An Exploration of Pedestrian Safety Through the Integration of HSIS and Emerging Data Sources: Case Study in Charlotte, NC

PUBLICATION NO. FHWA-HRT-21-087 OCTOBER 2021

U.S. Department of Transportation
Federal Highway Administration

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

Pedestrian and bicycle safety is a focus area under the Federal Highway Administration's Focused Approach to Safety. Pedestrian and bicycle crashes, along with roadway departure and intersection crashes, comprise almost 90 percent of U.S. traffic fatalities. One of the challenges for measuring pedestrian and bicycle safety is the availability of exposure data for pedestrians and bicycles. Agencies often do not have the resources to collect pedestrian and bicycle volumes. There is a need for innovative approaches to acquiring these data.

This study, conducted under the Highway Safety Information System Program, leveraged existing geospatial, pedestrian count, and integrated speed information from probe data to supplement other roadway and contextual transportation data from several agencies. The researchers developed a pedestrian count model to predict pedestrian volumes at locations without pedestrian counts. This study continued the exploration of expanding the utility of existing data resources, as started with *Photographic Data Extraction Feasibility and Pilot Study in Support of Roadside Safety and Roadway Departure Research* (Eigen, Valdivieso, and Ahrari 2014).

The results of this research will benefit safety professionals and State and local transportation engineers. This report provides a method for estimating pedestrian volumes that can be used for pedestrian exposure and demonstrates the use of nontraditional data sources when traditionally used data are not available.

Brian P. Cronin, P.E.
Director, Office of Safety and Operations
Research and Development

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation (USDOT) in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document.

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Recommended citation: Federal Highway Administration, *An Exploration of Pedestrian Safety Through the Integration of HSIS and Emerging Data Sources: Case Study in Charlotte, NC* (Washington, DC: 2021) https://doi.org/10.21949/1521675

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-21-087	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle An Exploration of Pedestrian Safety Through the Integration of HSIS and Emerging Data Sources: Case Study in Charlotte, NC		5. Report Date October 2021 6. Performing Organization Code:
7. Author(s) Ian Hamilton (ORCID: 0000-0003-0949-5495), Kristin Kersavage (ORCID: 0000-0002-3601-7766), R.J. Porter (ORCID: 0000-0001-8535-3451), Keith Smith (ORCID: 0000-0003-4520-8473), Josie Sanchez (ORCID: 0000-0002-6901-9899), Vikash Gayah (ORCID: 0000-0002-0648-3360), Kimberly Eccles (ORCID: 0000-0001-7522-9609)		8. Performing Organization Report No.
9. Performing Organization Name and Address Vanasse Hangen Brustlin 940 Main Campus Drive Suite 500 Raleigh, NC 27606		10. Work Unit No. 11. Contract or Grant No. DTFH61-11-C-00050
12. Sponsoring Agency Name and Address Office of Research and Design Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Report; February 2020–January 2021 14. Sponsoring Agency Code HRDS-20

15. Supplementary Notes
The Contracting Officer's Representative is Ana Maria Eigen (HRDS-20).

16. Abstract
This report built on a geospatial pilot effort by the Highway Safety Information System (HSIS) using data from Charlotte, NC. The main objective of this study was to spatially integrate HSIS data with multi-jurisdictional and emerging datasets to analyze two measures of pedestrian safety performance: the severity of a pedestrian crash that has occurred, and the probability that a pedestrian crash will occur. The study explored several high-priority research topics in safety data and analysis, including pedestrian crash analysis, probe data integration and analysis, and geospatial HSIS data integration. The project team developed a pedestrian count model to predict pedestrian volumes at locations without pedestrian counts and integrated speed information from probe data to supplement other roadway and contextual transportation data from several agencies. Demographic and socioeconomic characteristics, employment, land use, sidewalk presence, transit access, and roadway and intersection characteristics all significantly contributed to pedestrian volume predictions. The project team identified numerous significant factors that influenced pedestrian crash severity and probability. These factors included those identified in previous research, as well as new relationships between pedestrian volumes and vehicular traffic that have implications for pedestrian safety-in-numbers concepts. Results showed that higher pedestrian volumes resulted in both lower crash severities and probabilities, but the safety benefit was reduced by higher vehicle volumes. Higher speeds, higher traffic volumes, larger vehicles striking the pedestrian, pedestrian impairment, and older pedestrian ages were all indicative of higher probabilities of a pedestrian crash resulting in a fatality or serious injury. By adding a direct measure of speed from probe data (and given the known importance of speed to crash injury severity), the pedestrian crash severity model excluded commonly used speed surrogates without sacrificing model fit. The probability of a pedestrian crash occurring on a road segment was affected by segment length, interactions of pedestrian volumes and traffic volumes, and interactions of posted speed limit, median presence, and number of lanes. This study highlights the applicability of integrating HSIS with emerging safety data resources to inform data-driven and performance-based approaches to road safety management.

17. Key Words Highway Safety Information System, data integration, pedestrian safety, pedestrian exposure, probe data, speed		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. http://www.ntis.gov	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 63	22. Price N/A

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

SI* (MODERN METRIC) CONVERSION FACTORS				
APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1,000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2,000 lb)	0.907	megagrams (or “metric ton”)	Mg (or “t”)
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa
APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or “t”)	megagrams (or “metric ton”)	1.103	short tons (2,000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

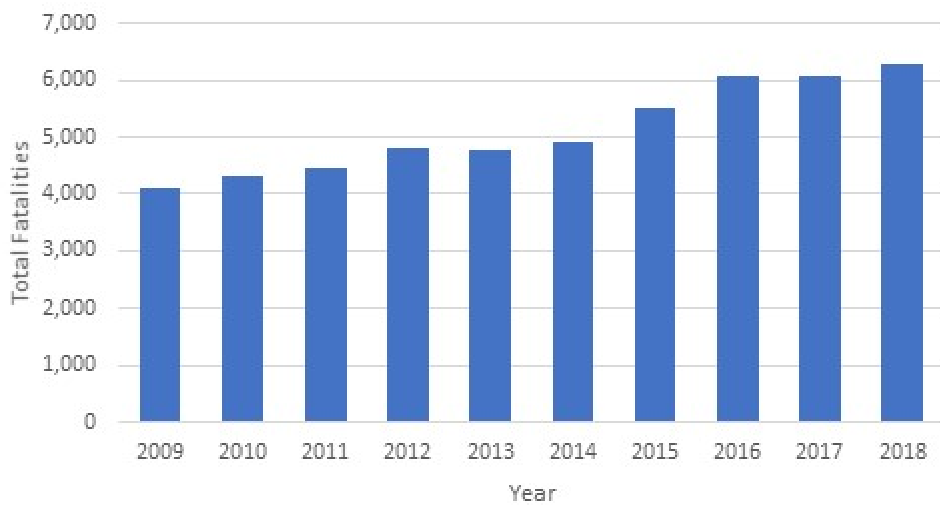
LIST OF ACRONYMS

AADT	annual average daily traffic
ACS	American Community Survey
CATT	Center for Advanced Transportation Technology
CURE	cumulative residual
FHWA	Federal Highway Administration
GIS	geographic information system
HSIS	Highway Safety Information System
LEHD	Longitudinal Employer-Household Dynamics
mph	miles per hour
NC	North Carolina
NCDOT	North Carolina Department of Transportation
NHTSA	National Highway Traffic Safety Administration
PHB	pedestrian hybrid beacon
RITIS	Regional Integrated Transportation Information System
TMC	traffic message channel
U.S.	United States

CHAPTER 1. INTRODUCTION

BACKGROUND AND OBJECTIVES

Pedestrian safety is a growing concern for transportation planners and safety engineers at both the local and State levels. The National Highway Traffic Safety Administration (NHTSA) reported that annual pedestrian fatalities grew by 50 percent between 2009 and 2018 (figure 1), far exceeding the rate of growth in motorist and bicyclist fatalities (NHTSA 2019). Previously published studies have explored contributing factors that lead to more severe or more frequent pedestrian crashes. These studies often focused on proxies or surrogates for some of the principle contributing factors, including speed (e.g., posted speed limit rather than actual vehicle speed) and exposure (e.g., land-use characteristics rather than actual or predicted counts).



Source: FHWA.

Figure 1. Graph. Annual pedestrian fatalities nationwide (based on data from NHTSA 2019).

Continued advancements in data availability, data integration abilities, and analysis methodologies offer new opportunities to identify factors influencing pedestrian safety and quantify their effects to inform data-driven road safety management. Direct measures of pedestrian activity, typically in the form of pedestrian counts, offer the most relevant metrics for pedestrian volumes and crash risk, but these are often expensive and difficult to implement over a large study area (e.g., a city or region). Research efforts in California and Michigan, as well as National Cooperative Highway Research Program project 17-84, indicate a growing interest in surrogate measures of pedestrian volumes and scalable crash risk assessments at the State level (Griswold et al. 2019; Hampshire et al. 2017). Furthermore, emerging data sources, such as vehicular probe data, have demonstrated their suitability as transportation performance monitoring tools. As these sources become more ubiquitous in transportation planning and safety modeling, the ability to integrate these data with more traditional safety data (e.g., crash, roadway) will be paramount for safety data systems. A recent proof-of-concept pilot study of a geospatial data delivery tool by the Federal Highway Administration (FHWA) Highway Safety Information System (HSIS) is an example of this evolution in practice.

Building on this initial geospatial pilot effort by HSIS using data from Charlotte, NC, the main objective of this study was to spatially integrate HSIS data with multi-jurisdictional and emerging datasets to analyze two measures of pedestrian safety performance: the severity of a pedestrian crash that has occurred, and the probability that a crash will occur. To meet the objective, this study explored several emerging areas of interest in the safety data and analysis research community:

- Estimation of pedestrian volumes and pedestrian crash analysis.
- Integration and analysis of probe data to characterize vehicular travel speeds.
- Use of geospatial HSIS data as a tool for addressing critical safety research priorities.

Pedestrian safety, particularly the relationship between a road's characteristics and the likelihood of a pedestrian crash, highly depends on pedestrian activity as a measure of exposure. In general, literature on the topic of pedestrian activity and safety reflects two possible trends: the likelihood of a crash involving a pedestrian is expected to increase with the number of pedestrian crossings (i.e., an exposure effect), or more pedestrian crossings may not increase the likelihood of a pedestrian crash and may actually decrease it when pedestrian activity increases above a certain level due to changes in driver behavior induced by the higher pedestrian activity (i.e., a safety-in-numbers effect). To explore pedestrian exposure effects as part of this study, the project team developed a pedestrian count model before investigating pedestrian safety models. This model allowed the project team to assess the effects of various features on pedestrian safety—such as infrastructure characteristics, traffic volumes, and vehicular speed—while simultaneously considering the level of expected pedestrian activity.

This study collected data from several sources that cover Charlotte, NC, including HSIS, the city of Charlotte, Mecklenburg County, North Carolina Department of Transportation (NCDOT), U.S. Census Bureau, and the Regional Integrated Transportation Information System (RITIS). The project team collected these data to address several potential contributing factors to pedestrian safety outcomes:

- Estimated vehicular speed.
- Pedestrian volumes.
- Roadway geometry and cross section.
- Intersection characteristics.
- Pedestrian infrastructure.
- Traffic volumes.
- Transit access.
- Land use.
- Demographic and socioeconomic profiles.

While traditional methods for analyzing pedestrian safety (e.g., crash trees or crash frequency) can examine the location and context of past crashes, they struggle to discern the magnitude of relationships between components of the transportation system and pedestrian safety. Vehicle speed at the time of impact is a leading contributing factor in determining the severity of a pedestrian crash, but other factors, such as the age of the pedestrian, crossing distance, and traffic volumes, may confound the influence of vehicle speed alone. Therefore, this study sought to integrate HSIS data with a broad spectrum of datasets to provide models that more fully describe the relationship between pedestrian safety and context.

ORGANIZATION OF THE REPORT

The report is structured as follows.

- Chapter 1 describes the motivation and objectives of this study.
- Chapter 2 includes a review of previous research related to the application of probe data in safety analyses, as well as the three primary model components of this study:

  1. Pedestrian volumes.
  2. Pedestrian crash severity.
  3. Pedestrian crash probability.

  These studies have documented key contributing factors to each model type, including factors that drive increased pedestrian activity, higher pedestrian crash severities, and an increased likelihood that a crash will occur at a location.

- Chapter 3 documents the data sources for this study, as well as the methods and assumptions used to integrate these data with HSIS data for analysis.
- Chapter 4 details the analysis methodology, variable selection, and modeling decisions.
- Chapter 5 examines the pedestrian volume model results.
- Chapter 6 describes the pedestrian crash severity and pedestrian crash probability model results.
- Chapter 7 summarizes this study, examines study implications, and makes recommendations for future research.

CHAPTER 2. LITERATURE REVIEW

This chapter highlights a selected set of studies that have investigated contributing factors to pedestrian activity and volumes, pedestrian crash severity, and pedestrian crash probability. The chapter also covers previous attempts to incorporate measures of speed derived from vehicle probe data into crash modeling and analysis.

PEDESTRIAN VOLUMES

Pedestrian activity is often determined by the built environment, proximity of origins and destinations, and the demographic and socioeconomic profile of a neighborhood. Past studies have shown that network factors, such as block length and street connectivity, contribute to higher levels of pedestrian activity and thus higher pedestrian volumes (Griswold et al. 2019; Hampshire et al. 2017). Additionally, proximity of origins and destinations, especially in combination with network connectivity, also contributes to increased pedestrian activity. This function has been measured through the development of origin and destination demand models (Hampshire et al. 2017), as well as metrics of land-use mixing that capture a diversity of land uses within a small geographic radius (Frank, Andresen, and Schmid 2004; Hankey et al. 2012). Nearly all studies also incorporate a socioeconomic profile to refine pedestrian volume estimates, including population and employment density, nonmotorized commuting patterns, median income, educational attainment, and vehicle availability (Griswold et al. 2019; Frank, Andresen, and Schmid 2004; Hampshire et al. 2017; Hankey et al. 2012).

PEDESTRIAN CRASH SEVERITY

Previous studies on factors influencing pedestrian crash severity have noted correlations with almost every dimension of context, including the built environment, traffic volumes, vehicle speed, vehicle size, person-level characteristics, and environmental conditions at the time of the crash.

Road characteristics associated with higher functional classification roads, such as wider vehicle traveled way widths and higher total number of lanes, are commonly associated with more severe pedestrian crashes (Aziz, Ukkusuri, and Hasan 2013; Rothman et al. 2012; Zajac and Ivan 2003). Furthermore, pedestrian crashes tend to be more severe at uncontrolled midblock locations than at intersections, particularly compared with signalized intersections (Aziz, Ukkusuri, and Hasan 2013; Eluru, Bhat, and Hensher 2008; Rothman et al. 2012).

In separate evaluations of unsignalized and signalized intersections, Haleem, Alluri, and Gan (2015) found that higher traffic volumes (especially truck volumes) and adverse weather conditions increased pedestrian crash severity at signalized intersections, whereas crosswalk type affected pedestrian crash severity at unsignalized locations. Posted speed limits, pedestrian ages, and lighting conditions were significant contributing factors at both intersection types, with higher speed limits, older ages, and dark conditions all correlated with more severe outcomes.

Measures of higher vehicle speeds and larger vehicle sizes were commonly associated with more severe pedestrian crashes (Aziz, Ukkusuri, and Hasan 2013; Haleem, Alluri, and Gan 2015; Li et al. 2017; Zajac and Ivan 2003). Posted speed limit is the most common proxy of vehicle speed used in pedestrian crash severity studies, with roadway width and functional classification also serving as surrogate measures of vehicle speed (Rothman et al. 2012).

Pedestrian age was a commonly cited contributing factor to crash severity: crashes involving older pedestrians were more likely to result in a fatality or serious injury compared with crashes involving younger pedestrians (Aziz, Ukkusuri, and Hasan 2013; Eluru, Bhat, and Hensher 2008; Haleem, Alluri, and Gan 2015; Li et al. 2017; Zajac and Ivan 2003). Pedestrian action during the time of the crash was also cited as a contributing factor to severity outcomes, with some studies assigning road user fault as a contributing factor (Haleem, Alluri, and Gan 2015). Other studies noted that different actions, by both the pedestrian and the driver, may result in different severity distributions, depending on context (Aziz, Ukkusuri, and Hasan 2013). Both pedestrian and driver alcohol consumption resulted in more severe pedestrian injury outcomes (Clifton, Burnier, and Akar 2009; Lee and Abdel-Aty 2005; Zajac and Ivan 2003).

Gender represented a point of disagreement in past research. Pitt et al. (1990) found no significant difference between male and female pedestrians in crash severity for pedestrians under the age of 20 yr, whereas other studies have noted that male pedestrian fatalities are proportionally higher in the total population (Rothman et al. 2012). Conversely, Lee and Abdel-Aty (2005) found a marginally significant increase in crash severity for female pedestrians in a study of Florida crashes.

Environmental circumstances have also been observed as significant contributors to pedestrian crash severity. Dark lighting conditions and adverse weather conditions have led to more severe injury outcomes in numerous contexts (Aziz, Ukkusuri, and Hasan 2013; Eluru, Bhat, and Hensher 2008; Haleem, Alluri, and Gan 2015; Lee and Abdel-Aty 2005; Li et al. 2017). Furthermore, studies by Eluru, Bhat, and Hensher (2008) and Lee and Abdel-Aty (2005) found that lighting conditions and other environmental factors could combine with other crash factors (e.g., alcohol consumption) to further increase the likelihood of a fatal or serious injury crash.

PEDESTRIAN CRASH PROBABILITY

Studies that have attempted to analyze contributing factors to pedestrian crash probability and frequency have typically looked at two different contexts: the geographic zone level (e.g., a census tract), or an individual road segment. These studies have shown differing contributing factors based on the chosen unit of analysis.

For zonal analyses, factors that contribute to the likelihood of a crash are typically associated with increased pedestrian activity. Street connectivity, proximity of origins and destinations, and socioeconomic characteristics have been associated with higher pedestrian crash frequency within geographic zones (Mansfield et al. 2018; Moradi et al. 2016; Siddiqui, Abdel-Aty, and Choi 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 and number of lanes were associated with an increase in pedestrian crash frequency. Mansfield et al. (2018) found a significant increase in the likelihood of a fatal pedestrian crash with higher traffic volume density (e.g., vehicle miles traveled per square mile).

Siddiqui, Abdel-Aty, and Choi (2012) noted a relationship between an increase in pedestrian crash frequency and higher posted speed limits.

At the road segment level, the relationship between pedestrian volumes and the number of pedestrian crashes is less direct. A "safety-in-numbers" effect has been observed in several studies, where the relative risk for pedestrian crashes is lower as pedestrian volume increases (Omer et al. 2017). Further study has suggested that specific design elements, such as block length and intersection configuration, may have more nuanced effects on pedestrian safety at the segment level (Dumbaugh, Li, and Joh 2013; Omer et al. 2017). In most studies, higher traffic volumes, measured in terms of annual average daily traffic (AADT), were associated with an increase in the number of pedestrian crashes (Omer et al. 2017; Torbic et al. 2010). Torbic et al. (2010) found that minor-leg AADT was a stronger predictor of crashes at intersections than major-road AADT.

Although sidewalks and other pedestrian infrastructure may produce crash-reduction benefits, they may not be strong predictors of reduced pedestrian crash probability (FHWA 2017). Pande and Abdel-Aty (2009) found that sidewalk presence was positively correlated with the number of pedestrian crashes, meaning that roads with sidewalks had more pedestrian crashes and roads without sidewalks had fewer pedestrian crashes. However, it should be noted that the negative correlation may be related to pedestrian volumes. One is likely to see a confounding exposure effect (e.g., better pedestrian facilities associated with more pedestrian crashes due to higher levels of activity), if a reliable measure of pedestrian exposure is not available.

PROBE DATA INTEGRATION

Probe data are an emerging asset to transportation practitioners and researchers. These promising new data have been applied primarily in a traffic operations context, with relatively few applications to safety research. Kersavage (2019) used INRIX® probe data to assess the relationship between speed metrics and crash risk and frequency (all crash types). Her research found that operating speeds higher than the road segment's average speed or road reference speed were associated with increased crash risk. The average speed refers to the average speed on a segment over the same weekday and hour, and the reference speed is the average speed for all periods and days. Additionally, Kersavage (2019) found that increases in the differences between operating speeds and average speed and differences between operating speed and inferred design speed were associated with an increase in crash frequency. To perform the crash frequency analysis, Kersavage (2019) disaggregated the probe speed data by year and season (i.e., one database row is one season of 1 yr).

In a study of rural roads in Ohio and Washington, Das et al. (2020) found that larger variability between weekday and weekend speeds was associated with higher numbers of crashes. Furthermore, larger hourly variability in operating speeds and higher average operating speeds were significant indicators of higher numbers of crashes on rural roads. The literature review did not uncover any study that had successfully integrated vehicular probe speed data into a pedestrian safety analysis.

SUMMARY

The literature review revealed several types of factors that contribute to the severity and likelihood of pedestrian crashes. These factors span roadway characteristics, vehicle travel, and descriptors of the persons involved in the crash. Most studies had access to variables characterizing the dimensions of the roadway, the presence of pedestrian infrastructure, and the traits of the vehicles, drivers, and pedestrians involved in the crashes, but they often relied on surrogates for vehicle speeds and pedestrian activity. The surrogates for speed were often posted speed limits and functional classification. The surrogates for pedestrian activity were often demographic and socioeconomic characteristics. As new data sources emerge and improve over time, more reliable and direct measures of these contributing factors to pedestrian safety may be obtainable. This study applies more direct measures of speed and pedestrian activity in the form of probe-derived vehicle speed and a pedestrian volume model developed with pedestrian count data to expand on the findings in previous studies.

CHAPTER 3. DATA ACQUISITION AND INTEGRATION

This chapter documents the data sources and elements, and it details the data integration and preparation processes for each type of modeling activity in this study: pedestrian volume modeling, pedestrian crash severity modeling, and pedestrian crash probability modeling.

SOURCES

HSIS is a multiagency database of transportation safety-related information, including road inventories, traffic volumes, and crash records. HSIS datasets span approximately 30 yr and cover seven States and one city:

- California.
- Illinois.
- Maine.
- Minnesota.
- North Carolina.
- Ohio.
- Washington.
- Charlotte.

Data collection for this research focused on Charlotte, building on a previous geospatial pilot effort by HSIS using data from Charlotte. The project team collected data from several sources and integrated these data with geospatial data for Charlotte, that were provide through HSIS (FHWA n.d.). Data sources included the City of Charlotte (2019), Mecklenburg County (n.d.), NCDOT (2021), U.S. Census Bureau (2021), and RITIS (The Eastern Transportation Coalition 2021). The project team collected data in geographic information systems (GIS) and tabular formats. Table 1 lists the data by source and relevant period.

Table 1. Data description and source.

Data Element	Category	Source(s)	Years Accessed
Pedestrian crashes	Crash	HSIS; NCDOT	2014–2018
AADT	Traffic	HSIS; NCDOT	2015–2018
Functional classification	Roadway	NCDOT	2019
One-way/two-way operation	Roadway	HSIS; Charlotte	2018
Number of lanes	Roadway	HSIS; Charlotte	2018
Median presence	Roadway	HSIS; Charlotte	2018
Pavement width	Roadway	Charlotte; NCDOT	2018
Posted speed limit	Roadway	HSIS; Charlotte	2018
Average vehicle speeds	Speed	RITIS	2015–2018
Intersection geometry	Intersection	HSIS; Charlotte	2018
Intersection traffic control	Intersection	HSIS; Charlotte	2018
Pedestrian counts	Pedestrian counts	Charlotte	2011–2020
Bus-stop locations	Transit	HSIS; Charlotte	2018
Light rail-stop locations	Transit	HSIS; Charlotte	2018
Sidewalks	Pedestrian/bicyclist infrastructure	HSIS; Charlotte	2018
Greenways	Pedestrian/bicyclist infrastructure	HSIS; Charlotte	2018
Bicycle lanes	Pedestrian/bicyclist infrastructure	HSIS; Charlotte	2014–2018
PHB	Pedestrian/bicyclist infrastructure	Charlotte	2014–2018
Other pedestrian beacons	Pedestrian/bicyclist infrastructure	Charlotte	2014–2018
Land use	Land use	Mecklenburg County	2011–2016
Total population	Demographic	U.S. Census Bureau	2014–2017
Total employment	Demographic	U.S. Census Bureau	2014–2017
Commute by mode	Demographic	U.S. Census Bureau	2014–2018

PHB = pedestrian hybrid beacon.
Note: The overall period for data analysis is 2014–2018.

COLLECTION

Crash Data

Pedestrian crash data provided by HSIS were compared with crash data made publicly available by NCDOT (2019). At the time that data were collected for this study, NCDOT-provided publicly available data contained some crash report information and a geographic location for all bicycle and pedestrian crashes that occurred between 2007 and 2018. For this study, the project team collected crashes occurring on "non-interstate" facilities between 2014 and 2018. The project team removed pedestrian crashes that occurred in parking lots, on private property, and on interstate highways.

HSIS and NCDOT pedestrian crash datasets were compared for consistency in the number of crash observations, and accuracy of geolocation information. On review, the project team learned that NCDOT data had gone through additional crash-by-crash scrutiny to refine the accuracy of the geolocation of each pedestrian crash. They investigated further instances in which the crashes could have involved a pedestrian, but the crash type field did not include pedestrian crash at the time the data were shared with HSIS. Based on this comparison and analysis of both pedestrian crash datasets, the project team proceeded with the NCDOT database. The project team offers a discussion at the conclusion of this report on whether it is practical for HSIS to monitor such targeted data efforts (which likely vary in topic and scope across HSIS participating agencies) and incorporate resulting changes to specific crashes into previously finalized HSIS data.

To match the availability of other traffic, speed, and roadway data, the project team narrowed the crash dataset to retain crashes that occurred on the following functional classification roads:

- Other freeways and expressways.
- Other principal arterials.
- Minor arterials.
- Major collectors.
- Minor collectors.

While previous studies have removed the other freeways and expressways functional classification due to access control restrictions, many of these facilities in Charlotte, NC (e.g., Independence Boulevard/US 74), have pedestrian accommodations, at-grade intersections, and a history of pedestrian crashes. Only crashes that occurred on interstates and local roads were excluded from this study.

Roadway Data

The project team obtained road centerline data from NCDOT and Charlotte via HSIS. NCDOT data were used to identify the relevant functional classification roads in the Charlotte area (excluding interstates and local roads). HSIS centerline data contained relevant operational, geometric, and posted speed-limit information. All geometric data were stored in a bidirectional format, allowing the project team to characterize a full cross section for each segment.

Traffic Data

Bidirectional AADT values were collected from NCDOT for all available years between 2015 and 2018. All available AADT values were assigned to each road segment for all available years. If AADT had not been collected on a road segment for a particular year, gaps were filled according to the following process:

- If a previous year of AADT on the road segment was available, the AADT value from the most recent previous year was applied to a gap year. For instance, if 2016 AADT were unavailable for a road segment, 2015 data for that road segment would be used instead.

- If no previous AADT was available, the AADT value from the most recent following year was applied to the gap year. For instance, 2015, 2016, 2017, or 2018 (whichever was the earliest year available) for a road segment would be applied to 2014 for that road segment.

Speed Data

Speed data consisted of probe data provided by HERE and accessed through RITIS. RITIS is housed and managed by the University of Maryland's Center for Advanced Transportation Technology (CATT) Laboratory (CATT Lab 2020). NCDOT has access to RITIS through its membership in the I–95 Corridor Coalition (I–95 Corridor Coalition 2018) and provided permission for the project team to use the data for this research.

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. The data can be queried based on road type or location, day of the week, and time of day. Average speed, calculated as the harmonic mean speed, is available at various frequencies, with 5 min being the most granular data available. The data are available via traffic message channel (TMC) roadway segmentation specific to the probe data. These TMC segments vary in length, with the beginning and ending of each segment corresponding to specific locations on the roadway. A confidence score is assigned to the average speed within each time interval, which is used to represent how much of the values reported are directly from vehicles observed traveling on the roadway versus calculated using historical trend data. The use of historical trend data may be necessary in rural locations and mountainous areas with poor signal penetration. It may also be necessary to assist with the removal of outliers.

In this study, the probe data downloaded from RITIS were aggregated at intervals of 1 h to remove random "noise" in the data while still representing a roadway's average speed during periods of interest. The project team used a threshold for confidence scores of over 70 to best represent data collected from vehicles during the periods that were queried. The project team matched the TMC segmentation to the NCDOT road segmentation. NCDOT's GIS team had previously conflated the probe data to the NCDOT roadway network.

Intersection Data

The project team obtained intersection locations from Charlotte via HSIS. These data also contained traffic control information. Intersection approach geometry (i.e., number of legs) was derived from the combination of roadway centerline data and intersection points using GIS geoprocessing tools.

Pedestrian Count Data

Charlotte collects pedestrian counts simultaneously with turning movement counts at intersections throughout the city. These counts disaggregate crossings by approach at each intersection. The project team obtained 13-h pedestrian counts (6 a.m. to 7 p.m.) using the city's traffic management system for 2011 through 2020 (City of Charlotte 2020). Since these data do not exist for every intersection in the city, the project team devised a data collection plan to obtain a representative sample of pedestrian counts according to the proportion of centerline miles in the city by functional classification (interstates and local roads excluded). Table 2 details the distribution of collected pedestrian counts by functional classification.

Table 2. Road centerline mileage and pedestrian count locations by functional classification in Mecklenburg County, NC.

Functional Classification	Total Mileage in Data (Percent)*	Pedestrian Count Observation Coverage (Percent)*
Major arterial	28.7	37.3
Minor arterial	36.7	50.1
Collector	34.5	40.2

*Interstate and local road mileage are not included in the data for this study.
Note: Major arterial classification includes other freeways and expressways and other principal arterials; collector classification includes major and minor collectors.

As table 2 details, minor arterials make up most of the study's centerline mileage and cover more of the study's pedestrian count observations than the other functional classifications. The percentages in the final column of table 2 will not sum to 100 percent, as more than one functional classification road can approach a single intersection.

Transit Data

Transit service in Charlotte consists of on-street bus service and light rail. The project team obtained Charlotte Area Transit System bus and light rail-stop information for 2018. The project team assumed that bus service experienced negligible changes over the 2014–2018 study period. However, Charlotte expanded its light rail service, the Lynx Blue Line, in March 2018. This change was noted for all 11 stations included in the Blue Line extension, and all datasets reflected the absence or presence of these stations for the appropriate periods.

Multimodal Infrastructure Data

The project team obtained bicycle and pedestrian infrastructure data from HSIS and Charlotte. These data included the presence of sidewalks, off-street greenways, bicycle lanes (both striped and separated), and pedestrian crossing beacons. Shared lane markings ("Sharrows") were not included in the dataset as a designated bicycle lane, and a distinction was made between pedestrian hybrid beacons (PHBs) and any other nonregulatory warning signals for crossing pedestrians. Charlotte does not typically install rectangular rapid flashing beacons, and these location data were unavailable.

Land-Use Data

The project team obtained land-use data from Mecklenburg County, NC, at the parcel level in GIS format.

Demographic Data

The project team obtained population, employment, and commuter mode share data from the U.S. Census Bureau. These data included two primary sources, the American Community Survey (ACS) and the Longitudinal Employer-Household Dynamics (LEHD) program. Population and commute mode share were collected from ACS 5-yr estimates for surveys between 2009–2013 and 2013–2017. Total employment was collected from the LEHD Origin-Destination Employment Statistics version 7 database (U.S. Census Bureau n.d.). The project team obtained all demographic data at the census block group level for Mecklenburg County and surrounding counties in North Carolina and South Carolina. To obtain 2018 values for population and employment, the project team determined the percent change in both datasets between 2016 and 2017 and applied that rate of growth to the 2017 observed values.

INTEGRATION

The project team prepared datasets for three modeling efforts: pedestrian volume modeling, pedestrian crash severity modeling, and pedestrian crash probability modeling. This section describes the methods and assumptions used to develop datasets for each modeling effort.

Pedestrian Volume

Pedestrian volume was quantified at the intersection level to match the unit of observation of existing pedestrian count data: total pedestrian counts at intersections over a 13-h period (6 a.m. to 7 p.m.) conducted by Charlotte. These 13-h counts were used as the dependent variable in the pedestrian volume modeling. Data were collected and aggregated to individual intersections with pedestrian count data in a GIS format. Table 3 is a description of all data collected to develop the pedestrian volume model. Model results produced a predicted average 13-h pedestrian count between 6 a.m. and 7 p.m. as a function of intersection characteristics, including surrounding context.

Table 3. Pedestrian volume data dictionary.

Data Element	Data Type	Description
Major arterial	Indicator	Intersection includes a leg classified as other freeways and expressway or other principal arterial.
Minor arterial	Indicator	Intersection includes a leg classified as a minor arterial.
Collector	Indicator	Intersection includes a leg classified as a major or minor collector.
Four-leg intersection	Indicator	Intersection geometry is a four-leg intersection.
Signalized intersection	Indicator	Intersection traffic control device is a traffic signal.
Speed 25 mph	Indicator	Intersection includes a posted 25-mph leg.
Speed 30 mph	Indicator	Intersection includes a posted 30-mph leg.
Speed 35 mph	Indicator	Intersection includes a posted 35-mph leg.
Speed 40 mph	Indicator	Intersection includes a posted 40-mph leg.

Data Element	Data Type	Description
Speed 45 mph	Indicator	Intersection includes a posted 45-mph leg.
Speed 50 mph	Indicator	Intersection includes a posted 50-mph leg.
Speed 55 mph	Indicator	Intersection includes a posted 55-mph leg.
Sidewalk present	Indicator	Sidewalk exists within 100 ft of the intersection.
Greenway present	Indicator	Greenway exists within 100 ft of the intersection.
Bus stops	Continuous number	Number of bus stops within 0.1-mi radius.
Light rail stops	Continuous number	Number of light rail stops within 0.1-mi radius.
Total population	Continuous number	Total population; ACS 5-yr estimates—year of count.
Total employment	Continuous number	Total employment; ACS 5-yr estimates—year of count.
Total households	Continuous number	Total households; ACS 5-yr estimates—year of count.
Total commuters	Continuous number	Total commuters; ACS 5-yr estimates—year of count.
Total walking commuters	Continuous number	Total walking commuters; ACS 5-yr estimates—year of count.
Total transit commuters	Continuous number	Total transit commuters; ACS 5-yr estimates—year of count.
Total zero-vehicle households	Continuous number	Total zero-vehicle households; ACS 5-yr estimates—year of count.
Total college-educated population	Continuous number	Total population over 25 yr of age with an Associate's degree or higher educational attainment; ACS 5-yr estimates—year of count.
Proportion of walking commuters	Continuous number	Proportion of total commuters who walk to work; ACS 5-yr estimates—year of count.
Proportion of transit commuters	Continuous number	Proportion of total commuters who take public transit; ACS 5-yr estimates—year of count.
Proportion of zero-vehicle households	Continuous number	Proportion of households without a motor vehicle; ACS 5-yr estimates—year of count.
Proportion of college-educated population	Continuous number	Proportion of population over 25 yr of age with an Associate's degree or higher educational attainment; ACS 5-yr estimates—year of count.
Land-use mix	Ratio	Modified version of land-use mixing variable first described by Frank, Andresen, and Schmid 2004 (figure 2).

mph = miles per hour.

Roadway Data

Centerline data were directly joined to each intersection using geoprocessing tools in a GIS format. This process allowed the project team to determine the functional classification, posted speed limit, and total number of approach legs for each intersection.

Socioeconomic Data

The project team collected demographic and socioeconomic data at the block group level and aggregated using ⅒-mi, ¼-mi, and ½-mi radii geographic buffer distances. Previous studies have suggested that these distances provide the most effective predictors of pedestrian traffic at the intersection level (Griswold et al. 2019). All socioeconomic variables were proportionally assigned to an intersection based on the overlap of the buffer area and the block group zone. For instance, if an intersection buffer overlapped 50 percent of a census block group, 50 percent of the socioeconomic data (population, employment, households) would be aggregated to that intersection area. The total of these overlaps produced the estimates of demographic and socioeconomic variables at each intersection.

Land-Use Mix

A dense mix of different land uses provides the proximity of origins and destinations that encourage pedestrian travel. Using parcel-level, land-use information from Mecklenburg County, NC, this study employed an approximation of land-use mix (figure 2) described by Frank, Andresen, and Schmid (2004).

$$Land\ Use\ Mix = -\sum_{i=1}^{n} p_i \frac{\ln p_i}{\ln n}$$

Figure 2. Equation. Land-use mix from Frank, Andresen, and Schmid (2004).

Where:

p_i = proportion of estimated square footage attributed to land use i.
n = number of land uses within 1 km.

This metric assesses the distribution of four land-use types—residential, commercial, office, and institutional—within a 1-km radius of an intersection. A totally uniform land use within the 1-km buffer would produce a value of “0,” whereas a completely even distribution of all four land uses would produce a value of “1.” Unlike the Frank, Andresen, and Schmid (2004) variable, this study used a straight-line, Euclidean distance buffer around each intersection rather than a network distance buffer. This practice matched the buffer methodology applied to socioeconomic and demographic data preparation. Figure 3 and figure 4 compare an example of a low-land-use mix location and a high-land-use mix location.

Original photo: © 2019 NC OneMap. Annotated by FHWA (see Acknowledgments section).

Figure 3. Photograph. Low-land-use mix example (land-use mix = 0.129) (NC OneMap 2019).

Original photo: © 2019 NC OneMap. Annotated by FHWA (see Acknowledgments section).

Figure 4. Photograph. High-land-use mix example (land-use mix = 0.955) (NC OneMap 2019).

Pedestrian Crash Severity

This section details the data preparation process for the pedestrian crash severity modeling. Table 4 presents the data dictionary for all independent variables included in the severity model. The pedestrian crash severity data are structured so that one row of the database represents one pedestrian crash, including characteristics associated with the crash that could influence the severity outcome. The pedestrian crashes in the crash severity dataset include both midblock- and intersection-related crashes. A predicted level of pedestrian activity is also included for each crash location, which is derived from the pedestrian volume model described in chapter 4.

Table 4. Pedestrian crash severity data dictionary.

Data Element	Data Type	Description
Highest AADT	Continuous number	Average AADT—highest AADT within 100 ft, year of the crash.
Lowest AADT	Continuous number	AADT—lowest AADT within 100 ft, year of the crash.
Closest AADT	Continuous number	AADT—closest road, year of the crash.
Predicted average pedestrian volume	Continuous number	Value after applying the pedestrian volume model to make a pedestrian volume prediction at the crash location.
Average speed	Continuous number	Average speed, in mph, based on hour of day of the week (RITIS; 2018 values).
Reference speed	Continuous number	Average speed, in mph, based on all times of every day of the week (RITIS; 2018 values).
Hourly speed-to-reference ratio	Continuous number	Ratio of average speed to reference speed.
One way	Indicator	Indicates the road is one way.
Divided/undivided	Indicator	Indicates the road is median divided.
Number of lanes	Continuous number	Total number of lanes.
Major arterial road	Indicator	Crash is within 100 ft of a road classified as other freeways and expressways or other principal arterial.
Minor arterial road	Indicator	Crash is within 100 ft of a road classified as minor arterial.
Collector road	Indicator	Crash is within 100 ft of a road classified as a major or minor collector.
Pavement width	Continuous number	Paved area of road segment (lane widths and shoulder widths).
Presence of a bicycle lane	Indicator	Crash occurred on a street with a designated bicycle lane.
Presence of a sidewalk	Indicator	Crash occurred near a sidewalk (<100 ft).
Presence of a greenway	Indicator	Crash occurred near a greenway (<100 ft).

Data Element	Data Type	Description
Intersection related (100 ft)	Indicator	Crash occurred within 100 ft of an intersection.
Signalized intersection related (100 ft)	Indicator	Intersection is signalized (if intersection related).
Four-leg intersection (100 ft)	Indicator	Intersection has four legs (if intersection related).
Intersection related (250 ft)	Indicator	Crash occurred within 250 ft of an intersection.
Signalized intersection related (250 ft)	Indicator	Intersection is signalized (if intersection related).
Four-leg intersection (250 ft)	Indicator	Intersection has four legs (if intersection related).
PHB	Indicator	Crash is within 100 ft of a PHB.
Other pedestrian crossing beacon	Indicator	Crash is within 100 ft of another pedestrian crossing beacon (over/under 12-inch signal head).
Bus stop adjacent	Indicator	Bus stop within 0.1-mi radius.
Posted speed limit	Continuous number	Posted speed limit on road.
Large-vehicle type	Indicator	Vehicle is flagged as a “light truck, commercial bus, sport utility, sport utility, van, pickup, single unit truck, other bus, tractor/trailer/truck” in the crash report.
Driver impairment	Indicator	Driver flagged for suspected or detected alcohol/drug impairment.
Age	Continuous number	Age of the pedestrian.
Gender	Indicator	Gender of the pedestrian (male = 1, female/unknown = 0).
Pedestrian impairment	Indicator	Pedestrian flagged for suspected or detected alcohol/drug impairment.
Lighting type—dark, not lighted	Indicator	Dark lighting conditions, unlit, or unknown lighting roadway.
Lighting type—dark, lighted	Indicator	Dark lighting conditions, lit roadway.
Lighting type—daylight	Indicator	Crash occurred during daylight, dusk, dawn, or unknown conditions.
Weather—rain	Indicator	Weather is indicated as “rain.”
Weather—snow	Indicator	Weather is indicated as “snow.”
Road condition	Indicator	Road condition indicated to be wet, snowy, or icy.
Work zone	Indicator	Crash occurred in a work zone.

Pedestrian Volume

Although not included in table 4, all variables that appear as predictor variables in the pedestrian volume model described in chapter 4 were also collected for the individual crash locations. These variables were then used to predict a pedestrian count at the crash location. The prediction represents an average 13-h pedestrian count between 6 a.m. and 7 p.m. as a function of intersection characteristics, including the surrounding context. For the purposes of pedestrian volume prediction, midblock crash locations were treated as unsignalized intersections.

Vehicular Speed

Vehicular speed was linked to crashes through the conflation of the RITIS probe data to NCDOT's road segment data. The project team derived three variables from available vehicle speeds: average speed, reference speed, and the ratio of average to reference speed.

- Average speed refers to the average vehicle speed for the specific hour of the day and day of the week that a crash occurred. For instance, if a crash occurred at 6:30 p.m. on a Thursday, the average speed variable refers to the average travel speed on that road segment between 6 p.m. and 7 p.m. on Thursdays.

- Reference speed refers to the average vehicle speed on that segment of road for all times of day for every day of the week.

- Speed ratio refers to the ratio of average speed to reference speed. Values greater than one indicate faster than longer-term average conditions, while values less than one would indicate more congested, lower than longer-term average speed conditions.

All speed data for the crash severity model data reference 2018 values. A total of 615 of 1,422 eligible pedestrian crashes (43 percent) received speed values from the probe data. To investigate the validity of the probe data, the project team compared the average speed ratio over the time of day (table 5), as well as reference speed values, for each functional classification (table 6).

Table 5. Speed ratio by period.

Period	Average Speed Ratio (mph)
Overnight: 10 p.m. to 6 a.m.	1.15
Morning/midday: 6 a.m. to 2 p.m.	1.00
Afternoon/evening: 2 p.m. to 10 p.m.	0.96

Table 6. Average reference speed by functional classification.

Functional Class	Average Reference Speed (mph)
Major arterial	30.55
Minor arterial	26.26
Collector	23.24

The speed ratios in table 5 show greater travel speeds during the overnight period, with lower than average speeds in the afternoon and evening. This result suggests that the probe data are generally a good indicator of congested and free-flow conditions during the time of a crash. Furthermore, table 6 shows a logical increase in average reference speeds based on higher functional class roads. Based on this assessment, the project team determined the RITIS probe speed data to be acceptable for exploration as part of crash severity modeling.

Roadway and Traffic

Roadway and traffic data were linked to crashes according to spatial proximity in GIS. Each crash received the roadway attributes from the closest road, which is intended to correlate with the cross section the pedestrian was crossing or walking along. Conversely, crashes received traffic information from both the closest road, as well as any other road within 100 ft. This action was intended to account for cross-street traffic information at intersections, as well as additional turning movement conflicts that do not exist midblock.

Pedestrian Crossing Beacons

Pedestrian crossing beacons were separated into PHB locations and all other pedestrian-specific crossing beacons. Installation dates were provided as a part of the Charlotte dataset. Beacon locations were only associated with crashes if the crash occurred after the installation date.

Crash Data

Crash report data related to person-level and environmental conditions at the time of the crash were included in the modeling process. If a crash report indicated a crash occurred in a work zone, that crash was dropped from the dataset before modeling.

Pedestrian Crash Probability

This section details the data preparation process for pedestrian crash probability modeling. The pedestrian crash probability data are structured so that one row of the database represents one road segment. Each road segment may or may not have had pedestrian crashes. Study area segments were defined as continuous road centerlines with homogenous characteristics. The criteria used to define homogenous segments were as follows:

- Functional classification.
- Number of lanes.
- Median presence.
- Speed limit.
- AADT.
- One-way vs. two-way traffic operation.

All centerlines on functional classification roads between principal arterials and minor collectors were collected, and eligible continuous segments had a minimum length of 300 ft. This practice allowed small downtown blocks to be included in the probability modeling while very small, ephemeral changes due to roadway and traffic conditions were removed. Less than 1 percent of potential study area road mileage within Charlotte was removed for not meeting the 300-ft threshold. Data were aggregated to study area segments in GIS.

Given the structure of the crash probability dataset, the project team applied the pedestrian volume model described in chapter 4 to predict average pedestrian volumes for each homogenous road segment. The project team used the pedestrian volume model to develop a pedestrian volume prediction for each intersection along a segment, then summed those predictions to arrive at a pedestrian volume for the entire segment.

Data were first collected for each road segment by year, creating a dataset of segment-years between 2014 and 2018. This action allowed the project team to incorporate changes to the built environment over time. Segment-years were then combined before modeling, and all variables characterizing each segment were averaged to generate the segment values. Table 7 presents the data dictionary for all independent variables included in the probability modeling process.

Table 7. Pedestrian crash probability data dictionary.

Data Element	Data Type	Description
Vehicle traffic (AADT)	Continuous Number	AADT by year (if year AADT missing, see the Collection section for assumptions).
Predicted average pedestrian volume	Continuous number	Predicted value after applying the pedestrian volume model to make pedestrian volume predictions at each intersection along a homogenous segment, then summing the predictions across all intersections along the segment.
Peak period speed—a.m.	Continuous number	Average speed for observation period (7–9 a.m.); uses speed for the direction of travel used to generate the speed ratio—a.m. variable.
Peak period speed—p.m.	Continuous number	Average speed for observation period (4–6 p.m.); uses speed for the direction of travel used to generate speed ratio—p.m. variable.
Overnight speed	Continuous number	Average speed for observation period (11 p.m. to 3 a.m.).
Speed ratio—a.m.	Continuous number	Ratio of peak period speed—a.m. to overnight speed.
Speed ratio—p.m.	Continuous number	Ratio of peak period speed—p.m. to overnight speed.
One way	Indicator	Indicates the road is one way.
Divided/undivided	Indicator	Indicates the road is median divided.
Number of lanes	Continuous number	Total number of lanes.
Major arterial road	Indicator	Road is classified as other freeways and expressway or other principal arterial.
Minor arterial road	Indicator	Road is classified as minor arterial.
Collector road	Indicator	Road is classified as major or minor collector.
Presence of a bicycle lane	Indicator	Crash occurred on a street with a bicycle lane; considered present during year of installation.

Data Element	Data Type	Description
		Segments with bicycle lanes installed during the study period would receive a value between 0 and 1.
Presence of a sidewalk—single coverage	Continuous number	Proportion of the continuous segment with a sidewalk on at least one side.
Presence of a sidewalk—dual coverage	Continuous number	Total length of sidewalk (both sides combined) divided by twice the segment length; complete sidewalk on one side only = 0.5, complete sidewalks on both sides = 1.
Presence of a greenway	Indicator	Segment is within 100 ft of a separated greenway facility.
Total intersection count	Continuous number	Total number of intersections along segment.
Total signalized intersection count	Continuous number	Total number of signalized intersections along segment.
PHB	Indicator	Crash is within 50 ft of a PHB; considered present if installed before July during year of installation. Segments with PHBs installed during the study period (i.e., indicator is 0 for some years, 1 for others) would receive a value between 0 and 1 when the segment-years are combined and variables averaged.
Other pedestrian crossing beacon	Indicator	Crash is within 50 ft of another pedestrian crossing beacon (over/under 12-inch signal head); considered present if installed before July during year of installation. Segments with other pedestrian crossing beacons installed during the study period (i.e., indicator is 0 for some years, 1 for others) would receive a value between 0 and 1 when the segment-years are combined and variables averaged.
Length of segment	Continuous number	Length of the segment in miles.
Posted speed limit	Continuous number	Posted speed limit on road.
Bus stop adjacent	Continuous number	Total number of bus stops (<100 ft).
Light rail stop adjacent	Continuous number	Total number of at-grade light rail stops (<100 ft).

Crash Data Linkage

Pedestrian crashes were joined to the respective study segments using a 100-ft buffer. If a pedestrian crash was located within 100 ft of more than one segment (e.g., at an intersection), that crash was joined to the closest continuous segment. The project team summed crash totals between 2014 and 2018 for each segment and then developed a binary variable representing the occurrence of no pedestrian crashes versus one or more pedestrian crashes.

Pedestrian Volume

Pedestrian counts were predicted for each intersection along a segment using the pedestrian volume model and then summed to produce a segment-wide estimate of pedestrian volume. The project team produced a predicted pedestrian count for each year between 2014 and 2018 based on changes to socioeconomic and demographic estimates. If more than one segment shared an intersection, the full predicted pedestrian count was applied equally to all associated segments; the predicted pedestrian counts were not divided between study area segments.

Vehicular Speed

The project team obtained speed data from RITIS for all study area segments for which the speed data were available. Average speeds were developed for three periods, the a.m. peak period (7 to 9 a.m.), the p.m. peak period (4 to 6 p.m.), and an overnight period (11 p.m. to 3 a.m.). Each segment received a speed value for these periods based on the average speeds obtained for the weekdays of Tuesday through Thursday during the second and third full weeks in March. These values were derived annually for each year between 2014 and 2018. The choice of study period was determined for two principle reasons:

1. Tuesdays, Wednesdays, and Thursdays are considered typical operational days for traffic studies.

2. The second and third full weeks in March represented a period in which school was in session and no meaningful breaks or holidays occurred just before these weeks, according to the Charlotte-Mecklenburg Schools academic calendar.

The project team derived five variables from speed data: average speeds for the morning, afternoon, and overnight periods, as well as the ratios of the morning and afternoon peak period speeds to the overnight speed. For data that only had a single direction of travel available, the project team collected data for the relevant periods and determined the ratios of peak period speed to the overnight condition. If data were available for both travel directions, the project team determined the travel direction that produced the lowest ratio of peak period speed to overnight speed and applied those values to the segment. The project team determined that the lowest ratio likely indicated the dominant direction of travel for a specific peak period. This period, therefore, was ideal for comparing differences in congested and free-flow conditions. The project team averaged overnight speeds for both directions on segments with bidirectional data.

The project team investigated the speed data for logical and consistent observations. During this process, several inconsistencies were noted:

- For some segments, a.m. and p.m. peak speed values were higher than the reported speed during overnight periods. This situation was questionable, as it inferred that peak hour traffic moved faster than traffic during overnight free-flow conditions.

- For roads with bidirectional speeds available, travel speeds in one direction would greatly exceed travel speeds in the opposing direction for all observed periods. Sometimes this discrepancy would represent a 40- to 50-percent difference in travel speeds. While large differences in travel speed by direction are to be expected, particularly during peak hour congestion, the consistency of this discrepancy by direction over all periods was questionable.

For these reasons, as well as a lack of all coverage for all segments, the project team had greater success incorporating measures of speed from the probe data into the pedestrian crash severity models than in the pedestrian crash probability models.

Sidewalk Presence

Sidewalk presence was measured by two variables, each intended to assess the coverage over the length of a segment. Single coverage determined the proportion of a segment with at least one sidewalk on either side; it did not discern whether one or both sides of the street had a sidewalk over this distance. Dual coverage makes a more refined measurement. This variable sums the total length of sidewalk on both sides of a segment and divides this number by twice the overall segment length. If a study area segment had a sidewalk on only one side of its entire length, it would produce a dual coverage value of 0.5; if the same segment had complete sidewalk coverage on both sides, it would have a dual coverage value of 1.

Bicycle Lane Presence

Bicycle lanes were assigned to segments based on the year of installation. If a bicycle lane were installed during the study period, the lane would be considered present during its installation year as well as for all subsequent years.

Pedestrian Crossing Beacons

Pedestrian crossing beacons, defined as PHB and other beacon types, were joined to each segment in GIS. Like bicycle lanes, these features were considered present if they were installed before July of their installation year as well as for all subsequent years.

Transit Stops

Transit stops were joined to study area segments using a 100-ft buffer. Only at-grade light rail stations were considered in the analysis, and LYNX Blue Line extension stations were only applied to 2018 segment-years.

FINAL DATA

This section provides summary statistics for all variables included in each model developed for this study. Table 8, table 9, and table 10 only provide the variables included in each model in the appropriate form.

Table 8. Pedestrian volume model summary statistics for variables in model.

Variable	Obs	Mean	SD	Min	Max
Indicator variable for year of count (1 = yr 2014; 0 otherwise)	485	0.14	0.3	0	1
Yr 2015	485	0.13	0.3	0	1
Yr 2016	485	0.13	0.3	0	1
Yr 2017	485	0.11	0.3	0	1
Yr 2018	485	0.21	0.4	0	1
Yr 2019	485	0.16	0.4	0	1
Indicator variable with value of 1 if lowest posted speed for roads at intersection is 25 or 30 mph; 0 otherwise	485	0.4	0.5	0	1
Indicator variable with value of 1 if sidewalk is present; 0 otherwise	485	0.8	0.4	0	1
Indicator variable with value of 1 if intersection is signalized; 0 otherwise	485	0.8	0.4	0	1
Indicator variable with value of 1 if a bus stop is present within 0.1-mi radius; 0 otherwise	485	0.7	0.5	0	1
Land-use mix ratio	485	0.6	0.2	0	1
Indicator variable with value of 1 if the intersection has an arterial (major and/or minor) approach leg; 0 otherwise	485	0.8	0.4	0	1
Indicator variable with value of 1 if the intersection has four approach legs; 0 otherwise	485	0.9	0.3	0	1
Total transit commuters within 0.5-mi radius	485	65	62.7	0	267
Natural log of total population within 0.5-mi radius	485	7.6	0.9	0	9
Total employment within 0.5-mi radius	485	5,882.6	11,145.6	23.6	65,187

Obs = number of observations; SD = standard deviation; Min = minimum; and Max = maximum.

Table 9. Crash severity model summary statistics for variables in model.

Variable	Obs	Mean	SD	Min	Max
Lowest AADT within 100 ft (vehicles per day divided by 1,000)	1,378	23.2	13.1	0.5	112
Indicator variable with value of 1 if crash involved a large vehicle type; 0 otherwise	1,378	0.3	0.5	0	1
Indicator variable with value of 1 if pedestrian impaired; 0 otherwise	1,378	0.1	0.3	0	1
Indicator variable with value of 1 if crash occurred in daylight; 0 otherwise	1,378	0.6	0.5	0	1
Indicator variable with value of 1 if pedestrian was 50 yr old or older; 0 otherwise	1,378	0.3	0.5	0	1
Average speed	590	27.5	8.2	7.4	57.8
Indicator variable with value of 1 if intersection is signalized (<250 ft) (if intersection related); 0 otherwise	1,378	0.5	0.5	0	1
Pedestrian volume	1,378	438	1,087	2.2	7,240.4
Indicator variable with value of 1 if crash is intersection related (<100 ft); 0 otherwise	1,378	0.7	0.5	0	1
Indicator variable with value of 1 if a crash occurred on a four-plus lane divided road; 0 otherwise	1,378	0.4	0.5	0	1
Indicator variable with value of 1 if the road had a posted speed limit 45 mph or higher; 0 otherwise	1,378	0.4	0.5	0	1

Table 10. Crash probability model summary statistics for variables in model.

Variable	Obs	Mean	SD	Min	Max
Predicted pedestrian volume (predicted average 13-h pedestrian counts)	1,619	523.2	1,498.7	3.3	16,735.9
Segment length (mi)	1,619	0.3	0.3	0.1	2.1
AADT (vehicles per day) and predicted pedestrian volume interaction	1,619	15,565.7	32,427.0	115.6	370,944.9
Speed limit 25–35 mph and median presence interaction	1,619	0.1	0.3	0	1
Indicator variable with value of 1 if posted speed limit is between 25 and 35 mph; 0 otherwise	1,619	0.5	0.5	0	1
Number of lanes	1,619	3.7	1.3	1	8

CHAPTER 4. PEDESTRIAN VOLUME RESULTS

This chapter describes the development of a pedestrian volume model. The pedestrian volume model predicts average 13-h pedestrian volumes at intersections as a function of intersection characteristics (including the context surrounding the intersection). The project team used the model to predict pedestrian volumes at intersections and other pedestrian crash locations and explore the predictions as part of the pedestrian crash severity and pedestrian crash probability modeling. The sections in this chapter describe the pedestrian volume modeling technique, variable selection process, and results.

MODEL DEVELOPMENT AND FORM

The pedestrian volume model was developed using a negative binomial count regression model. The dependent variable in the model is the pedestrian count over a 13-h period, making a count model appropriate for the data. Negative binomial regression is preferred to other count regression models as it can account for overdispersion, which occurs when the variance exceeds the mean of the observed data.

The functional form of the negative binomial regression model is shown in figure 5 (Lord and Mannering 2010).

$$\lambda_i = e^{\beta X_i + \varepsilon_i}$$

Figure 5. Equation. Negative binomial regression functional form.

Where:

e^{ε_i} = gamma distributed error term, where e^{ε_i} is gamma distributed with a mean equal to 1 and variance equal to α.
λ_i = expected number of pedestrians at location i.
β = vector of estimated parameters.
X_i = vector of independent variables that characterize location i and influence pedestrian volumes.

The variance in the number of pedestrians at location i is shown in figure 6 (Lord and Mannering 2010).

$$VAR[y_i] = E[y_i]\big[1 + \alpha E[y_i]\big] = E[y_i] + \alpha E[y_i]^2$$

Figure 6. Equation. Negative binomial regression variance calculation.

Where:

$E[y_i]$ = expected number pedestrians at location i.
$VAR[y_i]$ = variance in number of pedestrians at location i.
α = overdispersion parameter.

The form of the negative binomial distribution is shown in figure 7 (Washington, Karlaftis, and Mannering 2011).

$$P(y_i = y) = \frac{\Gamma\left((1/\alpha) + y\right)}{\Gamma(1/\alpha)y!}\left(\frac{1/\alpha}{(1/\alpha) + \lambda_i}\right)^{1/\alpha}\left(\frac{\lambda_i}{(1/\alpha) + \lambda_i}\right)^{y}$$

Figure 7. Equation. Negative binomial probability function.

Where:

$P(y_i = y)$ = probability of observing number of pedestrians equal to y at location i.

Γ = gamma function.

VARIABLE SELECTION

Variable selection for the pedestrian volume model included an examination of the descriptive statistics of available data to determine variability in the data. Sufficient variability needs to be present in the variables to be included in the models. For example, if a variable is categorical, enough observations should be available for each category to be considered for the model. Without enough observations, model results could be influenced by only a few observations in one or more categories. Similarly, continuous variables should show variability across ranges of values for which the model will apply.

After individual variable statistics were examined, the project team applied a forward selection process to develop the pedestrian volume model. Forward selection includes starting with a model with no variables, then gradually and progressively adding variables until the model no longer improves (Agresti 2007). After each variable is added to the model, it is tested to examine the effect. This test assesses how each variable improves model fit and the statistical significance of the model and remaining variables, including the magnitudes of the effects and significance of that effect size. The p-values were checked for variable statistical significance, and the pseudo R^2 and overdispersion values were checked for model fit and explanatory value. The project team also examined correlations between variables. Variable selection and model development are complete once the model no longer significantly improves with the addition of any other variables.

The variable selection process included the variables described in chapter 3 (Data Acquisition and Integration) and variable transformations. Variable transformations were reviewed and considered, including transforming continuous variables and adjusting categories for categorical variables.

PEDESTRIAN VOLUME MODEL RESULTS

The estimation results for the negative binomial regression model of pedestrian volumes are in table 11.

Table 11. Negative binomial regression model for pedestrian volume.

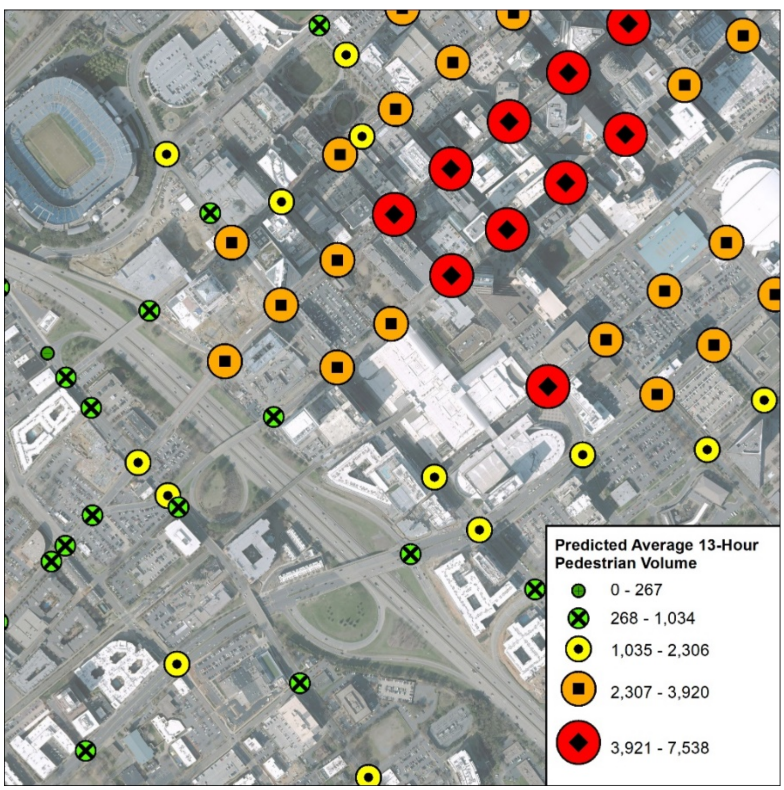
Variable	Coefficient	Standard Error	z-Value	$P>\|z\|$
Indicator variable for year of count (1 = yr 2014; 0 otherwise)	−0.0627	0.1503	−0.42	0.677
Yr 2015	−0.0838	0.1544	−0.54	0.587
Yr 2016	−0.1918	0.1567	−1.22	0.221
Yr 2017	0.0724	0.1617	0.45	0.654
Yr 2018	−0.0992	0.1463	−0.68	0.498
Yr 2019	−0.0404	0.1468	−0.27	0.783
Indicator variable with value of 1 if lowest posted speed for roads at intersection is 25 or 30 mph; 0 otherwise	0.3133	0.0915	3.43	0.001
Indicator variable with value of 1 if sidewalk is present; 0 otherwise	0.6682	0.1335	5.01	<0.001
Indicator variable with value of 1 if intersection is signalized; 0 otherwise	0.3523	0.1086	3.24	0.001
Indicator variable with value of 1 if a bus stop is present within 0.1-mi radius; 0 otherwise	0.4358	0.0916	4.76	<0.001
Land-use mix ratio	1.2306	0.2240	5.49	<0.001
Indicator variable with value of 1 if the intersection has an arterial (major and/or minor) approach leg; 0 otherwise	−0.6090	0.1058	−5.76	<0.001
Indicator variable with value of 1 if the intersection has four approach legs; 0 otherwise	0.6272	0.1778	3.53	<0.001
Total transit commuters within 0.5-mi radius	0.0034	0.0007	4.54	<0.001
Natural log of total population within 0.5-mi radius	0.3508	0.0551	6.37	<0.001
Total employment within 0.5-mi radius	0.0001	0.000005	12.03	<0.001
Constant	−0.8621	0.4335	−1.99	0.047
α	0.6728	0.0428	N/A	N/A

N/A = not applicable.

Note: Number of observations = 485; log likelihood = −2621.344; pseudo R^2 = 0.1367; likelihood ratio (LR) $\chi^2(16)$ = 830.41; probability > χ^2 < 0.0001.

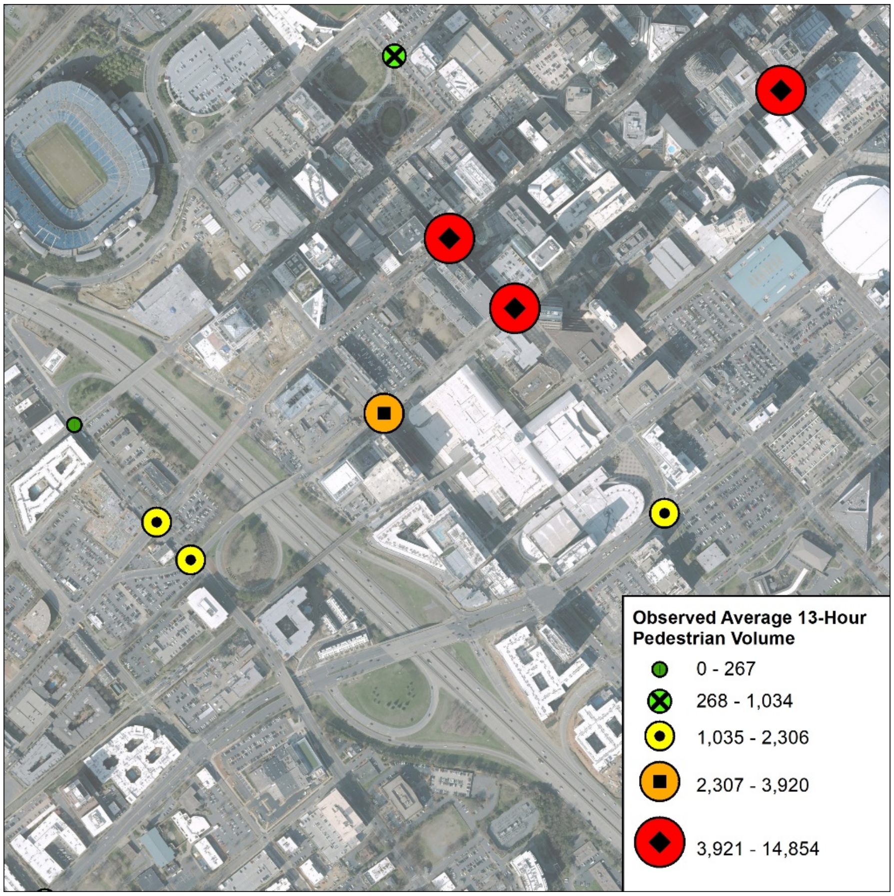
The pedestrian volume model in table 11 includes indicator variables for the year in which pedestrian counts are being predicted. Despite the higher p-values associated with the annual indicators, these variables capture "aggregate effects" of unknown or unmeasured variables that change from year to year and are not explicitly included in the model, but influence pedestrian volumes. Gayah et al. (2018) noted that a primary concern for statistical modeling is omitted variable bias. Including these yearly indicators is one way to account for confounding effects and thus more accurately estimate the influences of other variables that are included in the model. The yearly indicators also improved the accuracy of model predictions. For these reasons, the yearly indicators were kept in the model, regardless of statistical significance.

The model indicates that pedestrian volumes increase at intersections where the lowest posted speed limit on the approach roadways is equal to 25 or 30 miles per hour (mph), intersections with a sidewalk, signalized intersections, and intersections with at least one bus stop present. The model also suggests that pedestrian volumes increase as the land-use mix becomes more heterogeneous and as the total population and total employment within a 0.5-mi radius of the intersection increases. Conversely, one of the approach roadways being an arterial road is associated with a decrease in pedestrian volumes, compared with all approaches being collector roads. The impacts of the independent variables on pedestrian volumes are consistent with engineering and planning expectations. Figure 8 maps a sample subset of the pedestrian volume predictions near Charlotte. Comparison between these predicted values (figure 8) and observed pedestrian activity within this area (figure 9) suggests a good overall match between the pedestrian volume predictions and actual pedestrian travel demand patterns.



Original photo: © 2019 NC OneMap. Annotated by FHWA (see Acknowledgments section).

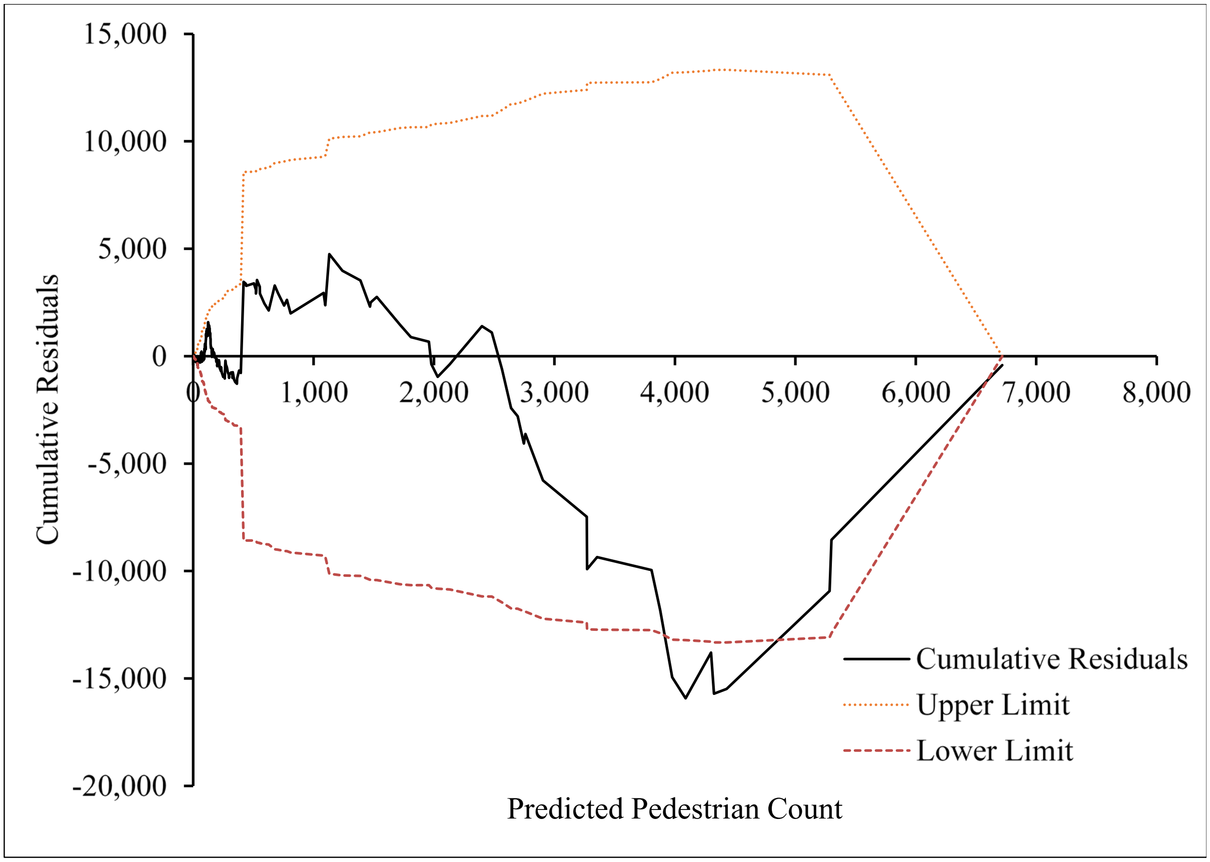
Figure 8. Graphic. Example pedestrian volume model output (NC OneMap 2019).



Original photo: © 2019 NC OneMap. Annotated by FHWA (see Acknowledgments section).

Figure 9. Graphic. Average observed 13-h counts at intersections during available years (Charlotte 2020; NC One Map 2019).

The project team developed a cumulative residual (CURE) plot to display the predicted pedestrian volume in terms of relation to cumulative residuals to assess how well the model fits the data (figure 10). Overall, the CURE plot shows that the model generally fits the data well, as just about 7 percent of the observations fall outside of the 95-percent confidence interval.



Source: FHWA.

Figure 10. Graph. CURE plot for the predicted pedestrian volume.

CHAPTER 5. SAFETY MODELING METHODOLOGY

This chapter describes the methodologies used for the crash severity and crash probability modeling. It includes descriptions of the statistical analysis approaches and variable selection process. Chapter 6 provides the model results and related discussions.

ANALYSIS APPROACH

This section describes the statistical analysis approaches that were applied to model pedestrian crash severity and pedestrian crash probability.

Pedestrian Crash Severity

The goal of the pedestrian crash severity model was to estimate the probability of a pedestrian crash resulting in a fatal or suspected serious injury for the pedestrian (as opposed to other injury outcomes—suspected minor injury, possible injury, or no apparent injury). The crash severity analysis is based on injury severity for the pedestrian involved in that crash and does not consider the injury level of vehicle drivers or passengers in the crash.

Pedestrian crash severity was modeled as a binary outcome: a value of 1 represented a fatal or suspected serious injury for the pedestrian, and a value of 0 represented all other severity levels (suspected minor injury, possible injury, or no apparent injury). Due to the binary nature of the crash severity outcome of interest, the project team used binary logistic regression. Binary logistic regression is a popular method to analyze binary data (Agresti 2007) where a binary outcome is modeled using predictors (Washington, Karlaftis, and Mannering 2011). The model developed here provides the probability of a pedestrian crash resulting in a pedestrian fatality or suspected serious injury. Figure 11 displays the functional form of the binary logistic regression (Agresti 2007), which applies a logit transformation to the probability of a fatality or suspected serious injury.

$$\text{logit}[P(FSI)] = \log\left(\frac{P(FSI)}{1 - P(FSI)}\right) = \alpha + \beta x$$

Figure 11. Equation. Binary logistic regression functional form.

Where:

- $P(FSI)$ = probability of fatality or suspected serious injury.
- α = regression model constant value.
- β = vector of estimated regression model coefficients.
- x = vector of independent variables influencing the injury outcome.

The resulting probability of a fatality or suspected serious injury as a function of the independent variables can be obtained using the equation in figure 12 (Agresti 2007).

$$P(FSI) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

Figure 12. Equation. Probability calculation using binary logistic regression.

Another method to interpret the binary logistic regression results is by computing the odds ratio, shown in figure 13.

$$\frac{Odds(x_j = m + 1)}{Odds(x_j = m)} = e^{\beta_j}$$

Figure 13. Equation. Odds ratio calculation for binary logistic regression.

Where:

m = a specific number or value for a variable x_j.

β_j = estimated regression model coefficient for a specific variable x_j.

The odds ratio associated with a specific independent variable represents the change in the likelihood of observing a fatality or suspected serious injury when that variable increases in value by 1 unit while all other independent variables are held constant. Odds ratios greater than 1 suggest an increase likelihood of a fatality or suspected serious injury outcome with an increase in the variable, while odds ratios less than 1 represent a reduced likelihood. Therefore, positive parameter estimates identify variables where an increase in those variables is associated with an increased likelihood of a fatality or suspected serious injury crash, whereas negative parameter estimates identify variables where an increase in those variables is associated with a decreased likelihood of a fatality or suspected serious injury crash.

Pedestrian Crash Probability

The goal of the crash probability was to estimate the probability of a pedestrian crash occurring on a given segment during the 5-yr analysis period. A binary outcome was used in the database to denote if a pedestrian crash occurred on a segment: a value of 1 was used for segments that experienced at least one pedestrian crash of any severity, whereas a value of 0 was used if no pedestrian crashes were observed on that segment. Due to the binary nature of this outcome variable, a binary logistic regression model was also used for the crash probability analysis.

VARIABLE SELECTION

The variable selection process considered the variables described in the pedestrian crash severity and pedestrian crash probability sections of chapter 3 (Data Acquisition and Integration) and was similar for development of both the pedestrian crash severity and pedestrian crash probability models. As with the development of the pedestrian volume model, the project team employed a forward selection process, testing statistical significance and model fit as variables were added to the models and considering the magnitude of the effects. Variable transformations and correlations were also considered and checked.

The project team developed two crash severity models: one not considering measures of speed derived from probe data, and the other considering speed variables from probe data. The crash severity model that excluded the probe speed variables was developed first and served as a "baseline" for the crash severity model that included probe speed variables. This approach allowed observations of the effects and explanatory contributions of the probe speed variables. Variables that were no longer statistically significant after the addition of the probe speed variables were removed, and correlations were rechecked.

SUMMARY

The pedestrian crash severity and pedestrian crash probability modeling both used binary logistic regression. The crash severity model predicts the probability of a pedestrian crash resulting in a pedestrian fatality or suspected serious injury. The crash probability model predicts the probability of a road segment experiencing one or more pedestrian crashes of any severity level over a 5-yr analysis period.

The development of model specifications used a forward selection process for determining which variables are appropriate to include in the models. Variable transformations and variable interactions were also considered, along with correlations between variables.

CHAPTER 6. SAFETY ANALYSIS RESULTS AND INTERPRETATION

This chapter provides estimation results and interpretations of the pedestrian crash severity and pedestrian crash probability models. As discussed throughout this report, the pedestrian crash severity model predicts the probability of a pedestrian crash resulting in a pedestrian fatality or suspected serious injury as a function of crash characteristics. The characteristics include a predicted level of average pedestrian activity at the crash location, which is derived from the pedestrian volume model described in chapter 4. The pedestrian crash probability model predicts the probability that one or more pedestrian crashes will occur on a road segment as a function of characteristics of that segment. As with the severity model, one of those characteristics is the level of pedestrian activity on the segment, which is also derived from the pedestrian volume model in chapter 4.

PEDESTRIAN CRASH SEVERITY

The project team performed two separate crash severity analyses: one without probe speed data and one with probe speed data. This approach allowed the project team to observe additional explanatory information from actual operating speeds from probe data that were added to the crash severity analysis compared with posted speed limit, a commonly used operating speed surrogate. Table 12 and table 13 display the estimation results for the binary logistic regression models excluding and including probe speed data, respectively. The models in table 12 and table 13 show that, when actual operating speeds are added, some variables are no longer significant based on magnitude of the effects, statistical significance, and correlations between other variables. This observation indicates that operating speeds capture the effects of those variables that may have been serving as surrogates for operating speed. These variables include the predicted pedestrian volume, the roadway cross section, and the speed limit.

Like the pedestrian volume model, the project team kept several variables in the model that were not statistically significant at the 95-percent confidence level. Variables kept in this way were those that were demonstrated in past research to be important contributing factors to crash severity and had a coefficient sign and magnitude that was reasonable. Note that recent safety research has mentioned that statistical significance of model coefficients is a secondary concern. Including statistically insignificant but informative variables can help reduce omitted variable bias and advance safety performance knowledge (Gayah et al. 2018). Variables that may not be statistically significant may still be important (Hauer 2004). Variables that remain in both models show similar effects, regardless of whether operating speed is included. Variables in both models that are associated with an increase in the odds of a pedestrian crash resulting in a fatality or suspected serious injury are as follows: whether there was an increase in the lowest AADT within 100 ft of the crash; whether a large vehicle struck the pedestrian (light truck, commercial bus, sport utility, van, pickup, single unit truck, other bus, or tractor/trailer/truck); whether the pedestrian was impaired; and whether the pedestrian was 50 yr old or older. If a pedestrian crash occurs during daylight, the odds of the crash resulting in a fatality or suspected serious injury decrease.

Both models also include a variable related to whether the pedestrian crash occurred near an intersection, with some variation. The project team developed two indicators of intersection proximity: the 100-ft buffer represents the area directly within the intersection itself, whereas the 250-ft buffer considers the intersection influence area. This larger window accounts for intersection-related traffic patterns, such as traffic queues and turn lanes. The model in table 12, which excludes probe speed, includes an indicator variable for whether a pedestrian crash is intersection related (i.e., defined in this model as a crash within 100 ft of any intersection). A pedestrian crash that is intersection related is associated with a decrease in the odds of a fatality or suspected serious injury. The intersection variable in table 13 also shows a decrease in fatality or suspected serious injury odds when a crash is intersection related, but the variable refers to whether the crash is intersection related and the intersection is signalized (i.e., defined in this model as a crash within 250 ft of a signalized intersection).

Table 12. Binary logit for crash severity not including probe speed variables.

Variable	Odds Ratio	Coefficient	Standard Error	*z*-Value	*P* > |*z*|
Lowest AADT within 100 ft (divided by 1,000)	1.0178	0.0176	0.0077	2.33	0.020
Pedestrian volume	0.9997	−0.0003	0.0002	−1.79	0.073
Indicator variable with value of 1 if crash is intersection related (<100 ft); 0 otherwise	0.5207	−0.6525	0.1012	−3.36	0.001
Indicator variable with value of 1 if crash involved a large vehicle type; 0 otherwise	1.715	0.5397	0.3316	2.79	0.005
Indicator variable with value of 1 if pedestrian impaired; 0 otherwise	3.0584	1.1179	0.6949	4.92	<0.001
Indicator variable with value of 1 if crash occurred in daylight; 0 otherwise	0.3896	−0.9425	0.0816	−4.50	<0.001
Indicator variable with value of 1 if pedestrian was 50 yr old or older; 0 otherwise	2.0235	0.7048	0.3985	3.58	<0.001
Indicator variable with value of 1 if a crash occurred on a four-plus lane divided road; 0 otherwise	0.57045	−0.5615	0.1254	−2.55	0.011
Indicator variable with value of 1 if the road had a posted speed limit 45 mph or higher; 0 otherwise	1.5058	0.4093	0.3209	1.92	0.055
Constant	0.0964	−2.3388	0.0279	−8.08	<0.001

Note: Number of observations = 1,378; log likelihood = −386.6432; pseudo R^2 = 0.1660; LR $\chi^2(9)$ = 153.87; probability > χ^2 < 0.0001.

Table 13. Binary logit for crash severity including probe speed variables.

Variable	Odds Ratio	Coefficient	Standard Error	*z*-Value	*P* > |*z*|
Lowest AADT within 100 ft (divided by 1,000)	1.0154	0.0153	0.0105	1.48	0.139
Indicator variable with value of 1 if crash involved a large vehicle type; 0 otherwise	1.6344	0.4913	0.4931	1.63	0.103
Indicator variable with value of 1 if pedestrian impaired; 0 otherwise	2.5989	0.9551	0.9498	2.61	0.009
Indicator variable with value of 1 if crash occurred in daylight; 0 otherwise	0.3633	−1.0125	0.1184	-3.11	0.002
Indicator variable with value of 1 if pedestrian was 50 yr old or older; 0 otherwise	2.2874	0.8274	0.7121	2.66	0.008
Average speed	1.0585	0.0569	0.0228	2.64	0.008
Indicator variable with value of 1 if intersection is signalized (<250 ft) (if intersection related); 0 otherwise	0.6500	−0.4308	0.2003	−1.4	0.162
Constant	0.0147	−4.2217	0.0107	−5.78	<0.001

Note: Number of observations = 590; log likelihood = −165.6118; pseudo R^2 = 0.1649; LR $\chi^2(7)$ = 65.39; probability > χ^2 < 0.0001.

PEDESTRIAN CRASH PROBABILITY

The pedestrian crash probability model predicts the probability that one or more pedestrian crashes will occur on a road segment as a function of characteristics of that segment. The project team used binary logistic regression to develop multiple pedestrian crash probability models, which are shown in table 14, table 15, and table 16. Table 14 provides a crash probability model applicable to all road segments, whereas table 15 and table 16 provide models for road segments with and without a median, respectively. Disaggregating the model by median presence allowed the project team to explore interactions that appeared to exist between median presence and the other variables included in the crash probability models.

Table 14. Binary logit for crash probability including all segments.

Variable	Odds Ratio	Coefficient	Standard Error	z-Value	$P > \|z\|$
Pedestrian volume	0.9994	−0.0006	0.0001	−4.18	<0.001
Segment length (mi)	2.2739	0.8215	0.4667	4.00	<0.001
AADT and pedestrian volume interaction	1.0001	0.0001	0.00001	7.07	<0.001
Speed limit 25 to 35 mph and median presence interaction	0.7691	−0.2626	0.1237	−1.63	0.102
Number of lanes	1.1762	0.1623	0.0547	3.49	<0.001
Constant	0.1914	−1.6534	0.0390	−8.12	<0.001

Note: Number of observations = 1,619; log likelihood = −996.6926; pseudo R^2 = 0.0867; LR $\chi^2(5)$ = 189.34; probability > χ^2 < 0.000.

Table 15. Binary logit for crash probability including segments with a median.

Variable	Odds Ratio	Coefficient	Standard Error	z-Value	$P > \|z\|$
Pedestrian volume	0.9967	−0.0033	0.00084	−3.88	<0.001
Segment length (mi)	1.9046	0.6443	0.8710	1.41	0.159
AADT and pedestrian volume interaction	1.0001	0.0001	0.00003	5.64	<0.001
Indicator variable with value of 1 if posted speed limit is between 25 and 35 mph; 0 otherwise	0.7509	−0.2865	0.1558	−1.38	0.167
Number of lanes	0.9709	−0.0295	0.0854	−0.34	0.737
Constant	0.3266	−1.1190	0.1387	−2.63	0.008

Note: Number of observations = 532; log likelihood = −320.36529; pseudo R^2 = 0.1111; LR $\chi^2(5)$ = 80.08; probability > χ^2 < 0.0001.

Table 16. Binary logit for crash probability including segments without a median.

Variable	Odds Ratio	Coefficient	Standard Error	*z*-Value	*P*> |*z*|
Pedestrian volume	0.9997	−0.0003	0.0001	−2.59	0.01
Segment length (mi)	2.2131	0.7944	0.5343	3.29	0.001
AADT and pedestrian volume interaction	1.00004	0.00004	0.000007	5.23	<0.001
Indicator variable with value of 1 if posted speed limit is between 25 and 35 mph; 0 otherwise	0.9298	−0.0727	0.1312	−0.52	0.606
Number of lanes	1.2225	0.2009	0.0737	3.33	0.001
Constant	0.1807	−1.7107	0.0483	−6.4	<0.001

Note: Number of observations = 1,087; log likelihood = −665.32705; pseudo R^2 = 0.0896; LR $\chi^2(5)$ = 131.00; probability > χ^2 < 0.0001.

Like the pedestrian volume and crash severity models, the project team included both statistically significant and less significant but informative variables; all variables included in these models have been demonstrated in past research to be important contributing factors to crash occurrence. The pedestrian crash probability model estimation results show that predicted pedestrian volumes are a statistically significant predictor of the probability of at least one pedestrian crash occurring on a segment. In all three models, pedestrian volume is statistically significant at a 95-percent confidence level. The odds ratios indicate that the odds of at least one pedestrian crash occurring on a segment decrease as pedestrian volumes increase. This finding is consistent with safety-in-numbers ideas and may be due to greater awareness of pedestrian activity and other contextual factors in more pedestrian-active areas.

The length of each segment is a statistically significant predictor of the probability of at least one pedestrian crash occurring on a segment. In table 14 and table 16, it is statistically significant at a 95-percent confidence level, whereas in table 15 (when a median is present), it is statistically significant at a 90-percent confidence level. The odds ratios, all greater than 1, indicate that the odds of at least one pedestrian crash occurring on a segment increase as the segment length increases. This result is consistent with engineering and planning expectations, as longer segments provide more opportunities for crashes to occur than shorter segments (i.e., an exposure effect).

The interaction between AADT and predicted pedestrian volume is a statistically significant predictor of the probability of a pedestrian crash at a 95-percent confidence level. The odds ratio is greater than 1, indicating that the odds of a least one pedestrian crash occurring on a segment increase as the value of this interaction increases. While the main effect of predicted pedestrian volume shows a safety-in-numbers effect, higher traffic volumes decrease or eliminate this safety-in-numbers benefit.

The results in table 14 also include an interaction between an indicator variable for whether the segment has a posted speed limit between 25 and 35 mph and a median. This interaction variable was not included in table 15 or table 16 because those models are already separated by median presence. This interaction variable in table 14 shows that the interaction between posted speed limit and median presence is a statistically significant predictor of the probability of a pedestrian crash at a 90-percent confidence level. The odds ratio, which is less than 1, indicates that, if a median is present and the segment has a posted speed limit between 25 and 35 mph, the odds of a pedestrian crash decrease. If there is either a median present or a low posted speed limit separately, the model indicates that there is no statistically significant effect. This result could indicate that there is some benefit to including both features to reduce the odds of a pedestrian crash. While table 14 showed the interaction between a low posted speed limit and median presence was statistically significant on all segments, table 15 and table 16 confirm the result that the presence of a low (25 to 35 mph) posted speed limit has a more significant impact (both in magnitude and statistical significance) when a median is present (table 15) than when it is not (table 16). Both odds ratios are less than 1, indicating that the odds of at least one pedestrian crash occurring on a segment decrease when the posted speed limit is between 25 and 35 mph compared with higher posted speed limits. While the variable is not statistically significant in either model, it is more statistically significant in table 15 (when a median is present) than in table 16 (when no median is present). This result reinforces the interpretation of the interaction term, showing that the combination of both a median and a lower posted speed limit provides a safety benefit, but one without the other may not have a quantifiable benefit.

The number of lanes is statistically significant at a 95-percent confidence interval when all segments table 14 and segments without a median (table 16) are considered; however, it is not statistically significant when a median is present (table 15). This finding indicates that the number of lanes is a predictor of the probability of at least one pedestrian crash occurring on a segment when the segment does not have a median. The odds ratios in table 14 and table 16 are both greater than 1, indicating that the odds of at least one pedestrian-involved crash increase when the number of lanes increases.

CHAPTER 7. DISCUSSION AND CONCLUSIONS

This section discusses the applicability of integrating HSIS data and other emerging data sources to answer high-priority research questions, details the findings from this study, and makes recommendations for future research.

APPLICATION OF HSIS DATA INTEGRATION

The Charlotte dataset provided by HSIS was a critical component of this study. The multimodal GIS dataset, the type of data piloted in the HSIS geospatial data delivery tool, allowed the project team to efficiently incorporate multiyear datasets from several different transportation agencies, as well as traffic and pedestrian count data from Charlotte directly. These other data sources included Mecklenburg County, NCDOT, U.S. Census Bureau, and RITIS. In chapter 3, the project team noted a difference between HSIS datasets and a more refined pedestrian crash dataset published by NCDOT. While the project team found the postprocessed NCDOT crash database to be more applicable to this study, future enhancements could allow HSIS to readily incorporate these spatial datasets into program workflows. The geospatial data delivery tool is an example of how the HSIS program could publish these supplementary datasets and allow users to explore them in parallel with more traditional HSIS data.

Access to HSIS data in a geospatial format significantly reduced the level of effort required to perform data collection and integration. With this advantage, the project team tested a broader spectrum of contributing factors to pedestrian activity and safety, as well as invested more time refining the models included in this study.

PEDESTRIAN VOLUME

The components of the pedestrian volume model showed highly intuitive results in both the effect and magnitude of each variable. The presence of pedestrian infrastructure, a higher mix of land uses (origins and destinations), transit access, lower vehicle speeds, and dense neighborhoods with higher numbers of residents and workers all contributed to higher pedestrian activity. The presence of at least one arterial approach roadway at an intersection, with its typically wider cross section and higher vehicle speeds, contributed to lower levels of pedestrian activity. Furthermore, the estimate provided by the pedestrian volume model resulted in intuitive results in the pedestrian crash probability and severity models.

PROBE SPEED DATA

The project team successfully integrated probe speed data with pedestrian crash location data and found that observed speeds were a significant predictor of pedestrian crash severity. This result is highly consistent with previous studies, and probe data could replace other, less direct indicators of vehicle speed as a preferred safety analysis metric. While the project team questioned the reliability of probe speed data at the segment level for crash probability modeling, particularly before 2018, this limitation could be overcome as the coverage of the dataset improves and bidirectional data become more reliable over time. Other than gaps in data coverage, possible limitations of the probe data at the segment level could include:

- Appropriateness of the study period (i.e., a period that produces an average reference speed that reflects typical operating conditions).

- Unobserved deviations in road conditions (i.e., temporary road work or closures affecting one or both directions of travel).

Future studies may find stronger links between crash probability at the segment level and probe speed data, thus allowing agencies to potentially use this resource as a pedestrian network screening tool. There are tradeoffs between the capacity of "big data" management required of an agency to conduct speed data analysis (e.g., minutes, hours, and days) and the time periods consistent with typical safety analysis (e.g., 3 to 5 yr). This practical implications of this trade-off for safety analysis and screening are an area that could benefit from further research.

CRASH SEVERITY

The crash severity model without probe speed data was highly consistent with expectations and previous studies (table 12). Higher posted speed limits, higher traffic volumes (indicative of larger, higher functional classification roads), larger vehicles striking the pedestrian, pedestrian impairment, and older pedestrian ages were all indicative of a higher probability of a pedestrian crash resulting in a fatality or serious injury. Conversely, proximity to an intersection and daytime lighting conditions were significant predictors of less serious injury outcomes. Proximity to an intersection likely indicates slower vehicle speeds and "more organized" crossings expected by drivers as opposed to the circumstances leading to midblock crashes, while daytime crashes likely relate to visibility. If pedestrians are more visible, drivers have the opportunity to reduce their speed, stop, or attempt to avoid a potential crash.

Two notable outcomes from the nonprobe speed model are the significance of pedestrian volumes and four-lane, median-divided roads as indicators of lower crash severity. Again, the concept of safety in numbers is a common theme throughout the literature, but its effect has not been frequently quantified due to lack of exposure information. This study showed that higher pedestrian volumes result in both lower crash severities and probabilities, but the safety benefit for crash probabilities is reduced by higher vehicle volumes. It is possible that pedestrian volumes are likely correlated with more urban environments, where city blocks are shorter, roads are narrower, and drivers are moving at more moderate speeds and anticipating pedestrian crossings, thus enhancing a safety-in-numbers effect. As noted in the pedestrian volume model in table 11, both minor and major arterials were negatively correlated with pedestrian volume (relative to collector streets). Furthermore, pedestrian crashes on four-lane, median-divided facilities had a lower probability of resulting in fatal and serious injuries compared with other facility types in the dataset (most of which were four-lane undivided). This finding may indicate a safety benefit provided by a median as a potentially moderating influence on vehicle speed or a refuge for pedestrians that allows them to wait for slower moving and more widely spaced traffic.

The model in table 13 that included observed vehicle speeds collected from probe data also showed highly intuitive and consistent results. All variables common to the nonprobe and probe data models show similar results in both overall effect and magnitude. Although the definition of the statistically significant intersection variable changes between both models, the effect of a signalized intersection in the probe speed model showing lower pedestrian crash severities is expected and consistent with previous studies. The statistical significance of traffic volumes and large vehicles is reduced somewhat, but this reduction may represent a strength of probe data capturing more of the speed effect. Rather than inferring road size and vehicle speed through traffic volume and posted speed limit, direct measures of speed may be a more nuanced indicator of potential pedestrian crash severity. By adding the direct measure of speed from probe data (and given the known importance of speed to crash injury severity), the probe speed model removed several surrogates of vehicle speed from consideration without sacrificing model fit. This outcome could indicate its applicability for existing conditions monitoring in future safety applications.

CRASH PROBABILITY

Like crash severity, the crash probability models showed highly intuitive results. In the all-segments model (table 14), wider roads, the length of the continuous segment, and the interaction of pedestrian volume with the average AADT were correlated with an increased likelihood of a pedestrian crash occurring on a particular segment. This finding is consistent with expectations and previous studies. Conversely, pedestrian volume alone and segments with a low posted speed and a median present were associated with a lower likelihood of a pedestrian crash occurring. The significance of the pedestrian volume variable alone is especially relevant compared with the interaction between pedestrian volume and AADT. Pedestrian volumes alone indicate a safety-in-numbers effect, but the interaction of pedestrian and traffic volumes shows an important corollary effect: while an increase in pedestrian activity alone may not lead to an increased likelihood of a pedestrian crash, the confluence of high pedestrian volumes and high vehicular traffic volumes does lead to an increased probability of a pedestrian crash.

The interaction of a median and low posted speed limits indicates a lower likelihood of a pedestrian crash and is an intuitive result. Lower speed limits and the presence of a potential crossing refuge appear to be especially important when combined. During the modeling process, the project team noted that median presence alone was not consistently an indicator of a decreased likelihood of a pedestrian crash. The separate models in table 15 and table 16 have some intuitive differences. While both exposure variables retain their effect, magnitude, and statistical significance, the low posted speed limit indicator and the number of lanes vary considerably between both models. Lower posted speed limits were much more significant to a lower likelihood of a pedestrian crash on median divided roads and were much less significant in the undivided model. Conversely, the presence of the median reduces the significance of the number of lanes a pedestrian must cross, whereas the number of lanes on an undivided road is a highly significant indicator of an increased likelihood of a pedestrian crash. While this finding may not definitively reinforce the efficacy of a median as a pedestrian safety device, it does suggest that medians may have some moderating influence on the risk posed by crossing distance; this effect may be enhanced through lower vehicle speeds.

LIMITATIONS

This section discusses some of the limitations of this study. One notable limitation is highlighted by the inconclusive results of medians alone in predicting crash probability. While the study took care to account for changes to the environment over time, the project team assumed that road infrastructure was constant throughout the study. Therefore, there is no comparison of a road before a median was installed relative to its safety performance after a median was installed. Furthermore, while continuous segments allowed the project team to aggregate highly similar segments, small deviations in the roadway geometry may not be adequately captured. For instance, medians may temporarily break to allow for turn lanes, but the segment may be predominantly median divided on either side of this gap. Finally, the project team did not have access to median type information. No distinction was made between a median designed to be a pedestrian refuge and one that merely separated bidirectional traffic. Some medians may not provide adequate protection, or be accessible, for pedestrians.

Related to safe crossing locations for pedestrians, the study did not take any pedestrian or vehicle action into account. For instance, the presence of a median does not provide protection for a pedestrian walking along a travel way or walking out between parked vehicles. While the study captured many of the commonly cited trends in pedestrian crash outcomes, especially for network screening purposes, prediction outcomes at certain locations could be refined by incorporating specific actions of persons involved in a crash. Finally, as noted in some of the previous studies, the efficacy of some pedestrian infrastructure may not be reflected in this type of study. Sidewalks, PHBs, and other crossing improvements may reduce crashes at specific locations, especially compared with conditions before a treatment was installed, but these locations may still experience more crashes (or more severe crashes) than peer locations with different site-specific qualities (e.g., sight distance, lighting, special events).

CONCLUSIONS

This study successfully integrated HSIS with numerous multi-jurisdictional and emerging datasets and identified promising applications of both direct measures of vehicle speed and estimated pedestrian counts based on observed pedestrian counts. Vehicle speed was a strong predictor of a fatal or serious injury, and pedestrian volume, especially when combined with high traffic volume, was a highly consistent predictor of the likelihood of a pedestrian crash on a given road segment. Both results are intuitive and consistent with findings in previous studies.

The objective of this study was to apply more direct measures of vehicle speed and pedestrian volume to assess pedestrian safety. The benefit of using these sources, rather than a collection of proxies and surrogates, is threefold. First, safety models developed using these direct measures (e.g., observed speed) could provide greater insight into the differences between two relatively similar sites. While two roads may have a posted speed limit of 45 mph, speeding may be a greater issue on one road than on the other. Observed speeds could allow agencies to make more nuanced interventions.

Second, the measures of speed and pedestrian volume applied in this study lend themselves to network screening and analysis. While speeds are typically measured for individual corridor studies and pedestrian counts are collected ad hoc at targeted locations, the RITIS data and the pedestrian volume model allow conditions to be assessed and compared over an entire road network efficiently. This process could enhance and supplement traditional screening for pedestrian safety issues.

Finally, while several proxies of pedestrian volume are most reliable at the zonal level (e.g., a census tract), specific countermeasure treatments are often only applicable at the segment or site level. Zonal analysis has the potential to be a strong planning tool to assess anticipated changes in a community, but it may lack the power to make targeted treatments in the present. Observed speeds and interpolated pedestrian volumes allow practitioners to make informed improvements immediately, rather than in the future.

ACKNOWLEDGMENTS

For figure 3 and figure 4, the original map is the copyright property of NC OneMap and can be accessed from https://www.nconemap.gov/datasets/c5b316f805ab4d74bf7b598220ac5558 (NC OneMap 2019). The map overlay showing the 1-km buffer around an intersection was developed because of this research project. The overlay includes a red circle showing the land-use study area for the intersection.

For figure 8 and figure 9, the original map is the copyright property of NC OneMap and can be accessed from https://www.nconemap.gov/datasets/c5b316f805ab4d74bf7b598220ac5558 (NC OneMap 2019). The map overlay showing symbolic circles indicating observed 13-h pedestrian counts was developed because of this research project. The overlay includes circles at intersections in uptown Charlotte, that increase in size, with larger estimated pedestrian counts produced by the pedestrian volume model.

The authors would like to thank Angela Berry and other city of Charlotte DOT staff for their time, data, and technical advisory contributions that helped shape the purpose and need for this study. The authors would also like to thank NCDOT staff for identifying recent road projects in the Charlotte area, as well as allowing the project team to access RITIS data through the State's participation in the I–95 Coalition (via email on February 19, 2020).

REFERENCES

Agresti, A. 2007. *An Introduction to Categorical Data Analysis*, 2nd ed. New York: John Wiley & Sons, Inc.

Aziz, H.A., S.V. Ukkusuri, and S. Hasan. 2013. "Exploring the Determinants of Pedestrian-Vehicle Crash Severity in New York City." *Accident Analysis & Prevention* 50: 1298-1309.

CATT Lab. 2020. "RITIS" (web page). http://www.cattlab.umd.edu/?portfolio=ritis, last accessed April 24, 2020.

City of Charlotte. 2020. "Transportation Data Management System" (web page). https://charlottenc.ms2soft.com/tcds/tsearch.asp?loc=Charlottenc&mod=TCDS, last accessed May 1, 2020.

City of Charlotte. 2019. "Charlotte Open Data Portal" (web page). https://data.charlottenc.gov/, last accessed July 1, 2021.

Clifton, K. J., C. V. Burnier, and G. Akar. 2009. "Severity of Injury Resulting from Pedestrian-Vehicle Crashes: What Can We Learn from Examining the Built Environment?" *Transportation Research Part D: Transport and Environment* 14, no. 6: 425-436.

Das, S., S. Geedipally, R. Avelar, L. Wu, K. Fitzpatrick, M. Banihashemi, and D. Lord. 2020. *Rural Speed Safety Project for USDOT Safety Data Initiative*. College Station, TX: Texas A&M Transportation Institute.

Dumbaugh, E., W. Li, and K. Joh. 2013. "The Built Environment and the Incidence of Pedestrian and Cyclist Crashes." *Urban Design International* 18, no. 3: 217–228.

Eigen, A. M., R. O. Valdivieso, and A. Ahrari. 2014. *Photographic Data Extraction Feasibility and Pilot Study in Support of Roadside Safety and Roadway Departure Research*. Report No. FHWA-HRT-13-088. Washington, DC: Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/safety/13088/index.cfm, last accessed May 1, 2020.

Eluru, N., C. R. Bhat, and D. A. Hensher. 2008. "A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes." *Accident Analysis & Prevention* 40, no. 3: 1033–1054.

FHWA. 2017. *Proven Safety Countermeasures: Walkways.* Report No. FHWA-SA-17-067. Washington, DC: Federal Highway Administration. https://safety.fhwa.dot.gov/provencountermeasures/walkways/, last accessed May 1, 2020.

FHWA. n.d. "What is HSIS?" (web page). https://www.hsisinfo.org/, last accessed July 1, 2021.

Frank, L. D., M. A. Andresen, and T. L. Schmid. 2004. "Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars." *American Journal of Preventive Medicine* 27, no. 2: 87–96.

Gayah, V. V, E. T. Donnell, Z. Yu, and L. Li. 2018. "Safety and Operational Impacts of Setting Speed Limits Below Engineering Recommendations." *Accident Analysis and Prevention* 121: 43–52.

Griswold, J. B., A. Medury, R. J. Schneider, D. Amos, A. Li, and O. Grembek. 2019. "A Pedestrian Exposure Model for the California State Highway System." *Transportation Research Record* 2673, no. 4: 941–950.

Haleem, K., P. Alluri, and A. Gan. 2015. "Analyzing Pedestrian Crash Injury Severity at Signalized and Non-Signalized Locations." *Accident Analysis & Prevention* 81: 14-23.

Hampshire, R. C., L. J. Molnar, A. Cao, Y. Cai, X. Li, and T. Fabusuyi. 2017. *Developing Michigan Pedestrian and Bicycle Safety Models*. Report No. SPR-1651. Ann Arbor, MI: University of Michigan Transportation Research Institute.

Hankey, S., G. Lindsey, X. Wang, J. Borah, K. Hoff, B. Utecht, and Z. Xu. 2012. "Estimating Use of Non-Motorized Infrastructure: Models of Bicycle and Pedestrian Traffic in Minneapolis, MN." *Landscape and Urban Planning* 107, no. 3: 307-316.

Hauer, E. 2004. "The Harm Done by Tests of Significance." *Accident Analysis & Prevention* 36: 495-500.

I–95 Corridor Coalition. 2018. "Regional Integrated Transportation System (RITIS)" (web page). http://i95coalition.org/projects/regional-integrated-transportation-information-system-ritis/, last accessed April 24, 2020.

Kersavage, K. 2019. "Relationship Between Speed Metrics and Crash Frequency and Severity: Implications for Roadway Design." Doctoral dissertation. Pennsylvania State University.

Lee, C., and M. Abdel-Aty. 2005. "Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida." *Accident Analysis & Prevention* 37, no. 4: 775–786.

Li, D., P. Ranjitkar, Y. Zhao, H. Yi, and S. Rashidi. 2017. "Analyzing Pedestrian Crash Injury Severity Under Different Weather Conditions." *Traffic Injury Prevention* 18, no. 4: 427-430.

Lord, D., and F. Mannering. 2010. "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives." *Transportation Research Part A: Policy and Practice* 44, no. 5: 291–305. https://doi.org/10.1016/j.tra.2010.02.001, last accessed July 1, 2021.

Mansfield, T. J., D. Peck, D. Morgan, B. McCann, and P. Teicher. 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.

Mecklenburg County. n.d. “Open Mapping” (web page). http://maps.co.mecklenburg.nc.us/openmapping/index.html, last accessed July 1, 2021.

Moradi, A., H. Soori, A. Kavousi, F. Eshghabadi, and E. Jamshidi. 2016. “Spatial Factors Affecting the Frequency of Pedestrian Traffic Crashes: A Systematic Review.” *Archives of Trauma Research* 5, no. 4: e30796.

National Highway Traffic Safety Administration. 2019. “National Statistics” (web page). https://www-fars.nhtsa.dot.gov/Main/index.aspx, last accessed April 29, 2020.

NCDOT. 2019. “NCDOT Bicyclist and Pedestrian Crash Map 2007–2019” (web page). https://www.arcgis.com/home/item.html?id=b4fcdc266d054a1ca075b60715f88aef, last accessed May 1, 2020.

NCDOT. 2021. “GIS Data Layers” (web page). https://connect.ncdot.gov/resources/gis/Pages/GIS-Data-Layers.aspx, last accessed July 1, 2021.

NC OneMap. 2019. “Latest Orthoimagery” (web page). https://www.nconemap.gov/datasets/c5b316f805ab4d74bf7b598220ac5558, last accessed May 1, 2020.

Omer, I., V. Gitelman, Y. Rofe, Y. Lerman, N. Kaplan, and E. Doveh. 2017. “Evaluating Crash Risk in Urban Areas Based on Vehicle and Pedestrian Modeling.” *Geographical Analysis* 49, no. 4: 387-408.

Pande, A., and M. Abdel-Aty. 2009. “A Novel Approach for Analyzing Severe Crash Patterns on Multilane Highways.” *Accident Analysis & Prevention* 41, no. 5: 985-994.

Pitt, R., B. Guyer, C. C. Hsieh, and M. Malek. 1990. “The Severity of Pedestrian Injuries in Children: An Analysis of the Pedestrian Injury Causation Study.” *Accident Analysis & Prevention* 22, no. 6: 549–559.

Rothman, L., A. W. Howard, A. Camden, and C. MacArthur. 2012. “Pedestrian Crossing Location Influences Injury Severity in Urban Areas.” *Injury Prevention* 18, no. 6: 365–370.

Siddiqui, C., M. Abdel-Aty, and K. Choi. 2012. “Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes.” *Accident Analysis & Prevention* 45: 382–391.

The Eastern Transportation Coalition. 2021. “RITIS & PDA Suite” (web page). https://tetcoalition.org/projects/ritis-pda-suite/, last accessed July 1, 2021.

Torbic, D. J., D. W. Harwood, C. D. Bokenkroger, R. Srinivasan, D. Carter, C. V. Zegeer, and C. Lyon. 2010. “Pedestrian Safety Prediction Methodology for Urban Signalized Intersections.” *Transportation Research Record* 2198, no. 1: 65–74.

Ukkusuri, S., L. F. Miranda-Moreno, G. Ramadurai, and J. Isa-Tavarez. 2012. ‘The Role of Built Environment on Pedestrian Crash Frequency.” *Safety Science* 50, no. 4: 1141–1151.

U.S. Census Bureau. 2019. “TIGER/Line with Selected Demographic and Economic Data” (web page). https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.html, last accessed July 1, 2021.

U.S. Census Bureau. n.d. “Longitudinal Employer-Household Dynamics” (web page). https://lehd.ces.census.gov/data/, last accessed June 24, 2021.

Washington, S. P., M. G. Karlaftis, and F. L. Mannering. 2011. *Statistical and Econometric Methods for Transportation Data Analysis*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.

Zajac, S. S., and J. N. Ivan. 2003. “Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut.” *Accident Analysis & Prevention* 35, no. 3: 369–379.

Recycled
Recyclable

HRDS-20/10-21(WEB)E