex, education level, employment status, income, poverty level, school enrollment status, English proficiency, primary commute mode, and disability status. Block groups can be aggregated to develop estimates at larger geographic levels, such as tracts, counties, or the state as a whole. The project team also obtained NCDOT's TDI scores for each block group. TDI inputs are derived from the ACS, but scores indicating the relative level of disadvantage for each block group are available and scaled to the state, NCDOT division, and county levels.

Employment data were also obtained from the Census's LEHD program[3]. This provides estimates of employment by industry (i.e., NAICS code) and place of work at the Census block level. Each roadway segment was assigned the data elements associated with the Census block group that covered the majority of the segment.

---

[2] https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.html

[3] https://lehd.ces.census.gov/data/

3.1.7 Parcel footprints

The research team obtained North Carolina land parcel data from NC One Map. Parcel size and density are surrogates for urbanism and land use intensity and may indicate pedestrian exposure and trip generation.

3.1.8 Land cover

The research team obtained land cover data from the MRLC[4]. This data set reflects the type of vegetation and human disturbance for a given area, and heavily urbanized areas stand out as developed relative to natural (e.g., vegetation or bare rock) surfaces. This includes a raster layer depicting annual land cover for 2023. The research team flagged segments based on three land cover classifications: Developed High Intensity, Developed Low Intensity, and Developed Medium Intensity. If a land cover classification was within 100 feet of a segment, the team assigned that classification to the segment.

3.1.9 Alcohol sales establishments

The research team has access to alcohol sales establishment location data obtained from Data Axel in December 2021. This dataset includes a list of business locations that are likely to sell alcohol based on the industry associated with that business (i.e., NAICS code). These include drinking places (7224); beer, wine, and liquor stores (4453); convenience stores (4451); full-service restaurants (722511); and limited-service restaurants (722513).

3.1.10 Building footprints

The research team obtained information on building footprints from the Federal Emergency Management Agency's (FEMA's) Geospatial Resource Center. Using spatial building footprints, the research team can calculate the total number of buildings (by land use type), as well as the square footage of buildings (by type) within a distance of a segment or intersection.

[4] https://www.mrlc.gov/data?f%5B0%5D=project_tax_term_term_parents_tax_term_name%3AAnnual%20NLCD

3.1.11 Pedestrian counts

Existing counts

Approximately 3,600 pedestrian exposure counts were obtained as part of *NCDOT 2022-11*. These counts were obtained from the following sources:

- Project-level counts collected as part of motor vehicle turning movement counts (TMCs), segment counts, zone counts, or other analyses.
- Counts obtained from the City of Charlotte as part of their TMC count program.
- Counts obtained as part of a downtown Raleigh pedestrian safety study conducted by NCDOT.
- Counts provided by Greensboro DOT/Greenville Urban Area Metropolitan Planning Organization.
- Counts provided by the Gaston-Cleveland-Lincoln Metropolitan Planning Organization.
- Counts provided by the City of Durham.

As part of this project, the research team identified additional counts to supplement these previous counts. The primary source of these additional counts was the NCDOT NMVDP. NCDOT also provided the research team with a previously conducted scan of municipalities and regional planning organizations, and the availability of pedestrian count data at each agency. As part of this project, the research team contacted each data owner included in this inventory to obtain pedestrian counts along segments or at intersections for potential inclusion. Of the 30 individual agencies contacted, the team received responses from 9. As a result, more than 1,200 new pedestrian counts were obtained for use as part of this project. These include:

- 90 counts from the NCDOT Pedestrian Corridor Crossing dataset
- Approximately 300 TMC pedestrian counts from the Town of Cary (from about 250 unique locations)
- Approximately 900 TMC pedestrian counts from the City of Greensboro (from about 550 unique locations)
- Approximately 320 counts obtained from the North Carolina Open Data Portal (https://ncdataportal.org/)
- Counts from approximately 80 unique locations from the NMVDP; see Figure 1. These data consist of approximately 370,000 individual daily counts spanning the years 2014 to 2024. Multiple data streams exist at many locations (each representing a unique counter), providing multiple count observations for a given day at a single location.

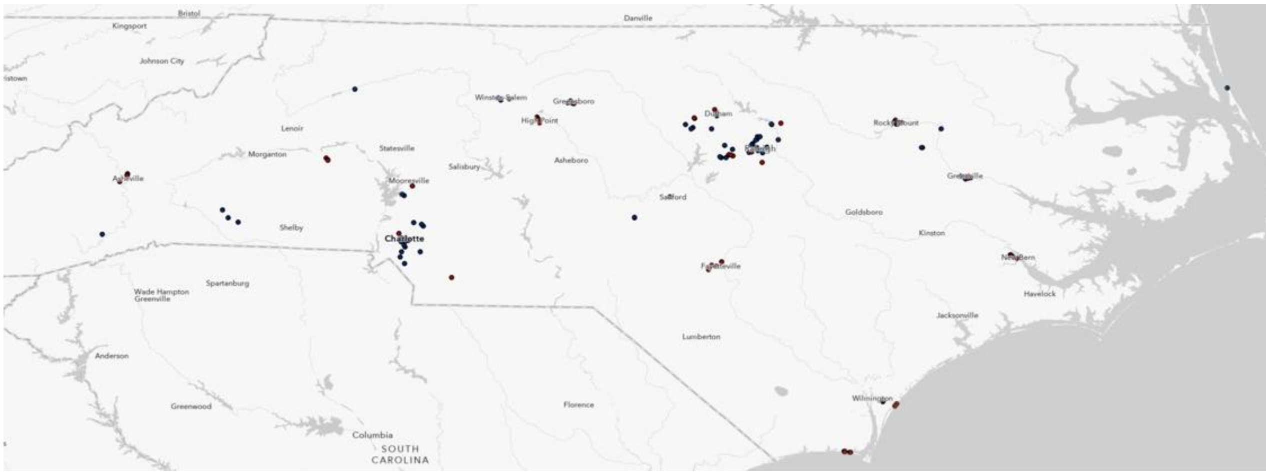

Figure 1. Location of pedestrian counts from NMVDP

Note that some counts – particularly those from the NMVDP – were obtained from trails or walking paths and did not reflect pedestrian activities near roadway segments. Counts that could not be directly attributed to a roadway segment were not used as part of the exposure modeling in this project.

<u>New counts as a part of this project</u>

The research team considered the statewide distribution of available pedestrian counts to identify where new counts should be performed. Specifically, the team compared the distribution of count locations with the distribution of population and arterial length by NCDOT division. The research team emphasized divisions that were underrepresented for additional counts; underrepresented divisions were those with a smaller share of available pedestrian counts than their share of the state's total population and arterial mileage.

Table 2 provides a summary of this analysis. The last two columns represent the difference between the percentage of pedestrian counts and the percent of population or arterial length, respectively, for a given NCDOT division. As shown, most differences were nearly zero, suggesting a balanced distribution of pedestrian counts by population or arterial length. However, some outliers existed. For example, NCDOT Division 7 contained 17 percent of the pedestrian count sites, but just nine percent of the total population and nine percent of the total arterial length. Therefore, adding additional counts in this division would only exacerbate existing imbalances across the state. Based on this summary, the research team recommended additional counts in NCDOT Divisions 9, 8, and 12 in particular.

Table 2. Relative representation of pedestrian counts vs. population and arterial length by NCDOT division

Division	# count sites	% count sites	Arterial length (mi)[5]	Arterial length (%)	Total pop.	% pop.	Relative rep. by arterial mileage	Relative rep. by pop.
1	133	2%	636.2	6%	251,096	2%	-4%	0%
2	256	5%	577.1	6%	485,823	5%	-1%	0%
3	523	9%	732.9	7%	770,186	7%	2%	2%
4	253	4%	722.6	7%	621,089	6%	-3%	-1%
5	840	15%	1073.9	11%	1,741,267	16%	4%	-1%
6	403	7%	667.3	7%	674,581	6%	1%	1%
7	980	17%	953.0	9%	989,398	9%	8%	8%
8	196	3%	708.9	7%	551,582	5%	-4%	-2%
9	128	2%	588.5	6%	804,287	8%	-4%	-5%
10	836	15%	854.1	8%	1,718,191	16%	6%	-1%
11	116	2%	535.9	5%	366,126	3%	-3%	-1%
12	370	7%	815.4	8%	828,486	8%	-2%	-1%
13	394	7%	589.2	6%	525,623	5%	1%	2%
14	204	4%	611.7	6%	377,668	4%	-2%	0%

The research team used a data-driven approach to identify candidate sites in those three Divisions for additional field pedestrian count data collection as a part of this project. The focus of this site identification was on urban arterial segments, which are the focus of this project. The following criteria were used to identify candidate sites for additional counts using this data-driven approach:

1. Sites with high pedestrian crash risk
2. Sites without existing pedestrian counts
3. Sites with large discrepancy in predicted exposure compared to surrounding locations

Criterion 1 is defined as arterial segments prioritized as being in the top 10 percent of pedestrian risk using the models developed in *NCDOT 2022-11*. The risk models were first developed for all urban arterials on the roadway network, but the risk scoring was applied to all non-access controlled public roads in the state. This criterion was used to prioritize and better understand pedestrian activity at high pedestrian risk locations.

[5] Includes Other Principal Arterials and Minor Arterials

Criterion 2 removed from consideration sites for which existing pedestrian counts were available (within approximately 1,000 ft), either through the counts available in *NCDOT 2022-11* or additional counts identified or collected as a part of this project. This criterion was used to ensure that additional counts were performed only at new locations where no existing pedestrian count information was available.

Criterion 3 utilized the pedestrian exposure model previously developed in *NCDOT 2022-11* to estimate pedestrian activity at each arterial segment. In addition, pedestrian activity was estimated for all non-arterial segments using the same suite of models. For each arterial segment, the absolute difference between the predicted exposure on that segment and the non-arterial segments in the same census tract was computed. Arterial sites were then ranked based on these absolute differences. This criterion was developed to help identify locations where pedestrian activity appears to decrease along an arterial in an otherwise highly active area (referred to by the research team as the arterial suppression effect). It is generally not known whether this arterial suppression is real in many areas or an artifact of the previous pedestrian exposure model, since pedestrian counts along arterials used to support the exposure model tended to be performed in conjunction with vehicular counts. Performing additional counts at these locations would help determine the true magnitude of the arterial suppression effect and likely support the development of a more robust and accurate pedestrian exposure model.

The research team provided the NCDOT technical panel with a list of the high-priority sites identified in this way. In consultation with the research team, the NCDOT technical panel helped identify a set of 170 sites for additional counts. Ultimately, pedestrian exposure counts were conducted at 163 intersection locations at these sites as a part of this project; 29 counts were funded as part of this research project, and 134 were funded through supplemental funds through NCDOT's Traffic Safety Unit. These counts were performed in March, April, and May 2025. Excluded sites were not counted due to ongoing construction or other unsuitable conditions. A full list of these additional count locations is provided in Appendix B.

3.2 Intersection-level pedestrian exposure database

A total of 6,979 pedestrian counts were available for use in developing the pedestrian exposure model. As shown in Table 3, most of these counts (over 3,500) were obtained within the last four years (2022-2025) and represent newer counts that were not available in the previous NCDOT Project 2022-11.

Table 3. Summary of counts by year performed

Count year	# counts
2010	7
2011	82
2012	118
2013	165
2014	169
2015	156
2016	136
2017	172
2018	564
2019	1,180
2020	113
2021	603
2022	896
2023	1,685
2024	668
2025	262
Missing	3

Table 4 provides a summary of count location by NCDOT division. As shown, the counts are not evenly distributed across NCDOT division. Divisions associated with larger urban areas (5 – Triangle, 7 – Greensboro, and 10 – Charlotte) represent approximately half of the available pedestrian counts in the state.

Table 4. Summary of count locations by NCDOT division

NCDOT Division	# counts
1	164
2	310
3	596
4	342
5	1,014
6	476
7	1,378
8	240
9	183
10	1,078
11	134
12	414
13	433
14	217

Table 5 provides a summary of counts by location type. The majority of counts are at intersection locations, which is expected since most come from turning movement counts.

Table 5. Summary of counts by location type

Location type	# counts
3-leg intersection	2,073
4-leg intersection	4,676
5-leg intersection	49
6-leg intersection	3
Roundabout	22
Midblock	112
U-Turn	37
Other	7

Table 6 provides a summary of on-road counts by duration length in hours.

Table 6. Summary of counts by count duration

Count duration (hours)	# counts
1	3
2	2
3	29
4	3
5	22
6	7
7	6
8	23
9	4
10	28
11	1
12	908
13	4,929
14	16
15	7
16	424
17	1
24	31
48	39
Missing	496

Table 7 provides a summary of counts by the observed pedestrian count value. The majority of counts are indicative of low pedestrian activity. Specifically, 21 percent of counts observed zero pedestrians during the entire count interval, 50 percent observed fewer than 11 pedestrians, and 76 percent observed fewer than 51 pedestrians. However, approximately 3.4 percent of the counts show pedestrian activity exceeding 500 pedestrians during the count interval, and about 1.8 percent show more than 1,000 pedestrians. Accounting for count duration (Table 8), approximately 76 percent of the counts observe fewer than 5 pedestrians per hour.

Table 7. Summary of counts by number of pedestrians

Pedestrian count	# counts within this range
0	2,774
1-10	690
11-50	1,801
51-100	722
101-500	756
501-1,000	108
1,001-5,000	107
5,001-10,000	14
10,000+	7

Table 8. Summary of counts by average hourly pedestrian volume

Hourly pedestrian volume	# counts
0	1,432
0-5	3,903
5-10	517
10-15	202
15-20	93
20-25	61
25-30	37
30+	237
Unknown	497

An analysis database was developed for estimating the pedestrian exposure model using data from locations where pedestrian counts were available. The majority (6,799 out of 6,979) of these counts were performed at intersection locations. Therefore, individual intersections were used as the primary unit of analysis for the development of the pedestrian exposure database to maintain consistency with future pedestrian count efforts that are expected to be undertaken by NCDOT. Each observation within the pedestrian exposure database represented a specific intersection at which the pedestrian count was performed and contained information about the specific count, such as location, facility type, when the count was performed, count duration, and number of pedestrians observed (summed across all individual legs of the intersection). The counts that were not performed at intersections were omitted from this database.

Various other data elements were appended to the count data that may serve as potential explanatory variables in the pedestrian exposure models. Pedestrian infrastructure elements were identified using a 100-ft radius around the intersection location and appended to each observation. Individual roadway segments present at each intersection were identified using GIS, and their characteristics were appended to each observation. The presence of K-12 schools, transit stops, alcohol sales establishments, and parks within a 0.25-mile radius of the intersection were also included in the database. University and college campus presence within a 0.5-mile radius was also identified and included; the larger radius reflects the larger catchment area of universities compared to K-12 schools. Finally, sociodemographic, socioeconomic, and land use data were appended using a 0.25-mile radius around the count location. The proportion of land coverage within each individual census tract was used to develop a weighted average for these metrics. Appendix C provides a data dictionary that contains a list of all specific data elements, along with a short description and how each was coded in this database.

3.3 Segment-level pedestrian crash risk database

A separate analysis database was developed to estimate a pedestrian crash risk model that could quantify the impact of various factors on the risk of pedestrian crashes occurring along arterial segments within North Carolina. Each observation in this database represented a unique roadway segment. Since the focus of this project was pedestrian safety, all full access-controlled roadway segments and ramps were removed from the database; however, partial access segments were retained due to knowledge of key locations in North Carolina (e.g., NC 54 in Carrboro and Independence Boulevard in Charlotte).

Segments were defined from intersection to intersection and, as a result, were not necessarily homogeneous. Characteristics associated with each segment represented the dominant traits for the segment (i.e., those that represent the longest homogeneous section of the segment). The segments only represented the "inventory" direction for all data, as opposed to "non-inventory" segments that represent the opposite direction of bifurcated centerlines. However, to ensure that the segment represented the actual roadway segment conditions, relevant data elements (e.g., number of lanes, AADT, etc.) were combined with the inventory direction. For example, some divided 6-lane two-way roadways were coded in the original NCDOT data as two 3-lane, unidirectional roadway segments. The research team combined these into a single 6-lane segment for this analysis. Socioeconomic and other data obtained at the Census tract level were appended to each segment based on the "most representative" census tract for the segment. This most representative census tract for each segment was identified as the tract within which the majority of the segment falls. Pedestrian exposure values were also appended to this database as described in the Model

Results section. Appendix D provides a data dictionary that contains a list of all specific data elements, along with a short description and how each was coded.

Chapter 4. Analysis

This section describes the analysis that was performed to develop the pedestrian exposure and risk models, which serve as the primary outcomes of this project.

4.1 Scope

The NCDOT technical panel indicated that the preferred scope for the models developed as part of this project was "target" roads, defined as arterial roadways with the following characteristics:

- 4+ lanes
- AADT greater than 12,500 vehicle/day
- 35+ mph speed limits
- Not fully access controlled

These thresholds were provided by the NCDOT technical panel and used to identify sites for risk modeling purposes. The research team identified 13,179 segments (representing 2,274.281 total miles) that met these definitions.

Further, the team identified all intersections with pedestrian exposure counts that of significant length (12, 13 or 16 hours, since this represented the majority of the available counts) and that were on target arterial roadways (i.e., at least one approach was a target arterial roadway segment). A total of 2,120 pedestrian exposure counts were identified; these were used for the development of the exposure model to improve its accuracy on these target arterials.

4.2 Statistical methodology

Two types of statistical models were developed: pedestrian exposure models and pedestrian risk models. For the exposure models, the team developed both direct exposure models and categorical exposure models. The former predicted the expected pedestrian count at a given location, while the latter classified counts into categories (low, medium, high) and predicted the probability that exposure fell within each category. This subsection describes the statistical methods used for each model in more detail.

4.2.1 Direct exposure model

NB regression was used to develop the direct pedestrian exposure models. This is a count regression technique used when the dependent variable (in this case, pedestrian count) takes count outcomes (0, 1, 2…). It is particularly flexible as it accounts for overdispersion that might be observed in the data, in which the variance exceeds the mean. The relationship between dependent and independent variables in this approach takes the following general form:

$$\ln \lambda_i = \beta X_i + \varepsilon_i \tag{1}$$

where λ_i is the expected number of crashes at location i (i.e., dependent variable), β are the set of estimable regression parameters, X_i is a collection of geometric design, traffic volume, and other site-specific data for location i (i.e., independent variables), and ε_i is a gamma-distributed error term.

Equation 2 shows the specific form of the pedestrian count models that were estimated to predict pedestrian activity at a given intersection:

$$N_{i,count} = Max\ AADT^{\beta_{Max\ AADT}} \times e^{\beta_0} \times e^{\sum x_{ij}\beta_j}$$

$$= Max\ AADT^{\beta_{Max\ AADT}} \times e^{\beta_0} \times e^{x_{i1}\beta_1} \times e^{x_{i2}\beta_2} \times \dots \times e^{x_{iJ}\beta_J} \tag{2}$$

where $N_{i,count}$ = predicted pedestrian count for intersection i [crashes/year]; $Max\ AADT$ = maximum annual average daily traffic observed across all approaches associated with intersection i [veh/day]; $\beta_{Max\ AADT}$ = estimated coefficient for maximum traffic volume observed at intersection i; β_0 = a regression constant; β_j = estimated coefficient for other variables, j; and, x_{ij} that describe roadway segment i for variable j

4.2.2 Categorical exposure model

Ordinal regression was used to predict the level of pedestrian exposure defined as a category (low, medium, high) rather than the actual count. This approach is appropriate when the dependent variable has a natural ranking (such as high, medium, or low), but the numeric spacing between categories is not assumed to be equal (in other words, low might be a value between 0-100, while a medium value might be between 100-500). The model relates a set of geometric design, traffic volume, and other site-specific factors to the probability that an intersection falls into each exposure category (i.e., a low, medium, or high number of pedestrian crossings), while explicitly accounting for the ordered nature of the response.

The proportional-odds (cumulative logit) model was used, specifically in this study. Letting Y_i denote the exposure category for intersection i where $Y_i \in \{1,2,3\}$ corresponds to {Low, Medium, High}, respectively, the model can be written as follows:

$$\text{logit}[P(Y_i \leq k)] = \alpha_k - \beta^\top X_i, k = 1,2 \quad (3)$$

where $P(Y_i \leq k)$ is the probability that intersection i is in category k or lower; α_k are category-specific threshold (cutpoint) parameters to be estimated; β is a vector of regression coefficients; and X_i is a vector of explanatory variables for intersection i (e.g., traffic volume measures such as maximum AADT, intersection geometry, surrounding land use, and other contextual characteristics).

The estimated coefficients β_j describes how each covariate shifts the likelihood of being in higher versus lower exposure categories. The proportional-odds structure implies that the effect of each predictor is assumed to be consistent across the two cumulative splits (low vs. medium/high and low/medium vs. high), which provides a parsimonious and interpretable model for ordered exposure outcomes.

4.2.3 Risk model

Consistent with the models developed as a part of NCDOT RP 2022-11, NB regression was used to develop pedestrian risk factors. This method is the most common and appropriate modeling type as identified in the literature review. NB regression is a count regression technique used when the dependent variable being modeled takes count or integer values (Shankar et al., 1998). It has been applied widely in safety modeling and preferred over other count regression techniques because it directly accounts for overdispersion that is often observed in crash data in which the variance exceeds the mean (Geedipally et al., 2012; Hilbe, 2011).

Equation 4 shows the general form of the crash frequency models that were estimated for roadway segments to obtain individual risk factors:

$$\begin{aligned} N_{i,risk} &= AADT^{\beta_{AADT}} \times L \times e^{\beta_0} \times e^{\sum x_{ij}\beta_j} \\ &= AADT^{\beta_{AADT}} \times L \times e^{\beta_0} \times e^{x_{i1}\beta_1} \times e^{x_{i2}\beta_2} \times \ldots \times e^{x_{iJ}\beta_J} \end{aligned} \quad (4)$$

Where $N_{i,risk}$ = predicted pedestrian crash frequency for roadway segment i [crashes/year]; $AADT$ = annual average daily traffic associated with roadway segment i [veh/day]; β_{AADT} = estimated coefficient for traffic volume on roadway segment i; L = length of roadway segment [mi]; β_0 = a regression constant; and, β_j = estimated coefficient for other variables, x_{ij}, that describe roadway segment i. These other variables include roadway features (e.g., number of lanes, speed limits, presence of a median), block length indicators, surrogates for pedestrian exposure (e.g., presence of high- or medium-intensity development within 100 ft, alcohol sales density, bus route presence, population density, K-12 school enrollment density), and socioeconomic characteristics associated with the location (e.g., median income, proportion of commuters that are non-motorized, proportion of the population with disabilities).

Please note that the form shown in Equation 4 and estimated in this project specifically treats segment length (L) as a proportional constant associated with predicted crash frequency; therefore, the resulting models can be used to compute the expected crash frequency per mile by dividing the output of Equation 4 by L. Also note that traffic volume was included in most models developed.

4.2.4 Model interpretation

The elasticity of each independent variable included in an NB model provides a measure of the responsiveness of the dependent variable (crash frequency) to a change in another. This elasticity can be used as a measure of the "risk" associated with each variable. For the continuous explanatory variables considered in this study (e.g., AADT), the elasticity is interpreted as the percent change in the expected roadway segment or intersection crash frequency given a one percent change in that continuous variable. In general, the elasticity of the expected crash frequency for continuous explanatory variable *k* on roadway segment *i* during time period *j* is defined as:

$$E_{X_{ijk}}^{\lambda_{ij}} = \frac{\partial\lambda_{ij}/\lambda_{ij}}{\partial X_{ijk}/X_{ijk}} = \frac{\partial\lambda_{ij}}{\partial X_{ijk}} \times \frac{X_{ijk}}{\lambda_{ij}} \tag{5}$$

Equation 5 reduces to the following expressions for the log-log (Equation 6) and log-linear (Equation 7) functional forms, respectively. These represent the two types of functional forms considered for continuous variables included in this paper. The first represents the relationship modeled between expected crash frequency and variables entered into the model in log form (AADT or estimated pedestrian count), and the second represents the relationship modeled between expected crash frequency and all other continuous variables in the crash frequency models.

$$E_{X_{ijk}}^{\lambda_{ij}} = \beta_k \tag{6}$$

$$E_{X_{ijk}}^{\lambda_{ij}} = \beta_k X_{ijk} \tag{7}$$

The elasticity for indicator variables (e.g., presence of a median), termed *pseudo-elasticity* (Lee and Mannering, 2002), is the percent change in expected crash frequency given a change in the value of the indicator variable from zero to unity. In general, the elasticity of the expected crash frequency for indicator variable *k* on roadway segment *i* during time period *j* is defined as:

$$E_{X_{ijk}}^{\lambda_{ij}} = \exp(\beta_k) - 1 \quad (8)$$

4.3 Model results

4.3.1 Direct exposure model

Numerous permutations of independent variables were tested to determine the best direct pedestrian exposure model. Table 9 summarizes the variables and coefficients that yielded the best overall results, based on model fit, interpretability, and discussions with the NCDOT technical panel. Positive coefficients represent factors that are associated with increased pedestrian activity at that intersection, while negative coefficients represent factors that are associated with decreased pedestrian activity. As shown in the table, pedestrian activity is expected to decrease at intersections with higher traffic volumes, on roadways classified as major or minor arterials, at signalized intersections, at intersections with higher posted speed limits, at locations with higher median incomes, and at locations with a higher percentage of older (over 65) and younger (under 18) populations. Pedestrian activity is expected to increase at intersections near bus stops, with a higher parcel count nearby, with more high-intensity land use, near a college or university, park, or greenway, at intersections with sidewalks or crosswalks, with a higher non-motorized population, at intersections near K–12 schools or alcohol sales locations, at locations with a higher proportion of zero-vehicle households, and at locations with a higher proportion of college-educated individuals aged 25 years or older. Sets of indicator variables were also included to account for regional differences across NCDOT engineering divisions, differences in count durations, and the month in which the count was taken.

Table 9. Summary of direct pedestrian exposure model for intersections on target arterials

Variable	Coefficient	p-value
Constant	*6.707*	*<0.001*
Log of maximum AADT at site	-0.536	<0.001
One approach is a major arterial	-0.374	<0.001
One approach is a minor arterial	-0.123	0.080
Presence of a bus stop within 1/4 mi	0.272	<0.001

Variable	Coefficient	p-value
Log of parcel count	0.350	<0.001
High intensity development	2.622	<0.001
Site near a university or college	0.436	<0.001
Site near a park	0.377	<0.001
Site near a greenway	0.569	<0.001
Log of total non-motorist population + 1	0.049	0.022
Sidewalk exists at the site	0.608	<0.001
Crosswalk exists at the site	0.850	<0.001
Signal exists at the site	-0.206	0.010
At least one K12 school near site	0.346	<0.001
Minimum posted speed limit is greater than or equal to 40 mph	-0.088	0.166
At least one alcohol sales location near site	0.403	<0.001
Proportion of zero car households	1.674	0.001
Proportion of college educated individuals 25 years or older	1.356	<0.001
Proportion of individuals less than 18 years old	-2.036	<0.001
Proportion of individuals greater than 65 years old	-2.029	<0.001
Median income	-9.16E-06	<0.001
Division = 2	-1.014	<0.001
Division = 3	-0.740	<0.001
Division = 4	-1.655	<0.001
Division = 5	-0.809	<0.001
Division = 6	-1.103	<0.001
Division = 7	-1.466	<0.001
Division = 8	-1.196	<0.001
Division = 9	-1.225	<0.001
Division = 10	-0.681	0.001
Division = 11	-1.014	<0.001
Division = 12	-1.002	<0.001
Division = 13	-0.922	<0.001
Division = 14	-1.044	<0.001
Count duration = 13 hours	0.153	0.337
Count duration = 16 hours	0.423	0.016
Count in August	0.132	0.360
Count in December	0.636	0.006
Count in February	0.166	0.175
Count in January	0.030	0.843
Count in July	0.279	0.082
Count in June	0.626	0.000
Count in March	0.177	0.127
Count in May	0.291	0.013
Count in November	0.136	0.318
Count in October	0.199	0.119
Count in September	0.310	0.011

Variable	Coefficient	p-value
Count after 2021	0.074	0.441
Count before 2020	0.069	0.493
Inverse of overdispersion parameter	*0.6552*	*0.0209*
2xLog Likelihood	*-17371.84*	

Figure 2 provides a plot of the predicted pedestrian counts obtained using the exposure model (y-axis) versus observed values (x-axis) for illustrative purposes; only observations less than 1,000 are shown in this plot for ease of visualization. The red solid line represents cases in which predicted and observed values are equal; for an ideal model with perfect prediction, plotted values would fall along this line. The green dashed lines, parallel to the red line, represent bounds for counts that are within 50 pedestrians of the predicted value. The blue dotted lines are bound for counts that are within 20 percent of the predicted value. These thresholds were identified through discussion with the NCDOT technical panel. The larger bounds defined by the green dashed and blue dotted lines represent counts that are "correct" for practical purposes. The model accurately captures 1,745 of the 2,120 observed counts using these measures, representing an accuracy of approximately 82 percent.

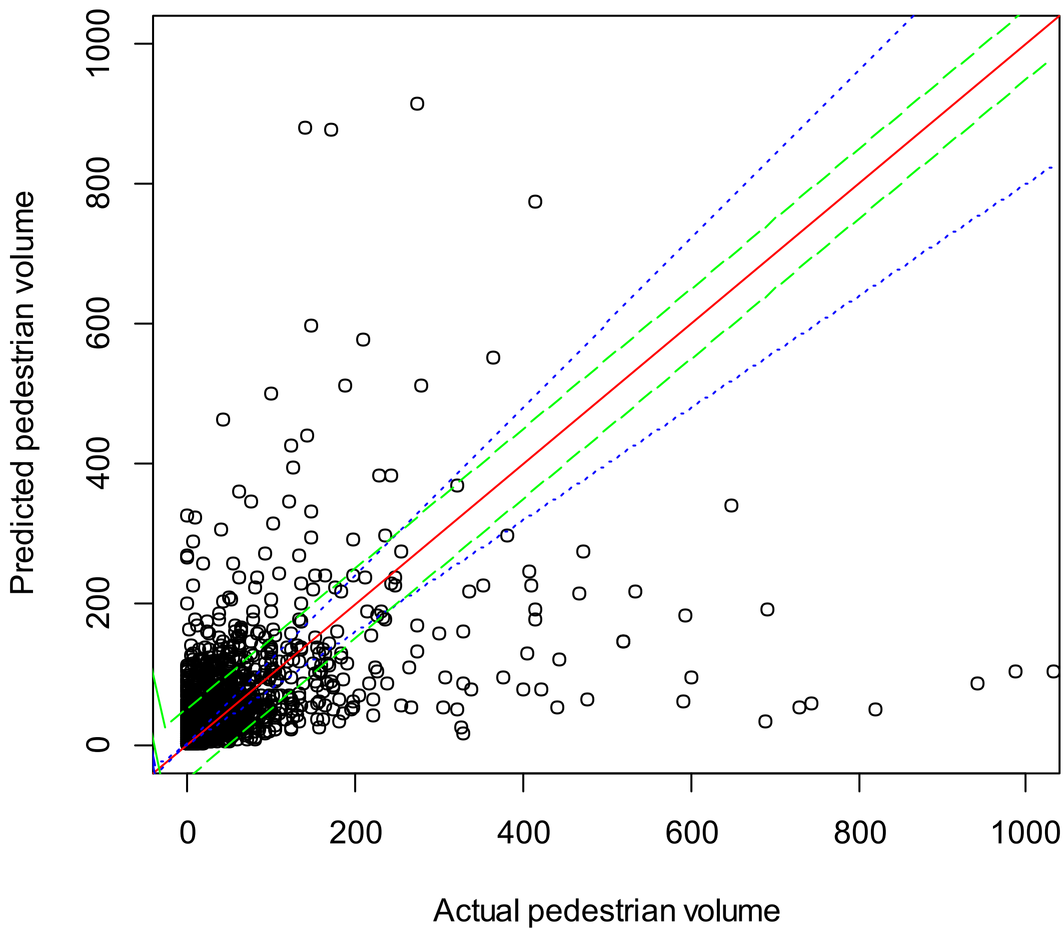

Figure 2. Predicted vs. observed pedestrian count values

4.3.2 Categorical exposure model

To develop the categorical exposure model, the research team first had to identify appropriate categories for pedestrian exposure values. After checking the distribution of the data, reviewing NCDOT guidance (Schroeder et al, 2015), and in consultation with the NCDOT technical panel, the research team defined the following categories:

- Low pedestrian activity: 0-30 pedestrians observed
- Medium pedestrian activity: 31-200 pedestrians observed
- High pedestrian activity: 201+ pedestrians observed

These specific thresholds aligned well with the observed pedestrian count data and were also meaningful values for NCDOT as they coincided with critical crossing guidelines for pedestrian activity. Approximately 69 percent of observed counts fell within the low activity category, 26 percent fell within the medium activity category, and 5 percent fell within the high activity category. Note that these counts represented different count durations – either 12, 13 or 16 hours. Indicator variables were included to adjust for the duration length in the modeling.

The results of the ordinal regression model are provided in Table 10. Like the direct exposure model, positive coefficients are variables associated with an increased likelihood of higher pedestrian activity, and negative coefficients are associated with an increased likelihood of lower pedestrian activity. The results are consistent with those of the direct exposure model.

Table 10. Summary of pedestrian exposure model for urban intersections

Variable	Coefficient	p-value
Log of maximum AADT at site	-0.7081	<0.001
One approach is a major arterial	-0.2996	0.029
Log of parcel count	0.6587	<0.001
High intensity development	3.2020	<0.001
Site near a university or college (but not a top 5 university by enrollment)	0.5579	0.013
Site near a top 5 university by enrollment	0.8630	0.081
Site near a park	0.2832	0.038
Site near a greenway	0.9433	<0.001
Log of total non-motorist population + 1	0.2213	<0.001
Sidewalk exists at the site	1.4970	<0.001
Crosswalk exists at the site	1.1840	<0.001
Signal exists at the site	-0.2852	0.101
Number of K12 schools near site	0.6176	0.001
Minimum posted speed limit is greater than or equal to 40 mph	-0.1972	0.162
Between 1 and 5 alcohol sales locations near the site	0.4821	0.002
More than 5 alcohol sales locations near the site	1.0740	<0.001
Proportion of zero car households	2.9450	0.004
Proportion of college educated individuals 25 years or older	1.3230	0.027
Proportion of individuals younger than 18 years old	-2.5100	0.007
Proportion of individuals 65 years old and older	-1.7710	0.061
Median household income	0.0000	0.027
Division = 2	-0.7467	0.082
Division = 3	0.2901	0.478
Division = 4	-1.8270	0.003
Division = 5	-0.0432	0.915

Variable	Coefficient	p-value
Division = 6	-0.5566	0.167
Division = 7	-1.0970	0.014
Division = 8	-2.2340	0.044
Division = 9	-0.3178	0.557
Division = 10	0.1188	0.763
Division = 11	-0.5485	0.355
Division = 12	-0.9776	0.029
Division = 13	-0.0762	0.859
Division = 14	-0.8573	0.087
Count duration = 13 hours	-0.1239	0.700
Count duration = 16 hours	0.4864	0.162
Count in August	0.0473	0.875
Count in December	1.1680	0.011
Count in February	0.1619	0.536
Count in January	-0.1107	0.749
Count in July	0.1220	0.714
Count in June	0.5334	0.037
Count in March	0.3134	0.200
Count in May	0.2846	0.243
Count in November	0.0284	0.925
Count in October	0.2369	0.357
Count in September	0.4102	0.100
Count after 2021	0.3426	0.110
Count before 2020	0.1478	0.502
Threshold between low and medium activity	*-0.444*	
Threshold between medium and high activity	*2.907*	

Table 11 provides a summary of the observed categories versus those predicted by the categorical exposure model; the same results obtained using the predicted direct exposure value to assign categories are provided in Table 12. The results reveal that the model does a fairly accurate job of predicting the pedestrian exposure category. Approximately 78 percent of the observations are correctly categorized. Of the remaining 22 percent of observations that are not correctly categorized, less than 0.5 percent are grossly incorrect (defined as a low value being categorized as high or a high value being categorized as low). By comparison, if the direct exposure model was used (i.e., if the count estimated by the model were categorized), the correct category was identified just 73 percent. While this represents just a small improvement, there are practical benefits of using the categorical model over the direct exposure model. One benefit is that model errors (sites that are known to be incorrect) are more easily corrected in the categorical model, as the NCDOT could simply reassign the category without having to use professional judgment to determine a specific pedestrian crossing value for use in the subsequent risk model. For example, review of the model results with the NCDOT technical panel revealed that the model performs poorly in predicting pedestrian exposure near university campuses due to a low number of actual counts at these locations. To correct for this in the categorical model, activity could simply be reassigned to "high," regardless of the model outputs. If the direct exposure model were used, an estimated exposure value would have to be determined to estimate pedestrian crash risk, which would be subject to greater error.

Table 11. Comparison of predicted vs. observed values for categorical exposure model

Actual category	Predicted category [number (percent)]		
	Low	Medium	High
Low	1,328 (62.6%)	145 (6.84%)	0 (0.0%)
Medium	239 (11.3%)	319 (15.1%)	5 (0.2%)
High	10 (0.5%)	64 (3.0%)	10 (0.5%)

Table 12. Comparison of predicted vs. observed values using direct exposure model

Actual category	Predicted category [number (percent)]		
	Low	Medium	High
Low	1,100 (51.9%)	365 (17.2%)	8 (0.4%)
Medium	111 (5.2%)	415 (19.6%)	37 (1.8%)
High	2 (0.1%)	49 (2.3%)	33 (1.6%)

Due to the superior performance and ease of application, the research team proposed using the categorical exposure model in the pedestrian risk estimates for the remainder of the project.

4.3.3 Pedestrian crash risk

The research team used the categorical exposure model to estimate the level of pedestrian activity for all intersections along target arterial roadways in North Carolina. The team then appended these estimates to the roadway segments for crash risk modeling. The exposure on a given segment was set equal to the highest exposure value of the two intersections bounding that segment. Of the 13,179 segments in the modeling database, 9,755 were classified as "low" pedestrian activity, 3,275 as "medium" pedestrian activity and 149 as "high" pedestrian activity. The team then estimated risk factors using total crash frequency as the dependent variable, considering roadway characteristics, vehicular and pedestrian exposure, and sociodemographic features as independent variables. The final risk model is provided in Table 13. The coefficient estimate for a given variable provides the relationship between that variable and pedestrian crash frequency: positive values (denoted by red or darker shades) represent factors associated with increased pedestrian crash risk, while negative values (denoted by green or lighter shades) represent factors associated with decreased pedestrian crash risk. As shown, the specific factors generally aligned with expectations. Factors associated with increased risk include:

- Vehicular traffic volume (i.e., AADT)
- Pedestrian exposure
- Number of lanes
- Presence of high-intensity (land use cover) development within 100 ft
- Density of alcohol sales establishments
- Density of bus stops along the segment
- Population density
- K12 school enrollment density

Factors associated with reduced risk include:

- Roadway segments with longer block lengths
- Median household income

The models also include indicator variables associated with the NCDOT engineering divisions. These indicators were used to account for geographic/regional differences across the state, as well as local differences associated with each NCDOT engineering division. Thirteen indicators are included in the model, one for each of Divisions 2 through 14, while NCDOT engineering Division 1 serves as the baseline condition. These division indicators are not color-coded to maintain focus on location-specific features.

The p-values associated with each coefficient were used to assess the statistical significance of the variables included in the model. Smaller values indicate stronger statistical significance; p-values less than 0.05 indicate variables that are statistically significant to the 95 percent confidence level. Note that most of the risk factors were statistically significant to the 95 percent confidence level. The research team still included variables that were not significant at this level, since the coefficient estimates were in line with expectations, and improved the overall model fit. P-values for the division indicators generally indicate that these variables are not statistically significant, suggesting that differences in pedestrian crash risk across engineering divisions are very small. However, they were retained in the model, as their inclusion increases the model fit and improves the usefulness of the model in identifying high-risk locations.

Table 13. Summary of crash frequency models developed for target arterials

Variable	Total crash frequency	
	Coefficient	**p-value**
Constant	*-5.195*	*0.000*
Natural log of AADT	0.511	<0.001
High pedestrian exposure	0.403	0.009
Medium pedestrian exposure	0.350	<0.001
5+ lane roadway	0.243	<0.001
Block length between 0.1-0.25 mi	-0.315	<0.001
Block length between 0.25-0.5 mi	-0.602	<0.001
Block length greater than 0.5 mi	-1.254	<0.001
High intensity development within 100 ft	0.626	<0.001
Alcohol sales density	0.010	<0.001
Bus stop density	0.020	<0.001
Population density	3.58E-05	0.001
K12 enrollment density	2.95E-04	<0.001
Median household income	7.50E-06	<0.001
NCDOT Division 2	0.378	0.021
NCDOT Division 3	0.212	0.186
NCDOT Division 4	0.527	0.002
NCDOT Division 5	0.410	0.008
NCDOT Division 6	0.617	<0.001
NCDOT Division 7	0.197	0.226
NCDOT Division 8	0.506	0.005
NCDOT Division 9	0.284	0.093
NCDOT Division 10	0.516	0.001
NCDOT Division 11	-0.199	0.337
NCDOT Division 12	0.393	0.016
NCDOT Division 13	0.306	0.078
NCDOT Division 14	0.039	0.839
Inverse of overdispersion parameter	*0.9858*	*<0.001*
2xlog-likelihood value	*-16054.762*	

Table 14 provides the elasticities for non-division indicator variables associated with the models in Table 13 computed using Equations 5 to 8. These elasticities quantify the amount of "risk" associated with each risk factor included in the model. Specifically, each value represents the relevant increase in total crash frequency associated with a change in a given variable, referred to hereafter as crash risk. Values greater than 0 represent an increase in crash risk associated with an increase in that variable (i.e., positive correlation), whereas values less than zero represent a decline in crash risk associated with an increase in that variable (i.e., negative correlation). Continuous variables that are not in a log form are assessed at the mean value observed in the dataset (provided in the table). The elasticity values would differ for other values of these continuous variables; however, these estimates provide a good indication of the strength of the relationship between that variable and pedestrian crash frequency. Despite being a continuous variable, AADT is entered in the log form; therefore, the elasticity values provided in this table hold for all AADT values.

Values in Table 14 can be interpreted as follows. Traffic volume is the only variable included in the model in log form. The elasticities suggest that a one percent change in traffic volume along a target arterial segment is associated with a 0.511 percent increase in total pedestrian crash frequency along that segment. For other continuous variables, the elasticity is provided at the mean value observed in the data. For example, a one percent change in population density—for the "average" roadway segment with population density of 1,653.2 people per square mile—would be associated with a 0.061 percent increase in total pedestrian crash frequency along that segment. Finally, indicator variables provide the percent change associated with the indicator being used. For example, the presence of 5 or more travel lanes is associated with a 27.5 percent increase in total pedestrian crash frequency along that segment. Other variables can be interpreted similarly.

Table 14. Elasticity values for risk factors on target arterials

Variable	Variable type	Elasticity for total crash frequency	Mean value (if applicable)
Natural log of AADT	log	0.511	N/A
High pedestrian exposure	I	0.496	N/A
Medium pedestrian exposure	I	0.419	N/A
5+ lane roadway	I	0.275	N/A
Block length between 0.1-0.25 mi	I	-0.270	N/A
Block length between 0.25-0.5 mi	I	-0.452	N/A
Block length greater than 0.5 mi	I	-0.715	N/A
High intensity development within 100 ft	I	0.870	N/A
Alcohol sales density	C	0.152	13.57
Bus stop density	C	0.222	9.971
Population density	C	0.061	1653.2
K12 enrollment density	C	0.073	237.8
Median household income	C	-0.361	59754

I – indicator variable; C – continuous variable; log – continuous variable included in log form

Chapter 5. Discussion

5.1 Summary

As part of this project, pedestrian exposure and crash risk models were developed for target arterials in North Carolina, defined as roadway segments that meet the following criteria:

- 4+ lanes
- AADT greater than 12,500 vehicle/day
- 35+ mph speed limits
- Not fully access controlled

5.1.1 Exposure model

Two types of exposure models were considered: a direct exposure model that predicted the actual daily pedestrian count and a categorical model that predicted the anticipated level of pedestrian activity, defined as low (0–30 pedestrians per day), medium (31–200 pedestrians per day), or high (greater than 200 pedestrians per day). The categorical model was shown to produce reasonable estimates and is easier to use than the direct exposure model; therefore, it is recommended for use by NCDOT to predict pedestrian activity at intersections along target arterial roadway segments. Factors that are found to influence pedestrian activity at these locations include:

- Traffic volume (i.e., AADT)
- Roadway classification as a major arterial
- Parcel count
- Amount of high-intensity development land use cover around the intersection
- Located near a college or university (including a top 5 university by enrollment), park, or greenway
- Presence of a sidewalk, crosswalk, or signal
- Non-motorized population
- Number of K-12 schools
- Posted speed limit
- Alcohol sales location
- Zero vehicle household proportion
- Percent of population aged 25 or older with an associate's degree or higher
- Proportion of younger (<18) or older (>65) population

- Median household income

The categorical exposure estimate for any arterial intersection can be computed as follows. First, a numerical exposure score for the intersection is computed using the following equation:

$$
\begin{aligned}
Z_{i,exposure} = {} & -0.7081 \times \log(AADTMax) - 0.2996 \times MajArt + 0.6587 \times \log(ParcelCount) \\
& + 3.2020 \times HI_{Dev} + 0.5579 \times UniMinus5 + 0.8630 \times Top5Uni + 0.2832 \times Park \\
& + 0.9433 \times Greenway + 0.2213 \times \log(NonMotor + 1) + 1.4970 \times Sidewalk \\
& + 1.1840 \times Crosswalk - 0.2852 \times Signal + 0.6176 \times K12Count \\
& - 0.1972 \times minPSL40p + 0.4821 \times AlcSales15 + 1.0740 \times AlcSales5p \\
& + 2.450 \times ZeroHHProp + 1.3230 \times College25pProp - 2.5100 \times Pop18Prop \\
& + 1.7710 \times Prop65Prop - 8.38 \times 10^{-6} \times \ Med\_Inc - 0.7467 \times Div2 \\
& + 0.2901 \times Div3 - 1.8270 \times Div4 - 0.0432 \times Div5 - 0.556 \times Div6 \\
& - 1.0970 \times Div7 - 2.2340 \times Div8 - 0.3178 \times Div9 + 0.1188 \times Div10 \\
& - 0.5485 \times Div11 - 0.9766 \times Div12 - 0.0762 \times Div13 - 0.8573 \times Div14
\end{aligned}
$$

where:

- $Z_{i,exposure}$ = the exposure score for the given intersection i;
- $AADTMax$ = maximum annual average daily traffic for all approaches at the intersection [veh/day];
- $MajArt$ = indicator variable for one of the intersection approaches having a functional class of Other Principal Arterial;
- $ParcelCount$ = count of the number of land use parcels within 0.25 miles [count];
- HI_{Dev} = fraction of land use within 1 km classified as high intensity [decimal];
- $UniMinus5$ = indicator variable for the intersection location being within 0.5 miles of a college of university that is not a top 5 university (by enrollment) within North Carolina;
- $Top5Uni$ = indicator variable for the intersection location being within 0.5 miles of a top 5 university (by enrollment) within North Carolina;
- $Park$ = indicator variable for the intersection being located within 0.25 miles of a park;
- $Greenway$ = indicator variable for the intersection being located within 100 ft of a greenway;
- $NonMotor$ = non-motorized population within 0.25 miles [count];
- $Sidewalk$ = indicator variable for a sidewalk present within 100 ft location;
- $Crosswalk$ = indicator variable for a crosswalk within 100 ft of a location;
- $Signal$ = indicator variable for the presence of a traffic signal at the intersection location;
- $K12Count$ = K-12 school count within 0.25 miles [count];

- $minPSL40p$ = indicator variable for minimum posted speed limit at the intersection greater than or equal to 40 mph;
- $AlcSales15$ = indicator variable for between 1-5 (inclusive) alcohol sales locations within 0.25 miles;
- $AlcSales5p$ = indicator variable for greater than 5 alcohol sales locations within 0.25 miles;
- $ZeroHHProp$ = proportion of households with zero vehicles within 0.25 miles [decimal];
- $College25pProp$ = percent of population within 0.25 miles aged 25 or older with an associate's degree or higher [decimal];
- $Pop18Prop$ = proportion of the population within 0.25 miles that is younger than 18 years old [decimal];
- $Prop65Prop$ = proportion of the population within 0.25 miles that is 65 years old or older;
- Med_Inc = median income of households with 0.25 miles of the segment [$]; and,
- $DivX$ = indicator variable for segment occurring within NCDOT engineering division X [1,0].

Next, the probability that a given location falls within a given category of exposure level (low, medium, high) is computed as follows:

$$Prob_{low,i} = \frac{1}{1+\exp\left(0.44+Z_{i,exposure}\right)} \tag{9}$$

$$Prob_{medium,i} = \frac{1}{1+\exp\left(-2.907+Z_{i,exposure}\right)} - Prob_{low,i} \tag{10}$$

$$Prob_{high,i} = 1 - Prob_{low,i} - Prob_{medium,i} \tag{11}$$

where $Prob_{low,i}$ = the probability that the given intersection is in the low pedestrian exposure category (i.e., expected count between 0-30); $Prob_{medium,i}$ = the probability that the given intersection is in the medium pedestrian exposure category (i.e., expected count between 30-200); and, $Prob_{high,i}$ = the probability that the given intersection is in the high pedestrian exposure category (i.e., expected count greater than 200).

The intersection is then assigned to the category (i.e., low, medium, or high) with the highest probability.

5.1.2 Risk model

The risk factors included the categorical measure of pedestrian activity, as well as the following additional factors:

- Vehicular traffic volume (i.e., AADT)
- Pedestrian exposure (i.e., high, medium, or low number of crossings)
- Number of lanes
- Presence of high-intensity development (land use cover) within 100 ft
- Density of alcohol sales establishments
- Density of bus stops along the segment
- Population density
- K12 school enrollment density
- Roadway segments with longer block lengths
- Median household income

Note that these are generally consistent with the previous NCDOR RP 2022-11 project. However, the incorporation of pedestrian exposure presents a significant improvement as the level of pedestrian activity can now be accounted for in the crash risk.

The final risk scores for any segment can be computed as follows:

$$N_{i,risk} = AADT^{0.511} \times L \times exp(-5.195) \times exp(0.403 \cdot HighPedExp) \times exp(0.350 \cdot MedPedExp) \times exp(0.243 \cdot 5p\_lanes) \times exp(-0.315 \cdot BL_{01025}) \times exp(-0.602 \cdot BL_{02550}) \times exp(-1.254 \cdot BL_{05p}) \times exp(0.626 \cdot HI\_Dev) \times exp(0.010 \cdot Alc\_Dens) \times exp(0.020 \cdot BusStop\_Dens) \times exp(3.58 \times 10^{-5} \cdot Pop\_Dens) \times exp(2.95 \times 10^{-4} \cdot K12\_Dens) \times exp(7.50 \times 10^{-6} \cdot Med\_Inc) \times e^{0.3779 \times Div2} \times e^{0.2123 \times Div3} \times e^{0.5274 \times Div4} \times e^{0.4101 \times Div5} \times e^{0.6173 \times Div6} \times e^{0.1965 \times Div7} \times e^{0.5061 \times Div8} \times e^{0.2836 \times Div9} \times e^{0.5161 \times Div10} \times e^{-0.1989 \times Div11} \times e^{0.3929 \times Div12} \times e^{0.3062 \times Div13} \times e^{0.0387 \times Div14}$$

where:

- $AADT$ = annual average daily traffic [veh/day];
- L = segment length [mi];
- $5p\_lanes$ = indicator variable for the road segment having 5 or more travel lanes [1,0];
- BL_{01025} = indicator variable for the road segment length between 0.1 and 0.25 miles [1,0];
- BL_{02550} = indicator variable for the road segment length between 0.25 and 0.5 miles [1,0];
- BL_{05p} = indicator variable for the road segment length greater than 0.5 miles [1,0];
- HI_{Dev}= indicator variable for the presence of high-intensity development within 100 ft of the road segment [1,0];

- Alc_Dens = density (per mile) of alcohol sales establishments within 0.25 miles of the segment [decimal];
- $BusStop\_Dens$= density (per mile) of bus stops along the route within 0.25 miles of the road segment [1,0];
- Pop_Dens = population density of the dominant census block group [decimal];
- $K12\_Dens$ = K-12 school enrollment density of the dominant census block group [decimal];
- Med_Inc = median income of households within the dominant census block group [$]; and,
- $DivX$ = indicator variable for segment occurring within NCDOT engineering division X [1,0].

5.2 Model application

The following subsections describe how NCDOT can practically apply the exposure model, as well as the associated revised risk factors and analysis. Furthermore, bicycle activity at a site does not directly correlate with levels of pedestrian activity, because bicycle trips may cover longer distances, are subject to different traffic laws, and the infrastructure dedicated bicycling may be different from pedestrian access routes. However, literature suggests that bicycling activity may be modeled based on similar characteristics as pedestrian activity, and bicycle crashes may also occur at similar locations as pedestrian crashes. Therefore, the pedestrian exposure model could be used as an interim measure of relative bicycle activity, in most planning-level studies.

5.2.1 Exposure model

The purpose of an exposure model is to understand the propensity and demand for pedestrian travel on North Carolina roads, especially where traditional counts or other data do not exist. The exposure model can contribute to planning and engineering needs at the statewide (e.g., NCDOT business cases), regional (e.g., MPO and RPO planning), and local (e.g., municipal corridor studies or county comprehensive plans) scales (Table 15).

Table 15. Potential Applications for Bicycle-Pedestrian Counts and Estimate Data

Application	Scale of Application		
	Statewide	Regional	Local
Current Activity / Activity Summaries	X	X	X
Long Range Transportation Plan Development		X	X
Mode and Trip Choice Models	X	X	
Bike-Ped Plans		X	X
Access to Transit/Micromobility Plans		X	X
Corridor Studies		X	X
Project Development/ Project Evaluation	X	X	X
Safety and Crash Risk	X		
Health and Social Impact	X	X	
Economic Impact	X	X	X
Environmental Impact	X		
Multimodal Access / Levels of Service	X		X
Project Prioritization	X	X	X
Grant Applications	X	X	X
Transportation Demand Management (TDM)		X	X

Statewide applications could most benefit from aggregate data to describe current trends and expected benefits of developing a robust and safe system that connects bicyclists and pedestrians to important destinations. Projects identified, prioritized, and developed by the NCDOT would also benefit from tools that estimate pedestrian traffic (as well as bicycle traffic, potentially) for many roadway contexts. NCDOT could use the model to describe impacts of increased pedestrian (and bicycle) activity to the environment, economy, public health and safety – and offer these findings to agencies applying for grants.

Regional applications, using bicycle and pedestrian estimates, could support the long-range transportation plans (i.e., Metropolitan Transportation Plans) developed by MPOs, RPOs and NCDOT. These plans would benefit from bicycle and pedestrian estimates to better describe mode choice and trip choice assumptions in travel demand models. Development and planning scenarios in these long-range plans would use projected bicycle and pedestrian estimates to determine potential benefits and impacts to the community.

Local applications can inform plans that local governments prepare for the preferred bicycling, pedestrian, micromobility, and transit networks. The gaps and needs identified in these multimodal transportation plans translate into projects that would benefit from data estimating or confirming expected pedestrian and bicycle activity at a site. Engineers and planners would use pedestrian and bicycle estimates to help select lower-stress bikeway networks, appropriate pedestrian crossing countermeasures, transit stop placement, and other multimodal decisions as part of corridor plans, feasibility studies, and Complete Streets project evaluations. Count and estimate data would improve decisions for multimodal access (or Levels of Service) as part of local Traffic Impact Analyses (TIAs).

5.2.2 Risk factors, arterial applications, and models

Similar to NCDOT 2022-11, the exposure-enhanced risk models can be directly applied to estimate the expected pedestrian crash risk at focus roadway segments within North Carolina. While these risk values are not useful on their own, they can be used to "rank" individual sites and identify those that have the highest pedestrian risk. These high-risk locations can then be considered for additional scrutiny, engineering studies for crossing improvements, or the application of systemic safety treatments.

Using the risk-based model to prioritize sites for review and describe expected pedestrian activity is particularly helpful because high-speed and high traffic volume arterials may suppress observed pedestrian activity (i.e., discourage pedestrians from crossing the arterial) and not provide a full picture of the benefits of potential safety improvements for pedestrians. Decisions to install countermeasures on arterials, such as Pedestrian Hybrid Beacons, require engineering studies and data about pedestrian activity that may not be adequately described by observed counts. Risk models and estimated exposure can "fill in the gaps" for expected pedestrian activity in the absence of reliable counts on arterials. The exposure and risk factor models are tools for describing this expected pedestrian activity, and the exposure model aligns with NCDOT guidance and thresholds regarding pedestrian crossing improvements.

5.3 Model updates

The exposure and risk factor models were developed using historical data and therefore represent observed relationships during these time periods. For example, crash data from 2015-2023 (inclusive) were used to develop these models; time periods for other explanatory variables are documented above. The models should be applicable and valid as long as the relationships between explanatory variables and crash frequency do not change. However, these relationships may no longer hold if there are significant changes in driving or pedestrian behavior or associated technologies (e.g., a complete overhaul of the vehicle fleet or substantially increased or reduced pedestrian activity). If such changes occur, updated data should be collected and the processes used in this project repeated to re-estimate the pedestrian crash risk factors. Provided that such changes do not occur, the relationships would be expected to subtly change over time. Accordingly, the models should be updated at regular intervals to capture these changing trends, as well as account to changes in input data. For example, future iterations of the model can leverage more contemporary Census population and employment estimates (i.e., updates to the ACS and LEHD), as well as updated NCDOT roadway, traffic, and intersection inventories.

In general, there is little to no formal guidance on how often such models should be updated if significant changes in behavior do not occur. However, the research team recommends that the models be updated every five years based on the relatively long data collection period (e.g., at least five years of crash data) and account for changes in input data sources over the entire state.

REFERENCES

Almasi, S. A., Behnood, H. R., 2022. Exposure based geographic analysis mode for estimating the expected pedestrian crash frequency in urban traffic zones; case study of Tehran. Accident Analysis & Prevention 168. https://doi.org/10.1016/j.aap.2022.106576

Arias, D., Ederer, D., Rodgers, M.O., Hunter, M.P., Watkins, K.E., 2021. Estimating the effect of vehicle speeds on bicycle and pedestrian safety on the Georgia arterial roadway network. Accid. Anal. Prev. 161, 106351. doi:10.1016/J.AAP.2021.106351

Arizona Department of Transportation, 2017. Pedestrian safety action plan.

Behnam, J., Patel, B.G., 1977. Pedestrian volume estimation by land-use variables. Transp. Eng. J. ASCE 103, 507–520. doi:10.1061/tpejan.0000649

Bergh, C., Griffin, A., Braughton, M., Boudart, J., 2015. Implementing a risk-based systemic safety prioritization method for pedestrian and bicycle crashes in Oregon, in: TRB 2015 Annual Meeting Compendium of Papers. Transportation Research Board.

"CDC/ATSDR Social Vulnerability Index (CDC/ATSDR SVI)." Centers for Disease Control and Prevention, 2024. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.

City of Pittsburgh, 2020. Pedestrian safety action plan.

Cooper, C.H. V, Harvey, I., Orford, S., Chiaradia, A.J.F., 2019. Using multiple hybrid spatial design network analysis to predict longitudinal effect of a major city center redevelopment on pedestrian flows. Transportation (Amst). 48, 643–672. doi:10.1007/s11116-019-10072-0

Cottrill, C.D., Thakuriah, P. (Vonu), 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accid. Anal. Prev. 42, 1718–1728. doi:10.1016/j.aap.2010.04.012

Devoe, E., 2019. Statewide bicycle/pedestrian suitability analysis.

Dolatsara, H.A., 2014. Development of safety performance functions for non-motorized development of safety performance functions for non-motorized traffic safety traffic safety.

Dumbaugh, E., Li, W., Joh, K., 2013. The built environment and the incidence of pedestrian and cyclist crashes. URBAN Des. Int. 18, 217–228. doi:10.1057/udi.2013.2

Elvik, R., Bjørnskau, T., 2017. Safety-in-numbers: A systematic review and meta-analysis of evidence. Saf. Sci. 92, 274–282. doi:10.1016/J.SSCI.2015.07.017

Environmental Justice / Transportation Disadvantage Index Tool, 2022. https://connect.ncdot.gov/projects/planning/Pages/EJ-TDI-maps.aspx.

Forrest, M., Heydari, S., Cherrett, T., 2022. Examining the impact of exposure, built environment and socio-demographics on pedestrian safety: A case study of Greater London. Safety Science. 159. https://doi.org/10.1016/j.ssci.2022.106015

Gates, T.J., Savolainen, P.T., Stapleton, S., Kirsch, T., Miraskar, S., 2016. Development of safety performance functions and other decision support tools to assess pedestrian and bicycle safety FINAL REPORT.

Gayah, V.V., Guler, S.I., Liu, H., Blackburn, L., Hamilton, I., 2022. Quantification of Systemic Risk Factors for Pedestrian Safety on North Carolina. Publication FHWA/NC/2022-11.

Georgia Department of Transportation, 2018. Pedestrian safety action plan 2018–2022.

Griswold, J.B., Medury, A., Schneider, R.J., Amos, D., Li, A., Grembek, O., 2019. A pedestrian exposure model for the California State highway system. Transp. Res. Rec. J. Transp. Res. Board 2673, 941–950. doi:10.1177/0361198119837235

Ha, H.-H., Thill, J.-C., 2011. Analysis of traffic hazard intensity: A spatial epidemiology case study of urban pedestrians. Comput. Environ. Urban Syst. 35, 230–240. doi:10.1016/j.compenvurbsys.2010.12.004

Hamilton, I., Himes, S., Wang, Y., Tanzen, R., & Zhou, Y., 2023. Planning-Level Crash Prediction Models in Southern California. Transportation Research Record: Journal of the Transportation Research Board, 2677(2), 431–442. https://doi.org/10.1177/03611981221106788

Hamilton, I., Kersavage, K., Porter, R.J., Smith, K., Sanchez, J., Gayah, V., Eccles, K., 2021. An Exploration of Pedestrian Safety Through the Integration of HSIS and Emerging Data Sources: Case Study in Charlotte, NC. FHWA Rep. No. FHWA-HRT-21-087.

Hampshire, R.C., Molnar, L.J., Cao, A., Cai, Y., Li, X., Fabusuyi, T., 2018. Developing Michigan pedestrian and bicycle safety models Final Report.

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

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., Xu, Z., 2012. Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. Landsc. Urban Plan. 107, 307–316. doi:10.1016/j.landurbplan.2012.06.005

Hankey, S., Lu, T., Mondschein, A., Buehler, R., 2017. Spatial models of active travel in small communities: Merging the goals of traffic monitoring and direct-demand modeling. doi:10.1016/j.jth.2017.08.009

Haynes, M., Andrzejewsk, S., 2010. GIS based bicycle & pedestrian demand forecasting techniques.

https://www.researchgate.net/publication/368572871_Risk_Factors_for_Pedestrian_Crashes_on_Utah_State_Highway_Segments_Results_from_Parametric_and_Non-Parametric_Approaches_Controlling_for_Pedestrian_Exposure

Islam, A., Mekker, M., Singleton, P.A., 2022. Examining Pedestrian Crash Frequency, Severity, and Safety in Numbers Using Pedestrian Exposure from Utah Traffic Signal Data. Journal of Transportation Engineering, Part A: Systems. 148(10). https://doi.org/10.1061/JTEPBS.0000737

Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. 9, 205–209. doi:10.1136/IP.9.3.205

Lagerwey, P.A., Hintze, M.J., Elliott, J.B., Toole, J.L., Schneider, R.J., and, 2015. Pedestrian and bicycle transportation along existing roadway--ActiveTrans Priority Tool Guidebook. Transportation Research Board. doi:10.17226/22163

Lee, K., Sener, I.N., 2020. Emerging data for pedestrian and bicycle monitoring: Sources and applications. Transp. Res. Interdiscip. Perspect. 4. doi:10.1016/j.trip.2020.100095

Lindsey, G., Han, Y., Wilson, J., Yang, J., 2006. Neighborhood correlates of urban trail use. J. Phys. Act. Heal. 3, S139--S157. doi:10.1123/jpah.3.s1.s139

Lindsey, G., Wilson, J., Rubchinskaya, E., Yang, J., Han, Y., 2007. Estimating urban trail traffic: Methods for existing and proposed trails. Landsc. Urban Plan. 81, 299–315. doi:10.1016/j.landurbplan.2007.01.004

Liu, X., Griswold, J., 2009. Pedestrian volume modeling: A case study of San Francisco. Assoc. Pacific Coast Geogr. 71, 164–181.

Lu, T., Mondschein, A., Buehler, R., Hankey, S., 2018. Adding temporal information to direct-demand models: Hourly estimation of bicycle and pedestrian traffic in Blacksburg, VA. Transp. Res. Part D Transp. Environ. 63, 244–260. doi:10.1016/j.trd.2018.05.011

Mansfield, T.J., Peck, D., Morgan, D., McCann, B., Teicher, P., 2018. The effects of roadway and built environment characteristics on pedestrian fatality risk: A national assessment at the neighborhood scale. Accid. Anal. Prev. 121, 166–176. doi:10.1016/J.AAP.2018.06.018

Mcarthur, A., Savolainen, P.T., Gates, T.J., 2014. Spatial analysis of child pedestrian and bicycle crashes development of safety performance function for areas adjacent to schools. Transp. Res. Rec. J. Transp. Res. 2465, 57–63. doi:10.3141/2465-08

Miranda-Moreno, L.F., Fernandes, D., 2011. Modeling of pedestrian activity at signalized intersections land use, urban form, weather, and spatiotemporal patterns. Transp. Res. Rec. J. Transp. Res. 2264, 74–82. doi:10.3141/2264-09

Monsere, C., Wang, Y., Wang, H., Chen, C., n.d. Risk factors for pedestrian and bicycle crashes FINAL REPORT.

Moradi, A., Soori, H., Kavousi, A., Eshghabadi, F., Jamshidi, E., 2016. Spatial factors affecting the frequency of pedestrian traffic crashes: a systematic review. Arch. trauma Res. 5. doi:10.5812/atr.30796

National Academies of Sciences, Engineering, and Medicine. 2023. Pedestrian and Bicycle Safety Performance Functions. Washington, DC: The National Academies Press. https://doi.org/10.17226/27294.

New York State, 2016. Pedestrian safety action plan.

North Carolina Department of Transportation. 2024. 2024 North Carolina Strategic Highway Safety Plan Update. https://connect.ncdot.gov/groups/echs/Documents/2024/2024%20NC%20SHSP.pdf

North Carolina Department of Transportation. 2024. Access in Appalachia: Pilot Study. https://www.arc.gov/wp-content/uploads/2024/12/Access-in-Appalachia-Pilot-Study-October-2024.pdf

Oh, J.-S., Kwigizile, V., Houten, R. Van, Mckean, J., Abasahl, F., Dolatsara, H., Wegner, B., Clark, M., 2013. Development of Performance Measures for Non-Motorized Dynamics FINAL REPORT.

Omer, I., Gitelman, V., Rofè, Y., Lerman, Y., Kaplan, N., Doveh, E., 2017. Evaluating crash risk in urban areas based on vehicle and pedestrian modeling. Geogr. Anal. 49, 387–408. doi:10.1111/gean.12128

Pande, A., Abdel-Aty, M., 2009. A novel approach for analyzing severe crash patterns on multilane highways. Accid. Anal. Prev. 41, 985–994. doi:10.1016/j.aap.2009.06.003

Pour, A.T.; Moridpour, S.; Rajabifard, A.; Tay, R. Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area. *Road Transp. Res.* **2017**, *26*, 4–20.

Pulugurtha, S.S., Repaka, S.R., 2008. Assessment of models to measure pedestrian activity at signalized intersections. Transp. Res. Rec. J. Transp. Res. Board 39–48. doi:10.3141/2073-05

Rahimi, A.; Azimi, G.; Asgari, H.; Jin, X. Injury Severity of Pedestrian and Bicyclist Crashes Involving Large Trucks. In Proceedings of the ASCE International Conference on Transportation and Development (ASCE ICTD), Seattle, WA, USA, 26–29 May 2020; pp. 110–122.

Rahman, M., Kockelman, K. M., and Perrine, K.A., 2022. Investigating risk factors associated with pedestrian crash occurrence and injury severity in Texas. Traffic Injury Prevention, 23(5): 283-289. https://doi.org/10.1080/15389588.2022.2059474

Rahman, R., Saheli, M. V., and Singleton, P., 2023. Risk Factors for Pedestrian Crashes on Utah State Highway Segments: Results from Parametric and Non-Parametric Approaches Controlling for Pedestrian Exposure. Conference: Transportation Research Board 102nd Annual Meeting. Washington D.C.

Saheli, M.V., Effati, M., 2021. Segment-based count regression geospatial modeling of the effect of roadside land uses on pedestrian crash frequency in rural roads. Int. J. Intell. Transp. Syst. Res. 19, 347–365. doi:10.1007/s13177-020-00250-1

Schneider, R.J., Arnold, L.S., Ragland, D.R., 2009. Pilot model for estimating pedestrian intersection crossing volumes. Transp. Res. Rec. J. Transp. Res. 2140, 13–26. doi:10.3141/2140-02

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. Transp. Res. Rec. 2299, 65–78. doi:10.3141/2299-08

Schroeder, B.J., O'Brien, S.W., and Findley, D.J. 2015. North Carolina pedestrian crossing guidance. NCDOT Project 2014-15.

Seattle Department of Transportation, 2016. City of Seattle bicycle and pedestrian safety analysis.

Seattle Department of Transportation, 2020. City of Seattle bicycle and pedestrian safety analysis Phase 2.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27, 371–389. doi:10.1016/0001-4575(94)00078-Z

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391. doi:10.1016/j.aap.2011.08.003

Singh, M., Cheng, W., Samuelson, D., Kwong, J., Li, B., Cao, M., Li, Y., 2021. Development of pedestrian- and vehicle-related safety performance functions using Bayesian bivariate hierarchical models with mode-specific covariates. J. Safety Res. 78, 180–188. doi:10.1016/j.jsr.2021.05.008

Singleton, P.A., Park, K., Lee, D.H., 2021. Varying influences of the built environment on daily and hourly pedestrian crossing volumes at signalized intersections estimated from traffic signal controller event data. J. Transp. Geogr. 93, 103067. doi:10.1016/j.jtrangeo.2021.103067

Tao W, Aghaabbasi M, Ali M, Almaliki AH, Zainol R, Almaliki AA, Hussein EE. An Advanced Machine Learning Approach to Predicting Pedestrian Fatality Caused by Road Crashes: A Step toward Sustainable Pedestrian Safety. *Sustainability*. 2022; 14(4):2436. https://doi.org/10.3390/su14042436

Thomas, L., Sandt, L., Zegeer, C., Kumfer, W., Lang, K., Lan, B., Horowitz, Z., Butsick, A., Toole, J., Schneider, R.J., 2018. NCHRP Research Report 893: Systemic Pedestrian Safety Analysis, National Cooperative Highway Research Program Project No. 17-73. doi:10.17226/25255

Torbic, D.J., Harwood, D.W., Bokenkroger, C.D., Srinivasan, R., Carter, D., Zegeer, C. V, Lyon, C., 2010. Pedestrian safety prediction methodology for urban signalized intersections. Transp. Res. Rec. J. Transp. Res. Board 2198, 65–74. doi:10.3141/2198-08

U.S. Department of Transportation Federal Highway Administration, 2017. Proven safety countermeasures: Walkways [WWW Document].

Ukkusuri, S., Hasan, S., Aziz, H.M.A., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transp. Res. Rec. J. Transp. Res. Board 2237, 98–106. doi:10.3141/2237-11

Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Saf. Sci. 50, 1141–1151. doi:10.1016/j.ssci.2011.09.012

Virginia Department of Transportation, 2018. Pedestrian safety action plan.

Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning 41, 137–145. doi:10.1016/j.aap.2008.10.001

Xie, Y.; Lord, D.; Zhang, Y. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. *Accid. Anal. Prev.* **2007**, *39*, 922–933.

Zhu, S.Y. Analyse vehicle-pedestrian crash severity at intersection with data mining techniques. *Int. J. Crashworthiness* **2021**, 1–9.

Zuniga-Garcia, N., Perrine, K. A., Kockelman, K. M., 2022. Predicting Pedestrian Crashes in Texas' Intersections and Midblock Segments. Sustainability 14, 7164. https://doi.org/10.3390/su14127164

Appendix A. Literature review

The first subsection provides a summary of statistical methods used to quantify pedestrian safety and corresponding contributing factors. This is followed by a summary of the pedestrian risk analysis approaches currently implemented across the U.S. The next section describes models to predict pedestrian exposure measures, which serve as an important factor in pedestrian safety modeling. Then, a variety of data sources are described. Finally, key takeaways are provided that summarize how the literature review findings were used to inform the present research project.

A.1. Quantification of pedestrian safety and risk factors

NCHRP Report 893: Systemic Pedestrian Safety Analysis provides an overview of methods that can be used to quantify pedestrian risk factors. These include development of statistical models, reliance on expert judgment, and simple frequency-based methods. However, one key limitation of non-statistical methods is that weights assigned to potential risk factors are based solely on expert judgment rather than data. The statistical modeling approach is generally preferred for identification and quantification of pedestrian risk factors. This method is data-driven and reflects the relationships between observed crash frequency and available explanatory factors. Relationships informed by this method have been found to be more reliable than those based on expert judgment or frequency-based methods. Since this project will develop statistical models to identify pedestrian risk factors in North Carolina, the literature review focuses on the statistical methods and results obtained from statistical analysis. *NCHRP Report 893* provides a summary of strengths and limitations of the alternative approaches.

A.1.1 Statistical models

Many statistical modeling approaches have been used to explore associations between pedestrian crash frequency or risk outcomes and potential explanatory variables. Models of crash frequency, typically called Safety Performance Functions (SPFs), relate the expected number of crashes observed at a given location during a specific time period (often annually) to a set of explanatory variables. A variety of statistical modeling methodologies have been proposed to estimate SPFs. In general, SPFs are estimated using count regression techniques that account for the count nature of crash frequency outcomes (i.e., observed crash frequencies take non-negative integer values). The most common count regression model applied for crash frequency prediction is the negative binomial (NB) regression model, which considers and accounts for over-dispersion commonly found in crash data in which the variance of the reported crash frequency exceeds the mean (Shankar et al., 1995). NB regression has been used to estimate the SPFs currently included in the *Highway Safety Manual* (HSM) (American Association of State Highway and Transportation Officials, 2010). The relationship between dependent and independent variables in this approach takes the following general form:

$$\ln \lambda_i = \beta X_i + \varepsilon_i \tag{12}$$

where λ_i is the expected number of crashes at location i (i.e., dependent variable), β are the set of estimable regression parameters, X_i is a collection of geometric design, traffic volume, and other site-specific data for location i (i.e., independent variables), and ε_i is a gamma-distributed error term.

SPFs for crash frequencies along roadway segments typically take the following general form:

$$N_{pred} = e^{\beta_0} \times L^{\beta_1} \times AADT^{\beta_2} \times e^{(\beta_3 X_3 + \cdots + \beta_n X_n)} \tag{13}$$

where N_{pred} is the predicted number of vehicle crashes on a segment, L is the segment length, $AADT$ is the annual average daily traffic (AADT), which is the typical measure of traffic volume on a roadway segment, $\{X_3, \ldots, X_n\}$ is a vector of geometric design and other site-specific data, and$\{\beta_0, \ldots, \beta_n\}$ is a vector of estimable regression coefficients. This form is achieved by using the natural logarithm of segment length and traffic exposure in the development of the statistical model. If available, this pedestrian exposure can be entered into the model in a form similar to vehicle traffic volume as follows:

$$N_{pred} = e^{\beta_0} \times L^{\beta_1} \times AADT^{\beta_2} \times PedEx^{\beta_3} \times e^{(\beta_4 X_4 + \cdots + \beta_n X_n)} \tag{14}$$

where N_{pred} is the predicted number of pedestrian crashes, and $PedEx$ is the pedestrian volume count over a pre-specified time period (e.g., 24-hours). Unfortunately, pedestrian exposure data needed for SPF development in systemic analysis is generally unavailable, especially at the system-wide scale. Instead, surrogate measures can be used as an alternative to pedestrian exposure, or a pedestrian exposure model can be developed to estimate pedestrian activity at a given site based on other characteristics. These surrogate measures and the exposure modeling approach will be described later.

In addition to developing SPFs that estimate pedestrian crash frequency, pedestrian crash risk model can be developed to relate the probability that one or more crashes will be observed at a given location during a specific time period (often several years) to a set of explanatory variables. In the modeling process, the dependent variable is binary, where a value of “1” represents at least one crash observed at a given location during that time period, while a value of “0” represents the lack of a crash. Binary logistic regression is typically used to estimate a model for this crash risk, which takes the following form:

$$\ln \frac{P}{1-P} = \beta X_i \tag{15}$$

where P is the probability that one or more crashes will be observed at location i during the analysis time period, β is the set of estimable regression parameters, and X_i is a collection of geometric design, traffic volume, and other site-specific data for location i (i.e., independent variables).

Other statistical modeling approaches have been used to model the pedestrian crash frequency or risk. For example, Wier et al. (2009) developed models of crash frequency using simple ordinary least squares regression (OLS). Positive counts were considered by using the natural log of crash frequency as the dependent variable; however, this approach did not account for the count nature of the crash frequencies. Poisson regression is a count regression modeling technique used to predict pedestrian crash frequency, but this approach cannot account for overdispersion that is usually observed in crash data (Cottrill and Thakuriah, 2010). Bayesian models were used to investigate the effects of spatial correlation in pedestrian crashes (Siddiqui et al., 2012) and common unobserved heterogeneity shared by pedestrian- and vehicle-related crashes at the same intersection. However, this modeling approach is complex to apply, requires substantial computational power, and produces results that are not always directly interpretable. This makes it a challenge for practitioners to apply the results. Pande and Abdel-Aty (2009) proposed a within stratum matched classification approach to study risk factors for severe pedestrian crashes. However, this approach is not well-suited for systemic safety analysis because network-wide data are not used, and locations where crashes have not yet occurred may not be identified as high-risk. Machine learning methods – such as regression trees (Pour et al., 2017; Rahimi et al., 2020), neural networks (Zhu, 2021), or random forest algorithms (Das et al., 2020; Rahimi et al., 2020; Zhu, 2021) – can be used to identify risk factors associated with increased pedestrian crash frequencies; however, they are more difficult to interpret, particularly when quantifying risk relationships or prioritizing among different risk factors. Machine learning methods are also vulnerable to overfitting data, which can lead to poor generalization and reduced applicability to new data (Tao et al., 2022).

A.1.2 Explanatory variables used as risk factors

In the research literature, SPFs and crash risk models have been developed for both specific roadway elements and regions/zones of a transportation network. The explanatory variables typically used in these models include transportation system attributes, built environment characteristics, demographics characteristics, and pedestrian safety outcomes. Transportation system-related variables include roadway characteristics (such as length of the road segment of interest, functional classification, number of lanes, travel width, number of intersection legs, speed limit), vehicular and pedestrian exposure (typically measured in vehicles or pedestrians per day), and measures of multimodal travel options (e.g., percentage of transit users in the vicinity or number of transit stations). These models also tend to incorporate measures that describe nearby land use characteristics, since the built environment greatly influences pedestrian behavior and activity. Commonly used explanatory variables representing built environment characteristics are land use type and mix, number of schools and parks, total area of parks, etc. Demographics included in these models include the percentage of the population in certain ethnic or racial groups, the percentage of the non-college-educated or high-school-graduate population, and the percentage of the population in different age groups.

NCHRP Report 893 provides a list of typical risk factors known to be associated with increased pedestrian crash frequency and higher pedestrian crash severity. The risk factors noted in the report for roadway segments, which are the focus of this project, are summarized in Table 16.

Table 16. List of pedestrian risk factors for roadway segments identified in NCHRP Report 893

Risk factor	**Crash frequency**	**Crash severity**
Traffic volume	Positive, non-linear	---
Higher functional classification	Positive	---
Proportion of truck/bus traffic	---	Positive
Pedestrian volume	Positive, non-linear*	---
Presence of median or pedestrian crossing island	Negative	---
Presence/number of transit stops	Positive	---
Presence of on-street parking	Positive	---
Presence of leading pedestrian interval	Negative	---
Higher number of lanes	Positive	---
Presence of two-way left-turn lane	Positive	---
Higher speed limits	Positive*	Positive
Vehicle speed	---	Positive
Dark lighting conditions	---	Positive

** denotes risk factor with ambiguous relationship in the research literature*

NCDOT Project 2022-11 applied similar methodology, developing pedestrian crash frequency models for urban roadway segments in North Carolina using NB regression techniques. The identified risk factors were estimated for total crash frequency and KA (fatal and suspected serious injury) crash frequency for the following functional classifications in urban areas:

- Principal Arterials – Others
- Minor Arterials
- Major + Minor Collectors (combined)
- Local

In addition to risk factors, each crash frequency model included a categorical variable for the NCDOT engineering division to account for regional differences in the model. The pedestrian crash risk factors from this project are summarized in Table 17.

Table 17. List of pedestrian risk factors for roadway segments identified in NCDOT Project 2022-11

Risk Factor	Principal Arterial - Others		Minor Arterials		Major + Minor Collectors		Local	
	Total	KA	Total	KA	Total	KA	Total	KA
Traffic volume	Positive	Positive	Positive	Positive	Positive	Positive	---	---
5+ lane roadway	Positive	Positive	Positive	Positive	---	---	---	---
Speed limit 35 mph or above	---	---	Positive	Positive	---	---	---	---
Speed limit 40 mph or above	---	---	---	---	---	---	Positive	Positive
Speed limit 40 or 45 mph	---	Positive	---	---	---	---	---	---
Speed limit 50 mph or above	---	Positive	---	---	---	---	---	---
Median present	Negative	Negative	Negative	Negative	Negative	Negative	---	---
Block length between 0.1-0.25 mi	Negative	Negative	Negative	---	Negative	---	Negative	Negative
Block length between 0.25-0.5 mi	Negative	Negative	Negative	---	Negative	Negative	Negative	Negative
Block length greater than 0.25 mi	---	---	---	Negative	---	Negative	---	---
Block length greater than 0.5 mi	Negative	Negative	Negative	---	Negative	---	Negative	Negative
High intensity development within 100 ft	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Alcohol sales density	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Bus route present	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
Population density	Positive	---	Positive	---	Positive	---	Positive	---
K12 enrollment density	Positive	Positive	Positive	Positive	---	---	Positive	Positive
Median household income	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative
Proportion of commuters non-motorized	Positive	---	Positive	---	Positive	---	---	---
Proportion of population disabled	Positive	Positive	Positive	Positive	Positive	Positive	---	---
Proportion of population age 65+	---	---	Negative	Negative	Negative	Negative	Negative	Negative
Proportion of zero vehicle households	---	---	Positive	Positive	---	---	Positive	---

NCHRP Research Report 1064 details the development of new pedestrian and bicycle SPFs to support the second edition of the HSM (HSM2). However, the crash prediction models from Part C: Predictive Method in the current edition of the HSM are primarily designed for predicting motor vehicle crashes. In Part C, pedestrian and bicycle crashes are predicted as a generalized proportion of motor vehicle crashes. The proposed approach for crash prediction in *NCHRP Research Report 1064* accounts for site-specific risk factors at urban and suburban intersections and roadway segments using traditional NB regression and U.S. Roadway Assessment Program (usRAP) methodology, with pedestrian exposure represented by annual average daily pedestrian crossing volumes and peak-hour volumes, respectively. Logistic regression was also employed to determine the effects of pedestrian risk factors on facilities without available pedestrian exposure data. Table 18 and Table 19 summarize the effects of risk factors as determined in *NCHRP Research Report 1064* using NB and logistic regression methods. The models in Table 18 were estimated considering only roadway type and exposure (reduced models) and additional roadway features (expanded models).

Table 18. Pedestrian risk factors for urban/suburban arterials from NCHRP Research Report 1064

	Urban 2-lane undivided roads		Urban 4-lane undivided and 4-lane divided roads		One-way roads	
Risk Factor	Reduced Model	Expanded Model	Reduced Model	Expanded Model	Reduced Model	Expande d Model
Traffic volume	Positive	Positive	Positive	Positive	Positive	Positive
Pedestrian volume	Positive	Positive	Positive	Positive	Positive	Positive
Indicator for divided road	---	---	Negative	Negative	---	---
Sidewalk buffer >0ft	---	Negative	---	---	---	Negative
Average lane width	---	Negative	---	---	---	Negative
Number of transit stops within 1,000ft	---	Positive	---	---	---	---
Number of schools within 1,000ft	---	---	---	Positive	---	---
Number of alcohol sales locations within 1,000ft	---	---	---	---	---	Positive
Indicator for 2-lane road	---	---	---	---	Negative	Negative
Indicator for 3-lane road	---	---	---	---	Negative	Negative

Table 19. Pedestrian risk factors from logistic regression in NCHRP Research Report 1064

Risk Factor	Rural roads	Rural two-lane roads	Urban 2-lane undivided roads	Urban 4-lane undivided and 4-lane divided roads	One-way roads
Traffic volume	Positive	Positive	Positive	Positive	Positive
Population	Positive	Positive	Positive	Positive	---
Segment length	Positive	Positive	Positive	Positive	Positive
Number of intersections	---	---	Positive	Positive	Positive
Lane width > 11ft	Negative	Negative	---	---	---
Lane width >= 12ft	---	---	---	---	---
Average shoulder width greater than 3ft	Negative	Negative	---	---	---
Outside shoulder width (ft)	---	---	Negative	Negative	---
Total roadway width (ft)	---	---	---	---	Positive
Divided segment	Negative	---	---	Negative	---

Table 20 and Table 21 (later) summarize explanatory variables identified in the literature for the segment/intersection-level models to estimate pedestrian crash frequency/risk and zone-level models, respectively. Positive and negative signs are included to show the direction of relationships between the different risk factors and pedestrian crash outcomes. The remainder of this section describes key findings and trends from these studies at both the individual segment and zone levels.

Table 20. A summary of explanatory variables in segment/intersection-level based models

Study Model	Independent variables[6]
NB model (Torbic et al., 2010)	• Total traffic volume entering intersection (+) • Minor road ADT as fraction of major road ADT (+) • Pedestrian volume (+) • Maximum number of lanes crossed by pedestrians (+) • Presence of bus stops within 1,000 ft of the intersection (+) • Presence of school within 1,000 ft of the intersection (+) • Number of alcohol sales within 1,000 ft of the intersection (+) • Average per capita income of all census block groups within 1,000 ft of the intersection (-) • Number of commercial structures on commercial land parcels within 0.5 mi of the intersection (+)
NB model (Omer et al., 2017)	• Log of hourly pedestrian traffic volume (PMV) (+, -) • Log of hourly vehicle traffic volume (VMV) (+) • Log of segment length based on vehicle's map (+) • Presence of commercial front (-) • Log of segment length based on pedestrian map (+)
Within stratum matched case-control sampling (Pande and Abdel-Aty, 2009)	• Percentage of trucks (PT) <8.75 (+) • PT >8.75 on weekday PM peak hours (-) • PT>8.75 on Friday or Saturday night (+) • PT>8.75 on weekday AM peak hours (-) • Sidewalk < 6ft (-) • Presence of horizontal curvature (+) • Presence of attenuators (+) • Presence of parking (+)
Geospatial count regression techniques (Saheli and Effati, 2021)	• Fraction of residential land use near segment (+) • Fraction of commercial land use near segment (+) • Fraction of governmental land use near segment (+) • Fraction of religious land use near segment (+) • Presence of a median crossover (+)
NB regression (Arias et al., 2021)	• Traffic volume (+) • Segment length (+) • Indicator for short TMC for probe speed data (+) • Local road (+) • Small town (+) • More than four lanes (+) • Urban area (+) • 85th percentile – median speed (+) • Median – 15th percentile speed (-)

[6] + indicates positive correlation with an increase in pedestrian crashes, while – indicates a negative correlation with an increase in pedestrian crashes.

Study Model	Independent variables[6]
Logistic regression (Hamilton et al., 2021)	• Segment length (+) • Pedestrian volume (-) • Interaction between traffic volume and pedestrian exposure (+) • Lower speed limit (-) • Number of lanes (+/-)
Zero-inflated NB regression (Islam et al., 2022)	• NB Regression Portion • Pedestrian volume (+) • Traffic volume (+) • Midblock crossing (-) • 3-leg intersection (-) • 5-leg intersection (-) • Diverging diamond interchange (-) • Single-point urban interchange (+/-) • Number of crosswalks (+) • Crosswalk length (+) • Right-turn on red (+) • Bike lanes (-) • Bus stops (+) • Vacant land use (+) • Employment density (-) • Percent population with disability (-) • Percent population non-white race/ethnicity (-) • Zero-inflated Portion • Pedestrian volume (-) • Population density (-) • Percent population non-white race/ethnicity (+)
NB regression (Rahman et al., 2022)	• Number of lanes (+) • Median width (-) • Shoulder width (-) • Roadway curvature (+/-) • Curve angle (-) • Speed limit (-) • Traffic volume (-) • Truck percentage (+) • Population density (+) • Job density (+) • Rural (-) • Small urban (-) • Large urban (+) • Rainfall (-) • Distance to nearest school (-) • Distance to nearest hospital (+/-) • Transit stops (+)

Study Model	Independent variables[6]
NB regression (Zuniga-Garcia et al., 2022)	• Walking density (+) • Signalized intersection (+) • Number of approaches (+) • Speed limit (-) • Number of lanes (+) • Lane width (+) • Median width (-) • One-way road (+/-) • AADT per lane (+) • Truck percentage (+) • Arterial (+) • Transit stops (+) • Distance to nearest hospital (-)
NB regression (Rahman et al., 2023)	• Pedestrian volume (AADP) (+) • Traffic volume (AADT) (+) • Truck proportion of AADT (+) • One-way segment (-) • Two-way left turn lanes (+) • Presence of pedestrian barrier (-) • Number of driveways (+) • Number of bus stops (+) • Employment density (-) • Percent zero-vehicle households (+) • Percent population with disability (-) • Percent population non-white race/ethnicity (-)

Many of the risk factors have expected unambiguous relationships. Pedestrian crash frequency generally increases with higher traffic volume, as more vehicle traffic creates more opportunities for crashes. Higher functional classification roadways (e.g., arterials) are also generally related to higher risk for severe pedestrian crashes, due to designs that favor vehicle movements. The presence of more vehicle travel lanes and two-way left-turn lanes is also associated with an increased risk of pedestrian crashes. Higher vehicle speeds and speed limits, a greater presence of large-sized vehicles, and dark lighting conditions are also associated with increased pedestrian crash severity. The presence of pedestrian refuge islands or medians, and leading pedestrian intervals at intersections, is associated with decreased pedestrian crash frequency. These features related to decreased pedestrian crash frequency generally help provide pedestrians with protection from vehicles and are often considered in transportation safety improvements.

Other features that show expected positive relationships with increased pedestrian crash risk include the presence and number of transit stops near the location of interest, on-street parking, alcohol sales near the location of interest, and the presence of schools. Land use patterns also play an important role; one study found that the number of commercial structures near a location was associated with increased pedestrian crash frequency, even after controlling for pedestrian volume, likely due to increased pedestrian exposure and vehicle traffic interactions (Torbic et al., 2010).

However, not all risk factors are unambiguous. *NCHRP Report 893* notes that pedestrian crash frequency increases with increased pedestrian volume, and NCDOT Project 2022-11 and *NCHRP Research Report 1064* had similar findings. While higher pedestrian activity is generally associated with more pedestrian crashes at a given location, several studies have found that increased pedestrian activity is not always associated with an increased risk for pedestrian crashes. Specifically, the increase in crash frequency is often less than proportional to the increase in pedestrian volume, which would reduce the overall crash risk. In some cases, increased pedestrian activity might also be associated with reduced pedestrian crash frequency. This implies a "safety in numbers" effect, in which higher pedestrian volumes at a given location may contribute to a reduced crash risk or frequency, potentially due to increased visibility or driver expectations (Elvik and Bjørnskau, 2017; Hamilton et al., 2022; Jacobsen, 2003). Rahman et al. (2023) found that pedestrian volumes increased pedestrian crash frequency in models estimated using NB regression; however, results from the same data using a marginal effects analysis suggest that increases are smaller for locations with more than 150-200 daily pedestrians, further supporting the safety in numbers effect. The presence of pedestrian improvements also has unclear impacts on pedestrian crash frequency or risk. For example, Pande and Abdel-Aty (2009) found that sidewalk presence was associated with increased pedestrian crash frequency, even though sidewalks are generally expected to make pedestrian movement safer. Similarly, Islam et al. (2022) found that pedestrian crash frequency increased with additional crosswalks, despite the expectation that crosswalks increase pedestrian visibility and reduce crash frequency. However, some of these findings may reflect correlation with exposure measures: locations with sidewalks and crosswalks tend to have higher pedestrian activity than those without, and the associated increase in crash frequency may reflect higher pedestrian exposure. While sidewalks and other pedestrian infrastructure may provide safety benefits, they may not always be strong predictors of pedestrian crash risk (U.S. Department of Transportation Federal Highway Administration, 2017). Understanding the circumstances under which these features are positively or negatively related to pedestrian crash frequency or risk is crucial for making more effective improvements to pedestrian safety.

The review of the literature suggests that other features used to explain pedestrian crash risk may also have nuanced effects on pedestrian crash frequency at the segment level. One of these "less studied" risk factors is speed. While speed (often measured by speed limit) is generally associated with increased crash severity (Hamilton et al., 2021), it is less clear how speed is related to crash frequency. Rahman et al. (2022) and Zuniga-Garcia et al. (2022) suggest that although crash frequency decreases with higher speed limits, higher speed limits are more likely to result in greater crash injury severity. However, locations with higher speed limits are often less pedestrian-friendly and the results may be skewed by lower pedestrian exposure. Arias et al. (2021) estimated a model of pedestrian and bicycle crash frequency that considered probe speed data at the segment level. The results revealed that the difference between the 85th percentile and median speed observed was associated with increased crash frequency, while the difference between median and 15th percentile speed was associated with decreased crash frequency. The latter could indicate that roadways with high levels of congestion—which would experience a larger difference between median and 15th percentile speeds—might be associated with reduced pedestrian and bicycle crash frequency. This may be because vehicles generally travel more slowly in congested conditions and may be better able to avoid conflicts with non-motorized roadway users. Hamilton et al. (2021) found that pedestrian crash risk was lower at sites with lower speed limits (25 to 35 mph) compared to those with higher speed limits. They also found that direct measures of vehicle speeds obtained from probe data could replace surrogates for speed (or speed limits directly) in models of pedestrian crash severity. These two studies suggest that probe speed data could be an important, but generally less studied, predictor in pedestrian crash risk models.

Table 21. A summary of explanatory variables in zone-level based models

Study	Independent Variables[7]
Multivariate, area-level regression model (Wier et al., 2009)	• Traffic volume (+) • % arterial streets without public transit (+) • % land area zoned commercial (+) • % residential-neighborhood commercial (+) • Land area (-) • Number of employees (+) • Population (+) • % living below poverty (+) • % older than 65 (-)
Exploratory analysis and Poisson regression (Cottrill and Thakuriah, 2010)	• Squared transit availability index (+) • Pedestrian accessibility index (+) • Squared sum of annual average daily traffic (+) • Squared total miles of roads (-) • Total number of schools (+) • Population density (+) • Crime rate (+) • Low pedestrian accessibility index (binary) (+) • Median household income (-) • Percent with no cars (+) • Percent commercial (+) • Percent children (-) • Percent who speak limited or no English (+)
Bayesian spatial analysis (Siddiqui et al., 2012)	• Total length of roadways with 35 mph posted speed limit (+) • Total number of intersections per TAZ (+) • Median household income per TAZ (-) • Total number of dwelling units (+) • Log of population per square mile of a TAZ (+) • Percentage of households with non-retired workers but zero auto (+) • Percentage of households with non-retired workers and one auto (+) • Long term parking cost (+) • Log of the total employment number in a TAZ (+)

[7] + indicates positive correlation with an increase in crashes, while – indicates a negative correlation with an increase in pedestrian crashes.

Study	Independent Variables[7]
NB model (Ukkusuri et al., 2011)	• Census tract population of 2000 (+) • Proportion of African-American population (+) • Proportion of Hispanic population (+) • Median-age population proportion (-) • Proportion of the population who are high school graduates (+) • Proportion of uneducated population (+) • Industrial land use proportion (+) • Open land use proportion (+) • Commercial land use proportion (+) • Total park area (-) • Total number of schools (+) • Total number of all-way stop intersections (+) • Total number of signalized intersections (+) • Number of three-approach intersections (-) • Number of five-approach intersections (+) • Number of subway stations in tract (+) • Number of bus stops in tract (+) • Primary roadway (with limited access) proportion of total roadway length (-) • Primary roadway (without access restriction) proportion of total roadway length (+) • Local rural road proportion of total roadway length (-) • Other thoroughfare roadway proportion of total roadway length (-) • Four-lane roadway proportion of total roadway length (+) • Five-lane roadway proportion of total roadway length (+) • Proportion of length of one-way streets to total roadway length (+) • Proportion of length of roads with widths less than 30 ft to total roadway length (-)
Zero inflated negative binomial (ZINB) model and ZINB mixed model (ZINBMM) (Mansfield et al., 2018)	• Vehicle-miles traveled (VMT) on highways (+) • VMT on principal arterials (+) • VMT on minor arterials (+) • VMT on major collectors (+) • Population density (-) • Employment density in office (-) • Employment density in retail (+) • Employment density in industry (-) • Employment density in general services (-) • Activity mix index (+) • Auto-oriented intersection density (+) • Non-auto-oriented intersection density (+)
Bayesian joint hierarchical approach (Singh et al., 2021)	• Override length on mainline (+) • Average daily traffic on mainline (+) • Average daily traffic on cross street (+) • Intersection rate group (-) • Estimated annual pedestrian volume (-) • Other categorical variables

Study	**Independent Variables**[7]
Regression modeling (Ha and Thill, 2011)	• Density of businesses (+) • Population density (+) • % under poverty level (+) • % of A2*A3 intersection (+) • % of A3*A4 intersection (+) • % population over 65 (-) • % of signalized intersection (+) • % African Americans (+) • % pedestrian commuters (-) • % of Caucasians (-)
Negative binominal model (Dumbaugh et al., 2013)	• Block group acreage (-) • Median household income (thousands) (-) • Population age (+) • Population age 65 and older (+) • Vehicle miles of travel (millions) (+) • Net population density (+) • # three-leg intersections (-) • # four or more leg intersections (+) • Miles of arterials (+) • # big box stores (+) • # strip commercial uses (+) • # pedestrian-scaled retail uses (-)
Review (Moradi et al., 2016)	• Number of schools (+) • Population density (+) • Traffic volume (+) • Improvement of socioeconomics (-) • Number of intersections (+) • Commercial land use (+) • Pedestrian volume (+)
Geographic weighted regression (Almasi and Behnood, 2022)	• Bus stop (+) • School (+) • Pedestrian overpass (+) • Intersection (+) • Total population (+) • Children population (+) • Elderly population (+) • Number of motorcycles (+) • Number of cars (+) • Residential land use (+) • Average width of roadways (+) • Average number of lanes (+) • Average length of median refuges (+) • Average width of sidewalks (+) • Average speed (+) • Average slope (+) • Speed cameras (+)

Study	Independent Variables[7]
NB regression (Forrest et al., 2022)	• Log of walking commuting population (-) • Number of schools (+) • Greenspace (-) • Cars per household (-) • Black, Asian, and minority ethnic population (+) • Log of crime rate (+) • Children in child benefit families (+) • Alcohol expenditure (+) • Job density (+)

Zone-level models have found that improved roadway connectivity, closer proximity of origins and destinations, and various socioeconomic characteristics are associated with higher pedestrian crash frequency within geographic zones (Mansfield et al., 2018; Moradi et al., 2016; Siddiqui et al., 2012; Ukkusuri et al., 2012). Zonal studies have also observed a relationship between direct measures or proxies of roadway and vehicular traffic characteristics and pedestrian crash frequency. Ukkusuri et al. (2012) noted that greater mileage of higher functional classification roadways, a higher number of lanes, transit ridership and subway stations, the presence of four- and five-way intersections, and a higher number of teenagers were associated with increased pedestrian crash frequency. Moreover, they found that a higher proportion of people aged 65 and older was correlated with higher fatal crash frequency. On the other hand, residential land use and all-way-stop and three-way intersections are correlated with lower pedestrian crash frequency. Mansfield et al. (2018) found several neighborhood, demographic, socioeconomic, employment, traffic, and infrastructure characteristics to be significant indicators of pedestrian crash occurrence. Specifically, they found a significant increase in the likelihood of a fatal pedestrian crash associated with higher traffic volume density (e.g., vehicle miles traveled per square mile), retail employment density, and multimodal intersection density.

Siddiqui et al. (2012) noted a relationship between an increased pedestrian crash frequency and higher posted speed limits, population and employment density, intersection density, dwelling unit density, and the percentage of households with fewer than two vehicles. On the other hand, they found a relationship between decreased pedestrian crash frequency and median household income. Cottrill and Thakuriah (2010) found that transit availability and pedestrian accessibility are positively correlated with increased pedestrian crash frequency. They also found that the number of schools, crime levels, children pedestrians, and the percentage of people without a car are positively related to increased pedestrian crash frequency. Ukkusuri et al. (2011) and Wier et al. (2009) found that residential neighborhoods, commercial land use, and a higher number of schools within an analysis zone are related to a higher number of pedestrian crashes. Ukkusuri et al. (2011) also found that total population, the percentage of high school graduates, and the percentage of people without formal education are positively correlated with increased pedestrian crashes. Similarly, Almasi and Behnood (2022) found that population characteristics (especially child and elderly populations), transit presence, intersections, schools, higher speeds, residential land use, and vehicle and motorcycle ownership are all correlated with increased pedestrian crash frequency. The study also identified that, across a zone, roadway slope, the presence of speed cameras, average roadway width, and the number of lanes are correlated with increased pedestrian crash frequency. The length of median refuges and width of sidewalks are also associated with increased pedestrian crash frequency. However, these results are counterintuitive and may reflect countermeasures implemented in zones with high pedestrian activity.

While these studies provide some clear relationships between land use, demographic characteristics, socioeconomic characteristics, and pedestrian crash risk, the relationships between factors that combine some of these features into a single metric are less clear. For example, few studies have examined the relationships between aggregate metrics such as social health metrics or food desert indicators and pedestrian crash frequency/risk. The closest social health metrics identified were in a 2022 study by Forrest et al., which found that crime rate, child benefit program eligibility, and alcohol expenditures were all positively related to increased pedestrian crash frequency. Beyond this study, little research has explored the relationship between pedestrian crash frequency and public health or social capital indicators, limiting understanding of how pedestrian safety measures and improvements relate to equity at a systematic level.

A.2. Models of pedestrian exposure

Pedestrian activity/exposure, measured either through direct counts, estimates of counts, or surrogate variables (e.g., population density), is one of the most critical factors that can be used to describe pedestrian crash frequency and risk. The level of pedestrian activity is related to some of the aforementioned factors, as pedestrian activity is often influenced by the built environment, proximity of origins and destinations, and the demographic and socioeconomic characteristics of a neighborhood. Unlike traffic volume, pedestrian volume data traditionally are not collected or available widely across roadway networks. In fact, pedestrian volumes tend to be collected sporadically, either as part of other traffic volume studies (i.e., when convenient to do so) or, if part of a pedestrian counting program, at a small subset of locations. Obtaining pedestrian counts widely across a roadway network at a scale sufficient to estimate pedestrian volumes would be extremely resource-intensive, due to high labor demands for manual counts or the expense of automated counts. Moreover, high variability, shorter trip lengths, and detection challenges make accurate pedestrian volume counts less feasible using either manual or automated methods (Lagerwey et al., 2015).

One alternative to these methods is to estimate pedestrian activity using pedestrian push-button information that is available in high-resolution traffic signal controllers (Singleton et al., 2021). However, the accuracy of this method is questionable, as pedestrian push-button actuations can vary significantly and may represent only a small fraction of pedestrian users, especially when large numbers of pedestrians are present. Therefore, the presence of destinations and the potential for new or enhanced walking connections are often more valuable pieces of information than directly counting the number of people who currently walk at a given location.

NCHRP Report 770: Estimating Bicycling and Walking for Planning and Project Development, A Guidebook provides a summary of pedestrian demand modeling research up to its publication date of 2014. The report identifies three general model types that have been used to estimate pedestrian activity: trip generation and flow models, network simulation models, and direct demand models.

The focus here is primarily on direct demand models, which estimate the level of pedestrian activity at a given location as a function of a set of explanatory variables. The most common statistical modeling technique used in direct demand models to estimate pedestrian activity is NB regression, since pedestrian counts take count outcomes, similar to crash observations. Several studies developed direct demand models using either log-linear OLS regression or NB regression to estimate pedestrian volumes along roadway segments and intersections using site and surrounding area characteristics (Behnam and Patel, 1977; Griswold et al., 2019; Hankey et al., 2012). These models assume that pedestrian count or volume is a function of the built environment and demographic attributes of the surrounding area. Hankey et al. (2012) developed OLS and NB models to identify factors related to increases in bicycle and pedestrian traffic volumes. However, linear regression models can produce unrealistic parameter estimates when negative count values for pedestrian exposure are included.

Other methods used in the literature to model pedestrian exposure include stepwise linear regression and supervised models. Stepwise linear regression has been used for pedestrian volume modeling to consider independent variables at varying spatial scales (Hankey et al., 2017, 2012; Hankey and Lindsey, 2016; Lu et al., 2018). Although the stepwise regression approach can select independent variables at different spatial scales, it is atheoretical, meaning it can result in the inclusion of variables that are counterintuitive or inconsistent with theory, complicating interpretation and limiting transferability. To address this issue, Hankey and Lindsey (2016) proposed two supervised approaches that were easier to interpret and simpler to apply. Hampshire et al. (2018) developed an origin-destination model with factors such as the built environment and the distance between origins and destinations to understand how they influence pedestrian travel demand.

Further, there have been modeling efforts using short-term count data to create expansion factors for predicting long-term pedestrian counts. Griswold et al. (2019) developed a log-linear regression model to calculate annual volume estimates using both short-term and long-term pedestrian count data. A study by the Wisconsin Department of Transportation (WisDOT) adjusted short-term count data to estimate annual values, which were then compared to characteristics from pedestrian demand models to assess which characteristics have a greater or lesser impact on pedestrian volumes (WisDOT Pedestrian Exposure Model, 2021). An NB regression model was developed based on key factors including population density, job density, bus stops, retail businesses, food and drink businesses, schools, and households without a vehicle.

Direct demand models focus more on individual links rather than individual movement within a transportation network or travel between or across links. To account for these, Cooper et al. (2019) proposed an assignment model using multiple variants of spatial network betweenness in a regression model. Moreover, instead of testing their model at a single point in time, they adopted a longitudinal evaluation that considers changes in response over time. However, this method requires complicated processes and data collection, making it less applicable for general planning policy recommendations.

NCDOT Project 2022-11 developed direct demand models for estimating pedestrian exposure at urban roadway intersections using an NB regression model with pedestrian count data from across the state. Pedestrian exposure in this model was estimated as a function of demographic characteristics, roadway characteristics, and land use data, with indicator variables to account for regional differences in NCDOT engineering divisions and count durations. The final pedestrian exposure model is summarized in Table 22.

Table 22. Summary of exposure model for North Carolina intersections from NCDOT Project 2022-11

Factor	Effect on Exposure Estimate
Traffic Volumes	Negative
Max Speed >= 45 mph	Negative
One leg classified as other freeway	Negative
One leg classified as major arterial	Negative
One leg classified as minor arterial	Negative
Bus stop within 1/4 mile	Positive
Land use mix	Positive
Parcel count	Positive
Proximate to college/university	Positive
Greenway present	Positive
1-5 Alcohol sales locations nearby	Positive
6+ Alcohol sales locations nearby	Positive
Total population	Positive
Total employment	Positive
K12 school count	Positive
Proportion of population age 65+	Negative
Proportion of population age 18 or younger	Negative
Proportion of non-motorized commuters	Positive
Median income	Positive

Findings from NCDOT Project 2022-11 indicate that the pedestrian exposure model could not accurately predict pedestrian counts due to highly variable data ranging from 0 to 14,854, with 94.5 percent of those counts below 500. Essentially, the model was skewed by exceptionally high outlier values; however, these were valid counts and therefore needed to be represented. The authors indicate that this model may be useful for identifying relative levels of pedestrian activity, but the specific predicted values may not be directly applicable.

Table 23 and Table 24 summarize the variables in segment/intersection-level and zone-level pedestrian exposure models and the associated data sources, respectively. Positive and negative signs are included to show the direction of relationships between the different variables and pedestrian exposure. The remainder of this section describes key findings and trends from these studies at both the individual segment and zone levels.

Table 23. A summary of variables in segment/intersection-level pedestrian exposure models

Model	Independent variables[8]
Multiple regression analysis through backward elimination (Pulugurtha and Repaka, 2008)	• Population (+) • Total employment (+) • Urban residential area (+) • Urban residential commercial area (-) • Number of transit stops (+) • Single-family residential area (-) • Mixed land use area (-)
OLS regression (Schneider et al., 2009)	• Population (within 0.5 mi of intersection) (+) • Total employment (within 0.25mi of intersection) (+) • Number of commercial retail properties (within 0.25mi of intersection) (+) • Number of regional transit stations (within 0.10 mi of intersection) (+)
Linear regression with backward elimination (Liu and Griswold, 2009)	• Population density (+) • Job density (-) • Residential land use (-) • Transit stop density (MUNI) (+) • Presence of bike lane (+) • Mean slope (-) • Patch richness density (+)
Linear regression (Haynes and Andrzejewsk, 2010)	• Employment density (+) • Neighborhood shopping district proximity (+) • Bus frequency (+) • Distance from the ocean (-) • Average speed limit of approaches (-)
Log-linear regression (Miranda-Moreno and Fernandes, 2011)	• Population (+) • Commercial space (+) • Open space (-) • Presence of a subway station (+) • Number of bus stations (+) • Number of schools (+) • % major arterials (-) • Number of street segments (+) • Presence of a four-way intersection (+) • Distance to downtown (-) • Max. temperature >32°C (-) • Min. temperature <-20°C (-)
Log-linear regression (Schneider et al., 2012)	• Total households within ¼ mi (+) • Total employment within ¼ mi (+) • Intersection is in a high-activity zone (+) • Maximum slope on any intersection approach leg (-) • Intersection is within ¼ mi of a university campus (+) • Intersection is controlled by a traffic signal (+)

[8] + indicates positive correlation with an increase in pedestrian crashes, while – indicates a negative correlation with an increase in pedestrian crashes.

Model	Independent variables[8]
Log-linear regression model (Griswold et al., 2019)	• Population (+) • Number of employees (+) • Number of street segments (+) • Walk commute mode share (+) • Number of schools (+) • Principal arterial (+) • Minor arterial (+) • Four-way intersection (+)
OLS (Hankey et al., 2012)	• Average number of violent crimes/year (+) • Measure of mixing of land uses (-) • Recorded precipitation (-) • Arterial street (+) • Collector street (+)
NB (Hankey et al., 2012)	• Percentage of non-white neighborhood residents (+) • Percentage of neighborhood residents with a college education (+) • Distance from nearest body of water (-) • Distance from the CBD (-) • Recorded precipitation (-) • Principal street (-) • Arterial street (+) • Collector street (+)
Stepwise linear regression (Hankey and Lindsey, 2016)	• Population density (+) • Job accessibility (+) • Retail area (+) • Industrial area (-) • Open space area (+) • Number of transit stops (+) • Number of major roads (+) • Number of off-street trails (+)
Stepwise linear regression (Hankey et al., 2017)	• Area-weighted average population density (+) • Number of residential addresses (-) • Length of sidewalks (+) • Length of off-street trails (-) • Number of bus stops (+) • Area-weighted average household income (-)

Model	Independent variables[8]
Stepwise linear regression (Lu et al., 2018)	• Population density (+) • Residential addresses (-) • Non-residential addresses (+) • Industrial area (-) • Number of bus stops (+) • Household income (-) • Length of major roads (+) • Length of local roads (+) • Number of intersections (-) • Number of sidewalks (+) • Time of day (dummy variables) • 0:00-4:00 (-) • 4:00-8:00 (-) • 8:00-12:00 (+) • 12:00-16:00 (+) • 16:00-20:00 (+)
Multiple regression model (Lindsey et al., 2007, 2006)	• Neighborhood population density (+) • Percent of neighborhood in commercial use (+) • Income (+) • Education (+) • Vegetative health (+) • Area of land in parking (+) • Mean length of street segments in access networks (+) • Percentage of neighborhood residents in age groups greater than 64 and less than 5 (-)
NB regression (Hamilton et al., 2021)	• Lowest posted speed limit on any approach of 25 or 30 mph (+) • Sidewalk present (+) • Intersection is signalized (+) • Bus stop present (+) • Higher land use mix (+) • Arterial approach present (-) • Four approach legs (+) • Total number of transit commuters within 0.5 mi (+) • Total population within 0.5 mi (+) • Total employment within 0.5 mi (+)

Table 24. A summary of variables in zone-level pedestrian exposure models

Model	Independent variables[9]
Stepwise multiple regression model (Behnam and Patel, 1977)	• Commercial space (+) • Office space (+) • Cultural and entertainment space (+) • Residential space (+) • Vacant space (+) • Storage and maintenance space (+)
Log-linear regression model (Singleton et al., 2021)	• Population density (+) • Employment density (+) • Household size (+) • Household income (-) • Vehicle ownership (-) • % residential land use (+) • % commercial land use (+) • Intersection density (+) • 4-way intersections (+) • # schools (+) • # places of worship (+) • # transit stops (+) • Park acreage (+)
NB regression and linear regression for home-based trips and non-home based trips (Hampshire et al., 2018)	• Household size (-) • Number of vehicles in each household (-) • Number of employees in each household (-) • Pedestrian index of environment (+)

In general, studies using direct exposure models suggest that higher population or population density, used as proxy variables, is associated with greater pedestrian activity or volume at roadway segments or intersections. Additionally, land use is found to be related to pedestrian activity; however, some studies using these models report contradictory results. Generally, urban residential areas are associated with decreased pedestrian activity, while commercial areas are associated with increased pedestrian activity (except for Pulugurtha and Repaka (2008) whose study found the opposite relationships). Employment, or employment density, is generally shown to have a positive impact on pedestrian activity (except for Liu and Griswold's 2009 study suggesting that job density had a negative impact on pedestrian activity). Further, roadway segments or intersections near universities or schools were found to have higher pedestrian activity (Griswold et al., 2019; Miranda-Moreno and Fernandes, 2011; Schneider et al., 2012; Singleton et al., 2021). Most studies also found that transit-related variables (e.g., number of transit stops or bus frequency) tend

[9] + indicates positive correlation with an increase in pedestrian crashes, while – indicates a negative correlation with an increase in pedestrian crashes.

to have a positive contribution to pedestrian activity (Hankey et al., 2017; Hankey and Lindsey, 2016; Haynes and Andrzejewsk, 2010; Liu and Griswold, 2009; Lu et al., 2018; Miranda-Moreno and Fernandes, 2011; Pulugurtha and Repaka, 2008; Schneider et al., 2009; Singleton et al., 2021). Roads on a slope are associated with lower pedestrian activity (Liu and Griswold, 2009; Schneider et al., 2012), while areas with longer roads (Lindsey et al., 2007, 2006; Lu et al., 2018), longer sidewalks (Hankey et al., 2017; Lu et al., 2018), and bike lanes (Liu and Griswold, 2009) are associated with higher pedestrian activity. Further, most studies found that minor arterials or collector roads have higher pedestrian activity compared to principal or major arterials (Griswold et al., 2019; Hankey et al., 2012; Miranda-Moreno and Fernandes, 2011). Regarding demographic characteristics, the educational attainment level of neighborhood residents is positively related to pedestrian activity (Hankey et al., 2012; Lindsey et al., 2006, 2007). However, it is unclear how household income impacts pedestrian activity (Lindsey et al., 2006, 2007; Lu et al., 2018). Finally, the models also found that weather conditions can affect pedestrian activity, with precipitation or cold weather associated with lower pedestrian activity (Hankey et al., 2012; Miranda-Moreno and Fernandes, 2011).

At the zone level, the studies suggest that population density, employment density, and commercial density are all positively related to pedestrian activity (Behnam and Patel, 1977; Singleton et al., 2021). Household size is shown to have both positive (Singleton et al., 2021) and negative (Hampshire et al., 2018) relationships with pedestrian activity, which may be due to correlations with other variables. Further, Singleton et al. (2021) found that pedestrian activity increases with more schools, more transit stops, and higher intersection density, which indicates shorter road segments, greater road connectivity, and potentially shorter paths to destinations. The study also found that pedestrian activity decreases with higher household income and vehicle ownership. The latter may reflect greater reliance on personal vehicles in areas with less developed pedestrian infrastructure or land uses less conducive to pedestrian activity.

A.3. State and local level practices for identifying pedestrian risk factors

This section describes existing practices used by various state and local agencies to identify and quantify factors related to pedestrian risk. Like the pedestrian frequency/risk and exposure models identified in the literature, these analyses have typically been done at two different levels: (1) the geographic zone level (e.g., a census tract) or (2) an individual roadway segment or intersection. Based on the type of analysis, the remainder of this section reviews guidance for pedestrian safety analysis used by different cities and states across the U.S.

A.3.1 Crash summary-based analysis

New York

The New York PSAP (New York State, 2016) mostly focused on summary statistics and reviews of historical crash trends to identify pedestrian risk factors. The purpose of this plan was to recommend a distinct set of engineering, education, and enforcement countermeasures after identifying pedestrian safety conditions on both state and locally owned roads. It outlined a systemic safety program for uncontrolled marked pedestrian crosswalks on urban state roads. Risk factors associated with pedestrian crashes were identified by reviewing historic crash data and other critical information that was unavailable or only partially available in the crash data. The plan found pedestrian crashes:

- Predominantly occur in urban areas;
- Are overrepresented on state roadways;
- Occur due to pedestrians crossing the road where no crosswalks or signals are available; and,
- Reflect behavioral factors such as inattention, failure to yield, alcohol and pedestrian errors.

Furthermore, the plan expanded NYSDOT's existing Pedestrian Safety Corridor Program and developed a Pedestrian Safety Corridor Evaluation Guide.

Georgia

The Georgia PSAP (2018-2022) (Georgia Department of Transportation, 2018) adopted a data-driven approach to improve statewide pedestrian safety. Crash data were reviewed and analyzed to identify factors associated with pedestrian crashes, including demographic characteristics of individuals involved, roadway characteristics, individual behavior. The plan outlined focus counties, cities, and corridors, but these were based on historic crash frequencies and did not necessarily reflect risk. Methods used by the Georgia Department of Transportation to determine focus counties and cities include:

- Focus County Metrics:
    - One of the top ten counties with the highest number of pedestrian crashes.
    - One of the top ten counties with the highest number of pedestrian injuries.
    - One of the top ten counties with the highest number of pedestrian fatalities.
- Focus City Metrics:
    - Averaged at least one death per year.

- Was in the top ten cities with the highest number of pedestrian crashes.
- Was in the top ten cities with the highest number of pedestrian injuries.
- Was in the top ten cities with the highest number of pedestrian fatalities.

Maryland

The Maryland PSAP (Maryland Department of Transportation, 2023) adopted a geospatial data-driven approach to address pedestrian safety on state-owned highways and identified areas of need. Crash data were reviewed and plotted on a map to show the density of crashes, which were found to occur at greater frequency in denser urban areas. The Maryland Department of Transportation identified areas of need by plotting the following metrics and selecting areas with the greatest overlap:

- Public Comment Density
- Equity Index
- Non-Fatal Crash Density
- Fatal and Serious Injury Crash Density
- Short Trip Opportunity Areas

State-owned corridors within areas of need were then prioritized for improvement through non-specified data-driven methods and public input. The prioritization process included factors of crash density (crashes/mile), equity, connections to destinations and essential services, consideration of Highway Safety Improvement Program (HSIP) eligibility, and density of pedestrian activity. The PSAP does not directly estimate risk factors; rather, it relies on historical crash frequency and observed trends crash locations.

A.3.2 Zone-level risk-based analysis

Michigan

Hampshire et al. (2018) combined concepts of pedestrian exposure, risk, and zonal analysis to develop SPFs for non-motorized users in Michigan. To facilitate this analysis, the researchers developed Pedestrian Analysis Zones (PAZs) and employed a modified four-step travel demand model to develop an origin/destination model. This model used several factors that influence pedestrian travel demand, including the built environment and distance between origins and destinations. Based on this model, the researchers assigned risk scores for PAZs throughout the state. The risk score was determined using an NB regression model and estimated as a function of traffic volume (i.e., AADT) and pedestrian exposure; other features were not considered in identifying pedestrian risk. Pedestrian exposure was modeled using a planning-level model and defined as the number of daily walking trips that originated or terminated in a PAZ. These trips were estimated as a function of zonal-level metrics, such as:

- Population;
- Job density;
- Transit access;
- Block size; and,
- Urban living infrastructure.

Minnesota

The Minnesota Department of Transportation (MnDOT; Devoe, 2019) overlaid a hexagonal grid (0.5-mile diameter) across the state and assigned a score to each area based on a suite of criteria. The use of the grid potentially avoided updates to the state linear referencing system (LRS) from skewing results over time. MnDOT only included data that were 1) spatial, 2) comprehensively available across the state, and 3) as localized as possible (i.e., smaller zones). Potentially insightful datasets were excluded as they did not meet these criteria, or they showed inconsistency in reporting. No details were provided on how the scoring system was developed. However, the variables that contributed to the score included:

- Proximity to a bus stop;
- Urban area;
- Fraction of population less than 0.5 miles from a supermarket;
- Percent of population below 185% of federal poverty line;
- Presence of state bicycle trail;
- Population density;
- Percent of population within specific age categories (between 5–17 and 65+);
- Percent of population with a disability;
- Percent of Native America population (or within a Native American boundary);
- Percent of population that is foreign born;
- Employment growth;
- Percent of workers with zero vehicles; and,
- Contains a high-risk intersection as identified through the District Safety Plan.

Each of the inputs were associated with a number of points in the scoring system. Using these points, the Suitability of Pedestrian and Cyclist Environment (SPACE) score was computed as follows:

$$SPACE_{Score} = 100 \times \frac{1}{19} \times \sum points \quad (16)$$

Los Angeles, CA

The Southern California Association of Governments (SCAG) metropolitan planning organization (MPO) implemented work derived from the NCHRP 17-81 project and resulting NCHRP 1044 research report (Hamilton et al., 2023). These areawide safety models, which predict crashes for different travel modes and severities at the Tier II Traffic Analysis Zone (TAZ) level, support a methodology for incorporating safety into SCAG's scenario planning, particularly when engaging county transportation commissions and other local stakeholders. These areawide models are most effective at predicting changes in crashes for places where people live, work, and play, and before design details are known (e.g., lane widths, median types, shoulder presence and width). The variables included:

- TAZ area (sq. mi)
- VMT
- Total population
- Total employment
- Median household income
- Transit stop count

A.3.3 Road segment/intersection-level risk-based analysis

U.S. Road Assessment Program

The International Road Assessment Program (iRAP) coordinates RAP efforts occurring in Europe, Australia, and the U.S., and provides software (*ViDA*) for assessing the safety of a given section of road. ViDA has two primary applications: generating star ratings for 100-meter sections of roadway and developing safer roads investment plans for networks of roadway segments. Star ratings are assigned to a roadway based on design features, traffic control, and other characteristics that can be observed through visual inspection of roadway images or video. The star ratings consider factors related to both crash likelihood and crash protection. Crash likelihood is influenced by road characteristics such as number of lanes, street lighting, and intersection type. Fatality estimates include traffic volumes and allow for the performance of roadway segments to be compared. The equations for fatality estimates are similar in form to the SPFs used in the upcoming HSM2.

North Carolina

The research team identified criteria that NCDOT has used for screening roadways in urban areas as part of pedestrian safety studies being performed in North Carolina. Table 25 provides a summary of these screening features. Note that several of these criteria are limited by data availability, but they can also be used as a starting point for developing systemic risk factors.

Table 25. North Carolina pedestrian section scoring

CATEGORY	*SCORING MEASURE*
Crash features	Severity Index of Ped crashes
	Frequency of Ped crashes
	Density of Ped crashes (cr/mile)
Infrastructure features	Speed Limit
	Crossing Length (w/ adjustment for adequate median refuge)
	Vehicle AADT
	Signal spacing
	Sidewalk Coverage
Pedestrian activity surrogates	Bus stop density
	Schools
	Shopping Centers
	Alcohol Establishments
	Population Density
	% Households with 1 or no vehicles

Washington

The Washington State Department of Transportation (WSDOT) developed a systemic analysis approach for prioritizing pedestrian crossing improvements as part of the State's Active Transportation Plan (WSDOT, 2020). This approach focused on the Level of Traffic Stress (LTS), accessibility, and network connectivity as priorities for investment. LTS helps determine the suitability of roads and connections to accommodate bicyclists and pedestrians. For low-stress pedestrian routes, WSDOT also used an evaluation framework involving criteria such as safety, equity, and demand to assess the potential need of pedestrian crossings. Connectivity, network permeability, and route directness are also considerations when assessing potential pedestrian improvements.

Ohio

The Ohio Department of Transportation (ODOT) piloted a traditional full systemic approach to identify priority locations for pedestrian safety countermeasures in District 8 in southwest Ohio (ODOT, 2020). This approach used focus crash and facility types to develop risk factors, prioritize network locations, and formulate a framework to apply relevant countermeasures. Risk factors include:

- Presence of lighting;
- Proximity to school or university;
- Presence of bus stop;
- Traffic volume;
- Percent of population with specific age categories;
- Posted speed limit;
- Number of lanes;
- Presence of pedestrian infrastructure;
- Percent of households with zero vehicles; and,
- Precent of non-motorized commuters.

City of Seattle

In 2016, the Seattle Department of Transportation (SDOT) conducted the first phase of the citywide Bicycle and Pedestrian Safety Analysis (Seattle Department of Transportation, 2016) focused on signalized intersections. Exploratory analysis included pedestrian and bicycle crash data and a wide range of roadway, land use, and environmental data. Additionally, multivariate statistical analysis was performed to better understand the importance of exposure estimates. This also led the city to develop a count optimization effort to fill specific gaps in pedestrian and bicyclist exposure knowledge. In 2020, the second phase of the plan (Seattle Department of Transportation, 2020) included significant advances by adding signal phasing data to the analysis. It also refined the exposure models developed in phase 1 by considering motor vehicle volumes along with pedestrian and bicycle volume data. The variables used for the refined exposure models included factors such as road speed limit, population, and number of households.

City of Pittsburgh

Pittsburgh's Department of Mobility and Infrastructure has developed a comprehensive plan (City of Pittsburgh, 2020) to better understand the causes and consequences of pedestrian crashes and improve pedestrian safety in Pittsburgh. Along with the historic crash data, this risk-based analysis included several factors, such as neighborhood connectivity, access to transit, lack of pedestrian infrastructure, and equity concerns (based on cost of living, age, race, ethnicity, access to a vehicle, and other individual and household characteristics reported in Census data). Moreover, the plan recommended four methods to identify and prioritize high-risk locations that would undergo Road Safety Audits (RSAs). The methods are:

- Hot Spot Analysis;

- High-Risk Corridors: Locations that may be more likely to have crashes in the future, based on a combination of physical and demographic traits;
- Network-Need Corridors: High-volume, high-speed streets that may lack sufficient infrastructure for pedestrians to navigate safely; and,
- Business Districts with High Frequency Transit: Streets within local business districts that have high-frequency transit access and provide an important contribution to the local economy.

Virginia

The PSAP (Virginia Department of Transportation, 2018) contains two complementary examples of systemic pedestrian screening: a historic crash frequency ("crash cluster") review and a more proactive corridor identification that relies less on historic crash frequency. Crash clusters are identified through a crash density analysis using geographic information systems (GIS), while the corridor analysis flags road segments that meet certain criteria (e.g., number of lanes, median presence, zero-vehicle household proportion, employment density, etc.). Crashes comprise less than 10 percent of the screening criteria value for these corridors. VDOT also used public health metrics, such as the Virginia Department of Health's Health Opportunity Index (HOI), to assess potential risk. The HOI provides Census tract-level indicators of public health outcomes, and VDOT found these to be highly correlated with pedestrian crashes. VDOT shares the results and locations with district and local staff to solicit projects and implement low-cost countermeasures. bit.ly/VDOTPBSAP

Tennessee

Based on VDOT's work, as well as federal and NCHRP guidance, TDOT developed separate indices for pedestrian safety at intersections and segments. Characteristics for segments include measures of equity, pedestrian demand (measured using a weighted composite of population density, employment density, mode split, lane use, and points of interest), historical crash frequency and severity, traffic volume, number of travel lanes, posted speed limit, pedestrian access, pedestrian protection, and a measure of crossing risk.

Massachusetts

As part of the USDOT's Safety Data Initiative, the Massachusetts Department of Transportation (MassDOT) developed a series of systemic analyses based on the state's SHSP emphasis areas. For pedestrian safety risk, MassDOT considered the following characteristics at the segment-level:

- Presence of a median;
- 3+ travel lanes in both directions of travel;

- Transit stop presence on a road segment, rail and/or bus;
- AADT;
- Median household income;
- Population density;
- Employment density;
- Ratio of employment in the accommodation, food services, or retail trades;
- Transit stop density;
- Two or more MassGIS Environmental Justice flags;
- Commuters that walk, bicycle, or take transit

MassDOT publishes priority locations based on systemic safety risk scores using the state's IMPACT tool (https://apps.impact.dot.state.ma.us/sat/landing).

Washington D.C.

The District Department of Transportation (DDOT) conducted numerous systemic data analyses to identify pedestrian access and safety needs and priorities based on the following attributes:

- Vision zero high-crash corridor;
- Walksheds to bike share and transit stations;
- Transit ridership by quarter-mile grid cells;
- Sidewalk gaps;
- Pedestrian Friendliness Index by Census block group that seeks to capture how comfortable walking is in a certain area. This is measured based on:
    - Street connectivity;
    - Sidewalk presence
    - Buildings set close to the street; and,
    - Intersections and street blocks that are easily navigable by pedestrians.
- Jobs within a 20-minute walk during the am peak period.

However, this approach has limited applicability to rural areas or agencies without comprehensive statewide data. Additional details are available at https://movedc-dcgis.hub.arcgis.com/pages/mobility-priority-networks and https://movedc-dcgis.hub.arcgis.com/pages/mapping-transportation-needs. The methods include a Pedestrian Friendliness Index that seeks to capture how comfortable walking is in a given area.

Arizona

The 2017 PSAP by the Arizona Department of Transportation (ADOT; Arizona Department of Transportation, 2017) developed a systemic process using available roadway, population, and land use data to identify locations with high pedestrian crashes and high-risk characteristics. ADOT also used crash trees, the Pedestrian and Bicycle Crash Analysis Tool (PBCAT), and crash frequency as part of its preliminary risk factor analysis; this analysis also noted that more than half of pedestrian crashes on the state highway system between 2011 and 2015 occurred where a sidewalk or crosswalk was present. The risk mapping procedure included GIS-based analysis for initial screening and visual review (e.g., Google Earth) for final screening. Furthermore, it developed an economic analysis approach that combined high-risk sites with high-crash sites proposed for the same treatment.

Michigan

Multiple studies have been conducted in Michigan to estimate pedestrian crashes. The Transportation Research Center for Livable Communities conducted research to develop SPFs for predicting pedestrian crashes along road segments and at intersections in Michigan (Gates et al., 2016). The models were developed for different segment types and levels of severity and were based solely on AADT data. Dolatsara (2014) developed SPFs using NB regression to estimate pedestrian crashes at urban signalized intersections in Michigan using pedestrian exposure data. In addition to the exposure, the model considered:

- Motor vehicle AADT;
- Number of left-turn lanes;
- Presence of on-street parking;
- Presence of speed signs; and,
- Presence of a bus stop within 0.1 mi of the intersection.

Oh et al. (2013) conducted a study for the Michigan Department of Transportation (MDOT) to develop a systematic approach to determine performance measures for non-motorized safety and to identify the need for countermeasures when designing facilities. NB and Poisson regression were used to estimate the SPFs. Data used in their models included pedestrian and bicycle volumes, non-motorized facility inventory, non-motorized improvement projects, activity locations, socioeconomic and demographic characteristics, crime rates, land use data, and traffic volume data. Different models were developed at the city level, census tract level, and corridor level. McArthur et al.(2014) investigated pedestrian crashed involving children aged 5 to 14 located within one mile of schools serving students from kindergarten through eighth grade in Michigan. In addition to crash data, demographic and socioeconomic characteristics were obtained from the U.S. Census Bureau, including:

- Child population;
- K-8 enrollment;
- If the school was located on a local roadway;
- Average family size;
- Population density;
- Median family income;
- Average number of parents per household; and,
- Portion of non-white households.

A random effects NB model was developed using these data elements.

Oregon

The Oregon Department of Transportation applied a risk-based network screening approach to prioritize corridors with the greatest potential for reducing pedestrian crashes (Bergh et al., 2015). Risk factors were identified first and then used to prioritize locations. Pedestrian volumes were not considered in the risk-based method due to the lack of consistent statewide data. There were no quantitative data associating the identified risk factors with crash frequency, so a subjective scoring system was developed to account for combinations of risk factors. Monsere et al. (n.d.) continued the work in Oregon to improve methods for identifying and prioritizing locations with increased or elevated pedestrian crash risk, with the objective of developing a risk-scoring method using weights derived from data analysis rather than best judgment or a subjective scoring system. Geometric, land use, volume, and crash data were collected from multiple sources. Logistic regression models were developed for both crash occurrence (crash or not) and crash severity.

Florida

TransPed is an interactive GIS-based tool designed to assist in the planning and analysis of pedestrian transportation (FDOT, 2017A). The tool includes a breadth of traditional transportation data, such as existing infrastructure, available routes, traffic counts, forecasts, and crashes, as well as information about land use and socioeconomic characteristics pertinent to travel by alternative modes. The data are combined into a Composite Ped Suitability index that shows the spectrum of opportunity for active transportation and a Ped Quality of Service grade that can be used to prioritize infrastructure improvements through spatial- or attribute-driven analyses.

City of Philadelphia

In 2021, the City of Philadelphia's PSAP (City of Philadelphia, 2021) prioritized intersections and corridors for pedestrian safety improvements as part of the City's Vision Zero initiative. In addition to identifying areas with the highest crash density for prioritization, the PSAP identified statistically significant pedestrian risk factors through non-specified statistical methods. Unique risk factors were identified for injury crashes and fatality crashes. Risk factors for injury crashes included:

- Intersections;
- Transit presence;
- Midblock locations;
- Turning movements;
- Urban arterial and auto-oriented commercial/industrial corridors;
- Population proportion under 19 years old;
- Population proportion over 50 years old;
- Night/evening; and
- Schools,
- While fatality crashes were associated with the following risk factors:
- Night/evening;
- Urban arterial and auto-oriented commercial/industrial corridors;
- Midblock locations;
- Intersections; and
- Population proportion over 50 years old.

<u>Texas</u>

The Texas Department of Transportation (TxDOT, 2023) used a dual-methodology systemic approach to prioritize corridors for improvement. The first approach was a systemic crash analysis that examined pedestrian crash data to identify risk factors by comparing roadway, traffic, and contextual attributes of crash sites where pedestrian crashes occurred. If an attribute accounted for a greater proportion of crashes than centerline miles, the attribute was deemed a possible risk factor. The following attributes of crash sites were identified as pedestrian crash risk factors:

- Average daily traffic;
- Truck percentage;
- Area type;
- Bus pad width and offset;
- TWLTL presence;

- Crosswalk presence;
- Curb presence;
- Curb cut presence and offset;
- Functional class;
- Highway division;
- Shoulder type, width, and use;
- Lane width;
- Median barrier presence;
- Median width;
- Median type;
- Minimum right-of-way;
- Number of lanes;
- Roadbed width;
- Sidewalk presence and condition; and
- Surface width.

The second approach was a targeted analysis which identified hotspot segments based on crash density. The greatest crash densities were found in densely populated urban areas.

A.3.4 Summary of state practices

The review of existing practices by state transportation agencies reveals general consistency in the types of features being used to quantify pedestrian crash risk. Roadway and traffic features that are common across these existing approaches include:

- Number of travel lanes;
- Posted speed limit;
- Presence of pedestrian infrastructure elements, such as marked crosswalks and sidewalks;
- Vehicular travel volumes; and,
- Measures or surrogates of pedestrian exposure.

Most states are not able to directly account for pedestrian exposure; however, surrogate variables are used to either estimate exposure for direct use in models or to account for exposure indirectly. Features that serve as surrogate variables for exposure include:

- Nearby population and/or population density;
- Employment and/or employment density (by sector);

- Modal split of nearby population;
- Land use; and,
- Other pedestrian attractors (e.g., presence of or proximity to a school or transit stop).

Most state agencies do not provide detailed information on the specific methods used to identify individual risk factors or the relative assigned weights. In most cases, the weights are round numbers (e.g., each factor is associated with some whole number of points that are then added together). This suggests the use of expert judgment in either directly determining these weights or modifying weights provided by a model. Those that did provide details generally used statistical techniques suggested by the literature, such as NB regression.

A.4. Data sources for pedestrian safety and pedestrian exposure models

A wide range of variables contribute to pedestrian crash risk and exposure modeling. Data on these variables come from different sources and they can be combined based on users' needs. This section summarizes the reviewed data sources for both collision records and contributing factors. Some of the data sources are maintained by national- and state-level agencies, while others are owned by private sector organizations.

A.4.1 Data collected by public agencies

Roadway data

A critical factor observed in both pedestrian crash frequency/risk and pedestrian exposure models is roadway features. These data include roadway functional classification, geometric characteristics (e.g., cross-sectional information and horizontal curvature), vehicular traffic volumes (typically provided in AADT), and the presence of pedestrian infrastructure features. Many state and local agencies maintain detailed roadway inventory databases that provide this information. Crash information is generally linked to specific roadway locations through a referencing system or spatial proximity, allowing these features to be associated with specific crash observations. Within North Carolina, the following resources are available:

- NCDOT's Road Characteristics file is a spatial representation of roadway and traffic data on all public roads in the State (where available). These data are stored using an LRS-enabled centerline, and this allows data to be locatable by physical location, as well as route and milepost information. The publicly released dataset is also dynamically segmented according to each attribute on the network (https://connect.ncdot.gov/resources/gis/pages/gis-data-layers.aspx).

- NCDOT's Pedestrian and Bicycle Infrastructure Network (PBIN) (https://connect.ncdot.gov/projects/BikePed/pages/pbin.aspx) provides existing and proposed bicycle and pedestrian facilities in North Carolina. The PBIN data is not comprehensive, however, and updates to the geodatabase are made continuously. GIS software is required to download, analyze and manipulate the data.
- NCDOT publishes a spatial file of all intersections statewide, based on a 2022 roadway centerline network (https://www.arcgis.com/home/item.html?id=a2c61baa9c8f4d2eb68ca65eb1b47bd9). This includes high-level information on intersection geometry, number of approaches, and traffic control (signalized and unsignalized).

Crash data

Reported crash information is also necessary to develop models to predict pedestrian crash frequency and crash risk. Most state and local agencies maintain detailed crash databases that store summarized or full versions of law enforcement officer crash reports (e.g., records of crashes that occur on state roads, including location, vehicles, people involved, and available injury outcomes). These reports typically include flags that identify crashes involving pedestrians.

Several national databases also have detailed crash information that can be used to identify national trends related to pedestrian safety performance. Examples include:

- The Collaborative Sciences Center for Road Safety (CSCRS) National Pedestrian and Bicycle Safety Data Clearinghouse (NPBSD) (https://pedbikedata.org/) is an online search tool that contains specific pedestrian safety data from a variety of agencies, usually police-reported collisions with motor vehicles. Data on roadway information such as speed limit, signs, street lights etc. can be downloaded from this source.
- NHTSA's Fatality Analysis Reporting System (FARS) (https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars) is a public database of annual fatal motor vehicle crashes. It provides roadway factors such as roadway function class, intersection, intersection leg, etc.

Within North Carolina, the following resources are available that provide this information:

- NCDOT Bicyclist and Pedestrian Crash Map (https://www.arcgis.com/home/item.html?id=b4fcdc266d054a1ca075b60715f88aef) is a web-based platform which contains spatial distribution of pedestrian crashes.
- The roadway information specific to each crash includes number of lanes, road characteristics/class/condition/configuration, road defects etc. These data can also be downloaded for detailed analysis and manipulation.

Pedestrian exposure data

While most agencies have well-developed vehicle counting programs for collecting or estimating vehicle AADT information, pedestrian traffic volumes are collected in a less systematic manner and at a much smaller scale. In some cases, pedestrian volumes are collected opportunistically when other counts are needed (e.g., only when volumes are required for a specific project). However, pedestrian volumes or exposure are a critical part of estimating pedestrian crash frequency or risk. The lack of broad pedestrian count coverage is one of the primary reasons pedestrian exposure models are needed.

Pedestrian volumes can be collected in a variety of ways:

- Manual counts: Pedestrian volumes can be collected using manual counts, but this is typically a resource-intensive process. As a result, manual counts are often collected for very short periods (e.g., 2 hours).
- Passive infrared sensors: This technology measures changes in ambient temperature compared to background radiation (heat) as the user moves through the detection zone. Counts both pedestrian and bicycle but cannot differentiate between these two types. A validation study was conducted on accuracy of this technology by MnDOT (2015).
- Slab sensors: These are sensors embedded in a sidewalk or pavement that use acoustic sensors to capture pedestrian movement.
- Camera technology: Cameras can be used to capture pedestrian movements at a particular location. These videos can be processed manually (i.e., by humans) or using video-processing information to count pedestrians.

The Institute for Transportation Research and Education (ITRE) at North Carolina State University manages the North Carolina Non-Motorized Volume Data Program (NMVDP) for NCDOT (https://itre.ncsu.edu/focus/bike-ped/nc-nmvdp/). This program continuously collects bicycle and pedestrian counts at fixed locations throughout the state. Although this provides a wealth of historic data, most counts are limited to greenways and other non-motorized user-specific facilities and may have limited applicability to road safety analysis. In North Carolina, road-based pedestrian counts are often collected in conjunction with vehicular volumes for individual project development.

Socioeconomic data

The previous sections indicate that socioeconomic characteristics can be valuable predictors of pedestrian crash risk and exposure. These data are generally available for several sources:

- The U.S. Census Bureau has databases with extensive demographic data that can be used in pedestrian safety analysis. The available tabulations include population size by sex, age, race, Hispanic origin, education status, employment status, occupation and industry, income, rent, and housing unit value. These tabulations are presented at many levels of observation, including regions, states, counties, metropolitan areas, places, county subdivisions, census tracts/block numbering areas, block groups, and blocks.
- The American Community Survey (ACS) is a demographic survey program conducted by the U.S. Census Bureau. The ACS releases social characteristics, economic characteristics, housing characteristic and demographic and housing estimates data annually.
- The Longitudinal Employer-Household Dynamics (LEHD) program is part of the Center for Economic Studies at the U.S. Census Bureau. LEHD Origin-Destination Employment Statistics (LODES), Job-to-Job Flows (J2J), and Post-Secondary Employment Outcomes (PSEO) are available online for public use.
- The North Carolina Department of Transportation (NCDOT) has developed tools such as the Transportation Disadvantage Index (TDI) to access transportation equity issues. The TDI was calculated as a composite score based on seven indicators (socioeconomic characteristics, automobile ownership, population with mobile impairment, etc.) at the census block group level for identifying the specific barriers and challenges different communities face. These data help inform pedestrian and bicycle access as part of the Access in Appalachia Pilot Project (NCDOT, 2024b; https://www.arc.gov/report/access-in-appalachia-north-carolina-dot-pilot-study/).
- The North Carolina Department of Health and Human Services (NC DHHS) publishes a suite of maps that profile socioeconomic and assess characteristics that have an effect on public health outcomes. The Social Determinants of Health (SDOH) provide Census tract-level characteristics that are largely collected from the U.S. Census Bureau, although additional datasets, such as presence of food deserts and rental cost burdens, are also available. The SDOH also compares individual tracts to state averages using z-scores:

$$Z - Score = \frac{Tract\ Value - State\ Value}{State\ Standard\ Deviation} \quad (17)$$

Land use data

The built environment also serves as a useful predictor of pedestrian exposure and (potentially) crash risk. Land use data can be obtained from local agencies. For example, land use data in Wier et al. (2009) were obtained from the San Francisco Planning Department. Data on the land use and demographic characteristics in Torbic et al. (2010) were assembled through analysis of planning data available in GIS format. Individual land use plans in the Twin Cities Metropolitan Area were used in several studies of pedestrian exposure in this region. Other examples in the literature include:

- Parcel boundaries are available through NC OneMap, a public repository geospatial information in North Carolina. Although land use information is incomplete, parcel size and density can be an indicator of local land use intensity and activity.
- The United States Geological Survey (USGS) publishes the National Land Cover Database (NLCD) through the Multi-Resolution Land Characteristics Consortium (MRLC). This nationwide raster dataset provides spatial land cover/use classification.
- The Federal Emergency Management Agency (FEMA) publishes a spatial inventory of building footprints (over 450 square feet). The data include a land use and occupancy classification, but this is estimated and may not necessarily be accurate for all locations in North Carolina.
- Torbic et al. (2010) collected intersection pedestrian and traffic count, vehicle-pedestrian collision data, intersection characteristics, such as number of intersection legs and posted speed limit, land use and demographic characteristics from The Charlotte Department of Transportation (CDOT) and they examined aerial photography and signal plans provided by CDOT to get intersection characteristics. This dataset is not publicly available.
- Cottrill and Thakuriah (2010) collected vehicle-pedestrian crash data and the environmental indicators at census tract-level, such as transit availability index, pedestrian accessibility index and crime rate, and behavioral indicators, such as median household income, percent with no cars and percent who speak English, from the Illinois Department of Transportation (IDOT). This dataset is not publicly available.
- Dumbaugh et al. (2013) examined how built environment impact pedestrian crash. They collected street network information from San Antonio–Bexar County Metropolitan Planning Organization, information on traffic volumes from City of San Antonio and TxDOT.

A.4.2 Non-traditional and alternative data

Smartphone-technique-based exposure data

Advancements in technologies and the proliferation of smartphones serve as an alternative for addressing challenges due to limited exposure data sources. This section lists examples of pedestrian exposure data derived from smartphone-related techniques or applications. A more comprehensive review can be found in Lee and Sener (2020).

- Cellular carriers collect time-stamped location data based on signaling between mobile phones and cell towers. From the positioning data, movements of users can be extracted. Secondary data vendors (e.g., Airsage) purchase raw data from telecom carriers and resell specific data after the data extraction. Released data types are origin-destination (OD) pairs (zone- and link-based), traffic speed and volume, but they are not yet customized for non-motorized modes.
- Many apps featuring location-based services (LBS; e.g., Yelp) lead data evolution in various fields. StreetLight Data and Cuebiq are two of the representative companies that provide aggregated multi-app LBS data sets. Cuebiq's database comes from hundreds of LBS apps, and StreetLight Data partnered with Cuebiq to integrate multi-app LBS data with other data. An easy-to-use-online platform, Bike Ped Essentials, was recently released by StreetLight. The on-demand analytic service provides a wide range of data types that can be used in safety analysis such as traffic attributes (e.g., volume, distance, time, and speed), geometry (e.g., zone, link, or city), and inferred context information (e.g., sociodemographic and trip purpose).
- Fitness-tracking apps use diverse built-in sensors to track users' physical activities, and some apps preprocess and commercialize the collected data. For example, Strava sells a license to allow access to walking datasets for research purposes and transportation planning. Strava's data service, Strava Metro, provides three licenses that can be purchased based on data aggregation units: node (point), street (segment), and OD (polygon). Strava publishes heat maps of user activity based on aggregated, public activities over the last year (https://www.strava.com/heatmap#8.84/-79.62832/35.92841/hot/all). This map is updated monthly; however, it is not an objective representation of all activity in any given location (i.e., there may be locations with trips that do not meet the threshold of activity to be visualized on the map).
- Volunteered geographic information (VGI) platforms enable community members to report localized knowledge and experiences. For example, OpenStreetMap (OSM) is one of the most

popular VGI platforms. Over one million individuals have contributed to a set of geographic data that include roads, cycle paths, and trails used.

Probe speed data

Probe data are generated by monitoring the position of individual vehicles (i.e., probes) over space and time. The individual probe data can be converted to performance measures such as speed and travel time, which are two of the commonly used contributing factors in pedestrian safety analysis. This section lists examples of probe vehicle data from different vendor companies:

- Probe sources of speed data can be accessed through the Regional Integrated Transportation Information System (RITIS), which is housed and managed by the University of Maryland's Center for Advanced Transportation Technology (CATT) Laboratory. Speed data can be accessed and downloaded from the RITIS platform using RITIS' Massive Data Downloader of archived data.
- HERE collects vehicle speeds using multiple real-time sources, including global positioning systems, probe vehicles, and cell phones.
- Inrix combines probe data from commercial GPS, DOT sensors, and other proprietary data sources and provides speed and travel time on Traffic Message Channels, which are defined road segments.
- Similarly to Inrix, TomTom provides speed and travel time on road segments; however, its data are collected from the Vodafone mobile phone network, governments and traffic control centers.
- NAVTEQ includes both point- and route-based data, and in addition to speed and travel time, volume data from its own sensors are also available.
- AirSage utilizes wireless signaling data and cell phone GPS, and travel mode information is available in this data source.
- TrafficCast leverages GPS tracking data, public sensors, accidents reports, roadwork data, and weather reports. Speed and travel time are available at the Traffic Message Channel level.

Appendix B. Additional pedestrian count locations

Table 26. Additional Pedestrian Count Locations

No.	County	City	Description
1	Aberdeen	Moore	S Sandhills Blvd @ W South St
2	Aberdeen	Moore	N Sandhills Blvd @ US 501
3	Asheboro	Randolph	US 64 Bus @ NC 42
4	Asheboro	Randolph	S Fayetteville St @ Worth St and Trade St
5	Asheboro	Randolph	Academy St @ S Fayetteville St
6	Bertie	Windsor	Water St @ S King St
7	Cabarrus	Concord	Harris Rd @ Skybrook Dr and Ellenwood Rd
8	Caldwell	Lenoir	Wright St @ Vaiden St
9	Caldwell	Lenoir	Virginia St @ Harper Ave
10	Catawba	Hickory	16th Ave NW @ NC 127
11	Catawba	Hickory	7th Ave SW @ 5th St SW
12	Catawba	Hickory	1st St SW @ 4th Ave SW
13	Catawba	Hickory	1st St SW @ 3rd Ave SW
14	Catawba	Hickory	1st St SW @ 5th Ave SW
15	Chowan	Edenton	Board St @ N Oakum St and Paradise Rd
16	Chowan	Edenton	Queen St @ S Broad St
17	Clayton	Johnston	US 70 Bus @ S Robertson St and Amelia Church Rd
18	Clayton	Johnston	E Main St @ Lombard St
19	Cleveland	Shelby	Charles Rd @ Wyanoke Ave and Dellinger Rd
20	Cleveland	Shelby	W Sumter St @ Cora St
21	Cleveland	Shelby	Buffalo St @ Jefferson St
22	Cleveland	Shelby	N Morgan St @ Wells St
23	Davidson	Lexington	Holly Grove Rd @ Raleigh Rd and Martin Luther King Junior Blvd
24	Davidson	Lexington	E Center St @ Cherry St
25	Davidson	Lexington	N Pugh St @ E 3rd St
26	Durham	Durham	Broad St @ Green St
27	Durham	Durham	Erwin Rd @ Douglas St and Research Dr
28	Durham	Durham	NC 55 @ Cecil St and S Alston Ave
29	Durham	Durham	NC 55 @ NC 54
30	Edgecombe	Tarboro	Western Blvd @ W Howard Ave
31	Fayetteville	Cumberland	Hay St @ Robeson St and Bragg Blvd
32	Fayetteville	Cumberland	Gillespie St @ Blount St and Cambell Ave
33	Fayetteville	Cumberland	US 401 @ Purdue Dr
34	Fayetteville	Cumberland	US 401 @ Glensford Dr and Hope Mills Rd
35	Fayetteville	Cumberland	US 401 Bus @ Hillsboro St
36	Forsyth	Clemmons	Clemmons Rd @ Lewisville Clemmons Rd and Middlebrook Dr
37	Forsyth	Walkertown	Reidsville Rd @ Old Hollow Rd

No.	County	City	Description
38	Forsyth	Winston Salem	N Trade St @ W 8th St and N Martin Luther King Jr Dr
39	Forsyth	Winston Salem	University Pkwy @ Coliseum Dr and W 27th St
40	Forsyth	Winston Salem	Peters Creek Pkwy @ W Academy St
41	Forsyth	Winston Salem	Stratford Rd @ Oakwood Dr
42	Forsyth	Winston Salem	Stratford Rd @ Westbrook Plaza Dr
43	Forsyth	Winston Salem	Reidsville Rd @ Rickard Dr
44	Forsyth	Winston Salem	Peters Creek Pkwy @ Brewer Rd
45	Forsyth	Winston Salem	Williamson St @ Bruce St
46	Forsyth	Winston Salem	W Academy St @ Marshall St
47	Forsyth	Winston Salem	Trade St @ 7th St
48	Forsyth	Winston Salem	S Broad St @ W St
49	Forsyth	Winston Salem	Fayetteville St @ Ehle Dr
50	Forsyth	Winston Salem	Adler St @ Powell St
51	Forsyth	Winston Salem	S Martin Luther King Jr Dr @ Junia Ave
52	Gaston	Gastonia	S Marrieta St @ Dixon Cir
53	Gaston	Gastonia	W Bradley Ave @ N Logan St
54	Gaston	Gastonia	N York St @ W Granite Ave
55	Guilford	Greensboro	Martin Luther King Jr Dr @ E McCulloch St and Julian St
56	Guilford	Greensboro	WB Randleman Rd @ South St
57	Guilford	High Point	Eastchester Dr @ Sutton Way
58	Henderson	Hendersonville	King St @ 1st Ave E
59	Henderson	Hendersonville	Spartanburg Hwy @ Main St/Greenville Hwy
60	Henderson	Hendersonville	Asheville Hwy @ Clairmont Dr
61	Henderson	Hendersonville	US 64 @ Blythe St
62	Hertford	Murfreesboro	US 258 @ US 158
63	Kill Devil Hills	Dare	US 158 @ W Carolista Dr and E Hollowell St
64	Kill Devil Hills	Dare	US 158 @ W Martin St
65	Kill Devil Hills	Dare	US 158 @ 3rd St
66	Kitty Hawk	Dare	US 158 @ W Eckner St
67	Lee	Sanford	Oddfellow St @ Price St
68	Lee	Sanford	S Vance St @ Church St
69	Lee	Sanford	S Steele St @ E Pearl St
70	Lenoir	Kinston	Dr. Martin Luther Kin Jr Blvd @ E Vernon Ave
71	Mecklenburg	Charlotte	Providence Rd @ Ardrey Kell Rd
72	Mecklenburg	Charlotte	Rea Rd @ Ardrey Kell Rd
73	Mecklenburg	Charlotte	Mt Holly Huntersville Rd @ Mount Holly Rd
74	Moore	Southern Pines	E Illinois Ave @ S Ashe St
75	Moore	Southern Pines	E Massachusetts Ave @ S Ashe St
76	Moore	Southern Pines	W Rhode Island Ave @ N Bennett St
77	Moore	Southern Pines	W New Hampshire @ N Bennett St
78	Nags Head	Dare	US 158 @ Seachase Dr
79	Nash	Rocky Mount	Nash St @ S Pearl St
80	Nash	Rocky Mount	Western Ave @ S Howell St

No.	County	City	Description
81	Nash	Rocky Mount	Western Ave @ S Pearl St
82	Nash	Rocky Mount	Braswell St @ Avent St
83	New Hanover	Wilmington	Eastwood Rd @ Cardinal Dr
84	New Hanover	Wilmington	Eastwood Rd @ Wrightsville Ave and Canal St
85	New Hanover	Wilmington	College Rd @ Wilshire Blvd
86	New Hanover	Wilmington	Market St @ N Kerr Ave
87	New Hanover	Wilmington	Dawson St @ S 16th St
88	New Hanover	Wilmington	S 17th St @ Glen Meade Rd
89	New Hanover	Wilmington	College Rd @ S 17th St and Waltmoor Rd
90	New Hanover	Wilmington	S 5th Ave @ Dock St
91	New Hanover	Wilmington	Oriole Dr @ College Acres Dr
92	Onslow	Jacksonville	College St @ Warlick St
93	Onslow	Jacksonville	Marine Blvd @ Henderson Dr
94	Onslow	Jacksonville	Western Blvd @ Country Club Rd
95	Orange	Carrboro	Greensboro St @ Main St
96	Orange	Chapel Hill	Columbia St @ Franklin St
97	Orange	Chapel Hill	Estes Dr @ E Franklin St
98	Pasquotank	Elizabeth City	E Ehringhaus St @ S Road St
99	Pasquotank	Elizabeth City	N Road St @ Burgess St
100	Pasquotank	Elizabeth City	MLK Jr Dr @ E Main St
101	Pasquotank	Elizabeth City	Herrington Rd @ White St
102	Pasquotank	Elizabeth City	Park St @ Harris Dr
103	Pasquotank	Elizabeth City	E Cypress St @ 2nd St
104	Pitt	Greenville	Memorial Dr @ W 5th St
105	Pitt	Greenville	Stantonsburg Rd @ W Arlington Blvd
106	Pitt	Greenville	Charles Blvd and E Fire Tower Rd
107	Pitt	Greenville	Evans St @ W Arlington Blvd
108	Pitt	Greenville	Arthur St @ Hopkins Dr
109	Pitt	Greenville	S Greene St @ W 5th St
110	Pitt	Greenville	W 14th Ave @ Dickinson Ave
111	Pitt	Greenville	Charles Blvd @ E 14th St
112	Plymouth	Washington	US 64 @ Washington St
113	Richmond	Rockingham	Wall St @ Ann St
114	Richmond	Rockingham	Ann St @ Foushee St
115	Richmond	Rockingham	E Greene St @ N Lawrence St
116	Rockingham	Richmond	US 74 Bus @ US 220
117	Rowan	Salisbury	W Horah St @ Partee St
118	Rowan	Salisbury	W Horah St @ Messner St
119	Rowan	Salisbury	Standish St @ Forney St
120	Rowan	Salisbury	S West St @ W Fisher St
121	Sanford	Lee	Main St @ Lee Ave
122	Sanford	Lee	Hawkins Ave @ Burns Dr
123	Sanford	Lee	Lee Ave @ Wilson Rd

No.	County	City	Description
124	Sanford	Lee	Carthage St @ S Horner Blvd
125	Sanford	Lee	Carthage St @ S Steele St
126	Southern Pines	Moore	US 501 @ Murray Hill Rd
127	Southern Pines	Moore	W Pennsylvania Ave @ N Bennett St
128	Surry	Mt Airy	Spring St @ Pender St
129	Surry	Mt Airy	Marshall St @ Roberts Rd
130	Surry	Mt Airy	S South St @ Durham St
131	Vance	Henderson	Garnett St @ Parham Rd
132	Wake	Morrisville	Davis Dr @ Parkside Valley Dr
133	Wake	Raleigh	Atlantic Ave @ E Millbrook Rd
134	Wake	Raleigh	Wade Ave @ Dixie Trl
135	Wake	Raleigh	Glenwood Ave @ W Johnson St
136	Wake	Raleigh	Hillsborough St @ Boylan Ave
137	Wake	Raleigh	Hillsborough St @ Dixie Trail and Friendly Dr
138	Wake	Raleigh	Hillsborough St @ Horne St
139	Wake	Raleigh	Western Blvd @ Whitmore Dr and Clanton St
140	Wake	Raleigh	Martin Luther King Jr Blvd @ Holmes St and Chavis Way
141	Wake	Raleigh	New Bern Ave @ Raleigh Blvd
142	Wake	Raleigh	New Bern Ave @ Tarboro Rd
143	Watauga	Boone	Blowing Rock Rd @ Rivers St
144	Watauga	Boone	Holmes Dr @ Hill St
145	Watauga	Boone	Holmes Dr @ Faculty St
146	Watauga	Boone	Faculty St @ Highland Ave
147	Watauga	Boone	Hilandale Dr. @ Edora St
148	Wayne	Goldsboro	N George St @ Murray St
149	Wayne	Goldsboro	N John St @ Swan St
150	Wayne	Goldsboro	E Holly St @ N Leslie St
151	Wilkes	North Wilkesboro	Sparta Rd @ Elkin Hwy
152	Williamston	Martin	US 17 @ Park St
153	Williamston	Martin	US 17 Bus @ West Blvd and East Blvd
154	Wilson	Wilson	Nash St N @ Jackson St
155	Wilson	Wilson	Hines St @ Tarboro St
156	Wilson	Wilson	Tarboro St W @ Barnes St
157	Wilson	Wilson	Hines St @ Pender St SE
158	Wilson	Wilson	Ward Blvd @ Tarboro St
159	Wilson	Wilson	Atlantic Christian College Dr @ Vance St
160	Wilson	Wilson	Nash St @ Goldsboro St
161	Wilson	Wilson	Broad St @ Jackson St
162	Wilson	Wilson	Broad St @ Moss St
163	Wilson	Wilson	Ward Blvd @ Black Creek Rd

Appendix C. Exposure modeling database dictionary

Table 27. Exposure Modeling Database Dictionary

Variable name	Type	Example values	Description
KeyIntersectionID	Text	TSUINT716991, TSUINT43242	Assigned unique identifier – acts as a unique tracking ID for the intersection feature.
IntersectionID	Text	40002542092, 40002544092, 50083677092	Comma-separated concatenation of Route IDs that converge at the principal node; this value applies to all nodes participating in the intersection. Route IDs are sourced from the NCDOT Route Arcs file.
KeyIntersectionGroupID	Text		Assigned unique identifier for a group of intersections (e.g., quadrant intersections). Values are null if intersection is not part of a larger system of intersections.
IntersectionGroupID	Text		Comma-separated concatenation of Route IDs that converge at the principal node of the intersection group. Values are null if intersection is not part of a larger system of intersections. Route IDs are sourced from the NCDOT Route Arcs file.
CrossStreet	Text		Comma-separated concatenation of common street names that converge at the principal node. Street names are sourced from the NCDOT Route Arcs file.
IntersectionGeometry	Text	1- T-Intersection	The type of geometric configuration that best describes the intersection/junction. Based primarily on number of principal approach legs in the intersection
IntersectionType	Text	2 Roadway/roadway (interchange ramp terminal)	Type of junction based on conflicting modes (e.g., roadway-to-roadway, railway-to-roadway, greenway-to-roadway). Calculated based on the source of the initial intersection node
NumberofApproaches	Long	3, 4	Count of principal approach legs at the intersection plus any applicable Exceptions. Note, manually updated where linework produced an incorrect number of approach legs
TrafficControlSource	Text	NCDOT	Source of traffic control information.

Variable name	Type	Example values	Description
SignalID	Text	05-0915	Signal ID of NCDOT-maintained signal at the intersection (if present). Signal ID sourced from NC State Signals Inventory. Includes flashers, as well as traffic signals.
InterchangeInfluence	Text	TSUINTC01098, TSUINTC01100	NCDOT TSU interchange ID sourced from NC Interchange Inventory.
IntSkewAngle	Long	64, 90	Intersection skew angle calculated as the minimum angle between two perpendicular principal approach legs at the intersection.
IntersectionAADT	Double	86.5, 13000	Average AADT of the intersection calculated as the mean of AADT values of the principal approach legs. AADT values are sourced from the NC AADT Inventory and NC Streetlight Traffic Counts.
IntersectionAADTRecent	Long	2020, 2021	Year of the most recent AADT value used in the intersection AADT calculation. Year values are sourced from the NC AADT Inventory. If source dataset is the NC Streetlight Traffic Counts, then the year is 2021.
IntersectionAADTOldest	Long	2020, 2021	Year of the oldest AADT value used in the intersection AADT calculation. Year values are sourced from the NC AADT Inventory. If source dataset is the NC Streetlight Traffic Counts, then the year is 2021.
UrbanAreaCode	Long	15670, 29440	Urban area code based on 2020 Census boundaries
RampTerminal	Binary	Yes, No	Indicator that intersection is a ramp terminal.
RailCrossingID	Text	Light Rail 6, Light Rail 45	USDOT rail crossing ID associated with a roadway/railroad grade crossing or the rail crossing present within the roadway/roadway intersection area of influence (150 ft). Rail crossing IDs are sourced from the NC Rail Crossings inventory.
CrossingGrade	Text	AT GRADE, RR OVER	Indicator that railroad or greenway crossing is present and if it is at grade. Current formats (based on NC Rail Crossings Inventory and NC At-Grade Trail Crossings)
Longitude	Double	-82.349569, -82.658842	Longitude coordinate of the principal node or Intersection point using NAD 1983 State Plane North Carolina FIPS 3200 (US Feet).
Latitude	Double	35.036433, 35.780733	Latitude coordinate of the principal node or Intersection point using NAD 1983 State Plane North Carolina FIPS 3200 (US Feet).

Variable name	Type	Example values	Description
CountyCode	Long	75, 92	Coded value of county in which the intersection is physically located (1-100, alphabetically).
CountyName	Text	Polk, Wake	Name of county in which the intersection is physically located.
CityCode	Long	438, 479	Coded value of municipality in which the intersection is physically located based on the data dictionary associated with the DMV-349 crash report. Values are null if intersection is not located within a municipal boundary.
CityName	Text	SALUDA, RALEIGH	Name of municipality in which the intersection is physically located. Values are null if intersection is not located within a municipal boundary.
MPOABBR	Text	Foothills RPO, French Broad River MPO	MPO or RPO in which the intersection is physically located.
FC_1_LEG_COUNT	Long	0, 1, 2, 3	Total number of FC 1 legs in the intersection
FC_2_LEC_COUNT	Long	0, 1, 2, 3	Total number of FC 2 legs in the intersection
FC_3_LEG_COUNT	Long	0, 1, 2, 3	Total number of FC 3 legs in the intersection
FC_4_LEC_COUNT	Long	0, 1, 2, 3	Total number of FC 4 legs in the intersection
FC_5_LEG_COUNT	Long	0, 1, 2, 3	Total number of FC 5 legs in the intersection
FC_6_LEC_COUNT	Long	0, 1, 2, 3	Total number of FC 6 legs in the intersection
FC_7_LEG_COUNT	Long	0, 1, 2, 3	Total number of FC 7 legs in the intersection
URBAN_FLAG	Binary	1, 0	Flag whether intersection is in an urban area. 1- yes, 0- no
CITY_FLAG	Binary	1, 0	Flag whether intersection is in a city. 1- yes, 0- no
GEOID	Text	371830540181	A unique identifier that distinguishes which census block group the majority of the intersection falls within
BG_AREA	Double	52.0751755, 49.6230077	Total area of the census block (sq. mi)
Proximate_University_College	Binary	0, 1	College or university present within HALF mile (0 if >HALF mile; 1 if <HALF mile) of intersection
Top5University	Binary	0,1	Top 5 (by enrollment) college or university present within HALF mile (0 if >HALF mile; 1 if <HALF mile) of intersection
K12_Count	Long	1, 3	Number of K-12 schools within QTR mile of intersection

Variable name	Type	Example values	Description
AlcSales_Count	Long	1, 8	Number of alcohol sales establishments within QTR mile of intersection
AlcSales_Density	Double	5.092958, 40.743665	Density of alcohol sales establishments within QTR mile of intersection
Park_prox	Binary	0, 1	Public greenspace within QTR mile (0 if >QTR mile; 1 if <QTR mile) of intersection
Urban	Binary	0, 1	Intersection is in a Census-designated urban area (urbanized area or urban cluster); 1 if in urban area, 0 if otherwise
Sidewalk	Binary	0, 1	Presence of sidewalk within 100 feet (0 if >100 feet; 1 if <100 feet) of intersection
Greenway	Binary	0, 1	Presence of greenway within 100 feet (0 if >100 feet; 1 if <100 feet) of intersection
Crosswalks	Binary	0, 1	Flag whether the intersection is within 100 ft of a crosswalk
Parcel_Count	Long	10, 22	Number of land parcels within QTR mile of intersection
Parcel_Denstiy	Double	249.554951, 524.574692	Density of land parcels within QTR mile of intersection
Speed_25	Binary	0, 1	Flag if the intersection has a leg with a speed of 25mph
Speed_30	Binary	0, 1	Flag if the intersection has a leg with a speed of 30mph
Speed_35	Binary	0, 1	Flag if the intersection has a leg with a speed of 35mph
Speed_40	Binary	0, 1	Flag if the intersection has a leg with a speed of 40mph
Speed_45	Binary	0, 1	Flag if the intersection has a leg with a speed of 45mph
Speed_50	Binary	0, 1	Flag if the intersection has a leg with a speed of 50mph
Speed_55	Binary	0, 1	Flag if the intersection has a leg with a speed of 55mph
Speed_60	Binary	0, 1	Flag if the intersection has a leg with a speed of 60mph
Speed_65	Binary	0, 1	Flag if the intersection has a leg with a speed of 65mph
Min_Speed_25	Binary	0, 1	Flag if the intersection has a minimum speed of 25mph
Min_Speed_30	Binary	0, 1	Flag if the intersection has a minimum speed of 30mph
Min_Speed_35	Binary	0, 1	Flag if the intersection has a minimum speed of 35mph
Min_Speed_40	Binary	0, 1	Flag if the intersection has a minimum speed of 40mph
Min_Speed_45	Binary	0, 1	Flag if the intersection has a minimum speed of 45mph
Min_Speed_50	Binary	0, 1	Flag if the intersection has a minimum speed of 50mph

Variable name	Type	Example values	Description
Min_Speed_55	Binary	0, 1	Flag if the intersection has a minimum speed of 55mph
Min_Speed_60	Binary	0, 1	Flag if the intersection has a minimum speed of 60mph
Max_Speed_25	Binary	0, 1	Flag if the intersection has a maximum speed of 25mph
Max_Speed_30	Binary	0, 1	Flag if the intersection has a maximum speed of 30mph
Max_Speed_35	Binary	0, 1	Flag if the intersection has a maximum speed of 35mph
Max_Speed_40	Binary	0, 1	Flag if the intersection has a maximum speed of 40mph
Max_Speed_45	Binary	0, 1	Flag if the intersection has a maximum speed of 45mph
Max_Speed_50	Binary	0, 1	Flag if the intersection has a maximum speed of 50mph
Max_Speed_55	Binary	0, 1	Flag if the intersection has a maximum speed of 55mph
Max_Speed_60	Binary	0, 1	Flag if the intersection has a maximum speed of 60mph
MAX_Undiv_1Lane	Binary	0, 1	Flag if one of the intersection legs is an undivided 1 lane road
MAX_Undiv_2Lane	Binary	0, 1	Flag if one of the intersection legs is an undivided 2 lane road
MAX_Undiv_3Lane	Binary	0, 1	Flag if one of the intersection legs is an undivided 3 lane road
MAX_Undiv_4Lane	Binary	0, 1	Flag if one of the intersection legs is an undivided 4 lane road
MAX_Undiv_5Lane	Binary	0, 1	Flag if one of the intersection legs is an undivided 5 lane road
MAX_Undiv_6Lane	Binary	0, 1	Flag if one of the intersection legs is an undivided 6 lane road
MAX_Div_1Lane	Binary	0, 1	Flag if one of the intersection legs is a divided 1 lane road
MAX_Div_2Lane	Binary	0, 1	Flag if one of the intersection legs is a divided 2 lane road
MAX_Div_3Lane	Binary	0, 1	Flag if one of the intersection legs is a divided 3 lane road
MAX_Div_4Lane	Binary	0, 1	Flag if one of the intersection legs is a divided 4 lane road
MAX_Div_5Lane	Binary	0, 1	Flag if one of the intersection legs is a divided 5 lane road
MAX_Div_6Lane	Binary	0, 1	Flag if one of the intersection legs is a divided 6 lane road
MAX_AADT_2015	Long	16000	MAX 2015 AADT for the entire intersection over all legs
MAX_AADT_2016	Long	4200	MAX 2016 AADT for the entire intersection over all legs
MAX_AADT_2017	Long	2000, 16500	MAX 2017 AADT for the entire intersection over all legs
MAX_AADT_2018	Long	3115	MAX 2018 AADT for the entire intersection over all legs
MAX_AADT_2019	Long	2300 , 24000	MAX 2019 AADT for the entire intersection over all legs
MAX_AADT_2020	Long	25000, 43200	MAX 2020 AADT for the entire intersection over all legs
MAX_AADT_2021	Long	18500, 22000	MAX 2021 AADT for the entire intersection over all legs
MAX_AADT_2022	Long	18500	MAX 2022 AADT for the entire intersection over all legs

Variable name	Type	Example values	Description
MAX_AADT_2023	Long	20500, 25500	MAX 2023 AADT for the entire intersection over all legs
MIN_AADT	Long	0, 200	Minimum AADT across all intersection legs
MAX_AADT	Long	22000, 35000	Maximum AADT across all intersection legs
POP_DENS	Double	135.84197, 110.25645	Population density within QTR mile of the intersection
EMP_DENS	Double	5664.194225, 6845.264875	Employment density within QTR mile of the intersection
NONMOT_DENS	Double	6.125451, 8.318231	Density of non-motorized commuters within QTR mile of the intersection
K12_DENS	Double	76.251547, 78.977889	K-12 enrollment density (by place of residence) within QTR mile of the intersection
POP_65_DENS	Double	110.26548, 212.52412	Density of persons 65 and older within QTR mile of the intersection
POP_18_DENS	Double	135.84197, 215.21254	Density of persons 18 and younger within QTR mile of the intersection
POP_PROP	Double	0.125773, 0.312547	Proportion of the population within QTR mile of the intersection
POP_65_PROP	Double	0.125773, 0.312547	Proportion of the population 65 and older within QTR mile of the intersection
POP_18_PROP	Double	0.147766, 0.256254	Proportion of population aged 18 and under within QTR mile of the intersection
NONMOT_PROP	Double	0.006873 , 0.005123	Proportion of non-motorized commuters within QTR mile of the intersection
UNEMP_PROP	Double	0.026, 0.125	Proportion of civilian labor force that is unemployed within QTR mile of the intersection
PROP_NONWHITE	Double	0.714777, 0.842542	Proportion of population that is non-white or 2 or more races within QTR mile of the intersection
DISABLE_PROP	Double	0, 0.01524	Proportion of the population with a disability within QTR mile of the intersection
COLLEGE_25PLUS_PROP	Double	0.373883, 0.514254	Proportion of the population over the age of 25 with an associate's degree or higher within QTR mile of the intersection
ZERO_HH_PROP	Double	0.023542, 0.036388	Proportion zero vehicle households within QTR mile of the intersection

Variable name	Type	Example values	Description
LEP_HH_PROP	Double	0, 0.087601	Proportion of households with limited English proficiency within QTR mile of the intersection
COLLEGE_25PLUS	Double	350, 540	Total population over the age of 25 with an associate's degree or higher within QTR mile of the intersection
LEP_HH	Double	25, 47	Total limited English proficiency households within QTR mile of the intersection
MED_INC	Double	16000, 52000	Median household income of the dominant census block group (i.e., covering most of the intersection)
NONWHITE	Double	1000, 1040	Total non-white or 2 or more races population within QTR mile of the intersection
TOT_16PLUS	Double	600, 1211	Total population in the civilian labor force over 16 within QTR mile of the intersection
POP_25_44	Double	250, 410	Total population between the ages of 25 and 44 within QTR mile of the intersection
POP_25PLUS	Double	750, 925	Total population over the age of 25 within QTR mile of the intersection
POP_POV	Double	100, 232	Total population living under the poverty line within QTR mile of the intersection
POP_POV_DET	Double	1455, 1625	Total population for which poverty status has been determined within QTR mile of the intersection
POP_65	Double	180, 220	Total population 65 and older within QTR mile of the intersection
POP_18	Double	200, 215	Total population 18 and younger within QTR mile of the intersection
UNEMP	Double	15, 30	Total unemployed persons in the civilian labor force within QTR mile of the intersection
WHITE	Double	400, 515	Total white population within QTR mile of the intersection
TOT_POP	Double	1325, 1455	Total population within QTR mile of the intersection
ZERO_HH	Double	12, 27	Total zero vehicle households within QTR mile of the intersection
TOT_EMP	Long	8500, 8965	Total employment within QTR mile of the intersection

Variable name	Type	Example values	Description
RETAIL	Long	2380, 2500	Total retail employment (NAICS 44-45) within QTR mile of the intersection
EDUCATION_61	Long	97, 112	Total educational services employment (NAICS 61) within QTR mile of the intersection
ARTSENTREC_71	Long	65, 68	Total arts, entertainment, and recreation services employment (NAICS 71) within QTR mile of the intersection
ACCFOOD_72	Long	617, 672	Total accommodation and food services employment (NAICS 72) within QTR mile of the intersection
TOT_COMM	Double	720, 850	Total commuters within QTR mile of the intersection
TOT_BIKE	Double	0, 10	Total bicycle commuters within QTR mile of the intersection
TOT_TRANS	Double	0, 5	Total transit commuters within QTR mile of the intersection
TOT_WALK	Double	15,32	Total walking commuters within QTR mile of the intersection
TOT_NONMOT	Double	3, 12	Total non-motorized commuters within QTR mile of the intersection
DISABLE_POP	Double	0, 15	Total population with a disability within QTR mile of the intersection
TOT_HH	Double	500, 720	Total households within QTR mile of the intersection
K12_ENROLL	Double	100, 316	Total K-12 enrollment residing within QTR mile of the intersection
Other_Freeway	Binary	0, 1	1 if one of the legs at the measurement location (<100 feet) is an Other Freeways or Expressway; 0 otherwise
Major_Arterial	Binary	0, 1	1 if one of the legs at the measurement location (<100 feet) is an Other Principal Arterial; 0 otherwise
Minor_Arterial	Binary	0, 1	1 if one of the legs at the measurement location (<100 feet) is a Minor Arterial; 0 otherwise
Major_Collector	Binary	0, 1	1 if one of the legs at the measurement location (<100 feet) is a Major Collector; 0 otherwise
Minor_Collector	Binary	0, 1	1 if one of the legs at the measurement location (<100 feet) is a Minor Collector; 0 otherwise
Local	Binary	0, 1	1 if one of the legs at the measurement location (<100 feet) is a Local road; 0 otherwise

Variable name	Type	Example values	Description
Ramp	Binary	0, 1	Flag whether the intersection has a Ramp leg
busStop_Count	Long	0, 4, 13	Total Bus Stops within a quarter mile of the intersection
busStop_Density	Double	0, 12.945753	Density of bus stops
Bus_1_4mi	Binary	0, 1	Flag whether the intersection is within a quarter mile of a bus route
Bus_100ft	Binary	0, 1	Flag whether the intersection is within 100 ft of a bus route
LightRail_1_4mi	Binary	0, 1	Flag whether the intersection is within a quarter mile of a light rail stop
LightRail_100ft	Binary	0, 1	Flag whether the intersection is within 100ft of a light rail stop
Developed_HI	Double	0.191193471, 13.9018331946759	Percentage (0-100) of land use within 1 km classified as "Developed, High Intensity" according to the 2023 National Land Cover Database.
Developed_MI	Double	2.08884510044468, 0.944817347280674	Percentage (0-100) of land use within 1 km classified as "Developed, Medium Intensity" according to the 2023 National Land Cover Database.
Developed_LI	Double	7.16388345328097, 2.06391264529122	Percentage (0-100) of land use within 1 km classified as "Developed, Low Intensity" according to the 2023 National Land Cover Database.
LU_Mix	Double	0.267950503905008, 0.920503350598568	Land use mix adapted from Frank et al.'s (2004) methodology. Four land uses included: Developed, High Intensity, Developed, Low Intensity, Developed, Medium Intensity, and all other land use classifications combined
Total_Buildings_SqFt_1km	Double	22.856522, 35.254524	Total area of buildings within 1 kilometer of the intersection (square feet)
Residential	Double	0.000001, .571804	Proportion of residential buildings within 1km of the intersection
Industrial	Double	0.000001, 0.312	Proportion of industrial buildings within 1km of the intersection
Commercial	Double	0.000001, 0.330998	Proportion of commercial buildings within 1km of the intersection
Institutional	Double	0.000001, 0.215252	Proportion of institutional buildings within 1km of the intersection
Tot_Residential_Buildings_1km	Long	1, 2, 32	Total number of residential of buildings within a kilometer of the intersection

Variable name	Type	Example values	Description
Residential_SqFt_1km	Double	22.856522, 35.254524	Total area of residential buildings within 1 kilometer of the intersection (square feet)
Total_Buildings_SqFt_1km	Double	22.856522, 35.254524	Total area of buildings within 1 kilometer of the intersection (square feet)
Total_Industrial_Buildings_1km	Long	1, 2, 32	Total number of industrial buildings within a kilometer of the intersection
Industrial_SqFt_1km	Double	22.856522, 35.254524	Total area of industrial buildings within 1 kilometer of the intersection (square feet)
Total_Commercial_Buildings_1km	Long	1, 2, 32	Total number of commercial buildings within a kilometer of the intersection
Commercial_SqFt_1km	Double	22.856522, 35.254524	Total area of commercial buildings within 1 kilometer of the intersection (square feet)
Total_Education_Buildings_1km	Long	1, 2, 32	Total number of education buildings within a kilometer of the intersection
Education_SqFt_1km	Double	22.856522, 35.254524	Total area of education buildings within 1 kilometer of the intersection (square feet)
Total_Government_Buildings_1km	Long	1, 2, 32	Total number of government buildings within a kilometer of the intersection
Government_SqFt_1km	Double	22.856522, 35.254524	Total area of government buildings within 1 kilometer of the intersection (square feet)
Total_Assembly_Buildings_1km	Long	1, 2, 32	Total number of assembly buildings within a kilometer of the intersection
Assembly_SqFt_1km	Double	22.856522, 35.254524	Total area of assembly buildings within 1 kilometer of the intersection (square feet)
BU_Mix	Double	0.4886847, 0.503649	Building use mix adapted from Frank et al.'s (2004) methodology.

Appendix D. Risk modeling database dictionary

Table 28. Risk Modeling Database Dictionary

Variable name	Type	Example values	Description
SegmentID	Long	1, 811, 74744	Unique (non-consecutive) integer representing the segment (i.e., intersection to intersection)
Division	Long	3, 6	NCDOT division in which the segment is physically located
LocCntyCode	Text	016, 067, 078	County in which the segment is physically located (coded value)
RouteClass	Text	1, 2, 3	NCDOT route class indicating the dominant route along the segment (e.g., Interstate, US Route, NC Route)
RouteNumber	Text	53, 71, 1312	Route number associated with the dominant route on the segment
RouteID	Text	40001312078, 30000072078	NCDOT route identifier based on NCDOT's route characteristics
County	Text	Onslow, Robeson, Wake	County in which the segment is physically located
FuncClass	Long	1, 2, 3, 4, 5, 6, 7	Functional classification of the road based on NCDOT's route characteristics
MedianType	Test	0, PM, Curb, Grass	Type of dividing median on the road (if applicable) based on NCDOT's route characteristics
SpeedLimit	Long	25, 35, 55	Posted speed limit on the road (if known) based on NCDOT's route characteristics
Total_Lanes	Long	2, 4, 6	Total number of through lanes in both directions based on NCDOT's route characteristics
AADT_2015	Long	115, 16000	Average annual daily traffic for the year 2015 (if available)
AADT_2016	Long	292, 4200	Average annual daily traffic for the year 2016 (if available)
AADT_2017	Long	2000, 16500	Average annual daily traffic for the year 2017 (if available)
AADT_2018	Long	0, 3115	Average annual daily traffic for the year 2018 (if available)
AADT_2019	Long	2300 , 24000	Average annual daily traffic for the year 2019 (if available)
AADT_2020	Long	25000, 43200	Average annual daily traffic for the year 2020 (if available)
AADT_2021	Long	18500, 22000	Average annual daily traffic for the year 2021 (if available)
AADT_2022	Long	243, 18500	Average annual daily traffic for the year 2022 (if available)

Variable name	Type	Example values	Description
AADT_2023	Long	20500, 25500	Average annual daily traffic for the year 2023 (if available)
Total_Segment_Length	Double	0.203174, 1.239075	Total length of the segment in miles
undiv	Binary	0, 1	Flag whether the segment is undivided
div	Binary	0, 1	Flag whether the segment is divided
Ramp	Binary	0, 1	Flag whether the segment is a Ramp
Other_Freeway	Binary	0, 1	Flag whether the segment is an Other Freeways or Expressway
Major_Arterial	Binary	0, 1	Flag whether the segment is an Other Principal Arterial
Minor_Arterial	Binary	0, 1	Flag whether the segment is a Minor Arterial
Major_Collector	Binary	0, 1	Flag whether the segment is a Major Collector
Minor_Collector	Binary	0, 1	Flag whether the segment is a Minor Collector
Local	Binary	0, 1	Flag whether the segment is a Local road
GEOID	Text	371830540181	A unique identifier that distinguishes which census block group the majority of the segment falls within
BG_AREA	Double	52.0751755, 49.6230077	Total area of the census block (sq. mi)
ACCFOOD_72	Long	617, 672	Total accommodation and food services employment (NAICS 72) within the dominant census block group (i.e., covering most of the segment)
ARTSENTREC_71	Long	65, 68	Total arts, entertainment, and recreation services employment (NAICS 71) within the dominant census block group (i.e., covering most of the segment)
EDUCATION_61	Long	97, 112	Total educational services employment (NAICS 61) within the dominant census block group (i.e., covering most of the segment)
RETAIL	Long	2380, 2500	Total retail employment (NAICS 44-45) within the dominant census block group (i.e., covering most of the segment)
TOT_EMP	Long	8500, 8965	Total employment within the dominant census block group (i.e., covering most of the segment)
EMP_DENS	Double	5664.194225, 6845.264875	Employment density within the dominant census block group (i.e., covering most of the segment)
UNEMP	Double	15, 30	Total unemployed persons in the civilian labor force within the dominant census block group (i.e., covering most of the segment)

Variable name	Type	Example values	Description
UNEMP_PROP	Double	0.026, 0.125	Proportion of civilian labor force that is unemployed within the dominant census block group (i.e., covering most of the segment)
TOT_HH	Double	500, 720	Total households within the dominant census block group (i.e., covering most of the segment)
MED_INC	Double	16000, 52000	Median household income of the dominant census block group (i.e., covering most of the segment)
ZERO_HH	Double	12, 27	Total zero vehicle households within the dominant census block group (i.e., covering most of the segment)
ZERO_HH_PROP	Double	0.023542, 0.036388	Proportion zero vehicle households within the dominant census block group (i.e., covering most of the segment)
LEP_HH	Double	25, 47	Total limited English proficiency households within the dominant census block group (i.e., covering most of the segment)
LEP_HH_PROP	Double	0, 0.087601	Proportion of households with limited English proficiency within the dominant census block group (i.e., covering most of the segment)
TOT_POP	Double	1325, 1455	Total population within the dominant census block group (i.e., covering most of the segment)
POP_DENS	Double	135.84197, 110.25645	Population density within the dominant census block group (i.e., covering most of the segment)
WHITE	Double	400, 515	Total white population within the dominant census block group (i.e., covering most of the segment)
NONWHITE	Double	1000, 1040	Total non-white or 2 or more races population within the dominant census block group (i.e., covering most of the segment)
PROP_NONWHITE	Double	0.714777, 0.842542	Proportion of population that is non-white or 2 or more races within the dominant census block group (i.e., covering most of the segment)
TOT_16PLUS	Double	600, 1211	Total population in the civilian labor force over 16 within the dominant census block group (i.e., covering most of the segment)
POP_18	Double	200, 215	Total population 18 and younger within the dominant census block group (i.e., covering most of the segment)
POP_18_DENS	Double	135.84197, 215.21254	Density of persons 18 and younger within the dominant census block group (i.e., covering most of the segment)

Variable name	Type	Example values	Description
POP_18_PROP	Double	0.147766, 0.256254	Proportion of population aged 18 and under within the dominant census block group (i.e., covering most of the segment)
POP_25PLUS	Double	750, 925	Total population over the age of 25 within the dominant census block group (i.e., covering most of the segment)
COLLEGE_25PLUS	Double	350, 540	Total population over the age of 25 with an associate's degree or higher within the dominant census block group (i.e., covering most of the segment)
COLLEGE_25PLUS_PROP	Double	0.373883, 0.514254	Proportion of the population over the age of 25 with an associate's degree or higher within the dominant census block group (i.e., covering most of the segment)
POP_25_44	Double	250, 410	Total population between the ages of 25 and 44 within the dominant census block group (i.e., covering most of the segment)
POP_65	Double	180, 220	Total population 65 and older within the dominant census block group (i.e., covering most of the segment)
POP_65_DENS	Double	110.26548, 212.52412	Density of persons 65 and older within the dominant census block group (i.e., covering most of the segment)
POP_65_PROP	Double	0.125773, 0.312547	Proportion of the population 65 and older within the dominant census block group (i.e., covering most of the segment)
DISABLE_POP	Double	0, 15	Total population with a disability within the dominant census block group (i.e., covering most of the segment)
DISABLE_PROP	Double	0, 0.01524	Proportion of the population with a disability within the dominant census block group (i.e., covering most of the segment)
K12_ENROLL	Double	100, 316	Total K-12 enrollment residing within the dominant census block group (i.e., covering most of the segment)
K12_DENS	Double	76.251547, 78.977889	K-12 enrollment density (by place of residence) within the dominant census block group (i.e., covering most of the segment)
TOT_COMM	Double	720, 850	Total commuters within the dominant census block group (i.e., covering most of the segment)
TOT_BIKE	Double	0, 10	Total bicycle commuters within the dominant census block group (i.e., covering most of the segment)
TOT_WALK	Double	15,32	Total walking commuters within the dominant census block group (i.e., covering most of the segment)

Variable name	Type	Example values	Description
TOT_TRANS	Double	0, 5	Total transit commuters within the dominant census block group (i.e., covering most of the segment)
TOT_NONMOT	Double	3, 12	Total non-motorized commuters within the dominant census block group (i.e., covering most of the segment)
NONMOT_DENS	Double	6.125451, 8.318231	Density of non-motorized commuters within the dominant census block group (i.e., covering most of the segment)
NONMOT_PROP	Double	0.006873 , 0.005123	Proportion of non-motorized commuters within the dominant census block group (i.e., covering most of the segment)
POP_POV	Double	100, 232	Total population living under the poverty line within the dominant census block group (i.e., covering most of the segment)
POP_POV_DET	Double	1455, 1625	Total population for which poverty status has been determined within the dominant census block group (i.e., covering most of the segment)
POV_PROP	Double	0.090722 , 0.256325	Proportion of the population living under the poverty line within the dominant census block group (i.e., covering most of the segment)
Developed_HI_100ft	Binary	0, 1	Segment is within 100 feet of land use classified as "Developed, High Intensity" according to the 2023 National Land Cover Database.
Developed_LI_100ft	Binary	0, 1	Segment is within 100 feet of land use classified as "Developed, Low Intensity" according to the 2023 National Land Cover Database.
Developed_MI_100ft	Binary	0, 1	Segment is within 100 feet of land use classified as "Developed, Medium Intensity" according to the 2019 National Land Cover Database.
K12_Count	Binary	0, 1	Primary or secondary schools present (grades K through 12) within QTR mile of segment
Proximate_University_College	Binary	0, 1	College or university present within HALF mile (0 if >HALF mile; 1 if <HALF mile) of segment
BusRoute_Present	Binary	0, 1	Bus route present on segment (excluding intercity routes)
GoldLineStop_1_4mi	Binary	0, 1	Gold Line light rail stop within QTR mile (0 if >QTR mile; 1 if <QTR mile) of segment (open July 2015)
BlueLineStop_1_4mi	Binary	0, 1	Blue Line light rail stop within QTR mile (0 if >QTR mile; 1 if <QTR mile) of segment (open March 2018)
busStop_Count	Long	0, 4, 13	Total Bus Stops within a quarter mile of the segment
busStop_Density	Double	0, 12.945753	Density of bus stops within a quarter mile of the segment

Variable name	Type	Example values	Description
AlcSales_Count	Long	1, 8	Number of alcohol sales establishments within QTR mile of segment
AlcSales_Density	Double	5.092958, 40.743665	Density of alcohol sales establishments within QTR mile of segment
Park_prox	Binary	0, 1	Public greenspace within QTR mile (0 if >QTR mile; 1 if <QTR mile) of segment
Sidewalk	Binary	0, 1	Presence of sidewalk within 100 feet (0 if >100 feet; 1 if <100 feet) of segment
Greenway	Binary	0, 1	Presence of greenway within 100 feet (0 if >100 feet; 1 if <100 feet) of segment
Parcel_Count	Long	10, 22	Number of land parcels within QTR mile of segment
Parcel_Denstiy	Double	249.554951, 524.574692	Density of land parcels within QTR mile of segment
Building_Count	Long	1, 2, 32	Total number of buildings within a quarter mile of the segment
Total_Build_Sqft	Double	22.856522, 35.254524	Total area of buildings within a quarter mile of the segment (square feet)
Building_Count_agriculture	Long	1, 2, 32	Total number of agriculture buildings within a quarter mile of the segment
Agriculture_Build_Sqft	Double	22.856522, 35.254524	Total area of agriculture buildings within a quarter mile of the segment (square feet)
Building_Count_assembly	Long	1, 2, 32	Total number of assembly buildings within a quarter mile of the segment
Assembly_Build_Sqft	Double	22.856522, 35.254524	Total area of assembly buildings within a quarter mile of the segment (square feet)
Building_Count_commercial	Long	1, 2, 32	Total number of commercial buildings within a quarter mile of the segment
Commercial_Build_Sqft	Double	22.856522, 35.254524	Total area of commercial buildings within a quarter mile of the segment (square feet)
Building_Count_education	Long	1, 2, 32	Total number of education buildings within a quarter mile of the segment
Education_Build_Sqft	Double	22.856522, 35.254524	Total area of education buildings within a quarter mile of the segment (square feet)
Building_Count_government	Long	1, 2, 32	Total number of government buildings within a quarter mile of the segment
Government_Build_Sqft	Double	22.856522, 35.254524	Total area of government buildings within a quarter mile of the segment (square feet)
Building_Count_industrial	Long	1, 2, 32	Total number of industrial buildings within a quarter mile of the segment

Variable name	Type	Example values	Description
Industrial_Build_Sqft	Double	22.856522, 35.254524	Total area of industrial buildings within a quarter mile of the segment (square feet)
Building_Count_residential	Long	1, 2, 32	Total number of residential buildings within a quarter mile of the segment
Residential_Build_Sqft	Double	22.856522, 35.254524	Total area of residential buildings within a quarter mile of the segment (square feet)
Building_Count_unclassified	Long	1, 2, 32	Total number of unclassified buildings within a quarter mile of the segment
Unclassified_Build_Sqft	Double	22.856522, 35.254524	Total area of unclassified buildings within a quarter mile of the segment (square feet)
Building_Count_utilityandmisc	Long	1, 2, 32	Total number of utility and misc. buildings within a quarter mile of the segment
Utilityandmisc_Build_Sqft	Double	22.856522, 35.254524	Total area of utility and misc. buildings within a quarter mile of the segment (square feet)
Urban	Binary	0, 1	Segment is in a Census-designated urban area (urbanized area or urban cluster); 1 if in urban area, 0 if otherwise
KABCO_2015	Long	0, 2, 4	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2015)
KA_2015	Long	0, 2, 5	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2015)
KABCO_2016	Long	0, 2, 6	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2016)
KA_2016	Long	0, 2, 7	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2016)
KABCO_2017	Long	0, 2, 8	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2017)
KA_2017	Long	0, 2, 9	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2017)
KABCO_2018	Long	0, 2, 10	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2018)
KA_2018	Long	0, 2, 11	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2018)

Variable name	Type	Example values	Description
KABCO_2019	Long	0, 2, 12	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2019)
KA_2019	Long	0, 2, 13	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2019)
KABCO_2020	Long	0, 2, 14	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2020)
KA_2020	Long	0, 2, 15	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2020)
KABCO_2021	Long	0, 2, 16	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2021)
KA_2021	Long	0, 2, 17	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2021)
KABCO_2022	Long	0, 2, 18	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2022)
KA_2022	Long	0, 2, 19	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2022)
KABCO_2023	Long	0, 2, 20	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2023)
KA_2023	Long	0, 2, 21	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2023)
KeyIntersectionID_1	Text	TSUINT716991, TSUINT43242	Unique identification of an intersection for the intersection at the beginning of the segment
KeyIntersectionID_2	Text	TSUINT716991, TSUINT43243	Unique identification of an intersection for the intersection at the end of the segment