RESEARCH & DEVELOPMENT

Quantification of Systemic Risk Factors for Pedestrian Safety on North Carolina

Vikash V. Gayah, S. Ilgin Guler, Hao Liu
Department of Civil and Environmental Engineering
The Pennsylvania State University

Lauren Blackburn, Ian Hamilton
Vanasse Hangen Brustlin, Inc (VHB)

NCDOT Project 2022-11

FHWA/NC/2022-11

October 2022

Technical Report Documentation Page

<table>
<tr><td>1. Report No.
2022-11</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle

Quantification of Systemic Risk Factors for Pedestrian Safety on North Carolina</td><td>5. Report Date
October 2022</td></tr>
<tr><td>6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s)

Vikash V. Gayah, S. Ilgin Guler, Hao Liu, Lauren Blackburn, Ian Hamilton</td><td>8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address

The Pennsylvania State University (Larson Transportation Institute)
Vanasse Hangen Brustlin, Inc</td><td>10. Work Unit No. (TRAIS)</td></tr>
<tr><td>11. Contract or Grant No.</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
North Carolina Department of Transportation
Research and Development Unit
104 Fayetteville Street
Raleigh, North Carolina 27601</td><td>13. Type of Report and Period Covered
Final Report

August 1, 2021 – October 31, 2022</td></tr>
<tr><td>14. Sponsoring Agency Code
2022-11</td></tr>
<tr><td colspan="3">Supplementary Notes:</td></tr>
</table>

16. Abstract

Pedestrian safety is a growing concern for transportation planners and safety engineers within North Carolina and across the country. Pedestrians are extremely vulnerable users of the transportation system and are particularly subject to serious injuries and fatalities in the event of a crash. For example, while pedestrians are involved in just 1% of all crashes in NC, they are represented in 10% of those crashes involving serious injuries or fatalities. The 2019 update to the North Carolina Strategic Highway Safety Plan (SHSP) identifies pedestrians, bicyclists, and other non-motorized users as a critical emphasis area within the broader objective of providing a safer transportation system to all North Carolinians. Furthermore, while the number of motor vehicle crashes has decreased significantly during 2020 due to reduced vehicular travel, pedestrian-related crashes have remained relatively steady throughout the year.

The goal of this project was to inform the systemic safety process for pedestrian safety on roadway segments in North Carolina. With this goal in mind, the specific objectives were to: 1) identify and quantify systemic risk factors for pedestrian safety on North Carolina roads; and, 2) develop guidance for analysts at NCDOT and local agencies within the state on how to apply these risk factors to proactively address potential safety concerns. Risk factors were developed for both fatal and severe pedestrian crashes, as well as all pedestrian crashes, on urban roadway segments. The products of this research can provide guidance on how to implement systemic pedestrian safety analysis in North Carolina, focusing on the identification and use of pedestrian risk factors in urban areas.

<table>
<tr><td colspan="2">17. Key Words
Systemic safety
Pedestrian safety
Pedestrian risk factors</td><td colspan="2">18. Distribution Statement</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
103</td><td>22. Price</td></tr>
</table>

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

DISCLAIMER

The contents of this report reflect the views of the author(s) and not necessarily the views of the University. The author(s) 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 either the North Carolina Department of Transportation or the Federal Highway Administration at the time of publication. This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Pedestrian safety is a growing concern for transportation planners and safety engineers within North Carolina and across the country. Pedestrians are extremely vulnerable users of the transportation system and are particularly subject to serious injuries and fatalities in the event of a crash. For example, while pedestrians are involved in just 1% of all crashes in NC, they are represented in 10% of those crashes involving serious injuries or fatalities. The 2019 update to the North Carolina Strategic Highway Safety Plan (SHSP) identifies pedestrians, bicyclists, and other non-motorized users as a critical emphasis area within the broader objective of providing a safer transportation system to all North Carolinians. Furthermore, while the number of motor vehicle crashes has decreased significantly during 2020 due to reduced vehicular travel, pedestrian-related crashes have remained relatively steady throughout the year.

The methods that have been taken to ensure safe walking for pedestrians in Pedestrian Safety Action Plans (PSAP) can be classified into two types: *crash summary-based methods* and *risk-based methods*. Crash summary-based methods are a reactive approach since they use historical crash data to identify hot spots with either high or higher than expected pedestrian frequency. Thus, crashes must occur at any location before underlying issues at that location may be addressed. Risk-based methods, on the other hand, are proactive since they aim to quantify the risk of a pedestrian crash occurring at a specific location based on its features, independent of its actual crash history. This is a systemic approach that seeks to identify sites with the highest risk of pedestrian-related safety issues across the entire transportation network s It relies on a system-wide analysis to identify common characteristics—or risk factors—of locations where pedestrian crashes frequently occur. Countermeasures can then be widely deployed at locations that have these risk factors to address potential concerns, even at locations that have these risk factors but has not experienced any crashes. In this way, potential concerns are addressed proactively at locations before crashes might have even occurred.

Different types of data are required by both methods. Generally, the reactive approach applies a crash density analysis (or crash frequency) in a geographic information systems (GIS) format to identify priority locations (VDOT, 2018). However, it does not necessarily focus on mapping high risk locations based on anticipated risk (as opposed to existing high crash locations). The proactive approach does not require previous crash occurrence. It identifies and quantifies the contributing factors of pedestrian crashes and locates sites with a higher possibility of pedestrian crashes. This proactive approach to traffic safety identifies circumstances along an entire network that may be potentially unsafe rather than retrofitting locations after crashes have occurred.

The goal of this project was to inform the systemic safety process for pedestrian safety on roadway segments in North Carolina. With this goal in mind, the specific objectives were to: 1) identify and quantify systemic risk factors for pedestrian safety on North Carolina roads; and, 2) develop guidance for analysts at NCDOT and local agencies within the state on how to apply these risk factors to proactively address potential safety concerns. Risk factors were developed for both fatal and severe pedestrian crashes, as well as all pedestrian crashes. The products of this research can provide guidance on how to implement systemic pedestrian safety analysis in North Carolina, focusing on the identification and use of pedestrian risk factors in urban areas.

The remainder of this document is organized as follows. First, a review of the literature on pedestrian risk factors and exposure modeling is provided. Then, the data used for this project are summarized. Next, the results of the pedestrian risk factor estimation are provided. Finally, discussion of how these factors can be used is provided.

REVIEW OF RESEARCH LITERATURE

As a first step, the research team gathered information about existing practices and guidance for the modeling and usage of pedestrian risk factors from state transportation agencies and journal articles. The remainder of this section summarizes the results of this 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.

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. The statistical modeling approach is generally preferred for identification and quantification of pedestrian risk factors. This method is data-driven and thus reflects the relationships between observed crash frequency and available explanatory factors. This method has been found to be more reliable than using expert judgment or frequency-based method. 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. However, one key limitation to these non-statistical methods is that weights assigned to potential risk factors would be solely based on expert judgment and would not be informed by data.

Statistical models

Many statistical modeling approaches have been used to explore associations between pedestrian crash frequency or risk outcomes and 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 specifically account for the count nature of crash frequency outcomes (i.e., that observed crash frequencies take non-negative integer values). The most common count regression model applied for crash frequency prediction is negative binomial (NB) regression model, which directly 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:

$$\ln \lambda_i = \beta X_i + \varepsilon_i \qquad (1)$$

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.

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

$$N_{pred} = e^{\beta_0} \times L^{\beta_1} \times AADT^{\beta_2} \times e^{(\beta_4 X_4 + \cdots + \beta_n X_n)} \qquad (2)$$

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, $\{X_4, \ldots, X_n\}$ is a vector of geometric design and other site-specific data, and $\{\beta_0, \ldots, \beta_n\}$ is a vector of estimable regression coefficients. However, for SPFs for pedestrian crashes it is critical to incorporate pedestrian exposure to account for the level of pedestrian activity. If available, this pedestrian exposure can be entered into the model in a form similar to vehicle traffic volume as follows:

$$N_{pred} = e^{\beta_0} \times L^{\beta_1} \times AADT^{\beta_2} \times PedEx^{\beta_3} \times e^{(\beta_4 X_4 + \cdots + \beta_n X_n)} \qquad (3)$$

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

In addition to the development of SPFs that provide estimates of pedestrian crash frequency, models of pedestrian crash risk can be developed that relate the probability that one or more crashes will be observed at a given location during a specific time period (often some number of years) to a set of explanatory variables. In the modeling process, the dependent variable takes a binary outcome where a value of “1” represents at least one crash observed at a given location during some 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:

$$\ln \frac{P}{1-P} = \beta X_i \quad (4)$$

where P is the probability that one or more crashes will be observed at location i during the analysis time period, β are 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 in the model; 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 in crash data (Cottrill and Thakuriah, 2010). Bayesian models were used to investigate the effect of spatial correlation of 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 large computational power, and leads to results that are not always directly interpretable. This makes it a challenge for practitioners to apply the outcomes. (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 as system-wide data are not used. Machine learning methods – such as regression trees – can be used to identify risk factors that are associated with increased pedestrian crash frequencies; however, they are more difficult to interpret, especial when quantifying the risk relationship or prioritizing between different risk factors.

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 models typically use as explanatory variables transportation system attributes, built environment characteristics and demographics and pedestrian safety outcomes. Transportation system-related variables generally include roadway features (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 percentage of population in certain ethnicity/racial groups, percentage of non-college educated/high school graduate population, percentage of population within certain age limit, etc.

NCHRP Report 893 provides a list of typical risk factors known to be associated with increased pedestrian crash frequency and increased pedestrian crash severity. The risk factors mentioned in the report for roadway segments, which are the focus of this project, are summarized in Table 1.

Table 1. 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*

Table 2 and Table 3 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/trends from these studies at both the individual segment and zone-levels.

Table 2. A summary of explanatory variables in segment/intersection-level based models

Study Model	Independent variables[1]
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 (-)
Logistic regression (Hamilton et al., 2021)	• Segment length (+) • Pedestrian volume (-) • Interaction between traffic volume and pedestrian exposure (+) • Lower speed limit (-) • Number of lanes (+/-)

Many of the risk factors have expected and unambiguous relationships. Pedestrian crash frequency generally increases with increased traffic volume, as more vehicle traffic relates to increased opportunity for crashes. Higher functional classification roadways (e.g., arterials) are also generally 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 are also associated with increased risk of pedestrian crashes. Higher vehicle speeds and speed limits, more large 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 are associated with decreased crash frequency. These features generally serve to help provide pedestrians with protection from vehicles and thus are associated with safety improvements.

Other features that have positive relationships as expected with increased pedestrian crash risk include: presence and number of transit stops near the location of interest, presence of on-street parking, alcohol sales near the location of interest, and 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 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. While it makes intuitive sense that more pedestrian activity would be associated with more pedestrian crashes at a given location, several studies have found that this might not necessarily be 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 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 large pedestrian volumes at a given location might be make pedestrian activity might contribute to a reduction in crash risk or frequency, perhaps due to increased visibility or driver expectations (Elvik and Bjørnskau, 2017; Hamilton et al., 2022; Jacobsen, 2003). The presence of pedestrian improvements also has ambiguous impacts on pedestrian crash frequency or risk; for example, (Pande and Abdel-Aty, 2009) found that sidewalk presence was actually associated with increased crash frequency even though sidewalks likely make pedestrian movement safer. However, some of these findings might suffer from correlation with exposure measures: locations with sidewalks tend to have higher pedestrian activity than those without and thus the increased crash frequency associated with sidewalk presence might reflect the increased pedestrian exposure. Thus, while sidewalks and other pedestrian infrastructure features may produce safety benefits, they might not always be strong predictors of crash risk (U.S. Department of Transportation Federal Highway

---

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

Administration, 2017). Care must be taken to control for as many of these related features to unveil the true underlying relationships/risk factors.

The review of the literature suggests that other features that can be 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 generally (often measured by speed limit) is associated with increased crash severity (Hamilton et al., 2021), less is known about the relationship between speed and crash frequency. (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 85^{th} percentile and median speed observed was associated with increased crash frequency, while the difference between median and 15^{th} 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 15^{th} percentile speed) might be associated with reduced pedestrian and bicycle crash frequency. This seems reasonable since vehicles generally travel slower in congestion and thus might be able to avoid conflicts with non-motorized roadway users. (Hamilton et al., 2021) found that pedestrian crash risk was lower at site with lower speed limits (25 to 35 mph) compared to those with higher speed limits. (Hamilton et al., 2021) 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 predictor that is generally lacking in pedestrian crash risk models.

Table 3. A summary of explanatory variables in zone-level based models

Study	Independent Variables[2]
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 (+)

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

Study	Independent Variables[2]
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[2]
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 (+)

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 road functional classifications, number of lanes, the transit ridership and subway stations, presence of four and five-way intersections and number of teenagers were associated with an increase in pedestrian crash frequency. Moreover, they found that proportion of people with age above 65 is correlated with higher fatal crash frequency. On the other hand, residential land use, all-way-stop and three-way intersections are correlated with lower pedestrian crash frequency. (Mansfield et al., 2018) found several neighborhood characteristics, demographic, socioeconomic, employment, traffic, and infrastructure, 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 increase in pedestrian crash frequency and higher posted speed limits, population and employment density, intersection density, dwelling unit density and percentage of household with less than two auto. On the other hand, they found a relationship between decrease in pedestrian crash frequency and median household income. (Cottrill and Thakuriah, 2010) found that transit availability and pedestrian accessibility is positively correlated with increase in pedestrian crash frequency. They also found that number of schools, crime, children pedestrian, percentage of people without car increases pedestrian crash frequency. (Ukkusuri et al., 2011; Wier et al., 2009) found that residential-neighborhood, commercial land use and higher number of school in an analysis zone result in higher number of pedestrian crashes. (Ukkusuri et al., 2011) also found that population, percentage of high school graduates and percentage of uneducated people is positive correlated with increase in pedestrian crashes.

While these studies provide some clear relationships between land use, demographics, socioeconomics and pedestrian crash risk, less clear are relationships between factors that might combine some of these features into a single metric. For example, little in the research literature was found on the relationships between aggregate metrics such as social health metrics or food desert indicators and pedestrian crash frequency/risk. This is an area that should be explored to better understand if these already at-risk areas also suffer from decreased pedestrian safety performance.

Models of pedestrian exposure

Pedestrian activity/exposure, measured either through direct counts, estimates of counts, or surrogate metrics (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 since pedestrian activity is often determined by the built environment, proximity of origins and destinations, and the demographic/ socioeconomic profile of a neighborhood. Traditionally, pedestrian volume data are not collected or available widely across a roadway network, like traffic volumes. Instead, pedestrian volumes tend to be collected sporadically either as a part of other traffic volume studies (i.e., collected when convenient to do so) or – if as a part of a pedestrian counting program – collected for a small subset of locations. To obtain pedestrian counts widely across a network at a scale that would provide sufficient coverage to either know or estimate pedestrian volumes would be extremely resource intensive; either due to high man-hours (manual counts) or expense (automated counts). Both manual and automatic counts of pedestrian are impractical due to large variability of pedestrian volume, shorter trip length and difficulty in detection (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: pedestrian push-button actuation can vary significantly and might represent just a small fraction of pedestrian users, especially when large numbers of pedestrians are present. Therefore, 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 are used to directly estimate the level of pedestrian activity at a given location as a function of a set of explanatory variables. The most common type of statistical modeling technique used to estimate direct demand models 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 determine the factors that give rise to higher number of bicycle and pedestrian traffic. However, a linear regression model can produce unrealistic parameter estimate including negative count value for pedestrian exposure.

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 with varying spatial scale (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, which means it can result in inclusion of variables that are counterintuitive or inconsistent with theory, complicating the interpretation and limiting the transferability. To address this issue, Hankey and Lindsey (2016) proposed two supervised approaches that are easier to interpret and simpler to apply. Hampshire et al. (2018) developed an origin-destination model that uses several factors that influence pedestrian travel demand, including the built environment and distance between origins and destinations.

Further, there have been modelling efforts to use short-term count data to predict long-term pedestrian counts, i.e., determine expansion factors. Griswold et al. (2019) developed a log-linear regression model to calculate annual volume estimates considering contributing factors 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 and the values were compared to characteristics from pedestrian demand models to see what characteristics have a greater/lesser impact on pedestrian volumes (WisDOT Pedestrian Exposure Model, 2021). A NB regression model was developed based on the key factors, including population density, job density, bus stops, retail businesses, food/drink businesses, schools, and households without a motor vehicle.

Direct demand models do not consider how individuals move throughout a transportation network and travel between/across individual links because the estimates are focused at individual links. To account for this, 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 on a single point at a specific time, they have adopted longitudinal evaluation which considers change in response with time. However, this method is complicated to apply for general planning purposes, such as in this work.

Table 4 and Table 5 summarize the variables in segment/intersection-level and zone-level pedestrian exposure models and the 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/trends from these studies at both the individual segment and zone-levels.

Table 4. A summary of variables in segment/intersection-level pedestrian exposure models

Model	Independent variables[3]
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 (+)

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

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 (-)

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 (-)

Table 5. A summary of variables in zone-level pedestrian exposure models

Model	Independent variables[4]
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, the direct exposure models suggest that for segments or intersections, population or population density is related to pedestrian activity since it shows up in most of the models as an independent variable and leads to higher pedestrian volumes. Additionally, land use is shown to be a contributor to pedestrian activity; however, some contradictory results are seen in the models. Urban residential areas are shown to generally decrease pedestrian activity while commercial areas are shown to increase pedestrian activity (except for Pulugurtha and Repaka (2008) which suggests the opposite finding). Employment, or employment density, is shown to have a positive impact on pedestrian activity (except for Liu and Griswold (2009), which finds that job density has a negative impact). Further, segments or intersections near universities or schools were found to have larger pedestrian activity as expected (Griswold et al., 2019; Miranda-Moreno and Fernandes, 2011; Schneider et al., 2012; Singleton et al., 2021). Most of the studies also find that transit related variables (e.g., number of transit stops or bus frequency) tend to increase 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,

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

2011; Pulugurtha and Repaka, 2008; Schneider et al., 2009; Singleton et al., 2021). Roads on a slope are found to have less 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 found to have more pedestrian activity. This perhaps reflects that in general these types of roads have more activity due to being near downtown areas. Further, most studies agree that minor arterial or collector roads have larger pedestrian activity as compared to principal or major arterials (Griswold et al., 2019; Hankey et al., 2012; Miranda-Moreno and Fernandes, 2011). Demographic information suggests that as the education level of the neighborhood residents increase, there is more pedestrian activity (Hankey et al., 2012; Lindsey et al., 2007, 2006), however the impact of income on pedestrian activity is inconclusive (Lindsey et al., 2007, 2006; Lu et al., 2018). Finally, the results suggest that, as expected, precipitation or cold weather reduces pedestrian activity (Hankey et al., 2012; Miranda-Moreno and Fernandes, 2011).

At the zone level, the results suggest that population density, employment density and commercial density all increase pedestrian activity (Behnam and Patel, 1977; Singleton et al., 2021). Household size shown to both increase (Singleton et al., 2021) and decrease (Hampshire et al., 2018) pedestrian activity. This result is likely due to the influence of other non-observed variables. Further, (Singleton et al., 2021) found that pedestrian activity increases with number of school, and transit stops, and intersection density (indicating shorter road segments). On the other hand, they found that pedestrian activity decreases with household income and vehicle ownership. The decrease in pedestrian activity with vehicle ownership appears to be counterintuitive and could indicate missing variables in this model.

State level practices for identifying pedestrian risk factors

This section describes existing practices that have been identified from various state agencies to identify and quantify pedestrian risk factors. Like the pedestrian frequency/risk and exposure models identified in the literature, these analyses have typically been done in two different contexts: (1) the geographic zone-level (e.g., a census tract) or (2) an individual road segment or intersection. Based on the type of the analysis, the rest of this section reviews the guidance for pedestrian safety analysis in different cities/states across the country.

Crash summary-based analysis

New York

The New York PSAP (New York State, 2016) mostly focuses on summary statistics and reviews of historical crash trends for identifying risk factors. The purpose of this plan is 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 critical information that are unavailable or 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 behavior 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 demographics of people hit, road types and features, individual behaviors etc. The plan outlined focus counties, cities, and corridors, but these were based on historic crash frequencies and not necessarily "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 highest number of pedestrian crashes.
  - One of the top ten counties with highest number of pedestrian injuries.
  - One of the top ten counties with 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.

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 a risk score for PAZs throughout the state. The risk score was determined using NB regression 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 planning-level model and considered as the number of daily walking trips that originate or terminate 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

Minnesota (Devoe, 2019) overlaid a hexagonal grid (0.5-mile diameter) across the State and scored each area based on a suite of criteria. The use of the grid potentially avoided updates to the state linear referencing system (LRS) 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 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:

$$SPACE_{Score} = 100 \times \frac{1}{19} \times \sum points \qquad (5)$$

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 United States, 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 these roadway segments. Star ratings are assigned to a roadway based on the design features, traffic control, and other characteristics of the roadway that can be observed by visual inspection of a picture or video of the roadway. The star ratings consider factors related to both crash likelihood and crash protection. Crash likelihood is impacted by road characteristics such as number of lanes, street lighting, intersection type, etc. 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 HSM.

North Carolina

The research team identified criteria that NCDOT has used for screening roadways as part of pedestrian safety studies being performed in North Carolina. Table 6 provides a summary of these screening features. Note that several of these criteria are limited by data availability, but these are factors that can also be used as a starting point for the development of systemic risk factors.

Table 6. 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

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 Level of Traffic Stress (LTS), accessibility, and network connectivity as priorities for investment. Level of traffic stress helps determine the suitability of roads and connections to accommodate bicyclists and pedestrians. For low stress pedestrian routes, WsDOT also uses an evaluation framework involving criteria such as safety, equity and demand, for assessing potential need of pedestrian crossing. 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. Screening results are also publicly available, along with data inputs: https://ana.gis.arcadis.com/apps/ODOTSTW/. 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, Seattle Department of Transportation (SDOT) conducted the first phase of the citywide Bicycle and Pedestrian Safety Analysis (Seattle Department of Transportation, 2016) 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 has also led the city to develop a count optimization effort to fill in specific gaps in pedestrian bicyclist exposure knowledge. In 2020, the second phase of the plan (Seattle Department of Transportation, 2020) included significant advances by adding signal phasing data in 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 include road speed limit, population, number of households etc.

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, they recommended four methods to identify and prioritize high-risk locations which would go through 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 in a geographic information systems (GIS) format, 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% of the screening criteria value for these corridors. VDOT also used the 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 low-cost countermeasure implementation.

Tennessee

Based on VDOT's work, as well as Federal and NCHRP guidance, TDOT developed separate indices for pedestrian safety at intersections and segments. Detailed explanations of each input are available in story maps published by TDOT at https://storymaps.arcgis.com/stories/9e5fc2e1c8a8487bbb1fdb585a18b4ed. 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 develop pedestrian access and safety needs and priorities based on 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 has limited applicability to areas with rural areas or agencies without comprehensive statewide data. The details can be found 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 certain 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 over half of all pedestrian crashes that occurred on the State Highway System between 2011-2015 occurred with a sidewalk or crosswalk 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 types of segment types and different levels of severity, and the models were based solely on AADT data. (Dolatsara, 2014) developed SPFs using NB regression to estimate pedestrian crashes using pedestrian exposure data for urban signalized intersections in Michigan. 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 include pedestrian and bicycle volumes, non-motorized facility inventory, non-motorized improvement projects, activity locations, socioeconomic and demographic data, crime rates, land use data and traffic volume data. Different models were developed for city level, census tract level, and corridor level. (Mcarthur et al., 2014) investigated pedestrian involving a child aged 5 to 14 located within one mile of a school that included students from kindergarten to eight grade in Michigan. In addition to the crash data, demographic and socioeconomic factors were obtained from the US 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 most potential for reducing pedestrian crashes (Bergh et al., 2015). Risk factors were identified first and then used to prioritize location. Pedestrian volumes were not considered in the risk-based method due to the lack of consistent statewide data. There was no quantitative data associating the identified risk factors and 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 to identify and prioritize locations with increased or elevated risk for pedestrian crashes with the objective to develop a risk-scoring method with weights derived from a data analysis, as compared to 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 socio-economic characteristics pertinent to travel by alternative modes. The data are amalgamated 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 for prioritization for infrastructure improvements through spatial or attribute driven analyses.

<u>Summary of state practices</u>

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 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 of pedestrian exposure.

Most states are not able to directly account for pedestrian exposure; however, surrogates are used to either estimate this exposure to directly apply within the model or indirectly account for this exposure. Features that serve as surrogates 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 generally provide detailed information on the specific methods used to identify individual risk factors or the relative weights assigned to each. 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.

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 summaries the reviewed data sources for both collision records and the contributing factors. Some of the data sources are maintained by national and state level agencies, while some are owned by private sector organizations.

<u>Data collected by public agencies</u>

Roadway data

A critical factor observed in both pedestrian crash frequency/risk and pedestrian exposure models are roadway features. These data include items like roadway functional classification, geometric characteristics (e.g., cross-sectional information and horizontal curvature), vehicular traffic volumes (typically provided in average annual daily traffic or AADT), and presence of pedestrian infrastructure features. More state and local agencies contained detailed roadway inventory databases that provide this information. Crash information are generally linked to specific roadway locations via a referencing system and in this way these features can 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 ongoing. ArcGIS is required to download, analyze and manipulate the data.
- NCDOT publishes a spatial file of State-owned and operated traffic signals and flashers throughout the State. This does not represent a comprehensive list of traffic signals, as it would exclude signals maintained by municipal or other agencies (https://ncdot.maps.arcgis.com/home/item.html?id=cd1fe92936ec44f8a3dbc002be2f68a3).
- Data are also available from the NCDOT Pavement Unit, which might contain more accurate values for some critical roadway features (e.g., number of lanes).

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., record of crashes that occur on state roads, including location, vehicles, and people involved, and available injury outcomes). These reports typically have flags available that identify crashes that involve 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 Fatal Accident Reporting System (FARS) (https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars) is a database providing public yearly data regarding injuries suffered in motor vehicle traffic 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:

- North Carolina Pedestrian and Bicycle Crash Data Tool (https://www.pedbikeinfo.org/pbcat_nc/_pedquery.cfm) is an interactive database of all pedestrian-motor vehicle crashes reported to the NC Division of Motor Vehicles. The user can query crash data on statewide-, region-, county- and city-level. It provides the roadway characteristics, road condition, road configuration, speed limit etc. where the crash happens.
- 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 conveniently when other counts are needed (e.g., collected only when volumes are needed for a specific project). However, pedestrian volumes or exposure is 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. Manual counts are thus 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 Minnesota DOT (2015).
- Slab sensors: These are sensors embedded in a sidewalk or pavement that use acoustic sensors to capture pedestrian movement.
- Camera technology: Video sensors 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 identify the count of pedestrians.

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

Socioeconomic data

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

- The U.S. Census databases have 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 demographics survey program conducted by the U.S. Census Bureau. The ACS releases social characteristics, economic characteristics, housing characteristic and demographic and housing estimates data every year.
- 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 Health and Human Services (NC DHHS) publishes a suite of maps that profile socioeconomic and access 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:

$$Z - Score = \frac{Tract\ Value - State\ Value}{State\ Standard\ Deviation} \tag{6}$$

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) was 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.
- 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.

<u>Non-traditional and alternative data</u>

Smartphone-technique-based exposure data

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

- Cellular carriers collect location data points by time based on the signaling of mobile phones and cell towers. From the positioning data, movements of users can be extracted. Secondary data vendors (e.g., Airsage) purchase the 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 released by StreetLight recently. 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 (sociodemographics 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 upon 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.
- Volunteered geographic information (VGI) platforms enable community members to report localized knowledge and experiences. For example, OpenStreeMap (OSM) is one of the most popular VGI platforms. Over 1 million individuals have contributed to a set of geographic data that include roads, cycle paths, and trails used.

Probe speed data

Probe data is 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, such as HERE, can be accessed through the Regional Integrated Transportation Information System (RITIS). This is housed and managed by the University of Maryland's Center for Advanced Transportation Technology (CATT) Laboratory. HERE collects vehicle speeds using multiple real-time sources, including global positioning systems, probe vehicles, and cell phones. The speed data can be accessed and downloaded from the RITIS platform using RITIS' *Massive Data Downloader* of archived data.

- Inrix combines probe data from commercial GPS, DOT sensors and other proprietary data sources, and it provides speed and travel time on Traffic Message Channels which are defined segments of road.
- Similarly to Inrix, TomTom provides speed and travel time on road segments. However, their data is collected from the Vodafone mobile phone network, governments and traffic control centers.
- NAVTEQ includes both point and route-based data, and besides speed and travel time, volume from own sensors are also available.
- AirSage utilizes wireless signaling data and cell phone GPS to collect data, and travel mode is available in this data source.
- TrafficCast leverages GPS tracking data, public sensors, accidents reports, road works and weather reports. Speed and travel time on Traffic Message Channel-level are available.

Key takeaways

The literature review reveals 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 crash occurring at a given location during some time period – and NB regression – which is used to estimate the number of pedestrian crashes that occur at a given location during some time period. These methods are particularly useful because they provide an interpretable relationship between model coefficient and change in crash risk/frequency. Both of these methods were considered to develop risk factors in this project, with NB regression ultimately being selected due to dependent variable being crash frequency (as opposed to the probability of a crash occurring on a segment).
- Common pedestrian risk factors that were identified in the literature (both academic literature and available state practices) and were 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 safety risk. However, this exposure is generally difficult to obtain for 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
    - Proximity to schools
    - Presence of transit stops
    - Land use mix
    - Presence of schools
    - 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 amount 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 information might lead to counterintuitive findings related to risk; e.g., sidewalks might be associated with increased pedestrian crash risk since the presence of sidewalk typically indicates higher pedestrian activity. Thus, the research team considered various means to account for pedestrian exposure. This included inclusion of pedestrian exposure surrogates, as well as the development of pedestrian exposure models to predict pedestrian activity at locations where sufficient information was available.

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 a part of this process and that were available for use in this project.

Data summary

Table 7 provides a summary of the relevant data items considered for this project, along with their source and if they were ultimately collected/used in this project. All data were collected using a GIS format so that individual data elements could be spatially joined for analysis purposes. The remainder of this section provides additional details on these data, including why some specific data elements were not collected.

Table 7. Summary of data collection status

Data category	Data element	Collected? (Y/N)	Source (if collected)	Version
Crash	Pedestrian crashes	Y	NCDOT	2007-2020 (2020 data are interim)
Roadway data	Functional classification	Y	NCDOT	2021 Q4
	Number of lanes	Y	NCDOT	2021 Q4
	Median presence	Y	NCDOT	2021 Q4
	Presence of traffic signals	Y	NCDOT	2021
	Traffic volume	Y	NCDOT	2019
	Posted speed limit	Y	NCDOT	2021 Q4
Pedestrian infrastructure	Sidewalk presence	Y	NCDOT	2021
	Greenway presence	Y	NCDOT	2021
	Crosswalk presence	Y	NCDOT	2021
Pedestrian exposure and exposure surrogates	Pedestrian volumes	Y	NCDOT; City of Charlotte	Various
	Transit stop presence	N	n/a	n/a
	Transit routes	Y	NCDOT	2019
	Parks nearby	Y	NCDOT	2019
	K-12 school nearby	Y	NCDOT	2019
	University campus	Y	NCDOT	2019
	Micromobility usage	N	n/a	n/a
Sociodemographic and socioeconomic data	Population by age, sex, and race	Y	US Census Bureau	2019 ACS 5-Year estimates – B01001; B02001
	Education level	Y	US Census Bureau	2019 ACS 5-Year estimates – B15001
	School enrollment	Y	US Census Bureau	2019 ACS 5-Year estimates – B14002
	Unemployment	Y	US Census Bureau	2019 ACS 5-Year estimates – B23001
	Median income	Y	US Census Bureau	2019 ACS 5-Year estimates – B19013
	Vehicle ownership	Y	US Census Bureau	2019 ACS 5-Year estimates – B25044
	Poverty rate	Y	US Census Bureau	2019 ACS 5-Year estimates – C17002
	English proficiency	Y	US Census Bureau	2019 ACS 5-Year estimates – C16002

Data category	Data element	Collected? (Y/N)	Source (if collected)	Version
	Commute mode	Y	US Census Bureau	2019 ACS 5-Year estimates – B08301
	Disability	Y	US Census Bureau	2019 ACS 5-Year estimates – B18101
	Employment by industry	Y	US Census Bureau	Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) v7 – Workplace Area Characteristics 2019
	Social health determinants	N	n/a	n/a
Land use	Parcel boundaries	Y	NC One Map	September 2021
	Alcohol sales locations	Y	Data Axle Reference Solutions	U.S. Business Database – November 2021
	Business locations by industry	N	n/a	n/a
Vehicle speed	Vehicle probe speed	N	n/a	n/a

Pedestrian crashes

Pedestrian crash information was obtained directly from NCDOT via ArcGIS Online. The research team identified all pedestrian crashes that occurred between years 2007 and 2020, along with information on the exact crash location, injury severity level, action, and relative location of each unit in the crash. The locations were geocoded and crashes then assigned to specific roadway segments. To ensure that pedestrian risk factors represent the most current conditions, only crashes from 2015-2020 (inclusive) were eventually considered in this analysis.

Roadway characteristics

Roadway characteristic information 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. Road characteristic data were primarily available for non-local roads (i.e., minor collector and above), although centerline and route information were available for all public roads. Specific data elements associated with each roadway segment included the functional classification, number of lanes, median presence (and type), and posted speed limit. Traffic signals and flashers owned by NCDOT were obtained from ArcGIS Online. Traffic volumes for non-local roadway segments were also obtained from NCDOT and merged with the individual roadway segments.

Pedestrian infrastructure

Information on pedestrian infrastructure was obtained from NCDOT's PBIN via ArcGIS Online. Specific data elements included presence of sidewalks, greenways, and crosswalks. However, while collected and incorporated into the analysis database, the research team and NCDOT technical panel eventually concluded that coverage of some data elements – such as crosswalks – were not reliable or sufficient for inclusion in this work.

Pedestrian exposure and surrogates for exposure

Several data elements were obtained as surrogates for pedestrian exposure. The research team obtained locations of parks, K-12 schools, and university campuses from NCDOT via the Advanced Transportation through Linkages, Automation and Screening (ATLAS) database. In addition, the research team obtained information on transit routes, since transit usage is typically associated with pedestrian activity; transit stop information were not available uniformly across the State. The research team tried to obtain information on micromobility usage (e.g., use of shared bikes, e-bikes and e-scooters), but these data were not available in a useful format. Thus, micromobility data were not obtained for this project.

In addition, the research team obtained actual pedestrian count information from 3,579 existing pedestrian counts provided by NCDOT and other local agencies within North Carolina. The project team obtained pedestrian volume counts from multiple sources:

1. Project-level counts collected as part of motor vehicle turning movement counts (TMCs), segment counts, zone counts, or other analyses. These were obtained as a convenience sample and thus do not necessarily reflect a randomized sample of intersections across North Carolina. Nevertheless, they provide an indication of the level of pedestrian activity that may be expected at various locations. These represent 1,993 counts or about 56 percent of the sample.
2. Counts obtained from the City of Charlotte as part of their TMC count program. These counts are similar to NCDOT's count program (i.e., as a component part of motor vehicle counts). The project team obtained these counts as part of a Federal Highway Administration (FHWA) research project and included counts for years between 2011 and 2020. These represent 496 counts or about 14 percent of the sample.
3. Counts obtained as part of a downtown Raleigh pedestrian safety study conducted by NCDOT. Although these counts were specifically collected for the purposes of pedestrian safety analysis, locations were not randomly selected. These represent 19 counts or less than 1 percent of the sample.

4. Counts provided by Greensboro DOT/Greenville Urban Area Metropolitan Planning Organization. These represent 539 counts or about 15 percent of the sample.
5. Counts provided by the Gaston-Cleveland-Lincoln Metropolitan Planning Organization. These represent 387 counts or 11 percent of the sample.
6. Counts provided by the City of Durham. These represent 184 counts or 5 percent of the sample.

Sociodemographic and socioeconomic data

The research team obtained sociodemographic and socioeconomic data from the US Census Bureau's ACS to serve as additional surrogates for 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. Employment data were also obtained from the Census's LEHD program. This provides estimates of employment by industry and place of work at the Census block level, and these data can be aggregated to larger Census geographies. The research team originally intended to obtain information on social health determinants for use in this project. However, these data were not included since they were generally older and overlapped with available Census data with very few exceptions.

Land use

Although there are no single sources of land use for the entire State of North Carolina and county tax assessors vary on the public availability of these data, parcel boundaries were obtained from the NC One Map open data portal. Parcel size and density are surrogates for urbanism and land use intensity and could be an indicator of pedestrian exposure and trip generation. The project team obtained a list of business locations that are likely to sell alcohol based on the industry associated with that business from Data Axle (i.e., North American Industry Classification System code or 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). Although the project team intended to obtain business locations for a broader range of potential pedestrian generating businesses, the size and limited accessibility to these data made this impractical. However, total employment by NAICS sector within Census geographies served as a surrogate for pedestrian trip generating economic activity, including:

- Retail Trade (44-45)

- Educational Services (61)
- Arts, Entertainment, and Recreation (71)
- Accommodation and Food Services (72)

Vehicle speed

The research team originally intended to obtain information on actual vehicle speeds on the roadway network using mobile probe or connected vehicle data. However, upon consultation with the NCDOT technical panel, these data were not included in the project due to complexity. Specifically, these data contain average vehicle speeds on individual roadway segments at a short temporal scale (every 5-minutes or 1-hour) and would need to be summarized into annual measures for use in this project. Likely, such information would not be readily available to NCDOT in an efficient manner and thus it was decided to exclude this information in the risk factor development for this project.

Segment-level pedestrian crash risk database

One analysis database was developed to estimate models that could quantify the impact of various factors on the risk of pedestrian crashes occurring along individual roadway 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 thus 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 one the majority of the segment falls within. Appendix A provides a data dictionary that contains a list of all specific data elements along with a short description and how each was coded.

Summary statistics that provide the geographic and functional classification distribution of the roadway segments are provided in Table 8 to Table 11. Note that some Route and Functional Classifications will likely be excluded from the analysis, including Interstates, Rest areas and Projected roadways. Furthermore, Table 11 reveals that most segments (approximately 80 percent) do not have AADT information. This is likely due to functional classification differences and how NCDOT performs routine traffic counts. Also, most segments did not have pedestrian crashes observed during the last ten years.

Table 8. Summary of risk data by NCDOT division

Division	Total Segments (#)	Relative Frequency by Segment (%)	Total Mileage (mi)	Relative Frequency by Mileage (%)
0	16	0	0.728	0
1	25,800	3.4	7,407.947	5.3
2	41,683	5.5	8,064.218	5.8
3	56,545	7.4	10,117.98	7.2
4	40,305	5.3	8,690.076	6.2
5	86,322	11.3	12,380.81	8.8
6	52,107	6.8	10,105.97	7.2
7	56,880	7.5	9,441.48	6.7
8	45,633	6	10,436.67	7.5
9	64,536	8.5	9,659.808	6.9
10	75,525	9.9	10,754.31	7.7
11	44,413	5.8	10,260.49	7.3
12	59,178	7.8	9,930.297	7.1
13	53,262	7	10,550.93	7.5
14	60,893	8	12,122.57	8.7
Total	**763,098**	**100**	**139,924.3**	**100**

Table 9. Summary of risk data by functional classification

Federal Functional Class	Total Segments (#)	Relative Frequency by Segment (%)	Total Mileage (mi)	Relative Frequency by Mileage (%)
Principal arterial – Other freeways and expressways	351	0.0	175	0.1
Principal arterials – Others	16,867	2.2	3,440	2.5
Minor arterial	33,340	4.4	6,265	4.5
Major collector	45,377	6.0	10,817	7.8
Minor Collector	19,162	2.5	6,488	4.7
Local	639,903	83.9	111,564	80.0
NA	7,306	1.0	774	0.6
Total	**762,305**	**100.0**	**139,523**	**100**

Table 10. Summary of risk data by route classification

Route Class	Total Segments (#)	Relative Frequency by Segment (%)	Total Mileage (mi)	Relative Frequency by Mileage (%)
US route	20,831	2.7	4,894	3.5
NC route	29,268	3.8	8,098	5.8
Secondary route	239,890	31.5	64,987	46.6
Non-system	469,609	61.6	61,201	43.9
Other state agency route	2,465	0.3	276	0.2
Federal route	40	0.0	40	0.0
Rest areas	186	0.0	25	0.0
NA	16	0.0	1	0.0
Total	**762,305**	**100**	**139,523**	**100**

Table 11. Summary statistics for risk data

Variable	Year	Mean	Std. Dev.	Min.	Max.	% non-zero or na
AADT	2015	5,813	7,392	10	107,000	19.96
AADT	2016	5,895	7,571	10	107,000	20.04
AADT	2017	6,016	7,620	10	110,000	20.08
AADT	2018	6,083	7,703	10	112,000	20.02
AADT	2019	6,137	7,738	10	113,000	20.01
AADT	2020	5,313	6,675	10	97,000	19.84
Pedestrian KABCO crashes	2020	0.001996	0.046273	0	4	0.19
Pedestrian KA crashes	2020	0.000567	0.024033	0	2	0.06
Pedestrian KABCO crashes	2019	0.002852	0.056008	0	3	0.27
Pedestrian KA crashes	2019	0.000641	0.025461	0	2	0.06
Pedestrian KABCO crashes	2018	0.003065	0.058459	0	4	0.29
Pedestrian KA crashes	2018	0.000575	0.024465	0	2	0.06
Pedestrian KABCO crashes	2017	0.002866	0.056043	0	5	0.27
Pedestrian KA crashes	2017	0.000533	0.023314	0	2	0.05
Pedestrian KABCO crashes	2016	0.002901	0.056472	0	4	0.28
Pedestrian KA crashes	2016	0.000448	0.02135	0	2	0.04
Pedestrian KABCO crashes	2015	0.002814	0.056092	0	5	0.27
Pedestrian KA crashes	2015	0.000418	0.020506	0	2	0.04
Pedestrian KABCO crashes	2014	0.002768	0.054807	0	4	0.27
Pedestrian KA crashes	2014	0.000396	0.020216	0	2	0.04
Pedestrian KABCO crashes	2013	0.002537	0.051844	0	3	0.25
Pedestrian KA crashes	2013	0.000353	0.01898	0	2	0.03
Pedestrian KABCO crashes	2012	0.002798	0.054697	0	3	0.27
Pedestrian KA crashes	2012	0.000459	0.021594	0	2	0.05
Pedestrian KABCO crashes	2011	0.002347	0.050353	0	3	0.23
Pedestrian KA crashes	2011	0.000341	0.018526	0	2	0.03
Pedestrian KABCO crashes	All	0.026943	0.218272	0	16	2.13
Pedestrian KA crashes	All	0.004731	0.07354	0	6	0.45

Intersection-level pedestrian exposure database

A second database was developed to predict the level of pedestrian exposure at locations where pedestrian counts were available. The majority (3,177 out of 3,579) of these counts were performed at intersection locations. Thus, 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). Summary statistics for key variables are provided in Table 12 to Table 15.

Table 12. Summary statistics for pedestrian counts and associated traffic volumes

Measure	Mean	Std. dev.	Minimum	Maximum	% Non-zero values
Pedestrian count	120.0871	558.7064	0	14,854	82
NCDOT estimate AADT[5]	19,546.14	14,025.93	0	104,000	55
Min. AADT of approaches[6]	7,395.976	6,397.200	10	51,000	48
Max. AADT of approaches[2]	18,398.176	11,081.002	400	77,000	93

Table 13. Summary of exposure data by count year

Year	Frequency	Relative Frequency
NA	6	0.3%
2011	55	2.8%
2012	101	4.4%
2013	157	7.6%
2014	273	6.3%
2015	225	4.6%
2016	163	1.9%
2017	69	15.9%
2018	569	31.9%
2019	1,140	4.0%
2020	142	19.0%
2021	679	1.5%
Total	**3,579**	**100%**

[5] Only applicable for counts and locations provided by NCDOT directly.

[6] Calculated based on spatial proximity in GIS using traffic count data provided by NCDOT (2015-2020).

Table 14. Summary of exposure data by duration

Duration (Hours)	Frequency	Relative Frequency
2.5	1	0.03%
12	540	15.09%
13	2,508	70.08%
16	129	3.60%
24	388	10.84%
48	13	0.36%
Total	**3,579**	**100%**

Table 15. Summary of exposure data by NCDOT division

Division	Frequency	Relative Frequency
1	54	1.5%
2	103	2.9%
3	254	7.1%
4	114	3.2%
5	362	10.1%
6	228	6.4%
7	815	22.8%
8	72	2.0%
9	45	1.3%
10	684	19.1%
11	60	1.7%
12	574	16.0%
13	131	3.7%
14	82	2.3%
Total	**3,578**	**100%**

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 that intersection were identified using GIS and their characteristics 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 presence within a 0.5-mile radius was also identified and included; note the larger radius for university presence was driven by 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 B 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.

PEDESTRIAN CRASH RISK FACTOR DEVELOPMENT

This section summarizes the model development process that was used to obtain pedestrian crash risk factors in North Carolina. The first subsection describes the scope of the risk factor development. This is followed by a description of the statistical methodology used in this project. Next, the risk factor estimates are provided for models that do not include direct pedestrian exposure estimates (but includes surrogates for pedestrian activity). Note that risk factor estimates were developed using models that include direct pedestrian exposure estimates; however, these results and the associated pedestrian exposure model, are included in an appendix to this report.

Scope

The first step in the risk factor quantification process was to determine the scope of the model development. As a part of this, the research team focused on two key aspects:

1. Roadway segment classification; and,
2. Spatial coverage (e.g., urban vs. rural).

The first aspect was necessary to determine how roadway segments would be categorized for risk factor development. Two classification schemes were readily available: NCDOT route class and functional classification. NCDOT route class included the following categories relevant to this project:

- US Route
- NC Route
- Secondary Route
- Non-System Route
- Other State Agency Route
- Federal Route

The following functional classifications were relevant to this project:

- Principal Arterial – Others
- Minor Arterial
- Major Collector
- Minor Collector

- Local

The second aspect, spatial coverage, was necessary to determine whether risk factors would be developed for (and thus could be applied to) all roadway segments (urban and rural) within North Carolina or only for roadway segments in urban areas.

The research team assessed the distribution of the number and mileage of roadway segments, the number of segments with average annual daily traffic (AADT) data available, as well as observed crash frequencies over 2015-2020 (inclusive), to help answer these questions. These distributions are provided in Table 16 to Table 19. Table 16 and Table 17 summarize the number of roadway segments, mileage, and pedestrian crash frequency when roadway segments are categorized by NCDOT route class for all roadways and for only urban roadways, respectively. Table 18 and Table 19 summarize the same information when roadway segments are divided by functional classification for all roadways and for only urban roadways, respectively.

Notice in Table 16 and Table 17 that only two NCDOT route class categories—US Route and NC Route—have traffic volume information available for a majority (>96%) of the roadway segments. The remaining categories only have traffic volume information available for a small subset of segments (between 1% to 39%). Thus, if NCDOT route classes were used to classify roadway segments for modeling, traffic volume could only be included as a potential risk factor for the first two categories (US Route and NC Route) and would not be available as a potential risk factor in the other categories (Secondary Route, Non-System Route, Other State Agency Route, Federal Route). This is not ideal since the Task 1 Literature Review shows vehicular traffic volume is a significant risk factor for pedestrian crash risk.

By contrast, when roadway segments are divided by functional classification, traffic volumes are available for most roadway segments (>87%) within each category. The lone exception is Local roadways, for which traffic volumes are only available for 2-5% of roadway segments. Thus, if functional classification is used to divide roadway segments for modeling, traffic volume could be included as a potential risk factor for most categories (Principal Arterials – Others, Minor Arterials, Major Collectors, and Minor Collectors) and would not be available as a potential risk factor for just one category (Local roads). For this reason, and with approval of the NCDOT technical panel, the research team decided to develop risk factors where roadway segments were divided using the functional classification categories.

Table 16. Distribution of roadway segments and crash frequencies by NCDOT route class (all segments)

Route class	Total number of segments	Segments with AADT in all years (2015-2020)	Total mileage	Number of total crashes (2015-2020)	Number of KA crashes (2015-2020)
US Route	20,575	19,968	4,695.92	1,857	532
NC Route	29,079	28,358	7,835.53	1,581	462
Secondary Route	239,454	91,362	27,720.45	3,069	691
Non-System Route	461,548	8,265	914.59	1,313	147
Other State Agency Route	2,465	2	1.69	0	0
Federal Route	39	13	15.87	0	0
Total	**753,160**	**147,968**	**41,184**	**7,820**	**1,832**

Table 17. Distribution of roadway segments and crash frequencies by NCDOT route class (urban segments only)

Route class	Total number of segments	Segments with AADT in all years (2015-2020)	Total mileage	Number of total crashes (2015-2020)	Number of KA crashes (2015-2020)
US Route	10,450	10,081	1,450.87	1,521	378
NC Route	10,281	10,029	1,423.00	1,146	278
Secondary Route	93,244	36,406	5,274.18	2,446	468
Non-System Route	295,249	8,096	829.97	1,310	146
Other State Agency Route	49	0	0.00	0	0
Federal Route	26	2	1.68	0	0
Total	**409,299**	**64,614**	**8,980**	**6,423**	**1,270**

Table 18. Distribution of roadway segments and crash frequencies by functional classification (all segments)

Functional classification	Total number of segments	Segments with AADT in all years (2015-2020)	Total mileage	Number of total crashes (2015-2020)	Number of KA crashes (2015-2020)
Principal Arterials – Others	16,674	16,030	3,268.42	2,508	647
Minor Arterial	33,205	32,105	6,021.52	2,854	594
Major Collector	45,225	44,027	10,585.98	1,711	367
Minor Collector	19,132	18,560	6,391.75	273	98
Local*	638,574	36,928	14,751.99 (111,272.60)	437 (4,783)	111 (635)
Total	752,810	147,650	41,020 (137,540)	7,783 (12,129)	1,817 (2,341)

** For local roads, values in parentheses represent those including all segments (not just segments with traffic volume information available)*

Table 19. Distribution of roadway segments and crash frequencies by functional classification (urban segments only)

Functional classification	Total number of segments	Segments with AADT in all years (2015-2020)	Total mileage	Number of total crashes (2015-2020)	Number of KA crashes (2015-2020)
Principal Arterials – Others	11,580	11,111	1,625.25	2,302	546
Minor Arterial	22,958	22,094	2,855.14	2,525	461
Major Collector	20,787	20,033	2,621.81	1,284	200
Minor Collector	2,715	2,376	330.85	63	16
Local*	351,030	8,782	1,479.60 (33,630.35)	216 (4,110)	35 (432)
Total	409,070	64,398	8,913 (41,063)	6,390 (10,284)	1,258 (1,655)

** For local roads, values in parentheses represent those including all segments (not just segments with traffic volume information available)*

With respect to spatial scale, a comparison of Table 18 and Table 19 reveals that the majority of pedestrian crashes during the analysis period (84.7% of all crashes and 70.7% of KA crashes) occurred on urban roadway segments, even though urban segments account for less than 30% of the total roadway mileage within North Carolina. Since the total number of pedestrian crashes observed during the analysis period is relatively low, focusing only on urban roadway segments would result in more reliable crash frequency models and risk factor estimates. Discussions with the NCDOT technical panel also suggested that the risk factors would have their highest application in urban areas. For these reasons, and with approval of the NCDOT technical panel, the research team decided to develop risk factors only for urban roadway segments as a part of this project.

Statistical modeling methodology

Two unique sets of statistical models were developed in this project: 1) crash frequency models used to quantify the impacts of individual factors of pedestrian crash risk at individual roadway segments; and, 2) pedestrian count models used to estimate pedestrian exposure at individual intersections. Both sets of models were estimated using NB regression, since this is the most common and appropriate modeling type as identified in the literature review. NB regression is a count regression technique that is 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 7 shows the general form of the crash frequency models that were estimated for roadway segments to obtain individual risk factors:

$$
\begin{aligned}
N_{i,risk} &= AADT^{\beta_{AADT}} \times L \times e^{\beta_0} \times e^{\sum x_{ij}\beta_j} \\
&= AADT^{\beta_{AADT}} \times L \times e^{\beta_0} \times e^{x_{i1}\beta_1} \times e^{x_{i2}\beta_2} \times \ldots \times e^{x_{iJ}\beta_J} \qquad (7)
\end{aligned}
$$

where $N_{i,risk}$ = predicted 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, K12 school enrollment density), and socioeconomic characteristics associated with the location (e.g., median income, proportion of commuters that are non-motorized, proportion of disabled population).

Please note that the form shown in Equation 7 and estimated in this project specifically treats segment length (L) as a proportional constant associated with predicted crash frequency; thus, the resulting models can be used to compute the expected crash frequency per mile by dividing the output of Equation 7 by L. Also note that traffic volume was included in most models developed. However, as previously mentioned, sufficient traffic volume information is not available for roadway segments classified as Local roads; thus, the traffic volume term in Equation 7 is not included in models for Local roads and thus not considered as a risk factor for Local roads.

The elasticity of each independent variable included in a NB model provides a measure of 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:

$$E_{X_{ijk}}^{\lambda_{ij}} = \frac{\partial\lambda_{ij}/\lambda_{ij}}{\partial X_{ijk}/X_{ijk}} = \frac{\partial\lambda_{ij}}{\partial X_{ijk}} \times \frac{X_{ijk}}{\lambda_{ij}} \qquad (8)$$

Equation 8 reduces to the following expressions for the log-log (Equation 9) and log-linear (Equation 10) 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 a 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.

$$E_{X_{ijk}}^{\lambda_{ij}} = \beta_k \qquad (9)$$

$$E_{X_{ijk}}^{\lambda_{ij}} = \beta_k X_{ijk} \qquad (10)$$

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:

$$E_{X_{ijk}}^{\lambda_{ij}} = \exp(\beta_k) - 1 \qquad (11)$$

Finally, Equation 12 shows the specific form of the pedestrian count models that were estimated to predict pedestrian activity at a given intersection:

$$N_{i,count} = Max\ AADT^{\beta_{Max\ AADT}} \times e^{\beta_0} \times e^{\Sigma x_{ij}\beta_j}$$
$$= Max\ AADT^{\beta_{Max\ AADT}} \times e^{\beta_0} \times e^{x_{i1}\beta_1} \times e^{x_{i2}\beta_2} \times \dots \times e^{x_{iJ}\beta_J} \quad (12)$$

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; and, x_{ij} that describe roadway segment i .

Pedestrian risk factor estimation

Two sets of pedestrian crash frequency models were developed as a part of this project to obtain pedestrian risk factors: 1) models that do not include direct exposure estimates for pedestrian activity; and, 2) models that include direct exposure estimates for pedestrian activity. The former uses only surrogates for pedestrian exposure activity (e.g., land use or socioeconomic characteristics), while the latter includes a direct estimate of pedestrian activity obtained as a function of these surrogates.

This section describes the models and risk factors obtained using the former approach, since the latter was deemed to be less reliable. However, the results of the pedestrian exposure model and risk factors developed using these exposure estimates are included in Appendix C. In both cases, unique models were developed (and thus unique risk factors were available) for the following roadway functional classifications:

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

Models were initially developed for Major Collectors and Minor Collectors individually; however, the models for Minor Collectors were not deemed reliable, likely due to relatively small number of crashes observed on roadway segments of this functional classification.

Within all functional classification groups, individual models were developed considering both total pedestrian crash frequency and only KA pedestrian crash frequency. Table 20 provides a summary of the crash frequency model developed for roadway segments categorized as Principal

Arterials – Other. Models were developed using both total pedestrian crash frequency and KA pedestrian crash frequency as the dependent variable and both models are summarized in the table. The coefficient estimate for a given variable provides the relationship between that variable and pedestrian crash frequency: values greater than 0 (denoted by red or darker shades) represent factors associated with increased pedestrian crash risk, while values less than 0 (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)
- Roadway segments with five or more lanes
- Roadway segments with higher speed limits (for KA crash frequency only)
- Presence of high-intensity (land use cover) development within 100 ft
- Density of alcohol sales establishments
- Presence of a bus route along the segment
- Population density
- K12 school enrollment density
- Proportion of commuters that are non-motorized
- Proportion of the population that is disabled

Factors associated with reduced risk include:

- Presence of a median
- Roadway segments with longer block lengths
- Median household income

It should be noted that models for total crash frequency exhibited a negative relationship between crash frequency and higher speed limits (i.e., higher speed limits were associated with reduced total crash risk). While this might be reasonable and somewhat expected due to pedestrians generally using high speed limit roads less, as well as lower statutory speed limits in more urban and city center environments, speed limits were removed from all total crash frequency models due to this observed relationship.

The models also include indicator variables associated with the NCDOT engineering divisions. These indicators are 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 focus more on the location-specific features.

The p-values associated with each coefficient are used to assess the statistical significance of the variable 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% confidence level. Note that most of the risk factors are statistically significant to the 95% confidence level. Those that are not (e.g., block length between 0.1-0.25 mi or proportion of the population disabled in the KA model) are still included since the coefficient estimate is in line with expectation, similar to the estimate in the total crash frequency model, and improves the overall model fit. P-values for the division indicators generally indicate that these are not statistically significant, which suggests differences in pedestrian crash risk across the engineering divisions are very small. However, they were retained in the model as their inclusion increase the model fit and would improve the use of the model in identifying high-risk locations.

Table 20. Summary of crash frequency models developed for Principal Arterials – Other

	Total crash frequency		KA crash frequency	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Constant	*-6.269E+00*	*0.000*	*-9.338E+00*	*0.000*
Natural log of AADT	5.597E-01	0.000	7.596E-01	0.000
5+ lane roadway	4.468E-01	0.000	3.503E-01	0.007
Speed limit 40 or 45 mph	---	---	2.436E-01	0.033
Speed limit 50 mph or above	---	---	2.614E-01	0.105
Median present	-3.086E-01	0.000	-1.456E-01	0.203
Block length between 0.1-0.25 mi	-4.380E-01	0.000	-4.042E-01	0.001
Block length between 0.25-0.5 mi	-6.262E-01	0.000	-4.617E-01	0.001
Block length greater than 0.5 mi	-1.155E+00	0.000	-1.134E+00	0.000
High intensity development within 100 ft	5.102E-01	0.000	5.186E-01	0.000
Alcohol sales density	1.236E-02	0.000	8.729E-03	0.000
Bus route present	5.236E-01	0.000	4.408E-01	0.000
Population density	1.152E-04	0.000	---	---
K12 enrollment density	4.705E-04	0.007	8.149E-04	0.000
Median household income	-8.185E-06	0.000	-5.666E-06	0.025
Proportion of commuters non-motorized	1.007E+00	0.041	---	---
Proportion of population disabled	2.412E+00	0.000	1.703E+00	0.127
NCDOT Division 2	*2.155E-01*	*0.402*	*2.457E-02*	*0.951*
NCDOT Division 3	*-5.273E-02*	*0.836*	*-1.748E-01*	*0.656*
NCDOT Division 4	*5.280E-01*	*0.044*	*2.265E-01*	*0.587*
NCDOT Division 5	*5.904E-01*	*0.016*	*2.553E-01*	*0.504*
NCDOT Division 6	*5.869E-01*	*0.019*	*5.753E-01*	*0.131*
NCDOT Division 7	*2.422E-01*	*0.335*	*-2.179E-01*	*0.589*
NCDOT Division 8	*5.871E-01*	*0.023*	*3.489E-01*	*0.387*
NCDOT Division 9	*6.279E-01*	*0.016*	*6.739E-01*	*0.091*
NCDOT Division 10	*5.762E-01*	*0.019*	*-9.014E-02*	*0.814*
NCDOT Division 11	*2.631E-01*	*0.359*	*-4.199E-02*	*0.930*
NCDOT Division 12	*2.447E-01*	*0.324*	*-4.579E-01*	*0.255*
NCDOT Division 13	*3.929E-01*	*0.130*	*-1.355E-01*	*0.748*
NCDOT Division 14	*3.034E-02*	*0.916*	*-5.884E-01*	*0.260*
Inverse of overdispersion parameter	*0.970*	*0.000*	*0.878*	*0.000*
2xlog-likelihood value	*-10028.006*		*-3779.246*	

Table 21 provides the elasticities for non-division indicator variables associated with the models in Table 20, computed using Equations 9 to 11. These elasticities quantify the amount of "risk" associated with each risk factor included in the model. Specifically, each value represents the relevant increase in 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 0 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 median 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, the AADT is entered in the log form and hence the elasticity values provided in this table would hold for all AADT values.

Values in Table 21 can be interpreted as follows. Traffic volume is the only variable included in the model in a log form. The elasticities suggest that a one percent change in traffic volume along a Principal Arterial – Other roadway segment is associated with a 0.560 percent increase in total pedestrian crash frequency and 0.760 percent increase in KA pedestrian crash frequency along that segment. For other continuous variables, the elasticity is provided at the median value observed in the data. For example, a one percent change in population density—for the "average" roadway segment with population density of 1,077.2 people per square mile—would be associated with a 0.124 increase in total pedestrian crash frequency and no observable change in KA 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 56.3 percent increase in total pedestrian crash frequency and 41.9 percent increase in crash frequency along that segment. Other variables can be interpreted similarly.

Table 21. Elasticity values for Principal Arterials – Other

Variable	Variable type	Elasticity for total crash frequency	Elasticity for KA crash frequency	Median value (if applicable)
Natural log of AADT	log	0.560	0.760	N/A
5+ lane roadway	I	0.563	0.419	N/A
Speed limit 40 or 45 mph	I	---	0.276	N/A
Speed limit 50 mph or above	I	---	0.299	N/A
Median present	I	-0.266	-0.135	N/A
Block length between 0.1-0.25 mi	I	-0.355	-0.332	N/A
Block length between 0.25-0.5 mi	I	-0.465	-0.370	N/A
Block length greater than 0.5 mi	I	-0.685	-0.678	N/A
High intensity development within 100 ft	I	0.666	0.680	N/A
Alcohol sales density	C	0.099	0.070	8.02
Bus route present	I	0.688	0.554	N/A
Population density	C	0.124	---	1077.2
K12 enrollment density	C	0.073	0.127	155.4
Median household income	C	-0.381	-0.264	46506
Proportion of commuters non-motorized	C	0.019	---	0.019
Proportion of population disabled	C	0.352	0.249	0.146

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

Table 22 to Table 24 provide a summary of the crash frequency model developed for roadway segments categorized as Minor Arterials, Major + Minor Collectors, and Local roads, respectively. Factors are similar to those included in the model for Principal Arterial – Others and align with expectations. Those associated with increased risk include:

- Vehicular traffic volume (i.e., AADT)
- Roadway segments with five or more lanes
- Roadway segments with higher speed limits
- Presence of high-intensity (land use cover) development within 100 ft
- Density of alcohol sales establishments
- Presence of a bus route along the segment
- Population density
- K12 school enrollment density
- Proportion of commuters that are non-motorized
- Proportion of the population that is disabled
- Proportion of households with zero vehicles

Factors associated with reduced risk include:

- Presence of a median
- Roadway segments with longer block lengths
- Higher median household income near the roadway segment
- Proportion of the population that is 65 and above

Some assumptions were necessary in the estimation of these models. For Local roads, a large fraction (81.65%) of the roadway segments did not have associated speed limits. Based on conversations with the NCDOT technical panel, a statutory value of 35 mph was used to replace all missing values. Similarly, a significant fraction (22.45%) of roadway segments classified as Major or Minor Collectors did not have speed limit information. For these segments, a statutory value of 35 mph was assumed for segments within city limits and a value of 45 mph was assumed for segments outside of city limits. However, speeds limits did not emerge as a statistically significant risk factor in the crash frequency models for Major + Minor Collector roadways.

Table 25 to Table 27 provide a summary of the elasticities/crash risks estimates from the Minor Arterials, Major + Minor Collectors, and Local Roads, respectively. These values were obtained in a similar manner to those provided in Table 21 for Principal Arterial – Others.

Table 22. Summary of crash frequency models developed for Minor Arterials

	Total crash frequency		KA crash frequency	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Constant	*-7.867E+00*	*0.000*	*-1.237E+01*	*0.000*
Natural log of AADT	7.868E-01	0.000	1.043E+00	0.000
5+ lane roadway	2.671E-01	0.075	6.896E-01	0.005
Speed limit 35 mph or above	---	---	6.637E-01	0.155
Median present	-4.527E-01	0.000	-4.376E-01	0.031
Block length between 0.1-0.25 mi	-3.054E-01	0.000	---	---
Block length between 0.25-0.5 mi	-7.140E-01	0.000	---	---
Block length greater than 0.5 mi	-8.610E-01	0.000	---	---
Block length greater than 0.25 mi	---	---	-2.497E-01	0.041
High intensity development within 100 ft	3.592E-01	0.000	5.140E-01	0.088
Alcohol sales density	1.317E-02	0.000	7.121E-03	0.024
Bus route present	5.349E-01	0.000	2.271E-01	0.077
Population density	1.061E-04	0.003	---	---
K12 enrollment density	2.943E-04	0.156	5.965E-04	0.030
Median household income	-6.711E-06	0.000	-5.691E-06	0.080
Proportion of commuters non-motorized	9.682E-01	0.090	---	---
Proportion of population disabled	3.405E+00	0.000	5.421E+00	0.000
Proportion of population 65+	-2.508E+00	0.000	-4.296E+00	0.000
Proportion of zero vehicle HHs	1.117E+00	0.024	1.324E+00	0.156
NCDOT Division 2	*-3.159E-01*	*0.216*	*-8.394E-01*	*0.078*
NCDOT Division 3	*-1.076E-01*	*0.674*	*-5.427E-01*	*0.242*
NCDOT Division 4	*-2.719E-02*	*0.910*	*-4.806E-01*	*0.270*
NCDOT Division 5	*1.956E-02*	*0.934*	*-5.832E-01*	*0.166*
NCDOT Division 6	*1.263E-01*	*0.602*	*-2.377E-01*	*0.578*
NCDOT Division 7	*-1.120E-01*	*0.634*	*-2.939E-01*	*0.483*
NCDOT Division 8	*-3.769E-02*	*0.884*	*-2.378E-01*	*0.602*
NCDOT Division 9	*-2.921E-02*	*0.902*	*-2.183E-01*	*0.603*
NCDOT Division 10	*7.522E-02*	*0.750*	*-5.838E-01*	*0.170*
NCDOT Division 11	*-6.559E-01*	*0.033*	*-8.321E-01*	*0.126*
NCDOT Division 12	*-1.123E-01*	*0.634*	*-4.467E-01*	*0.288*
NCDOT Division 13	*1.075E-01*	*0.658*	*-5.883E-01*	*0.204*
NCDOT Division 14	*-1.344E-01*	*0.649*	*-6.488E-01*	*0.273*
Inverse of overdispersion parameter	*1.078*	*0.000*	*2.91*	*0.165*
2xlog-likelihood value	*-10750.636*		*-3316.294*	

Table 23. Summary of crash frequency models developed for Major + Minor Collectors

	Total crash frequency		KA crash frequency	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Constant	*-7.013E+00*	*0.000*	*-9.146E+00*	*0.000*
Natural log of AADT	7.213E-01	0.000	8.857E-01	0.000
Median present	-2.679E-01	0.251	---	---
Block length between 0.1-0.25 mi	-3.577E-01	0.000	-2.517E-01	0.133
Block length between 0.25-0.5 mi	-5.494E-01	0.000	---	---
Block length greater than 0.5 mi	-6.796E-01	0.000	---	---
Block length greater than 0.25 mi	---	---	-3.218E-01	0.085
High intensity development within 100 ft	3.726E-01	0.000	2.732E-01	0.071
Alcohol sales density	8.861E-03	0.000	7.558E-03	0.000
Bus route present	4.128E-01	0.000	3.119E-01	0.073
Population density	1.723E-04	0.000	---	---
Median household income	-7.301E-06	0.000	-1.377E-05	0.001
Proportion of commuters non-motorized	1.416E+00	0.024	---	---
Proportion of population disabled	3.114E+00	0.001	3.460E+00	0.084
Proportion of population 65+	-3.082E+00	0.000	-3.298E+00	0.031
NCDOT Division 2	*-5.876E-02*	*0.854*	*-1.085E+00*	*0.112*
NCDOT Division 3	*-3.057E-01*	*0.307*	*-1.019E+00*	*0.081*
NCDOT Division 4	*1.292E-01*	*0.683*	*-1.592E-01*	*0.780*
NCDOT Division 5	*6.132E-02*	*0.830*	*-1.112E-02*	*0.982*
NCDOT Division 6	*1.293E-01*	*0.663*	*-8.535E-02*	*0.869*
NCDOT Division 7	*2.085E-01*	*0.464*	*4.683E-02*	*0.926*
NCDOT Division 8	*-1.559E-01*	*0.626*	*-1.236E+00*	*0.070*
NCDOT Division 9	*1.021E-02*	*0.971*	*-3.612E-01*	*0.475*
NCDOT Division 10	*1.341E-01*	*0.636*	*-7.737E-01*	*0.139*
NCDOT Division 11	*-3.897E-01*	*0.246*	*-1.141E+00*	*0.094*
NCDOT Division 12	*-2.369E-01*	*0.439*	*-5.824E-01*	*0.292*
NCDOT Division 13	*1.065E-01*	*0.720*	*-4.153E-01*	*0.445*
NCDOT Division 14	*2.247E-01*	*0.483*	*-1.864E-01*	*0.755*
Inverse of overdispersion parameter	*0.684*	*0.000*	*1.085*	*0.123*
2xlog-likelihood value	*-8615.23*		*-2175.880*	

Table 24. Summary of crash frequency models developed for Local Roads

	Total crash frequency		KA crash frequency	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Constant	*-2.796E+00*	*0.000*	*-3.571E+00*	*0.000*
Speed limit 40 mph or above	*1.831E-01*	*0.011*	*8.207E-01*	*5.260*
Block length between 0.1-0.25 mi	-2.030E-01	0.000	-1.876E-01	0.082
Block length between 0.25-0.5 mi	-4.523E-01	0.000	-4.120E-01	0.010
Block length greater than 0.5 mi	-5.188E-01	0.000	-5.786E-01	0.035
High intensity development within 100 ft	8.625E-01	0.000	8.297E-01	0.000
Alcohol sales density	9.722E-03	0.000	8.076E-03	0.000
Bus route present	1.353E+00	0.000	1.704E+00	0.000
Population density	1.523E-04	0.000	---	---
Employment density	2.228E-05	0.000	1.419E-05	0.078
K12 enrollment density	3.863E-04	0.000	1.171E-03	0.000
Median household income	-1.110E-05	0.000	-1.471E-05	0.000
Proportion of population 65+	-1.484E+00	0.000	-2.211E+00	0.015
Proportion of zero vehicle HHs	2.391E+00	0.000	---	---
NCDOT Division 2	*2.858E-01*	*0.157*	*-2.514E-01*	*0.541*
NCDOT Division 3	*1.641E-01*	*0.409*	*-6.106E-01*	*0.136*
NCDOT Division 4	*2.547E-01*	*0.209*	*-6.955E-01*	*0.110*
NCDOT Division 5	*7.907E-01*	*0.000*	*-2.470E-01*	*0.523*
NCDOT Division 6	*2.852E-01*	*0.152*	*-4.047E-01*	*0.315*
NCDOT Division 7	*4.054E-01*	*0.036*	*-5.043E-01*	*0.197*
NCDOT Division 8	*1.209E-01*	*0.571*	*-8.987E-01*	*0.059*
NCDOT Division 9	*3.530E-01*	*0.069*	*-3.183E-01*	*0.411*
NCDOT Division 10	*9.240E-01*	*0.000*	*-2.853E-01*	*0.460*
NCDOT Division 11	*2.084E-01*	*0.348*	*-3.964E-01*	*0.389*
NCDOT Division 12	*3.131E-01*	*0.112*	*-9.858E-01*	*0.020*
NCDOT Division 13	*-1.875E-01*	*0.376*	*-1.099E+00*	*0.019*
NCDOT Division 14	*1.248E-01*	*0.585*	*-4.163E-01*	*0.383*
Inverse of overdispersion parameter	*0.343*	*0.061*	*0.369*	*0.027*
2xlog-likelihood value	*-37380.214*		*-5987.926*	

Table 25. Elasticity values for Minor Arterials

Variable	Variable type	Elasticity for total crash frequency	Elasticity for KA crash frequency	Median value (if applicable)
Natural log of AADT	log	0.787	1.043	N/A
5+ lane roadway	I	0.306	0.993	N/A
Speed limit 35 mph or above	I	---	0.942	N/A
Median present	I	-0.364	-0.354	N/A
Block length between 0.1-0.25 mi	I	-0.263	---	N/A
Block length between 0.25-0.5 mi	I	-0.510	---	N/A
Block length greater than 0.5 mi	I	-0.577	---	N/A
Block length greater than 0.25 mi	I	---	-0.221	N/A
High intensity development within 100 ft	I	0.432	0.672	N/A
Alcohol sales density	C	0.122	0.066	9.299
Bus route present	I	0.707	0.255	N/A
Population density	C	0.131	---	1231.5
K12 enrollment density	C	0.039	0.078	131.21
Median household income	C	-0.332	-0.282	49531
Proportion of commuters non-motorized	C	0.013	---	0.013
Proportion of population disabled	C	0.490	0.781	0.144
Proportion of population 65+	C	-0.414	-0.709	0.165
Proportion of zero vehicle HHs	C	0.058	0.069	0.052

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

Table 26. Elasticity values for Major + Minor Collectors

Variable	Variable type	Elasticity for total crash frequency	Elasticity for KA crash frequency	Median value (if applicable)
Natural log of AADT	log	0.721	0.886	N/A
Median present	I	-0.235	---	N/A
Block length between 0.1-0.25 mi	I	-0.301	-0.223	N/A
Block length between 0.25-0.5 mi	I	-0.423	---	N/A
Block length greater than 0.5 mi	I	-0.493	---	N/A
Block length greater than 0.25 mi	I	---	-0.275	N/A
High intensity development within 100 ft	I	0.452		N/A
Alcohol sales density	C	0.065	0.055	7.283*
Bus route present	I	0.511	0.366	N/A
Population density	C	0.128	---	741.8
Median household income	C	-0.370	-0.697	50610
Proportion of commuters non-motorized	C	0.018	---	0.013
Proportion of population disabled	C	0.433	0.481	0.139
Proportion of population 65+	C	-0.512	-0.547	0.166

*I – indicator variable; C – continuous variable; log – continuous variable included in log form; * note that mean value used for alcohol sales density because median value was 0.*

Table 27. Elasticity values for Local Roads

Variable	Variable type	Elasticity for total crash frequency	Elasticity for KA crash frequency	Median value (if applicable)
Speed limit 40 mph or above	I	0.201	1.272	N/A
Block length between 0.1-0.25 mi	I	-0.184	-0.171	N/A
Block length between 0.25-0.5 mi	I	-0.364	-0.338	N/A
Block length greater than 0.5 mi	I	-0.405	-0.439	N/A
High intensity development within 100 ft	I	1.369	1.293	N/A
Alcohol sales density	C	0.049	0.041	5.028*
Bus route present	I	2.869	4.496	N/A
Population density	C	0.168	---	1104.9
Employment density	C	0.006	0.004	286
K12 enrollment density	C	0.065	0.197	168.49
Median income	C	-0.610	-0.809	54968
Proportion of population 65+	C	-0.233	-0.347	0.157
Proportion of zero vehicle HHs	C	0.103	---	0.043

*I – indicator variable; C – continuous variable; log – continuous variable included in log form; * note that mean value used for alcohol sales density because median value was 0.*

DISCUSSION

The previous section provides the crash risk factor estimates that are developed for total and KA pedestrian crash frequency. However, after review of the models and discussion with the NCDOT technical panel, the research team ultimately recommends the use of risk factors developed for KA pedestrian crash frequency. While both sets of risk factors are useful, the KA risk factors were found to better identify high-risk locations that are known to the research team and NCDOT technical panel across North Carolina.

Risk factor summary

Pedestrian risk factors were developed for four roadway functional classification types: Principal Arterial – Others, Minor Arterials, Major + Minor Collectors, and Local Roads. For ease of reading, the risk factor models are summarized together in Table 28 for KA crash risk. The equations that can be used to compute the risk factor score of urban roadway segments of each functional classification are also provided after the tables.

Table 28. Summary of KA crash risk factor model coefficients

Coefficients	Principal Arterials - Other	Minor Arterials	Major + Minor Collectors	Local Roads
Constant	*-9.338E+00*	*-1.237E+01*	*-9.146E+00*	*-3.571E+00*
Natural log of AADT	7.596E-01	1.043E+00	8.857E-01	---
5+ lane roadway	3.503E-01	6.896E-01	---	---
Speed limit 40 or 45 mph	2.436E-01	---	---	---
Speed limit 50 mph or above	2.614E-01	---	---	---
Speed limit 35 mph or above	---	6.637E-01	---	---
Speed limit 40 mph or above	---	---	---	8.207E-01
Median present	-1.456E-01	-4.376E-01	---	---
Block length between 0.1-0.25 mi	-4.042E-01	---	-2.517E-01	-1.876E-01
Block length between 0.25-0.5 mi	-4.617E-01	---	---	-4.120E-01
Block length greater than 0.5 mi	-1.134E+00	---	---	-5.786E-01
Block length greater than 0.25 mi	---	-2.497E-01	-3.218E-01	---
High intensity development within 100 ft	5.186E-01	5.140E-01	2.732E-01	8.297E-01
Alcohol sales density	8.729E-03	7.121E-03	7.558E-03	8.076E-03
Bus route present	4.408E-01	2.271E-01	3.119E-01	1.704E+00
Employment density	---	---	---	1.419E-05
K12 enrollment density	8.149E-04	5.965E-04	---	1.171E-03
Median household income	-5.666E-06	-5.691E-06	-1.377E-05	-1.471E-05
Proportion of commuters non-motorized	---	---	---	---
Proportion of population disabled	1.703E+00	5.421E+00	3.460E+00	---
Proportion of population 65+	---	-4.296E+00	-3.298E+00	-2.211E+00
Proportion of zero vehicle HHs	---	1.324E+00	---	---
NCDOT Division 2	*2.457E-02*	*-8.394E-01*	*-1.085E+00*	*-2.514E-01*
NCDOT Division 3	*-1.748E-01*	*-5.427E-01*	*-1.019E+00*	*-6.106E-01*
NCDOT Division 4	*2.265E-01*	*-4.806E-01*	*-1.592E-01*	*-6.955E-01*
NCDOT Division 5	*2.553E-01*	*-5.832E-01*	*-1.112E-02*	*-2.470E-01*
NCDOT Division 6	*5.753E-01*	*-2.377E-01*	*-8.535E-02*	*-4.047E-01*
NCDOT Division 7	*-2.179E-01*	*-2.939E-01*	*4.683E-02*	*-5.043E-01*
NCDOT Division 8	*3.489E-01*	*-2.378E-01*	*-1.236E+00*	*-8.987E-01*
NCDOT Division 9	*6.739E-01*	*-2.183E-01*	*-3.612E-01*	*-3.183E-01*
NCDOT Division 10	*-9.014E-02*	*-5.838E-01*	*-7.737E-01*	*-2.853E-01*
NCDOT Division 11	*-4.199E-02*	*-8.321E-01*	*-1.141E+00*	*-3.964E-01*
NCDOT Division 12	*-4.579E-01*	*-4.467E-01*	*-5.824E-01*	*-9.858E-01*
NCDOT Division 13	*-1.355E-01*	*-5.883E-01*	*-4.153E-01*	*-1.099E+00*
NCDOT Division 14	*-5.884E-01*	*-6.488E-01*	*-1.864E-01*	*-4.163E-01*

The model coefficient estimates in Table 28 can be directly incorporated into Equation (7) to obtain the risk factor score for a given segment. The resulting equation for **KA crash risk score for Principal Arterial – Others** is:

$$N_{i,risk} = AADT^{7.596E-01} \times L \times e^{-9.338} \times e^{3.503E-01 \times 5p\_lanes} \times e^{2.436E-01 \times SL\_40\_45} \times e^{2.614E-01 \times SL\_50p} \times e^{-1.456E-01 \times Median} \times e^{-4.042E-01 \times BL\_01\_025} \times e^{-4.617E-01 \times BL\_025\_050} \times e^{-1.134 \times BL\_050p} \times e^{5.186E-01 \times HI\_Dev} \times e^{8.729E-03 \times Alc\_Dens} \times e^{4.408E-01 \times Bus\_Route} \times e^{8.149E-04 \times K12\_Dens} \times e^{-5.666E-06 \times Med\_Inc} \times e^{1.703 \times Disabled\_Prop} \times e^{2.457E-02 \times Div2} \times e^{-1.748E-01 \times Div3} \times e^{2.265E-01 \times Div4} \times e^{2.553E-01 \times Div5} \times e^{5.753E-01 \times Div6} \times e^{-2.179E-01 \times Div7} \times e^{3.489E-01 \times Div8} \times e^{6.739E-01 \times Div9} \times e^{-9.014E-02 \times Div10} \times e^{-4.199E-02 \times Div11} \times e^{-4.579E-01 \times Div12} \times e^{-1.355E-01 \times Div13} \times e^{-5.884E-01 \times Div14} \quad (13)$$

The equation for **KA crash risk score for Minor Arterials** is:

$$N_{i,risk} = AADT^{1.043} \times L \times e^{-1.237E+01} \times e^{6.896E-01 \times 5p\_lanes} \times e^{6.637E-01 \times SL\_35p} \times e^{-4.376E-01 \times Median} \times e^{-2.497E-01 \times BL\_025p} \times e^{5.140E-01 \times HI\_Dev} \times e^{7.121E-03 \times Alc\_Dens} \times e^{2.271E-01 \times Bus\_Route} \times e^{5.965E-04 \times K12\_Dens} \times e^{-5.691E-06 \times Med\_Inc} \times e^{5.421 \times Disabled\_Prop} \times e^{-4.296 \times A\_65p\_Prop} \times e^{1.324 \times zeroHHs\_Prop} \times e^{-8.394E-01 \times Div2} \times e^{-5.427E-01 \times Div3} \times e^{-4.806E-01 \times Div4} \times e^{-5.832E-01 \times Div5} \times e^{-2.377E-01 \times Div6} \times e^{-2.939E-01 \times Div7} \times e^{-2.378E-01 \times Div8} \times e^{-2.183E-01 \times Div9} \times e^{-5.838E-01 \times Div10} \times e^{-8.321E-01 \times Div11} \times e^{-4.467E-01 \times Div12} \times e^{-5.883E-01 \times Div13} \times e^{-6.488E-01 \times Div14} \quad (14)$$

The equation for **KA crash risk score for Major + Minor Collectors** is:

$$N_{i,risk} = AADT^{8.857E-01} \times L \times e^{-9.146} \times e^{-2.517E-01 \times BL\_01\_025} \times e^{-3.218E-01 \times BL\_025p} \times e^{2.732E-01 \times HI\_Dev} \times e^{7.558E-03 \times Alc\_Dens} \times e^{3.119E-01 \times Bus\_Route} \times e^{-1.377E-05 \times Med\_Inc} \times e^{3.460 \times Disabled\_Prop} \times e^{-3.298 \times A\_65p\_Prop} \times e^{-1.085 \times Div2} \times e^{-1.019 \times Div3} \times e^{-1.592E-01 \times Div4} \times e^{-1.112E-02 \times Div5} \times e^{-8.535E-02 \times Div6} \times e^{4.683E-02 \times Div7} \times e^{-1.236 \times Div8} \times e^{-3.612E-01 \times Div9} \times e^{-7.737E-01 \times Div10} \times e^{-1.141 \times Div11} \times e^{-5.824E-01 \times Div12} \times e^{-4.153E-01 \times Div13} \times e^{-1.864E-01 \times Div14} \quad (15)$$

The resulting equation for **KA crash risk score for Local Roads** is:

$$N_{i,risk} = L \times e^{-3.571} \times e^{8.207E-01 \times SL\_40p} \times e^{-1.876E-01 \times BL\_01\_025} \times e^{-4.120E-01 \times BL\_025\_050} \times e^{-5.786E-01 \times BL\_050p} \times e^{8.297E-01 \times HI\_Dev} \times e^{8.076E-03 \times Alc\_Dens} \times e^{1.704 \times Bus\_Route} \times e^{1.419E-05 \times Emp\_Dens} \times e^{1.171E-03 \times K12\_Dens} \times e^{-1.471E-05 \times Med\_Inc} \times e^{-2.211 \times A\_65p\_Prop} \times e^{-2.514E-01 \times Div2} \times e^{-6.106E-01 \times Div3} \times e^{-6.955E-01 \times Div4} \times e^{-2.470E-01 \times Div5} \times e^{-4.047E-01 \times Div6} \times e^{-5.043E-01 \times Div7} \times e^{-8.987E-01 \times Div8} \times e^{-3.183E-01 \times Div9} \times e^{-2.853E-01 \times Div10} \times e^{-3.964E-01 \times Div11} \times e^{-9.858E-01 \times Div12} \times e^{-1.099 \times Div13} \times e^{-4.163E-01 \times Div14} \quad (16)$$

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]; SL_40_45 = indicator variable for the road segment speed limit of 40 or 45 mph [1,0]; SL_50p = indicator variable for the road segment speed limit of 50 mph or greater [1,0]; SL_35p = indicator variable for the road segment speed limit of 35 mph or greater [1,0]; SL_40p = indicator variable for the road segment speed limit of 40 mph or greater [1,0]; $Median$ = indicator variable for presence of a median on the road segment [1,0]; BL_01_025 = indicator variable for the road segment length between 0 and 0.25 miles [1,0]; BL_025_050 = indicator variable for the road segment length between 0.25 and 0.5 miles [1,0]; BL_050p = indicator variable for the road segment length greater than 0.5 miles [1,0]; BL_025p = indicator variable for the road segment length greater than 0.25 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]; Bus_Route = indicator variable for the presence of a bus route within 100 ft of the road segment [1,0]; Emp_Dens = employment density within 0.25 miles of the segment [decimal]; $K12\_Dens$ = K-12 school enrollment density within 0.25 miles of the segment [decimal]; Med_Inc = median household income [$]; $Disabled\_Prop$ = proportion of the population with a disability (as defined by the ACS) within 0.25 miles of the segment [decimal]; A_65p_Prop =proportion of the population 65 and older within 0.25 miles of the segment [decimal]; $zeroHHs\_Prop$ = proportion zero vehicle households with 0.25 miles of the segment [decimal]; and, $DivX$ = indicator variable for segment occurring within NCDOT Engineering Division X [1,0].

Risk factor application

The risk factors can be used in a variety of ways. For one, the models can be directly applied to estimate the expected pedestrian crash risk at individual 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 or the application of systemic safety treatments. The research team has performed these calculations and developed an interactive GIS-based map that identifies the riskiest roadway segments within North Carolina. These risky segments are those that have calculated risk factors that are in the top 1% or top 5% of all segments. This interactive map is available at this link: https://vhb.maps.arcgis.com/apps/mapviewer/index.html?webmap=4acb41d58d39436592d560f6d4ac2903. Examples of the results on this platform are shown below in Figure 1, Figure 2, Figure 3, and Figure 4.

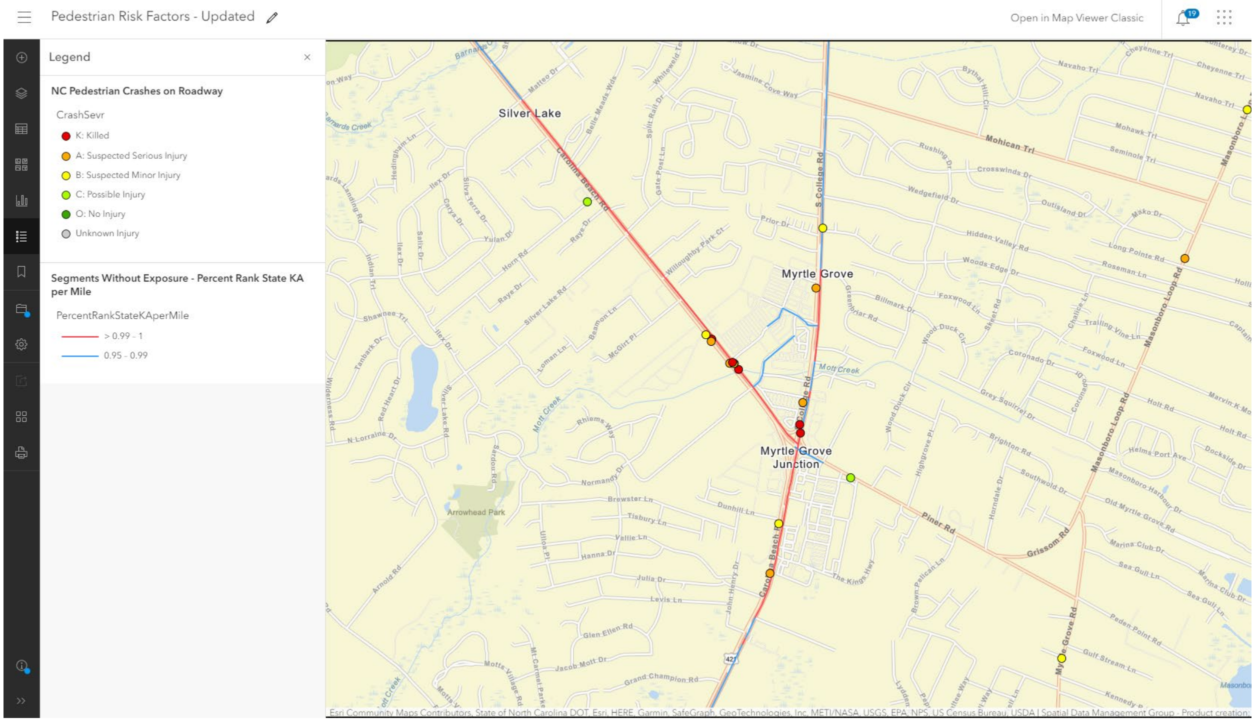

Figure 1. Example of KA Crash per Mile Screening – Carolina Beach Rd and S College Rd in Wilmington.

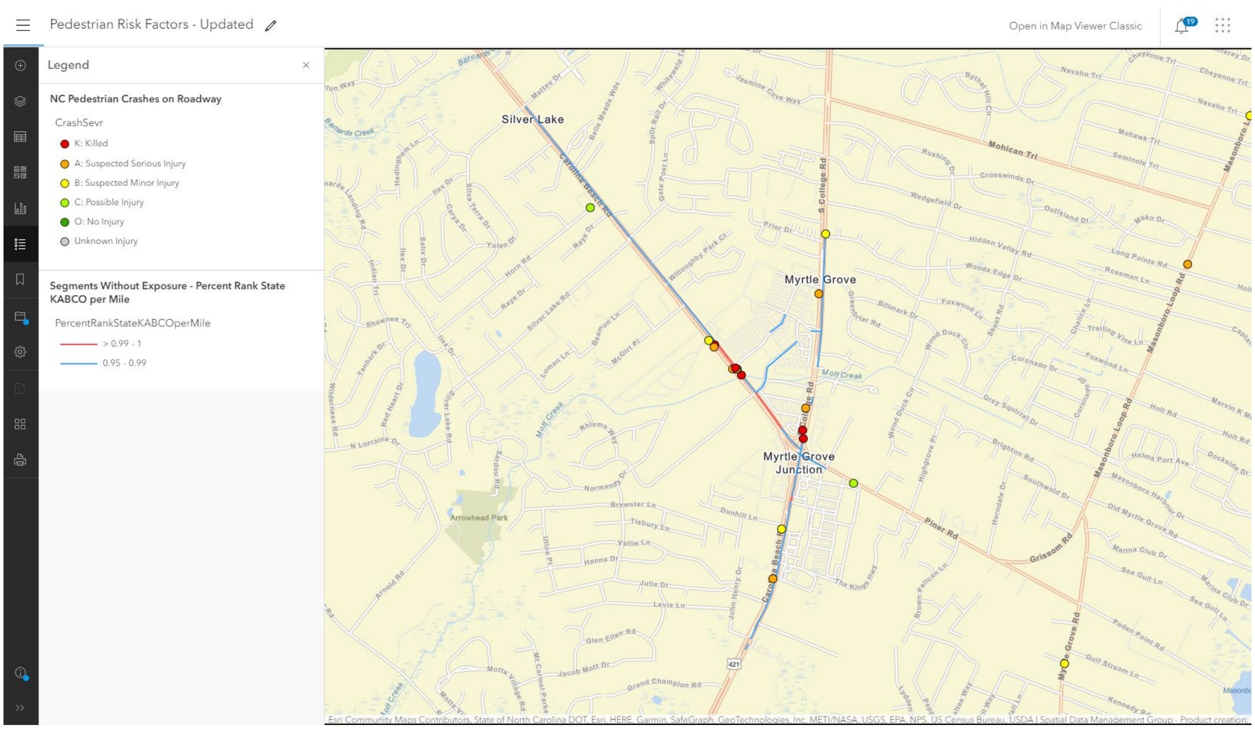

Figure 2. Example of KABCO Crash per Mile Screening – Carolina Beach Rd and S College Rd in Wilmington.

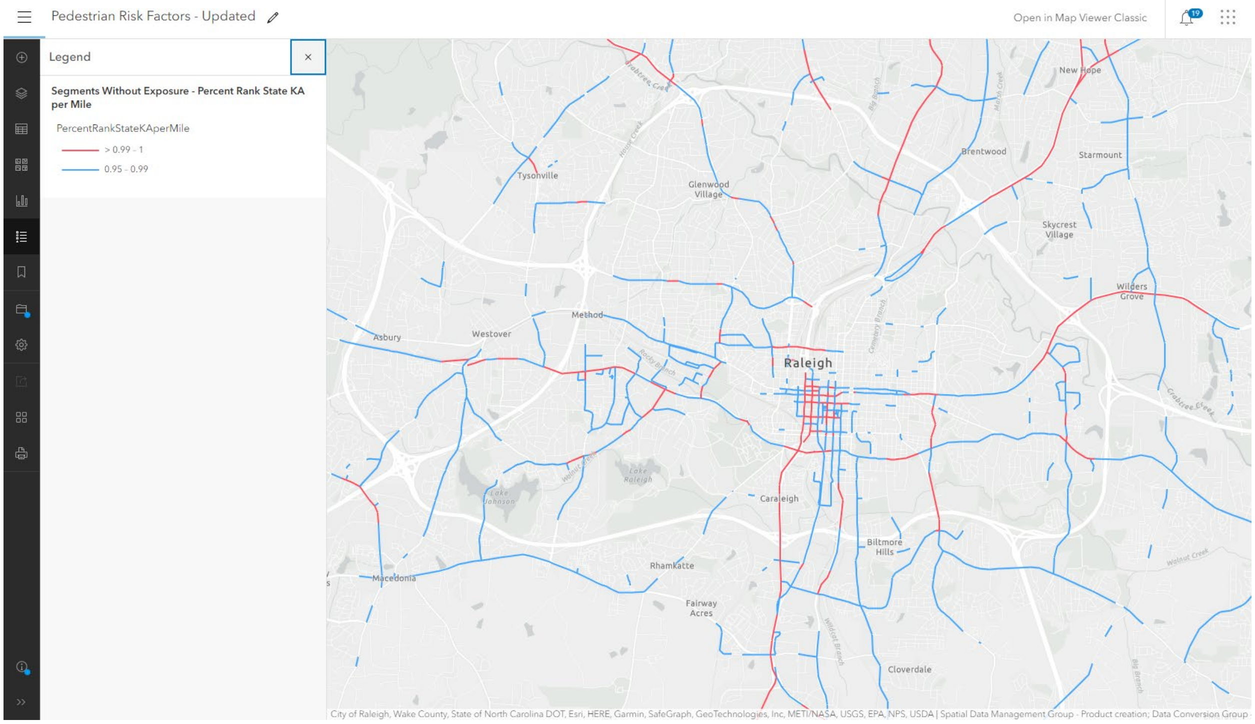

Figure 3. Example of KA Crash per Mile Screening – Downtown Raleigh.

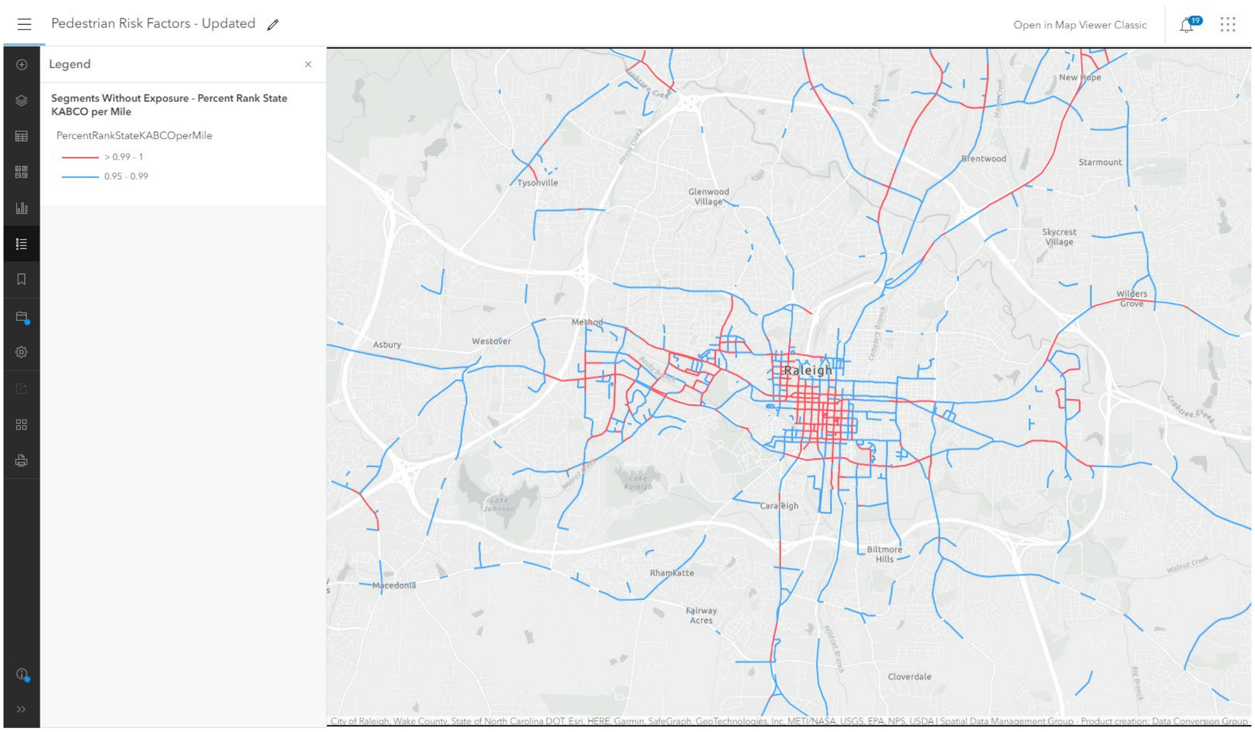

Figure 4. Example of KABCO Crash per Mile Screening – Downtown Raleigh.

These figures illustrate key differences between KA and KABCO model outputs. Although there is considerable overlap and consistency between both models, Figure 1 and Figure 2 highlight a high crash intersection south of Wilmington and north of Carolina Beach (informally referred to as “Monkey Junction”). At least 8 fatal or suspected serious injury crashes have occurred since 2015; although the KABCO model results show some segments of the intersecting roads in the top 5 percent statewide (blue), the KA model results show the entire intersection in the Top 1 percent of segments statewide (red). This one example is illustrative of broader trends in the model results. KA model results will tend to focus on higher speed, urban, commercial arterials (Figure 3), while KABCO model results skew toward urban cores and central business districts where exposure (i.e., pedestrian crossing and foot traffic) is higher but vehicular speeds tend to be lower (Figure 4).

The risk factors themselves can also be used to understand the relative impacts of various features on pedestrian crash risk. Specifically, the elasticities quantify the observed relationship between individual features on pedestrian crash frequency. For example, the elasticity values in Table 21 suggest that roadway segments with speed limits of 40 or 45 mph have 27.6% higher expected crash frequency than segments with lower speed limits. Segments with higher speed limits (50 mph or more) are associated with 29.9% higher crash frequency than speed limits lower than 40 mph. Such information can be used to compare the relative safety of different locations or when making changes to their features (e.g., setting speed limits). It should be noted, however, that these are observed relationships and the causal relationship between these features and crash risk may not always be clear. Nevertheless, they can be used to help identify risky sites or identify how pedestrian crash risk might change from one context to another.

Model updates

The risk factor models were developed using historical data and thus represent observed relationships during these time periods. For example, crash data from 2015-2020 (inclusive) were used to develop these models; time periods for other explanatory variables are included in the report 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 not longer hold if there are significant changes in driving and/or pedestrian behavior or associated technologies; e.g., complete overhaul of the vehicle fleet or drastically 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. Thus, the models should be updated at regular intervals to capture these changing trends. In general, there is little to no 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 5 years based on the relatively long data collection period (six years of crash data) and broad coverage of data availability (over the entire state).

Pedestrian exposure modeling

The research team recommends the use of risk factors developed without direct exposure estimates due to the relative inaccuracy of the pedestrian exposure model; see Appendix C for more details. One reason for the reduced accuracy of the pedestrian exposure model is data availability. As mentioned, pedestrian exposure counts were generally performed in conjunction with traffic volume counts and thus may not be representative of the entire roadway network. For this reason, there is likely an overrepresentation of the types of sites associated with extremely low pedestrian activity. Figure 5 provides a comparison between locations with pedestrian counts and roadway segment centerline mileage by roadway functional classification. Notice that there is an overrepresentation of counts at higher functional classification levels (Other Freeway, Arterials, Major Collectors) compared to the actual roadway mileage. A more representative and balanced dataset might help improve the pedestrian exposure model. For these reasons, the research team recommends the implementation of a designated pedestrian count program designed to supplement the existing counts with new counts at locations to improve the spatial and temporal coverage of pedestrian and provide a more representative sample for the development of a statewide pedestrian exposure model. This includes counts at arterials that are based on anticipated pedestrian exposure – as opposed to at locations where vehicular counts are being performed – and additional counts at lower functional classification roadways (Local and Minor Collectors) that are more representative pedestrian activity in the vicinity.

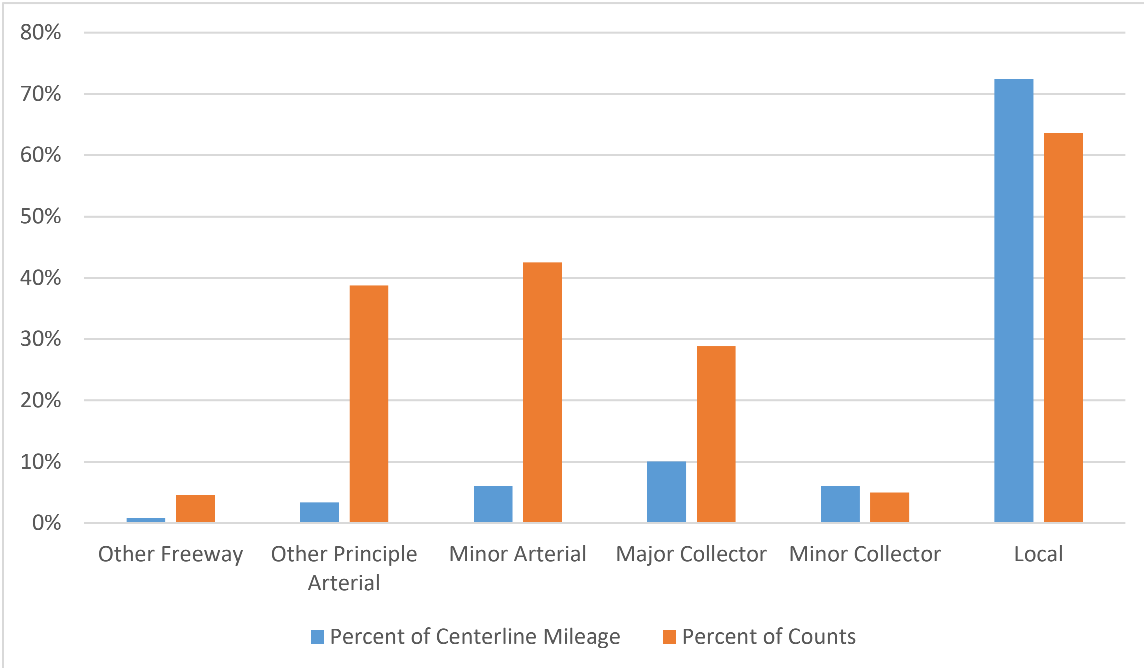

Figure 5. Distribution of Statewide Roadway Mileage and Number of Sample Counts by Functional Class.[7]

Additionally, there are several extremely high pedestrian exposure counts that might bias the estimates of the pedestrian exposure model. However, these high counts are not necessarily outliers as they represent locations with extremely high pedestrian activity. The research team suggests considering alternative approaches to pedestrian exposure modeling. One opportunity could be to classify locations based on categories: little to no pedestrian activity, low activity, medium activity, high activity, and extreme activity. These categories could then be predicted based on site-specific contexts and then used in the development of future pedestrian crash risk factors.

[7] Note that percentage of counts do not add up to 100% as they were performed at intersections; more than one functional class may be associated with each count.

References

American Association of State Highway and Transportation Officials, 2010. Highway Safety Manual.

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. Accident Analysis & Prevention 161, 106351. https://doi.org/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. Transportation Engineering Journal of ASCE 103, 507–520. https://doi.org/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.

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 centre redevelopment on pedestrian flows. Transportation 48, 643–672. https://doi.org/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. Accident Analysis & Prevention 42, 1718–1728. https://doi.org/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 DESIGN International 18, 217–228. https://doi.org/10.1057/udi.2013.2

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

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.

Geedipally, S.R., Lord, D., Dhavala, S.S., 2012. The negative binomial-Lindley generalized linear model: Characteristics and application using crash data. Accident Analysis & Prevention 45, 258–265. https://doi.org/10.1016/j.aap.2011.07.012

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. Transportation Research Record: Journal of the Transportation Research Board 2673, 941–950. https://doi.org/10.1177/0361198119837235

Ha, H.-H., Thill, J.-C., 2011. Analysis of traffic hazard intensity: A spatial epidemiology case study of urban pedestrians. Computers, Environment and Urban Systems 35, 230–240. https://doi.org/10.1016/j.compenvurbsys.2010.12.004

Hamilton, I., Kersavage, K., Porter, R.J., Gayah, V., Sanchez, J., Smith, K., Tan, C., Eigen, A.M., 2022. Application of Emerging Data Sources for Pedestrian Safety Analysis in Charlotte, NC. Transportation Research Record 03611981221093330. https://doi.org/10.1177/03611981221093330

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 report 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. Transportation Research Record: Journal of the Transportation Research 2586, 48–58. https://doi.org/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. Landscape and Urban Planning 107, 307–316. https://doi.org/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. https://doi.org/10.1016/j.jth.2017.08.009

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

Hilbe, J.M., 2011. Negative Binomial Regression. Cambridge University Press.

Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Injury Prevention 9, 205–209. https://doi.org/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. https://doi.org/10.17226/22163

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis & Prevention 34, 149–161. https://doi.org/10.1016/S0001-4575(01)00009-4

Lee, K., Sener, I.N., 2020. Emerging data for pedestrian and bicycle monitoring: Sources and applications. Transportation Research Interdisciplinary Perspectives 4. https://doi.org/10.1016/j.trip.2020.100095

Lindsey, G., Han, Y., Wilson, J., Yang, J., 2006. Neighborhood correlates of urban trail use. Journal of Physical Activity and Health 3, S139--S157. https://doi.org/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. Landscape and Urban Planning 81, 299–315. https://doi.org/10.1016/j.landurbplan.2007.01.004

Liu, X., Griswold, J., 2009. Pedestrian volume modeling: A case study of San Francisco. of the Association of Pacific Coast Geographers 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. Transportation Research Part D: Transport and Environment 63, 244–260. https://doi.org/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. Accident Analysis & Prevention 121, 166–176. https://doi.org/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. Transportation Research Record: Journal of the Transportation Research 2465, 57–63. https://doi.org/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. Transportation Research Record: Journal of the Transportation Research 2264, 74–82. https://doi.org/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. Archives of trauma research 5. https://doi.org/10.5812/atr.30796

New York State, 2016. Pedestrian safety action plan.

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. Geographical Analysis 49, 387–408. https://doi.org/10.1111/gean.12128

Pande, A., Abdel-Aty, M., 2009. A novel approach for analyzing severe crash patterns on multilane highways. Accident Analysis & Prevention 41, 985–994. https://doi.org/10.1016/j.aap.2009.06.003

Pulugurtha, S.S., Repaka, S.R., 2008. Assessment of models to measure pedestrian activity at signalized intersections. Transportation Research Record: Journal of the Transportation Research Board 39–48. https://doi.org/10.3141/2073-05

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. International Journal of Intelligent Transportation Systems Research 19, 347–365. https://doi.org/10.1007/s13177-020-00250-1

Schneider, R.J., Arnold, L.S., Ragland, D.R., 2009. Pilot model for estimating pedestrian intersection crossing volumes. Transportation Research Record: Journal of the Transportation Research 2140, 13–26. https://doi.org/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. Transportation Research Record 2299, 65–78. https://doi.org/10.3141/2299-08

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

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

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

Shankar, V.N., Albin, R.B., Milton, J.C., Mannering, F.L., 1998. Evaluating Median Crossover Likelihoods with Clustered Accident Counts An Empirical Inquiry Using the Random Effects Negative Binomial Model. Transportation Research Record: Journal of the Transportation Research Board 1635, 44–48.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis & Prevention 45, 382–391. https://doi.org/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. Journal of Safety Research 78, 180–188. https://doi.org/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. Journal of Transport Geography 93, 103067. https://doi.org/10.1016/j.jtrangeo.2021.103067

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. https://doi.org/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. Transportation Research Record: Journal of the Transportation Research Board 2198, 65–74. https://doi.org/10.3141/2198-08

Ukkusuri, S., Hasan, S., Aziz, H.M.A., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transportation Research Record: Journal of the Transportation Research Board 2237, 98–106. https://doi.org/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. Safety Science 50, 1141–1151. https://doi.org/10.1016/j.ssci.2011.09.012

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

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. https://doi.org/10.1016/j.aap.2008.10.001

APPENDIX A: SEGMENT RISK DATA DICTIONARY.

Variable name	Type	Description
SegmentID	Integer	Unique (non-consecutive) integer representing the segment (i.e., intersection to intersection)
Division	Integer	NCDOT division in which the segment is physically located
LocCntyCode	Integer	County in which the segment is physically located (coded value)
RouteClass	Integer	NCDOT route class indicating the dominant route along the segment (e.g., Interstate, US Route, NC Route). Defined using NCDOT definitions
RouteNumber	Integer	Route number associated with the dominant route on the segment
RouteID	Integer	NCDOT route identifier based on NCDOT's route characteristics
County	String	County in which the segment is physically located
FuncClass	Integer	Functional classification of the road based on NCDOT's route characteristics. Defined using NCDOT definitions
MedianType	Integer	Type of dividing median on the road (if applicable) based on NCDOT's route characteristics. Defined using NCDOT definitions
SpeedLimit	Integer	Posted speed limit on the road (if known) based on NCDOT's route characteristics
Total_Lanes	Integer	Total number of through lanes in both directions based on NCDOT's route characteristics
AADT_2015	Integer	Average annual daily traffic for the year 2015 (if available)
AADT_2016	Integer	Average annual daily traffic for the year 2016 (if available)
AADT_2017	Integer	Average annual daily traffic for the year 2017 (if available)
AADT_2018	Integer	Average annual daily traffic for the year 2018 (if available)
AADT_2019	Integer	Average annual daily traffic for the year 2019 (if available)
AADT_2020	Integer	Average annual daily traffic for the year 2020 (if available)
ACCFOOD_72	Integer	Total accomodation and food services empoyment (NAICS 72) within the dominant census tract (i.e., covering most of the segment)
AREA	Float	Total area of the census tract (sq. mi)
ARTSENTREC_71	Integer	Total arts, entertainment, and recreation services empoyment (NAICS 71) within the dominant census tract (i.e., covering most of the segment)
COLLEGE_25PLUS	Integer	Total population over the age of 25 with an associates degree or higher within the dominant census tract (i.e., covering most of the segment)
COLLEGE_25PLUS_PROP	Float	Proportion of the population over the age of 25 with an associates degree or higher within the dominant census tract (i.e., covering most of the segment)
DISABLE_POP	Integer	Total population with a disability within the dominant census tract (i.e., covering most of the segment)
DISABLE_PROP	Float	Proportion of the population with a disability within the dominant census tract (i.e., covering most of the segment)
EDUCATION_61	Integer	Total educational services empoyment (NAICS 61) within the dominant census tract (i.e., covering most of the segment)
EMP_DENS	Integer	Employment density within the dominant census tract (i.e., covering most of the segment)
K12_DENS	Float	K-12 enrollment density (by place of residence) within the dominant census tract (i.e., covering most of the segment)

K12_ENROLL	Integer	Total K-12 enrollment residing within the dominant census tract (i.e., covering most of the segment)
LEP_HH	Integer	Total limited English proficiency households within the dominant census tract (i.e., covering most of the segment)
LEP_HH_PROP	Float	Proportion of households with limited English proficiency within the dominant census tract (i.e., covering most of the segment)
MED_INC	Integer	Median household income of the dominant census tract (i.e., covering most of the segment)
NONMOT_DENS	Float	Density of non-motorized commuters within the dominant census tract (i.e., covering most of the segment)
NONMOT_PROP	Float	Proportion of non-motorized commuters within the dominant census tract (i.e., covering most of the segment)
NONWHITE	Integer	Total non-white or 2 or more races population within the dominant census tract (i.e., covering most of the segment)
POP_18	Integer	Total population 18 and younger within the dominant census tract (i.e., covering most of the segment)
POP_18_DENS	Float	Density of persons 18 and younger within the dominant census tract (i.e., covering most of the segment)
POP_18_PROP	Float	Proportion of population aged 18 and under within the dominant census tract (i.e., covering most of the segment)
POP_25_44	Integer	Total population between the ages of 25 and 44 within the dominant census tract (i.e., covering most of the segment)
POP_25PLUS	Integer	Total population over the age of 25 within the dominant census tract (i.e., covering most of the segment)
POP_65	Integer	Total population 65 and older within the dominant census tract (i.e., covering most of the segment)
POP_65_DENS	Float	Density of persons 65 and older within the dominant census tract (i.e., covering most of the segment)
POP_65_PROP	Float	Proportion of the population 65 and older within the dominant census tract (i.e., covering most of the segment)
POP_DENS	Float	Population density within the dominant census tract (i.e., covering most of the segment)
POP_POV	Integer	Total population living under the poverty line within the dominant census tract (i.e., covering most of the segment)
POP_POV_DET	Integer	Total population for which poverty status has been determined within the dominant census tract (i.e., covering most of the segment)
POV_PROP	Float	Proportion of the population living under the poverty line within the dominant census tract (i.e., covering most of the segment)
PROP_NONWHITE	Float	Proportion of population that is non-white or 2 or more races within the dominant census tract (i.e., covering most of the segment)
RETAIL	Integer	Total retail empoyment (NAICS 44-45) within the dominant census tract (i.e., covering most of the segment)
TOT_16PLUS	Integer	Total population in the civilian labor force over 16 within the dominant census tract (i.e., covering most of the segment)
TOT_BIKE	Integer	Total bicycle commuters within the dominant census tract (i.e., covering most of the segment)
TOT_COMM	Integer	Total commuters within the dominant census tract (i.e., covering most of the segment)
TOT_EMP	Integer	Total employment within the dominant census tract (i.e., covering most of the segment)

TOT_HH	Integer	Total households within the dominant census tract (i.e., covering most of the segment)
TOT_NONMOT	Integer	Total non-motorized commuters within the dominant census tract (i.e., covering most of the segment)
TOT_POP	Integer	Total population within the dominant census tract (i.e., covering most of the segment)
TOT_TRANS	Integer	Total transit commuters within the dominant census tract (i.e., covering most of the segment)
TOT_WALK	Integer	Total walking commuters within the dominant census tract (i.e., covering most of the segment)
UNEMP	Integer	Total unemployed persons in the civilian labor force within the dominant census tract (i.e., covering most of the segment)
UNEMP_PROP	Float	Proportion of civilian labor force that is unemployed within the dominant census tract (i.e., covering most of the segment)
WHITE	Integer	Total white population within the dominant census tract (i.e., covering most of the segment)
ZERO_HH	Integer	Total zero vehicle households within the dominant census tract (i.e., covering most of the segment)
ZERO_HH_PROP	Float	Proportion zero vehicle households within the dominant census tract (i.e., covering most of the segment)
Total_Segment_Length	Float	Total length of the segment in miles
Developed_HI_100ft	Float	Segment is within 100 feet of land use classified as "Developed, High Intensity" according to the 2019 National Land Cover Database.
Developed_LI_100ft	Float	Segment is within 100 feet of land use classified as "Developed, Low Intensity" according to the 2019 National Land Cover Database.
Developed_MI_100ft	Float	Segment is within 100 feet of land use classified as "Developed, Medium Intensity" according to the 2019 National Land Cover Database.
Proximate_University_College	Binary	College or university present within HALF mile (0 if >HALF mile; 1 if <HALF mile) of segment
K12_Count	Binary	Primary or secondary schools present (grades K through 12) within QTR mile of segment
BusRoute_Present	Binary	Bus route present on segment (excluding intercity routes)
GoldLineStop_1_4mi	Binary	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	Blue Line light rail stop within QTR mile (0 if >QTR mile; 1 if <QTR mile) of segment (open March 2018)
AlcSales_Count	Integer	Number of alcohol sales establishments within QTR mile of segment
AlcSales_Density	Float	Density of alcohol sales establishments within QTR mile of segment (per mile)
Park_prox	Binary	Public greenspace within QTR mile (0 if >QTR mile; 1 if <QTR mile) of segment
Sidewalk	Binary	Presence of sidewalk within 100 feet (0 if >100 feet; 1 if <100 feet) of segment
Greenway	Binary	Presence of greenway within 100 feet (0 if >100 feet; 1 if <100 feet) of segment
Parcel_Count	Integer	Number of land parcels within QTR mile of segment
Parcel_Density	Float	Density of land parcels within QTR mile of segment
Pedestrian_KABCO_2020	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2020)

Pedestrian_KA_2020	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2020)
Pedestrian_KABCO_2019	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2019)
Pedestrian_KA_2019	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2019)
Pedestrian_KABCO_2018	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2018)
Pedestrian_KA_2018	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2018)
Pedestrian_KABCO_2017	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2017)
Pedestrian_KA_2017	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2017)
Pedestrian_KABCO_2016	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2016)
Pedestrian_KA_2016	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2016)
Pedestrian_KABCO_2015	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2015)
Pedestrian_KA_2015	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2015)
Pedestrian_KABCO_2014	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2014)
Pedestrian_KA_2014	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2014)
Pedestrian_KABCO_2013	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2013)
Pedestrian_KA_2013	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2013)
Pedestrian_KABCO_2012	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2012)
Pedestrian_KA_2012	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2012)
Pedestrian_KABCO_2011	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - all collision severities (2011)
Pedestrian_KA_2011	Integer	Total number of pedestrian crashes (assigned as on-road) that occurred on the segment (<250 ft) - fatal and serious injury collisions (2011)

APPENDIX B: PEDESTRIAN EXPOSURE DATA DICTIONARY

Variable name	Type	Description
RequestID	Integer	A unique (non-consecutive) integer unique to the count's location in GIS format
County	String	County location - County
City	String	Count location - City
Division	Integer	Count location - NCDOT Division
Loc Description	String	Textual description of the count location
Latitude	Float	Count location - Latitude
Longitude	Float	Count location - Longitude
Alt_RequestID	Integer	Alternate ID used to merge NCDOT pedestrian count data with HSIS data
Facility Type	String	Facility type description of the count location
Duration	Float	Count duration in hours
Increments	String	Time increment for the ped count
Begin Time	Timestamp	Timestamp of the count beginning
Begin Date	Date	Count date - Beginning of the count
Begin Year	Integer	Count year - Beginning of the count
Begin Month	String	Count month - Beginning of the count
Begin DayWeek	String	Count day of the week - Beginning of the count
Begin Hour	Integer	Count hour - Beginning of the count
End Time	Timestamp	Timestamp of the count end
End Date	Date	Count date - End of the count
End Year	Integer	Count year - End of the count
End Month	String	Count month - End of the count
End DayWeek	String	Count day of the week - End of the count
End Hour	Integer	Count hour - End of the count
Pedestrian Count	Integer	Ped count of each record
Intersection	Binary	Intersection indicator (0 if it is NOT an intersection; 1 if it is an intersection)
RCUT_Flag	Binary	Reduced conflict intersection indicator (0 if it is NOT part of an RCI - either at the u-turn or primary intersection; 1 if it is part of an RCI)
Proximate to University/College	Binary	College or university present within HALF mile (0 if >HALFmile; 1 if <HALF mile)
K12_Count	Integer	Primary or secondary school (grades K through 12) within QTR mile (0 if >QTR mile; 1 if <QTR mile)
Bus_1_4mi	Binary	Bus route within QTR mile (0 if >QTR mile; 1 if <QTR mile)
Bus_100ft	Binary	Bus route within 100 feet (0 if >100 feet; 1 if <100 feet)
LightRail_1_4mi	Binary	Light rail route within QTR mile (0 if >QTR mile; 1 if <QTR mile)
LightRail_100ft	Binary	Light rail route within 100 feet (0 if >100 feet; 1 if <100 feet)
Park_prox	Binary	Public greenspace within QTR mile (0 if >QTR mile; 1 if <QTR mile)
Open Area_Acres	Float	Acreage of public greenspace within QTR mile
Sidewalk	Binary	Presence of sidewalk within 100 feet (0 if >100 feet; 1 if <100 feet)
Greenway	Binary	Presence of greenway within 100 feet (0 if >100 feet; 1 if <100 feet)
Crosswalk	Binary	Presence of crosswalk within 100 feet (0 if >100 feet; 1 if <100 feet)
Signal	Binary	Presence of traffic signal within 100 feet (0 if >100 feet; 1 if <100 feet)
AlcSales_Count	Integer	Number of alcohol sales establishments within QTR mile
Parcel_Count	Integer	Number of land parcels establishments within QTR mile

Variable name	Type	Description
SPEED_25	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 25 mph; 0 otherwise
SPEED_30	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 30 mph; 0 otherwise
SPEED_35	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 35 mph; 0 otherwise
SPEED_40	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 40 mph; 0 otherwise
SPEED_45	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 45 mph; 0 otherwise
SPEED_50	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 50 mph; 0 otherwise
SPEED_55	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 55 mph; 0 otherwise
SPEED_60	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 60 mph; 0 otherwise
SPEED_65	Binary	1 if one of the legs at the measurement location (<100 feet) has a speed limit of 65 mph; 0 otherwise
Undiv_1Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 1 thru thru lane undivided road; 0 otherwise
Undiv_2Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 2 thru lane undivided road; 0 otherwise
Undiv_3Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 3 thru lane undivided road; 0 otherwise
Undiv_4Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 4 thru lane undivided road; 0 otherwise
Undiv_5Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 5 thru lane undivided road; 0 otherwise
Undiv_6Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 6 thru lane undivided road; 0 otherwise
Div_1Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 1 thru lane divided road; 0 otherwise
Div_2Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 2 thru lane divided road; 0 otherwise
Div_3Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 3 thru lane divided road; 0 otherwise
Div_4Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 4 thru lane divided road; 0 otherwise
Div_5Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 5 thru lane divided road; 0 otherwise
Div_6Lane	Binary	1 if one of the legs at the measurement location (<100 feet) is 6 thru lane divided road; 0 otherwise
Ramp	Binary	1 if there is a ramp at the measurement location (<100 feet); 0 otherwise
Other_Freeway	Binary	1 if one of the legs at the measurement location (<100 feet) is an Other Freeways or Expressway; 0 otherwise
Major_Arterial	Binary	1 if one of the legs at the measurement location (<100 feet) is an Other Principal Arterial; 0 otherwise
Minor_Arterial	Binary	1 if one of the legs at the measurement location (<100 feet) is a Minor Arterial; 0 otherwise

Variable name	Type	Description
Major_Collector	Binary	1 if one of the legs at the measurement location (<100 feet) is a Major Collector; 0 otherwise
Minor_Collector	Binary	1 if one of the legs at the measurement location (<100 feet) is a Minor Collector; 0 otherwise
Local	Binary	1 if one of the legs at the measurement location (<100 feet) is a Local road; 0 otherwise
AADT_max	Integer	Maximum annual average daily traffic (AADT) among all legs (<100 feet); null if no AADT available
AADT_min	Integer	Minimum AADT among all legs; null if no AADT available
AADT_Form	Integer	AADT derived from NCDOT count documentation; null if unavailable
TOT_POP	Float	Total population within QTR mile; calculated by % of land coverage within Census tract(s)
TOT_EMP	Float	Total employment within QTR mile; calculated by % of land coverage within Census tract(s)
AREA	Float	Total area within QTR mile (sq. mi)
POP_POV	Float	Total population living under the poverty line within QTR mile; calculated by % of land coverage within Census tract(s)
POP_POV_DET	Float	Total population for which poverty status has been determined within QTR mile; calculated by % of land coverage within Census tract(s)
TOT_HH	Float	Total households within QTR mile; calculated by % of land coverage within Census tract(s)
ZERO_HH	Float	Total zero vehicle households within QTR mile; calculated by % of land coverage within Census tract(s)
POP_65	Float	Total population 65 and older within QTR mile; calculated by % of land coverage within Census tract(s)
POP_18	Float	Total population 18 and younger within QTR mile; calculated by % of land coverage within Census tract(s)
LEP_HH	Float	Total limited English proficiency households within QTR mile; calculated by % of land coverage within Census tract(s)
MED_INC	Float	Median household income; calculated by % of land coverage within Census tract(s) - no data not included in calculation
TOT_COMM	Float	Total commuters within QTR mile; calculated by % of land coverage within Census tract(s)
TOT_BIKE	Float	Total bicycle commuters within QTR mile; calculated by % of land coverage within Census tract(s)
TOT_TRANS	Float	Total transit commuters within QTR mile; calculated by % of land coverage within Census tract(s)
TOT_WALK	Float	Total walking commuters within QTR mile; calculated by % of land coverage within Census tract(s)
TOT_NONMOT	Float	Total non-motorized commuters within QTR mile; calculated by % of land coverage within Census tract(s)
UNEMP	Float	Total unemployed persons in the civilian labor force within QTR mile; calculated by % of land coverage within Census tract(s)
TOT_16PLUS	Float	Total population in the civilian labor force over 16 within QTR mile; calculated by % of land coverage within Census tract(s)
WHITE	Float	Total white population within QTR mile; calculated by % of land coverage within Census tract(s)

Variable name	Type	Description
NONWHITE	Float	Total non-white or 2 or more races population within QTR mile; calculated by % of land coverage within Census tract(s)
K12_ENROLL	Float	Total K-12 enrollment residing within QTR mile; calculated by % of land coverage within Census tract(s)
DISABLE_POP	Float	Total population with a disability within QTR mile; calculated by % of land coverage within Census tract(s)
SOME_COLL_25_44	Float	Total population between the ages of 25 and 44 with some college education within QTR mile; calculated by % of land coverage within Census tract(s)
COLLEGE_25PLUS	Float	Total population over the age of 25 with an associates degree or higher within QTR mile; calculated by % of land coverage within Census tract(s)
POP_25_44	Float	Total population between the ages of 25 and 44 within QTR mile; calculated by % of land coverage within Census tract(s)
POP_25PLUS	Float	Total population over the age of 25 within QTR mile; calculated by % of land coverage within Census tract(s)
RETAIL	Float	Total retail empoyment (NAICS 44-45) within QTR mile; calculated by % of land coverage within Census tract(s)
EDUCATION	Float	Total educational services empoyment (NAICS 61) within QTR mile; calculated by % of land coverage within Census tract(s)
ARTS_ENT_REC	Float	Total arts, entertainment, and recreation services empoyment (NAICS 71) within QTR mile; calculated by % of land coverage within Census tract(s)
ACC_FOOD	Float	Total accomodation and food services empoyment (NAICS 72) within QTR mile; calculated by % of land coverage within Census tract(s)
POP_DENS	Float	Population density within QTR mile; calculated by % of land coverage within Census tract(s)
EMP_DENS	Float	Employment density within QTR mile; calculated by % of land coverage within Census tract(s)
POP_PROP	Float	Proportion of the population living under the poverty line within QTR mile; calculated by % of land coverage within Census tract(s)
POP_65_PROP	Float	Proportion of the population 65 and older within QTR mile; calculated by % of land coverage within Census tract(s)
POP_18_PROP	Float	Proportion of population aged 18 and under within QTR mile; calculated by % of land coverage within Census tract(s)
NONMOT_DENS	Float	Density of non-motorized commuters within QTR mile; calculated by % of land coverage within Census tract(s)
NONMOT_PROP	Float	Proportion of non-motorized commuters within QTR mile; calculated by % of land coverage within Census tract(s)
UNEMP_PROP	Float	Proportion of civilian labor force that is unemployed within QTR mile; calculated by % of land coverage within Census tract(s)
PROP_NONWHITE	Float	Proportion of population that is non-white or 2 or more races within QTR mile; calculated by % of land coverage within Census tract(s)
K12_DENS	Float	K-12 enrollment density (by place of residence) within QTR mile; calculated by % of land coverage within Census tract(s)
DISABLE_PROP	Float	Proportion of the population with a disability within QTR mile; calculated by % of land coverage within Census tract(s)

Variable name	Type	Description
COLLEGE_25PLUS_PROP	Float	Proportion of the population over the age of 25 with an associates degree or higher within QTR mile; calculated by % of land coverage within Census tract(s)
ZERO_HH_PROP	Float	Proportion zero vehicle households within QTR mile; calculated by % of land coverage within Census tract(s)
POP_65_DENS	Float	Density of persons 65 and older within QTR mile; calculated by % of land coverage within Census tract(s)
POP_18_DENS	Float	Density of persons 18 and younger within QTR mile; calculated by % of land coverage within Census tract(s)
LEP_HH_PROP	Float	Proportion of households with limited English proficiency within QTR mile; calculated by % of land coverage within Census tract(s)
Developed_HI	Float	Percentage (0-100) of land use within 1 km classified as "Developed, High Intensity" according to the 2019 National Land Cover Database.
Developed_LI	Float	Percentage (0-100) of land use within 1 km classified as "Developed, Low Intensity" according to the 2019 National Land Cover Database.
Developed_MI	Float	Percentage (0-100) of land use within 1 km classified as "Developed, Medium Intensity" according to the 2019 National Land Cover Database.
LU_Mix	Float	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

APPENDIX C: RISK FACTORS WITH DIRECT EXPOSURE ESTIMATES

Risk factors estimated including direct pedestrian exposure estimates involved a two-step process: 1) estimate a model to predict pedestrian exposure for individual roadway segments; and, 2) estimate crash frequency models using the pedestrian exposure estimates as explanatory variables.

Pedestrian exposure model

A NB model of the form shown in Equation 12 was estimated to predict pedestrian exposure at intersections as a function of site-specific features. The resulting exposure model is summarized in Table 29. 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 speeds, higher vehicle volumes, at intersections with legs classified as other freeway, major arterial or minor arterials, and at locations with a higher percentage of older (over 65) and younger (under 18) population. Pedestrian activity is expected to increase at intersections near bus stops, with a higher mix of land use, with more parcels nearby, near a college or university, near a greenway, near alcohol selling establishments, at locations with more population, employment and K12 schools, and at locations with a higher proportion of non-motorized commuters. These results are in line with expectations. Sets of indicator variables were also included to account for regional differences across NCDOT engineering divisions and to control for the differences in count durations.

Table 29. Summary of pedestrian exposure model for urban intersections

Factor	Coefficient	p-value
Constant	3.834E+00	0.0000
Max speed >= 45 mph	-5.673E-01	0.0000
Log of max AADT	-2.572E-01	0.0000
One leg classified as other freeway	-4.526E-01	0.0004
One leg classified as major arterial	-1.731E-01	0.0121
One leg classified as minor arterial	-1.356E-01	0.0225
bus stop within 1/4 mile	4.253E-01	0.0000
Land use mix	1.211E+00	0.0000
Log of parcel count	4.425E-01	0.0000
Proximate to college/university	5.109E-01	0.0000
Greenway present	4.517E-01	0.0001
1-5 Alcohol sales locations nearby	2.599E-01	0.0001
>6 Alcohol sales locations nearby	7.629E-02	0.0000
Total population	7.643E-04	0.0000
Total employment	6.516E-05	0.0003
K12 school count	2.238E-01	0.0001
Proportion of population 65+	-2.339E+00	0.0000
Proportion of population 18-	-2.832E+00	0.0000
Proportion of zero vehicle HHs	3.625E+00	0.0000
Proportion of non-motorized commuters	1.980E+00	0.0016
Median income	7.648E-06	0.0000
NCDOT division = 2	-1.195E+00	0.0000
NCDOT division = 3	-8.377E-01	0.0004
NCDOT division = 4	-2.188E+00	0.0000
NCDOT division = 5	-1.852E+00	0.0000
NCDOT division = 6	-1.351E+00	0.0000
NCDOT division = 7	-1.828E+00	0.0000
NCDOT division = 8	-1.577E+00	0.0000
NCDOT division = 9	-1.947E+00	0.0000
NCDOT division = 10	-1.549E+00	0.0000
NCDOT division = 11	-1.378E+00	0.0000
NCDOT division = 12	-1.397E+00	0.0000
NCDOT division = 13	-1.216E+00	0.0000
NCDOT division = 14	-7.367E-01	0.0090
Duration = 13	4.268E-01	0.0011
Duration = 16	3.223E-01	0.0604
Duration = 24	-3.305E-02	0.8958
Overdispersion parameter	0.691	0.000
2xLog Likelihood value	-22626.683	

Figure 6 provides a plot of the predicted pedestrian counts obtained using the exposure model values (y-axis) vs. observed values (x-axis) for illustrative purposes. The red 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. Note that while observations are generally clustered around the line, there is still significant variation and scenarios in which predicted and observed values are not in agreement. These discrepancies are attributed to several reasons. First, while the observed pedestrian counts range from 0 to 14,854, the majority of observations (94.5%) are less than 500. The combination of very small average value per observation (143.2), low median value (19) and presence of extreme outliers (72 values greater than 10,000) leads to inaccurate model predictions. Furthermore, not all pedestrian counts were obtained in the same way: some were convenient samples taken when vehicular traffic studies were being performed, while others were likely done specifically at high pedestrian locations/time periods. These differences would also contribute to inaccurate model predictions. Thus, while the model can be used to identify locations with higher or lower pedestrian activity, the specific value of the predictions might not be as useful.

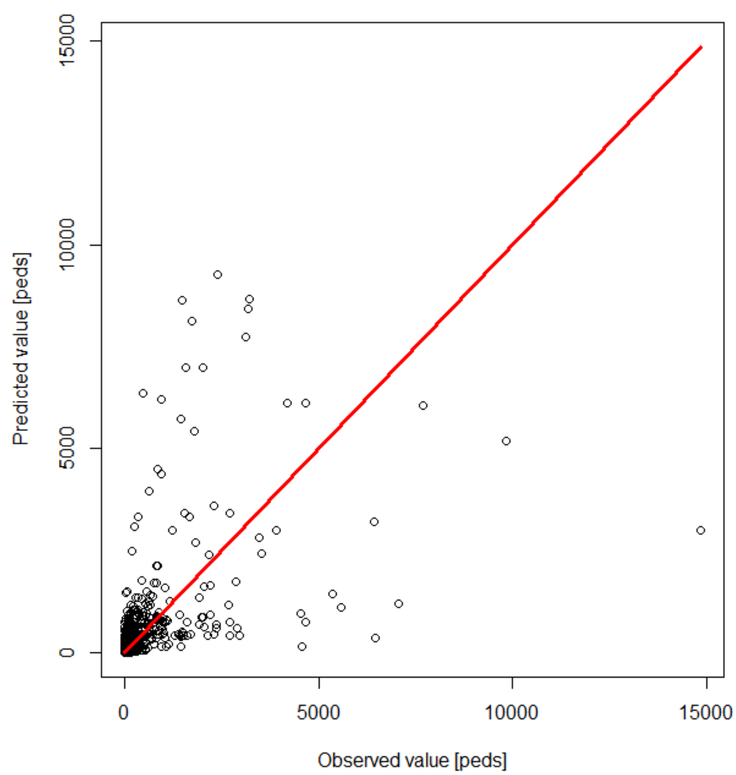

Figure 6. Predicted vs. observed pedestrian count values

Pedestrian risk factors

Risk models were re-estimated after including the predicted pedestrian counts obtained from the pedestrian exposure model developed above as a potential risk factor. Since the exposure model was estimated for intersections and risk models were estimated for roadway segments, exposure estimates were first computed for each intersection (estimated as both terminal ends of a roadway segment where "intersecting" centerlines meet) in the NCDOT roadway network. Then, exposure values were assigned to individual roadway segments by taking the average of the exposure counts for the intersections that a specific roadway segment touched. This average exposure was then used as the potential risk factor for each roadway segment.

The direct exposure estimates were statistically significant and improved the total and KA crash frequency models for roadway segments classified as Principal Arterial – Others and Minor Arterials, but were only statistically significant in the total crash frequency models for Collectors and Local Roads. The lack of impact on the KA crash frequency models for roadway segments classified as Collectors (Major + Minor) or Local Roads may be attributed to the lack of accuracy of the exposure models and relatively small number of KA crashes on these functional classification types. Additionally, the exposure models provided larger maximum observed pedestrian count estimates on Collectors (maximum value of 6,181) and Local Roads (maximum value of 10,028), than on Principal Arterial – Others (maximum value of 3,776) and Minor Arterials (maximum value of 2,025). Further refinements to the exposure model – including considering additional explanatory variables and estimating ranges of pedestrian activity (e.g., low, medium, high, ultra high) as opposed to precise counts might help alleviate this in the future.

Table 30 to Table 33 provide a summary of the crash frequency model with direct exposure estimates developed for roadway segments that include the pedestrian exposure estimate. The pedestrian count estimate is a significant parameter to estimating crash frequency and increases model fit. As expected, larger pedestrian count estimates lead to larger crash frequencies. The impact of other parameters on crash frequency remain similar to the models without direct exposure estimates.

Table 30. Summary of crash frequency models developed for Principal Arterials – Other (with direct exposure estimates)

	Total crash frequency		KA crash frequency	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Constant	*-8.217E+00*	*0.000*	*-1.027E+01*	*0.000*
Natural log of AADT	6.626E-01	0.000	7.904E-01	0.000
Natural log of 13-hr pedestrian count estimate	2.617E-01	0.000	1.169E-01	0.092
5+ lane roadway	4.387E-01	0.000	3.577E-01	0.006
Speed limit 40 or 45 mph			3.238E-01	0.009
Speed limit 50 mph or above			3.615E-01	0.038
Median present	-2.735E-01	0.000	-1.378E-01	0.231
Block length between 0.1-0.25 mi	-3.696E-01	0.000	-3.737E-01	0.002
Block length between 0.25-0.5 mi	-5.006E-01	0.000	-4.042E-01	0.004
Block length greater than 0.5 mi	-9.873E-01	0.000	-1.053E+00	0.000
High intensity development within 100 ft	4.581E-01	0.000	5.118E-01	0.000
Alcohol sales density	1.042E-02	0.000	7.822E-03	0.000
Bus route present	3.966E-01	0.000	3.713E-01	0.002
Population density	4.742E-05	0.100		
K12 enrollment density	4.835E-04	0.002	7.182E-04	0.002
Median income	-8.502E-06	0.000	-5.089E-06	0.044
Proportion of commuters non-motorized	1.830E+00	0.004	1.826E+00	0.102
Proportion of population disabled	*4.769E-01*	*0.067*	*1.554E-01*	*0.702*
NCDOT Division 2	*1.328E-01*	*0.605*	*-6.913E-02*	*0.862*
NCDOT Division 3	*1.061E+00*	*0.000*	*5.095E-01*	*0.252*
NCDOT Division 4	*9.614E-01*	*0.000*	*4.536E-01*	*0.257*
NCDOT Division 5	*8.790E-01*	*0.001*	*7.182E-01*	*0.066*
NCDOT Division 6	*6.388E-01*	*0.013*	*-2.168E-02*	*0.959*
NCDOT Division 7	*1.031E+00*	*0.000*	*5.871E-01*	*0.168*
NCDOT Division 8	*1.072E+00*	*0.000*	*9.147E-01*	*0.031*
NCDOT Division 9	*8.788E-01*	*0.000*	*4.750E-02*	*0.904*
NCDOT Division 10	*5.625E-01*	*0.052*	*1.287E-01*	*0.791*
NCDOT Division 11	*5.731E-01*	*0.024*	*-3.056E-01*	*0.464*
NCDOT Division 12	*7.082E-01*	*0.007*	*2.538E-02*	*0.953*
NCDOT Division 13	*2.615E-01*	*0.369*	*-4.416E-01*	*0.403*
NCDOT Division 14	*-8.217E+00*	*0.000*	*-1.027E+01*	*0.000*
Inverse of overdispersion parameter	1.0232	0.000	0.899	0.000
2xlog-likelihood value	-9904.523		-3747.590	

Table 31. Summary of crash frequency models developed for Minor Arterials (with direct exposure estimates)

	Total crash frequency		KA crash frequency	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Constant	*-9.767E+00*	*0.000*	*-1.383E+01*	*0.000*
Natural log of AADT	8.639E-01	0.000	1.080E+00	0.000
Natural log of 13-hr pedestrian count estimate	3.030E-01	0.000	2.418E-01	0.000
5+ lane roadway	3.484E-01	0.002	7.825E-01	0.002
Speed limit 35 mph or above	---	---	8.410E-01	0.073
Median present	-4.555E-01	0.000	-4.715E-01	0.023
Block length between 0.1-0.25 mi	-2.297E-01	0.000	---	---
Block length between 0.25-0.5 mi	-5.568E-01	0.000	---	---
Block length greater than 0.5 mi	-6.380E-01	0.000	---	---
Block length greater than 0.25 mi	---	---	-1.636E-01	0.194
High intensity development within 100 ft	2.658E-01	0.000	3.989E-01	0.189
Alcohol sales density	1.194E-02	0.000	5.273E-03	0.105
Bus route present	0.3755	6.86E-09		
K12 enrollment density	4.913E-04	0.000	4.063E-04	0.148
Median income	-8.654E-06	0.000	-6.684E-06	0.020
Proportion of population disabled	3.218E+00	0.000	5.692E+00	0.000
Proportion of population 65+	-2.240E+00	0.000	-4.015E+00	0.001
NCDOT Division 2	*7.316E-02*	*0.775*	*-5.584E-01*	*0.243*
NCDOT Division 3	*2.562E-01*	*0.321*	*-2.304E-01*	*0.621*
NCDOT Division 4	*6.861E-01*	*0.007*	*1.389E-01*	*0.762*
NCDOT Division 5	*6.220E-01*	*0.010*	*-9.579E-02*	*0.823*
NCDOT Division 6	*6.271E-01*	*0.011*	*1.722E-01*	*0.694*
NCDOT Division 7	*4.907E-01*	*0.044*	*2.002E-01*	*0.641*
NCDOT Division 8	*4.948E-01*	*0.061*	*2.030E-01*	*0.667*
NCDOT Division 9	*6.227E-01*	*0.012*	*3.147E-01*	*0.477*
NCDOT Division 10	*5.731E-01*	*0.018*	*-1.449E-01*	*0.738*
NCDOT Division 11	*-2.153E-01*	*0.491*	*-4.668E-01*	*0.398*
NCDOT Division 12	*3.757E-01*	*0.122*	*-3.981E-02*	*0.927*
NCDOT Division 13	*4.912E-01*	*0.046*	*-2.773E-01*	*0.553*
NCDOT Division 14	*1.527E-01*	*0.607*	*-4.153E-01*	*0.485*
Inverse of overdispersion parameter	*1.110*	*0.000*	*2.98*	*0.173*
2xlog-likelihood value	-10692.394		*-3302.625*	

Table 32. Summary of crash frequency models developed for Major + Minor Collectors (with direct exposure estimates)

	Total crash frequency	
	Coefficient	**p-value**
Constant	*-8.96E+00*	*0.000*
Natural log of AADT	7.89E-01	0.000
Natural log of 13-hr pedestrian count estimate	3.04E-01	0.000
Median present	-2.10E-01	0.364
Block length between 0.1-0.25 mi	-2.90E-01	0.000
Block length between 0.25-0.5 mi	-3.66E-01	0.000
Block length greater than 0.5 mi	-4.14E-01	0.009
High intensity development within 100 ft	2.87E-01	0.000
Alcohol sales density	7.59E-03	0.000
Bus route present	2.20E-01	0.006
Population density	7.35E-05	0.018
Median income	-7.29E-06	0.000
Proportion of population disabled	2.76E+00	0.003
Proportion of population 65+	-3.06E+00	0.000
NCDOT Division 2	*3.32E-01*	*0.306*
NCDOT Division 3	*2.29E-02*	*0.940*
NCDOT Division 4	*8.38E-01*	*0.012*
NCDOT Division 5	*6.52E-01*	*0.029*
NCDOT Division 6	*6.18E-01*	*0.042*
NCDOT Division 7	*8.01E-01*	*0.007*
NCDOT Division 8	*4.36E-01*	*0.188*
NCDOT Division 9	*6.68E-01*	*0.026*
NCDOT Division 10	*6.29E-01*	*0.031*
NCDOT Division 11	*1.63E-01*	*0.635*
NCDOT Division 12	*2.80E-01*	*0.375*
NCDOT Division 13	*5.45E-01*	*0.073*
NCDOT Division 14	*5.48E-01*	*0.089*
Inverse of overdispersion parameter	*0.740*	0.000
2xlog-likelihood value	*-8547.026*	

Table 33. Summary of crash frequency models developed for Local Roads (with direct exposure estimates)

	Total crash frequency	
	Coefficient	**p-value**
Constant	*-3.703E+00*	*0.000*
Natural log of 13-hr pedestrian count estimate	1.685E-01	0.000
Speed limit 40 mph or above	*3.346E-01*	*0.000*
Block length between 0.1-0.25 mi	-1.774E-01	0.000
Block length between 0.25-0.5 mi	-3.686E-01	0.000
Block length greater than 0.5 mi	-3.815E-01	0.001
High intensity development within 100 ft	8.610E-01	0.000
Alcohol sales density	9.173E-03	0.000
Bus route present	1.293E+00	0.000
Population density	8.918E-05	0.000
Employment density	2.059E-05	0.000
K12 enrollment density	5.119E-04	0.000
Median income	-1.200E-05	0.000
Proportion of population 65+	-1.249E+00	0.000
Proportion of zero vehicle HHs	1.633E+00	0.000
NCDOT Division 2	*4.706E-01*	*0.022*
NCDOT Division 3	*3.289E-01*	*0.101*
NCDOT Division 4	*6.408E-01*	*0.003*
NCDOT Division 5	*1.104E+00*	*0.000*
NCDOT Division 6	*5.377E-01*	*0.008*
NCDOT Division 7	*7.121E-01*	*0.000*
NCDOT Division 8	*4.432E-01*	*0.044*
NCDOT Division 9	*7.004E-01*	*0.001*
NCDOT Division 10	*1.180E+00*	*0.000*
NCDOT Division 11	*4.789E-01*	*0.034*
NCDOT Division 12	*5.780E-01*	*0.004*
NCDOT Division 13	*6.585E-02*	*0.760*
NCDOT Division 14	*3.027E-01*	*0.188*
Inverse of overdispersion parameter	*0.373*	*0.000*
2xlog-likelihood value	*-37154.047*	