



TECHNICAL REPORT 0-7048-1
TxDOT PROJECT NUMBER 0-7048

Developing Countermeasures to Decrease Pedestrian Deaths

Kara Kockelman, Ph.D.
Kenneth A. Perrine
Natalia Zuniga-Garcia, Ph.D.
Maximillian Pleason
Maxwell Bernhardt
Mashrur Rahman
Jooyong Lee, Ph.D.
Alex Karner, Ph.D.

April 2022; Published August 2022

<http://library.ctr.utexas.edu/ctr-publications/0-7048-1.pdf>



Technical Report Documentation Page

1. Report No. FHWA/TX-22/0-7048-1	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Developing Countermeasures to Decrease Pedestrian Deaths		5. Report Date Submitted: December 2021; Revised: April 2022; Published: August 2022	
		6. Performing Organization Code	
7. Author(s) Kara Kockelman, Ph.D. http://orcid.org/0000-0003-4763-1304 ; Kenneth A. Perrine; Natalia Zuniga-Garcia, Ph.D.; Maximillian Pleason; Maxwell Bernhardt; Mashrur Rahman; Jooyong Lee, Ph.D.; Alex Karner, Ph.D.		8. Performing Organization Report No. 0-7048-1	
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 3925 W. Braker Lane, 4 th Floor Austin, TX 78759		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. 0-7048	
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Division 125 E. 11th Street Austin, TX 78701		13. Type of Report and Period Covered Technical Report January 2020–December 2021	
		14. Sponsoring Agency Code	
15. Supplementary Notes Project performed in cooperation with the Texas Department of Transportation.			
16. Abstract In efforts to reverse the rising rate of pedestrian-related crashes in Texas, researchers analyze crash data to determine common causes and effective countermeasures. Extensive use of visualizations, regression models, advanced tree-based analysis, and GIS-based data management adds insight. Significant variables include roadway characteristics, transit stops, population, job density, time of day, as well as regional average income and homelessness. Researchers demonstrate a methodology for identifying Texas' corridors with the highest crash rates, strategically choosing proven treatments, and calculating benefit-cost ratios. This is documented in a guidebook that is approachable for practitioners and managers at all levels of government. Examples of treatments proposed for the 10 most crash-prone corridors in Texas indicate favorable benefit-cost ratios. This warrants comprehensive use of the analysis methodology across Texas in expectation of significant reduction in pedestrian-related fatalities and injuries.			
17. Key Words crash, pedestrian, injury, fatality, roadway treatment, countermeasure, accident, cost-benefit analysis, crash reduction factor, crash modification factor		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Alexandria, Virginia 22312; www.ntis.gov .	
19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 176	22. Price



**THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH**

Developing Countermeasures to Decrease Pedestrian Deaths

Kara Kockelman, Ph.D.
Kenneth A. Perrine
Natalia Zuniga-Garcia, Ph.D.
Maximillian Pleason
Maxwell Bernhardt
Mashrur Rahman
Jooyong Lee, Ph.D.
Alex Karner, Ph.D.

CTR Technical Report:	0-7048-1
Report Date:	Submitted: December 2021; Revised: April 2022
Project:	0-7048
Project Title:	Identifying Risk Factors that Lead to Increase in Fatal Pedestrian Crashes and Develop Countermeasures to Reverse Trend
Sponsoring Agency:	Texas Department of Transportation
Performing Agency:	Center for Transportation Research at The University of Texas at Austin

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Center for Transportation Research
The University of Texas at Austin
3925 W. Braker Lane, 4th floor
Austin, TX 78759

<http://ctr.utexas.edu/>

Disclaimers

Author's Disclaimer: The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation.

Patent Disclaimer: There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine manufacture, design or composition of matter, or any new useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

Engineering Disclaimer

NOT INTENDED FOR CONSTRUCTION, BIDDING, OR PERMIT PURPOSES.

Project Engineer: Kara Kockelman

Professional Engineer License State and Number: Texas No. 93443

P.E. Designation: Research Supervisor

Acknowledgments

The authors express appreciation to key supporters within TxDOT, including project manager Shelley Pridgen and project monitoring committee members Carol Campa, Khalid Jamil, Noah Heath, Sharlotte Teague, and Solomon Thomas. Special thanks go to Mark Dickerman, Larbi Hanni, and the CRIS Support Team for insight on proper usage of CRIS and coordinated assistance in retrieving crash records. The Network Modeling Center at UT Austin's Center for Transportation Research provided computational resources. Alex Wu contributed to vehicle obscenity measures. The authors also appreciate all to practitioners in roadway safety who interacted with this project and who have a vested interest in ensuring the well-being of pedestrians.

Products

0-7048-P1: Developing Countermeasures to Decrease Pedestrian Deaths: Guidebook

0-7048-P2: Workshop: Pedestrian Protection via Cost-effective Countermeasures

Table of Contents

Executive Summary	1
Chapter 1. Literature Review	4
1.1. Background and Geography	4
1.2. Examining Crash Factors	7
1.2.1. Age and Other Demographic Variables	7
1.2.2. Distracted Drivers and Pedestrians	8
1.2.3. Presence of Signals, Crosswalks, and Other Facilities	9
1.2.4. Speed	10
1.3. Countermeasures	11
1.3.1. Effectiveness of Physical Countermeasures	12
1.3.2. Effectiveness of Non-Physical Countermeasures	13
1.3.3. Crash Costs	14
1.3.4. Obtaining Pedestrian Counts and Usage	15
1.4. Conclusions	16
Chapter 2. Assemble and Prepare Relevant Crash Data	18
2.1. Explaining Data Sources	18
2.1.1. CRIS: Crash Records Information System	18
2.1.2. Detailed Crash Records	19
2.1.3. Road Inventory Data	20
2.1.4. Other Data Sources and Walk-Miles Traveled	21
2.2. Summary Statistics	26
2.3. Other Data Processes and Challenges	35
2.4. Conclusions	36
Chapter 3. Identify Analysis Methodology and Key Results	38
3.1. Summary Statistics for Pedestrian Crash Variables	38
3.1.1. Data Details for CRIS and TxDOT Roadway Inventory	39
3.1.2. OLS Regression Data Description	40
3.2. Key Crash Factors	42
3.2.1. Pedestrian Crash and Fatality Trends	42
3.2.2. Time of Day	42
3.2.3. Speed	43
3.3. Heat Maps and Lat-Long Analysis	44

3.3.1. Investigating the Factors for Missing Geographic Location of Pedestrian Crash Events.....	51
3.4. Methodologies for Prediction of Crash Counts and Injury Severity.....	54
3.4.1. Analysis of Pedestrian Crash Counts – NB Models	54
3.4.2. Analysis of Pedestrian Injury Severity – OP and HOP Models.....	55
3.4.3. Ordinary Least-Squares Regression.....	56
3.4.4. Limitations	56
3.5. Results from Crash Count and Injury Severity Models.....	56
3.6. Key Results	64
3.6.1. NB & OP Models Results Discussion	64
3.6.2. Pedestrian Injury Severity.....	65
3.6.3. OLS Regression Model Results Discussion	68
3.7. Conclusions.....	69
Chapter 4. Develop Decision Trees to Classify Crashes across Contributing Factors	71
4.1. Decision Trees	71
4.1.1. Data and Method Description	73
4.1.2. Data and Method Description	76
4.1.3. Summary	83
4.2. Tree-based Methods.....	83
4.2.1. Background.....	83
4.2.2. Model Fit Evaluation	85
4.2.3. Hyperparameter Tuning.....	86
4.2.4. Sensitivity Analysis	87
4.3. Results.....	87
4.3.1. Pedestrian Crash Occurrence Prediction.....	87
4.3.2. Pedestrian Crash Injury Severity Prediction.....	91
4.3.3. Summary	96
4.4. Conclusions.....	97
Chapter 5. Evaluate Value and Cost-Effectiveness of Treatments.....	99
5.1. Segment Ranking and Creation Methodology.....	99
5.1.1. Dataset Preparation	99
5.1.2. Finding Corridors.....	100
5.2. Crash Count Estimation Model.....	101
5.2.1. Negative Binomial Count Model	102
5.2.2. Model Results	102

5.3. Treatments by Category	105
5.4. Cost and Benefit Calculation Models	112
5.4.1. BCR Calculation Example	113
5.5. Recommendations for Texas' Highest-Crash Corridors	114
5.5.1. I-35 Southbound Frontage Road – Martin Luther King Jr. Boulevard to Holly Street – Austin.....	114
5.5.2. Tomball Parkway (SH-249) – Sam Houston Tollway (SL-8) to Breen Road – Houston	114
5.5.3. Westheimer Road – Fondren Road to Chimney Rock Road – Houston.....	115
5.5.4. Congress Avenue – 12th Street to Barton Springs Road – Austin	116
5.5.5. Lamar Boulevard – Masterson Pass to Payton Gin Road – Austin	118
5.5.6. Congress Avenue – Woodward Street to St. Elmo Road - Austin.....	119
5.5.7. E. Riverside Drive – Pleasant Valley Road to Faro Drive – Austin	119
Zarzamora Street – Cincinnati Street to Delgado Street – San Antonio.....	121
5.5.8. Fannin Street – Commerce Street to Jefferson Street – Houston.....	121
5.5.9. Milam Street – McGowan Street to Alabama Street – Houston.....	122
5.6. Recommendations for Pedestrian Safety in Texas.....	123
5.6.1. Pedestrian Leading Interval	123
5.6.2. Speed Hump.....	124
5.6.3. Raised Median	124
5.6.4. Speed Limit 10% Reduction	124
5.7. Policy Implementations for Vehicles Sold in the US	125
5.8. Conclusions.....	126
References.....	128
Appendix A. Additional Examples of BCR Calculations.....	137
Appendix B. CRIS Online Hot Spot Analysis Tool's Intersection Evaluations.....	143
Appendix C. Vehicle Make/Model Obscurity	150
Appendix D. CR-3 Training Results.....	160
Appendix E. Datasets and Technical Documentation.....	164

List of Tables

Table 1.1 Fatality and Injury Rates per Person-Mile Traveled (US & TX, 2020)	4
Table 1.2 Survey Results of Los Angeles Elementary School Students.....	14
Table 1.3 Breakdown of Crash Severity Ratings.....	15
Table 2.1 Mileage by Road Segments	20
Table 2.2 Ordinary Least-Squares Model Results for Predicting $Y = WMT$	23
Table 2.3 Number of Crashes Over Time (2010–2019)	28
Table 2.4 Pedestrian Deaths in Texas Cities (of Population >100,000)	30
Table 2.5 Pedestrian Crashes and Fatalities under Different Daylight Conditions in Texas.....	32
Table 2.6 Alcohol Involvement in Crashes that Resulted Pedestrian Fatalities	34
Table 3.1 Summary Statistics of Variables for Road Segments across Texas.....	40
Table 3.2 Summary Statistics for 254 Texas Counties, OLS Regression.....	41
Table 3.3 Summary statistics of variables used in binary logit model to investigate contributing factors in records lacking geographic location (lat/long) details	52
Table 3.4 Estimation Results of NB for All Pedestrian Crashes and Fatal Pedestrian Crashes ...	57
Table 3.5 Injury Severity Results: OP versus HOP Models	58
Table 3.6 Marginal Effects of Risk Factors Associated with Pedestrian Injury Severity (HOP Model).....	60
Table 3.7 OLS Regression – Results for All and Fatal-only Pedestrian Crashes Per 1 M VMT for the 254 Texas Counties.....	62
Table 3.8 OLS Regression – Results for All and Fatal-only Pedestrian Crashes per WMT for Texas’ 254 Counties	63
Table 4.1 Variable Description.....	75
Table 4.2 Model Accuracy.....	83
Table 4.3 Comparison of Model Performance and Computation Time.....	88
Table 4.4 Summary of Model Performance on Injury Severity Prediction	91
Table 5.1 Estimation Results of NB for Pedestrian Crashes, Texas.....	103
Table 5.2 Estimation Results of NB for Pedestrian Crashes, City of Austin.....	104
Table 5.3 Basic Roadway Treatments	107
Table 5.4 Traffic Calming Treatments.....	108
Table 5.5 Pedestrian-specific Treatments.....	108
Table 5.6 Street Furniture Treatments	110
Table 5.7 New Sidewalk Treatments	110

Table 5.8 Education Treatments	111
Table 5.9 Homelessness-centric Treatments, Direct Outreach to Pedestrians.....	111

List of Figures

Figure 1.1 Pedestrian Fatality Risk.....	11
Figure 2.1 Road Segment Types.....	21
Figure 2.2 Demographic and Spatial Data.....	24
Figure 2.3 WMT by Texas Counties.....	25
Figure 2.4 Number of Pedestrian Fatalities (2010–2019).....	29
Figure 2.5 Walk-Miles Traveled per Resident per Year.....	31
Figure 2.6 Injury Types Suffered in Texas’ Pedestrian Crashes	31
Figure 2.7 Most Severe Injury Suffered by Pedestrians in Texas.....	32
Figure 2.8 Pedestrians Fatalities by Time of Day.....	33
Figure 2.9 Injury Outcomes for Pedestrians on Roadways with Different Speed Limits.....	33
Figure 2.10 Light-Duty Vehicles Sales across US by Body Type, over Time	34
Figure 2.11 Pedestrian Killed in Texas Traffic Crashes by Age Group	35
Figure 3.1 Distribution of Pedestrian Fatalities in Texas by Time-of-day, 2010–2018	43
Figure 3.2 Distribution of Injury Severity and Fatalities in Texas by Roadway Speed Limit, 2010–2018.....	44
Figure 3.3 Heatmap of Pedestrian Crashes in Downtown Houston.....	45
Figure 3.4 Heatmap of Pedestrian Crashes in Central Houston.....	46
Figure 3.5 Heatmap of Pedestrian Crashes in Downtown Austin	46
Figure 3.6 Heatmap of Pedestrian Crashes in Greater Austin	47
Figure 3.7 Heatmap of Pedestrian Crashes in Downtown Dallas.....	48
Figure 3.8 Heatmap of Pedestrian Crashes in Central Dallas.....	49
Figure 3.9 Heatmap of Pedestrian Crashes in Downtown San Antonio	50
Figure 3.10 Heatmap of Pedestrian Crashes in Greater San Antonio.....	51
Figure 3.11 Status of Recorded Geographic Location of the Crash Events for Selected Factors	53
Figure 4.1 Example of a Regression Tree.....	72
Figure 4.2 Example of a Classification Tree	73
Figure 4.3 Texas’ Annual Crash Counts.....	74
Figure 4.4 Pedestrian Injury Severity Distribution.....	74

Figure 4.5 Decision Tree for Highway Design Characteristics	78
Figure 4.6 Decision Tree for Pedestrian and Driver Characteristics	79
Figure 4.7 Decision Tree for Environmental and Other Conditions.....	80
Figure 4.8 Decision Tree for Vehicle Characteristics.....	80
Figure 4.9 Model with All Variables	81
Figure 4.10 Variable Importance	82
Figure 4.11 Sensitivity Analysis for Crash Occurrence Predictions.....	89
Figure 4.12 Margins of Classification Models on Different Classes.....	92
Figure 4.13 ROC Curve of Different Classification Models	93
Figure 4.14 Parallel Coordinate Plots of Marginal Effects for Fatal Crashes	96
Figure 5.1 Corridors Among the Texas “top 100” That Exist within the Central Austin Area..	101

Executive Summary

Pedestrian crashes are a rising issue in Texas. While the total walk-miles traveled (WMT) is estimated to have risen 16% (BTS, 2019) between 2009 and 2017, the number of (reported) pedestrian deaths rose 46% (GHSA, 2020). Texas averaged 1.14 pedestrian deaths per 100,000 residents in 2019, which is 26% higher than the US average of 0.90.

This project used TxDOT's Crash Records Information System (CRIS) and Roadway Inventory files—combined with other statewide databases on land use, climate, hospital locations, travel behaviors, and other sources—to understand and ultimately predict crash counts and to design treatments and public policies that can decrease those counts. This report explains the data manipulation, analysis methods, decision-tree models, and benefit-cost analysis results of the team's research into pedestrian crashes in Texas.

To assemble and prepare relevant crash data (Chapter 2), various datasets including CRIS, Roadway Inventory files, and police-recorded crash reports are used. From 2010 through 2019, 5.6 million CRIS entries were recorded across the state, which contains over 73,000 centerline miles of inventoried roadways. The research team has also obtained a strategic sample of 300 detailed police-recorded crash reports for pedestrian-related crashes in recent years. The capabilities and nuances of each dataset are analyzed, including mapping crash points across the network for hot-spot identification, walk-miles traveled per county, rainfall, population and job densities, and other factors.

The team also analyzed the walking distances in the two most recent National Household Travel Surveys and assembled many other spatial and demographic datasets for this project, in order to identify risk factors that are leading to Texas' and the US' rise in pedestrian deaths and to develop countermeasures to reverse this trend. Results observed include a disproportionately high percentage of pedestrians among all Texas crash fatalities, with elevated risk of injury at night, along higher-speed roadways, and in high-density (in terms of jobs and population) settings. Alcohol and/or drugs were involved in 37.6% of pedestrian deaths. 72% of all the pedestrians killed in Texas' traffic crashes were male. Higher speed limits were associated with more severe injuries and deaths, where median speed limit was 45 mph in locations of fatal pedestrian crashes versus 30 mph for non-fatal crashes.

As seen in Chapter 3, analysis methods used in this report include negative binomial (NB), ordered probit (OP), heteroskedastic ordered probit (HOP), and ordinary least-square (OLS) regression models. NB model results show how total and fatal pedestrian-crash rates and counts rise with a segment's number of lanes, transit stops, population, and job density, as well as proximity to schools and hospitals, while greater median and shoulder widths provide some protection. A HOP model for injury severity demonstrates that pedestrian crashes are more likely to be severe and fatal at night (8 PM–5 AM), without overhead lighting, and when the pedestrians or drivers are

intoxicated. The OLS regression analyzes pedestrian crash rates (per vehicle-miles and walk-miles traveled) across Texas' 254 counties. At the county level, there is a moderately positive relationship between job density and pedestrian crash rates, but a practically significant and negative relationship with population density, while income-related variables and homelessness rates have substantial impacts on pedestrian crash and fatality rates.

Next, in the later part of Chapter 3 and all of Chapter 4, the report discusses various crash counts and severity estimation methods to predict pedestrian crashes. With 2010–2019 CRIS data, decision trees are used to classify pedestrian injury severity reported by Texas police officers into the severity levels found in the CRIS data. Decision tree structures illuminate predictive results that are intuitive for readers and users (with binary yes/no decisions at each branch), but they are less accurate and robust than more complex methods in terms of prediction. To address this limitation, ensemble tree-based models are also calibrated and evaluated. By aggregating many decision trees (into an “ensemble” or “forest”) using methods like bagging, boosting, and random forests, tree-based methods can substantially improve the predictive performance of standard, singular decision trees and can provide more robust estimates. In addition to the decision tree analysis, random forests (RF), gradient boosting (Light GBM and XGBoost), and Bayesian additive regression trees (BART) were applied. Findings underscore the importance of campaigns against driving and walking while intoxicated, installation of streetlights in pedestrian-active areas, improved roadway design, and enforcement of safety countermeasures in areas where pedestrians are more vulnerable such as near schools and hospitals.

Using insight gained from prior analyses and intensive data processing to derive intersections and uniform-length segments from the TxDOT Roadway Inventory, researchers devised an algorithm (shown in the first part of Chapter 5) to identify the top pedestrian crash-prone corridors across Texas and rank them according to crash severity. This allows for quick identification of locations deserving of immediate analysis for possible countermeasures. Appendix E describes this process and links to downloadable datasets. The Texas “Top 100” corridors roughly represents 111 centerline miles of roadway, or just 0.035% of Texas' total.

Lastly, the results of various treatments and the benefit-cost ratio (BCR) of applying suitable and cost-effective treatments for segments and intersections in Texas over the next 10-year period are presented in Section 5.4. An overview of the corridor ranking and creation methodology is provided, which creates a list of ranked corridors by pedestrian crash and fatality by total comprehensive costs. A breakdown is conducted of estimated costs and crash modification factor (CMF) estimates for each treatment within seven categories. An overview is presented in Section 5.5 of the 10 corridors with the highest fatal crash rates as provided by the clustering model, with BCRs for sample treatments to its intersections.

The results include the types and specifications of treatments to reduce pedestrian crash counts, identification of the 10 highest-crash corridors, and application of the treatments to the corridors and intersections to provide crash reduction estimates, benefit-cost analysis results, and

detailed step-by-step calculation results. The BCR analyses show that the treatments considered—such as prohibited right-turn on red, speed limit reductions, pedestrian leading interval, road diet, pedestrian refuge islands, and streetlights—have BCRs ranging from 1.67 to 5.38 when the delay costs are added, and from 428 to 6,689 when delay costs are ignored. (Note that a BCR greater than 1 indicates that the suggested treatments are cost-effective.)

Policy implementations for promoting pedestrian safety in the US, compared to the European regulations, found that a transition from testing the vehicles' autobrake to testing the actual ability to protect pedestrians are needed. A higher speed limit than 40 kph (24.85 mph) assumed in the vehicle test should be incorporated, and OEM's responsibilities to protect pedestrians should be clearly defined.

The primary product proceeding from this work is the "Developing Countermeasures to Decrease Pedestrian Deaths: Guidebook." It was created to readily introduce the roadway treatment strategy selection methodology exercised within this project to a practitioner audience. The guidebook's methodology was piloted through a workshop presented to an audience of relevant professionals and enthusiasts at state and local levels of government. Researchers recommend further work to find opportunities to more closely tie the Guidebook into existing TxDOT road safety guidance and BCR methodologies, such as those documented within the Highway Safety Improvement Program Manual (TxDOT, 2015).

In short, this report is organized according to the sequence of tasks that were conducted within the entire project. Chapter 1 introduces relevant prior research and current industry practices as a literature review. This is followed by Chapter 2 that introduces the datasets that were crucial for all subsequent analysis within the project. Chapter 3 then documents the beginning of the analysis process, showing overall trends found in the data, along with sensitivity analysis and "heat-map" visualizations of crash activity in Texas. Chapter 4 covers the project's exploration of decision trees and demonstrates their benefits and limitations. Chapter 5 then documents intersection and corridor-level analysis in Texas, ranking the most crash-prone corridors in the state and introducing a methodology for identifying treatments and their benefit-cost ratios. The appendices further exemplify BCR and treatment analysis on the worst corridors (Appendices A and B), explore an experiment that analyzes vehicle body style against crash severity (Appendix C), describe lessons learned from crash record training (Appendix D), and introduce the project's technical data processing documentation (Appendix E).

Chapter 1. Literature Review

This chapter is a review of recent literature related to pedestrian-vehicle crashes, their distribution, trends, contributing factors, and countermeasures taken to mitigate injuries and deaths. This review brings together academic literature, technical documents, reports from public and nonprofit agencies, media, and other practitioner-involved documents to enable greater understanding of factors impacting pedestrian crash counts, their severity, trends, and countermeasure benefits. This review emphasizes relatively recent sources, and was compiled via conferencing, communications with experts (including practitioners, first responders, and public agencies), and online searches using the Web of Science, Transport Research International Documentation (TRID) database, Google Scholar, JSTOR, and the University of Texas online library system.

1.1. Background and Geography

Pedestrian miles-traveled are less than 1% of total person-miles traveled in the US, but pedestrian crash fatalities remain a key component of crash fatalities (National Household Travel Survey, 2017). Their share of total crash deaths rose from 12% in 2009 to 17% in 2018, with the total number of pedestrian crash fatalities rising 53% between 2009 and 2018 (NHTSA, 2019). Around 10% of the nation’s pedestrian fatalities happen in just four counties: Los Angeles, CA; Miami-Dade, FL; Maricopa, AZ; and Harris, TX—home to around 7% of the US population. Texas’s four largest metropolitan areas (Dallas-Fort Worth, Houston, Austin, and San Antonio) are all in the nation’s top 25 metro areas for pedestrian fatalities (NHTSA, 2019). San Antonio has the highest crash fatality rate of all major metros in Texas, with 2.46 pedestrian fatalities per 100,000 people, followed by Austin at 2.21, Dallas-Fort Worth at 1.94, and Houston at 1.9 (Webb, 2019). Overall, pedestrian crash deaths have risen in most US states, particularly in the South and West, including Texas in recent years, even as the numbers of crash deaths as a whole are decreasing (GHSA, 2018).

Table 1.1 shows the fatality and injury rates per person-miles traveled (PMT) by modes for both US and Texas. In this table, the PMT made by airline, ferry, rail, or boat are not considered. Per PMT, the walk mode (exhibited by pedestrians) carries the second fatality rate and third highest injury rate (per PMT) for US, but Texas experiences more severe pedestrian fatality and injuries than US per 100 million PMT. These results emphasize the previous paragraph’s statistics on pedestrian crash fatalities, where pedestrians are much less shielded and thus more vulnerable than those using other transportation modes, given their very low total PMT.

Table 1.1 Fatality and Injury Rates per Person-Mile Traveled (US & TX, 2020)

US	Annual PMT (2016 NHTS)	# Deaths in 2019	# Injuries in 2020	Deaths per 100M PMT	Injuries per 100M PMT
Bike	8,500 M	846	49,000	9.95	576.54

US	Annual PMT (2016 NHTS)	# Deaths in 2019	# Injuries in 2020	Deaths per 100M PMT	Injuries per 100M PMT
Car	3,020 Billion	22,215	2,448,000	0.74	81.06
Motorcycle	17,632 M	5,015	84,000	28.44	476.41
Walk	33,651 M	6,205	76,000	18.44	225.85
Total	3,079 Billion	34,281	2,657,000	1.11	86.27
Texas State	Annual PMT (2016 NHTS)	# Deaths in 2019	# Injuries in 2020	Deaths per 100M PMT	Injuries per 100M PMT
Bike	347 M	79	1,969	22.77	567.43
Car	202 Billion	2,472	189,558	1.22	93.52
Motorcycle	3,681 M	489	6,339	13.28	172.21
Walk	1,921 M	717	4,203	37.31	218.71
Total	208,636 M	3757	202,069	1.80	96.85

Source: TxDOT (2020b), Bureau of Transportation Statistics (2020).

Note: Walk PMT values do not include walking to and from one's car or most bus stops, or between stores while shopping, etc. Much of that walking is not along public roadways, however, and it is not many more miles when adding a ¼ or ½ mile distance to all car trips in the US.

In the period 2017–2018, US pedestrian deaths rose 3.4%, against a 2.4% decline across all crash fatalities (NHTSA 2019). This increase is concentrated in urban areas, with a 69% increase since 2009 compared to a 0.1% increase in rural areas. The Insurance Institute for Highway Safety (IIHS) noted that two-thirds of US-reported pedestrian crashes occur in urban areas, with the most significant increases (of 7.5% between 2009 and 2019) in urban areas occurring on arterial roads (defined as nonresidential corridors of two or more lanes in each direction) and a 4.9% increase on freeways. Male and female pedestrian fatalities rose 3% and 4.8%, respectively. Nighttime fatalities rose 4.6%, and fatalities from alcohol-impaired driving rose 2.2%, showing across-the-board increases in pedestrian fatalities as opposed to all crash fatalities (NHTSA, 2019).

Pedestrians' crashes tend to be more severe in rural areas, due in part to higher speeds and lack of sidewalks and/or protective longitudinal barriers (including medians). Stoker (2015) showed that risk for pedestrian injury on Dutch roads increased about 140% at night when lights were present, and 340% when lighting was not present. Urban areas, by contrast, typically have lower speeds, more sidewalks, and denser street networks, leading to lower rates of pedestrian death per person-mile walked due to the lower speeds there (Stoker, 2015; Zegeer & Bushell, 2012).

Data from IIHS (2018), as well as the Governor's Highway Safety Association (GHSA, 2018) point to SUVs having a higher rate of involvement in pedestrian crash fatalities in recent years, with a 50% increase in SUV-caused fatalities in the period 2009–2016, and a 7.9% year-over-year increase in SUV-caused fatalities from 2017 to 2018. This comes as the market share of 'light-duty trucks,' which encompasses SUVs, pickup trucks, minivans, and crossover utility vehicles

has increased, from less than 50% of vehicle sales in 2009 to 70% of new passenger-vehicle sales in the first half of 2019 (Ulrich, 2019). According to the GHSA study (2018), pedestrians struck by SUVs were about twice as likely to die as those struck by standard passenger cars, with significant increases in power-to-weight ratios at all vehicle weight benchmark percentiles (IIHS, 2018).

The WHO's Global Status Report on Road Safety (WHO 2018) notes that a 1% increase in average speed translates to a 4% higher likelihood of deadly crash outcomes (across all manner of collision, not just pedestrians) and a 3% increase in serious crash risk. Thus, "a 5% reduction in average speed can reduce the number of fatalities by 30%". The 5 "best practices" around the globe center on local authorities being able to set speed limits, city speed limits not exceeding 30 mph (and approximately 19 mph in residential areas), and the presence of national speed limits (which the US no longer has, alongside Russia, Brazil, and India, and unlike Mexico, Canada, Australia, Europe, and China). The US also does not demonstrate best practices for blood alcohol content (BAC) limits for ticketing "drunk drivers". China, Australia, Canada, Brazil and most of Europe do demonstrate best practices on that metric, with BAC having to be below 0.05 gm/deciliter (of blood) for adults and less than 0.02 (!) for young and other new drivers.

As the World Health Organization's Global Status Report (WHO 2018) notes, a stunning 25% of the world's public-roadway deaths are pedestrians (versus 17% in the US and 19% in Texas). Road traffic crashes are the 8th leading cause of death globally, and the leading cause of death for the world's youth (i.e., those ages 5 to 29 years). Road crashes take more than 1.35 million lives a year, along with 50 million injuries. As the report authors note, "people are less likely to walk, cycle, or use public transportation when conditions are unsafe", which then contributes to other leading causes of death, like "ischemic heart disease, stroke, chronic obstructive pulmonary disease, and diabetes" (WHO 2018, p. 5) Vehicle emissions (especially diesel emissions) also contribute to lung dysfunction and early death (claiming tens of thousands of American lives early every year, or over 100,000 Americans when coupled with others sources of fine particulate matter, like coal-fired power plants) (Goodkind et al. 2019).

They also note that "every one of those (crash) deaths and injuries is preventable" (WHO 2018, p. ix). Moreover, Canada, Australia, EU nations, and many others are quite far ahead of the US on various critical indicators for greater road safety. In other words, the US and Texas lag the world's nations and cities in many ways, not just on pedestrian design and death statistics (per capita, per VMT, and per vehicle owned), especially given Americans' relatively high income status (and decades of transportation system investment and vehicle safety testing). Americans' culture of high-speed, motorized travel, with drivers focused on other cars and trucks, rather than other travelers, and low-density land use probably needs to change if we are ever going to achieve anything close to our state's and nation's Vision of Zero Deaths on public roadways.

1.2. Examining Crash Factors

Many factors affect the likelihood of a pedestrian crash and pedestrian injury or death. These include demographic variables, whether or not a driver or pedestrian is distracted, roadway design details, traffic, lighting, weather conditions, and so forth.

1.2.1. Age and Other Demographic Variables

A pedestrian's age typically appears as the most significant factor affecting his/her road-crossing speed, with slower speeds thus increasing exposure to traffic while crossing. An observational study in Tel Aviv, Israel, found that at a 10-meter-wide crossing, persons over 65 walked across at 1.05 meters per second (mps) (3.45 feet per second [fps]), compared to about 1.45 mps (4.75 fps) for those aged 18 to 35, corresponding to a 28% decrease in walking speed for the older group. A fear of falling among older people has particularly worrisome affects. When controlled for age, interviewed crossers who reported a fear of falling spent more time looking at the pavement (and thus away from traffic) than those who did not report a fear of falling (26.4% vs. 14%) (Avineri et al., 2012). A study of crossing behavior in Utah also found a slower walking speed among seniors, especially those with canes or walkers (Berrett et al., 2020). The authors noted that the Utah Department of Transportation recommends a 3 or 3.5 fps crossing speed as opposed to the typical 4.0 fps crossing speed recommended in the 2009 *Manual on Uniform Traffic Control Devices* (MUTCD) (Berrett et al., 2020).

An analysis completed by the Massachusetts Department of Transportation (Dugan, 2019) found increases in pedestrian crashes among 55- to 74-year-olds in the period 2006–2015, with the proportion of deaths in this age group increasing from 18 to 27% in the same period, with people of color having higher death rates than white pedestrians. This study found that deaths peak during the evening rush hour for pedestrians aged 55 to 74 however, for those 75 or over, rates remained relatively flat, suggesting that older working adults of age group in 55-to-74 years are more at risk. Among drivers, the most frequently issued driver citation for those 25 to 54 and 55+ is for inattention, with 55+ drivers more frequently cited for attention and health-related issues. A hotspot analysis of crashes involving seniors in New Bedford, MA, showed that the majority of incidents happened on two-lane roads with a parking lane and were close to significant retail spaces/businesses. Two-thirds of these hotspots had crosswalks with faded striping, and several of these were the closest intersection to major apartment buildings catering to senior citizens (Dugan, 2019).

Furthermore, older adults are at a higher risk of death if they are involved in a crash. A 2013 study found that in any given crash scenario, a 70-year-old that was hit had the same risk of death as a 30-year-old who was hit by a car going 11.8 mph faster (Tefft, 2013). Older adults in New York City were also overrepresented in pedestrian crash deaths, comprising 38% of crash fatalities but only 12% of NYC's population (NYCDOT, 2010). These variables create a picture of

disproportionate vulnerability for older adults, with appropriate countermeasures needed to reduce vehicle speeds and increase visibility for older adults through dedicated crossing infrastructure.

Lower-income people, people of color, and younger children living in urban areas are broadly at a higher risk of being involved in a crash as a pedestrian. A longitudinal study in Canada found that for every quintile (20% of population) decrease in income, a person's crash risk rose 13% (GHSA, 2018). Stoker (2015) estimated that Americans making \$20,000 per year are around 7 times more likely to be injured as a pedestrian than someone making \$30,000 per year. Furthermore, analyses of crash data have found urban schoolchildren of color to be at a disproportionate risk for crash deaths, resulting in targeted educational programs to improve pedestrian safety around these schools (Bachman et al., 2015; McLaughlin et al., 2019).

1.2.2. Distracted Drivers and Pedestrians

Distracted driving as well as distracted pedestrians can be a significant factor in the prevalence of pedestrian crash injuries and fatalities. A 2014 study in Australia found that pedestrians under 31 are more likely to be involved in automobile accidents related to their mobile phone use that result in injuries. Pedestrians observed in this age group were found to have a 16% higher propensity to answer a text while crossing the street and a 24.5% higher propensity to “monitor” or look at their phone while crossing the street. The study noted that pedestrians holding phones while crossing the street tended to cross more slowly, look around less, acknowledge others less, and look out towards vehicles less (Lennon et al., 2016).

Erratic pedestrian behavior along with distracted driving together formed 67% of determined reasons for crash types that involve a vehicle going straight while the pedestrian is crossing the road. While causation patterns are heterogeneous overall, distracted driving was found to be the most common contributing factor (Yue, 2019). A broader pedestrian crash study conducted by the New York City Department of Transportation (NYCDOT) (2010) found crossing against a walk signal to be about 56% deadlier than crossing while the walk signal was activated. Overall, driver distraction was identified as a factor in 36% of crashes.

The basic idea of “distracted pedestrians” remains contested, as is the threshold of external stimulation at which a pedestrian would be considered distracted. Existing literature on distracted pedestrians generally finds no significant difference in instances of looking both ways before crossing the street between pedestrians that were using a phone and those that were not, particularly pedestrians were talking on the phone or listening to music (Simmons et al., 2020). Furthermore, Simmons et al. (2020) found no significant link between distraction and walking speed, or between decision-making processes when timing crossing the street between passing cars.

Ralph et al. (2020) examined broad trends in the literature and surveyed medical, planning, and engineering professionals at the 2019 Transportation Research Board (TRB) annual conference on their conceptions of distracted pedestrians and how large of a role they play in crash fatalities. The

survey of practitioners conducted by Ralph et al. (2020) finds differences between medical, engineering, and planning professionals in terms of attitudes surrounding distracted pedestrians and potential countermeasures. A “windshield bias” was displayed among those who used private car transportation to get to work, with that group on the whole believing that distracted walking was a large problem, coupled with a propensity to support lower-impact countermeasures, such as educational campaigns, rather than structural changes in the way infrastructure is developed. Ralph et al. (2020) attributed professional differences and windshield biases to two phenomena: 1) “signature pedagogies,” or the distinct personality and values of a discipline/academic area, and 2) an “illusory truth effect” that stems from media framing of distracted pedestrians as a legitimate issue.

Finally, one’s walking direction (either with or against traffic) seems to influence the frequency and severity of pedestrian crash injuries and fatalities. A 2013 study in Finland suggested a 77% decrease in fatal and non-fatal accidents when walking against traffic versus walking with traffic (Luoma & Peltola, 2013). Pai et al. (2019) found a similar pattern when analyzing 5 years of crash data and about 14,000 incidents in Taiwan. Pedestrians walking with traffic were about 2.21 times more likely to sustain fatal injuries than those walking against traffic. Furthermore, the percentage of non-fatal head and neck injuries was significantly higher among individuals that were walking with traffic, as opposed to head-on (Pai et al., 2019).

1.2.3. Presence of Signals, Crosswalks, and Other Facilities

Multiple studies examine the presence of signals and crosswalks to help understand how pedestrian and driver behavior changes with the presence or absence of control for the pedestrian or the driver. The literature mostly seeks to compare crossing behavior with certain facilities (such as a signal) to those without facilities in similar contexts.

Attitudes surrounding crossing at a crosswalk or crossing in the absence of a crosswalk are influenced by a variety of factors, including sex and age. A 2020 study from New Zealand found that 95% of that country’s pedestrian fatalities took place at uncontrolled intersections, and accompanying surveys about the issue found that the majority of the population saw nothing wrong with crossing at a location lacking pedestrian infrastructure of any kind if it seemed safe to do so. Additionally, respondents in this same group were more likely to agree that they crossed instinctively, while checking for cars on multiple occasions (Saethong, 2020). Further exploratory and observational surveys conducted in Wisconsin showed both pedestrians’ low propensity to believe that drivers would stop for them in the crosswalk as well as a low percentage of observed drivers yielding to someone crossing in the crosswalk. For unmarked crosswalks and marked crosswalks, about 22% and 36%, respectively, of those surveyed believed that a driver would yield; in observation studies, the average proportion for all intersections regardless of crosswalk status was 16% (Schneider et al., 2019).

The safety of unsignalized crosswalks seems dependent on which treatments they are combined with, rather than conditions such as the width of the road, presence or absence of a raised median, and presence of older pedestrians who cross more slowly. At large arterial roads, in excess of 12,000 annual average daily traffic (AADT), unsignalized crosswalks that were marked had higher pedestrian crash rates when paired with no other treatments compared to those that were unmarked (Zegeer & Bushell, 2012). Treatments that improve upon unsignalized crosswalks also tend to involve changing road design in such a way that roadway speeds are reduced, further decreasing risk (Stoker, 2015).

1.2.4. Speed

Average traffic speeds and speed limits play an outsized role in pedestrian crashes and, in particular, fatalities. A 2013 study of crash fatality records in the United States, normalized to 2007–2009 risk levels, found that the median impact speed for injured pedestrians was 14 mph, and for pedestrian fatalities, the median impact speed was 35 mph. The speed range at which the probability of a fatality increased most sharply was found to be between 25 to 40 mph, with a ~3% increase in the likelihood of a fatality with every 1 mph increase in speed. Furthermore, fatality risk at 54.6 mph was 90% (Tefft, 2013).

Higher speeds manifest themselves in a variety of crash scenarios. NYCDOT concluded that pedestrians hit by cars turning left are significantly more likely to be killed than those hit by cars making right turns, due in part to higher average speeds involved in left turns (NYCDOT, 2010). Data collected in Washington D.C.’s Vision Zero report puts the survival likelihood for a pedestrian struck at 20 mph at around 94%; that likelihood falls to around 25% for pedestrians hit at 50 mph (dc.gov, 2019). While these data differ as to the exact likelihood of fatalities, they show consistently that higher speeds generally lead to higher rates of pedestrian fatalities.

Analysts very rarely have access to actual speed (at time of collision) data when analyzing severity (or crash likelihood) data. The UK’s Department for Transport (DfT 2010) compiled a series of analyses with equations for predicting pedestrian crash outcomes across collision speeds, while controlling for pedestrian age, with the following results, for frontal collisions (the most common type of pedestrian-vehicle impact): “In all of the pedestrian datasets, the risk of fatality increases slowly until impact speeds of around 30 mph. Above this speed, risk increases rapidly—the increase is between 3.5 and 5.5 times from 30 mph to 40 mph. Although the risk of pedestrians being killed at 30 mph is relatively low, approximately half of pedestrian fatalities occur at this impact speed or below.” It should be noted that “fatal” means death within 30 days of the collision, due to being struck by the vehicle. Serious injury outcomes include “fractures, concussion, internal injury, crushing, severe cuts and lacerations, severe shock requiring medical treatment, or any casualty who was detained as an in-patient in hospital” (DfT 2010, p. 10), and slight injuries include minor sprains and non-serious bruises or lacerations.

Using Ashton and McKay’s 1970’s data set, with relatively current weight factors, the DfT (2010) report suggests a 9% of death for the pedestrian at 30 mph, rising to 50% at 40 mph, and 90% at 50 mph, as shown below, in Figure 1.1. Current NHTSA safety expert Peter Martin (2022) suggests that many or most pedestrians are “vaulted over” the striking vehicle at speeds of 35 mph and above, and then probably most harmed by the ground impact (since pavement and curbs are very unforgiving obstacles). Nevertheless, the type of vehicle striking the pedestrian (e.g., high or low bumper, tall or low, soft or hard hood), and his/her health condition (e.g., young or old, weak or strong bones) and motion (e.g., forward leg raised or lowered) during the collision are often extremely important for injury outcomes. Those age 0 to 14 years have estimated to have probabilities of fatality of 5, 30, and 75% at 30, 40 and 50 mph (as compared to 30, 50, and 90%, across all pedestrian-vehicle combinations), while those age 60 and up have fatal outcome probabilities of 60, 97, and 100% at those same 3 speeds!

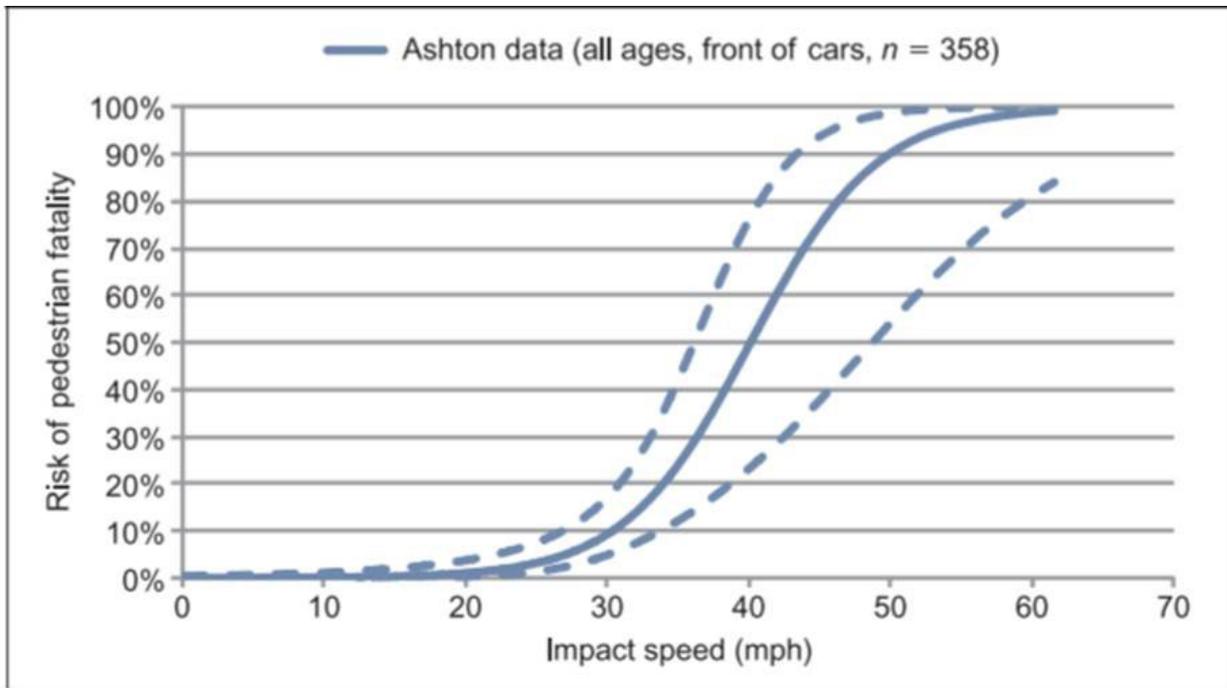


Figure 1.1 Pedestrian fatality risk
(and 95% confidence interval on those probabilities) at various vehicle approach speeds (in frontal collisions) (Source: DfT 2010 Figure 2.1)

1.3. Countermeasures

Countermeasures for pedestrian safety are broad and range from the installation of facilities for pedestrians to roadway improvements that enhance pedestrian visibility to education and automation. When understanding and evaluating the effectiveness of countermeasures, cost-benefit analyses are important to understand the relative impact of a given countermeasure; therefore, methods of cost-benefit analysis as well as pedestrian count estimation are included.

1.3.1. Effectiveness of Physical Countermeasures

Individual road treatments can be highly effective in determining pedestrian crash rates. A 2010 NYCDOT report documented physical improvements made at intersections throughout the city, starting with the highest risk intersections and working towards those with fewer crashes and fatalities. This included prioritizing pedestrian countdown signals at the riskiest 1,500 intersections throughout the city, with the aim to provide treatments to 60 miles of road per year, beginning with arterial roads. This priority list was developed after finding that 47% of all fatalities happen on 12% of roadways throughout the city (NYCDOT, 2010).

In the New York City study, streets with bike lanes implemented were around 40% less deadly, with speed bump treatments in certain areas reducing speeds in those areas by around 19%. A Safe Routes to School (SRTS) program was rolled out to around 135 schools, instituting permanent school zones around them to reduce speeds (NYCDOT, 2010). As a result, New York City has seen the sharpest decline in pedestrian crash fatalities in the United States between 2009 and 2018 (GHSA, 2018).

Schools have introduced a combination of physical and non-physical countermeasures through the Safe, Efficient and Flexible Transportation Equity Act (SAFETEA-LU) of 2005, which had made improvements to crossings and signage at about 10,400 elementary schools as of 2012. In Texas, these interventions included significant additions to the SRTS program, which culminated in a 42.5% decline in annualized rates of school-age pedestrian injuries in the pre- to post-SRTS intervention time period (2009–2010). This program specifically attempted to introduce crosswalk and signage improvements in the areas that had the highest risk, mainly at urban elementary schools (DiMaggio et al., 2015).

Similarly, studies that model demand changes show that creating safer conditions for pedestrians will lead to an increase in the use of pedestrian facilities. A study of the Great Dublin areas in Ireland by Carroll et al. (2018) showed that widening footpaths, increasing street lighting, and reducing the speed of the adjacent road to 30 km/h (18 mph) would result in a 25% increase in walking speed and a 5% increase in walking trips. A level-of-service regression model found that vehicle turning and turning radii had the largest impact on pedestrian level of service, suggesting a higher level of protection is needed at intersections to meaningfully improve perceptions of pedestrian safety (Carroll et al., 2018).

As noted earlier, speed kills. And cars and trucks can turn right (and left) at higher speeds, into crossing pedestrians, for example, when corner radii are longer. Fitzpatrick et al.'s (2022) recent FHWA report assigns a CMF of 1.0 to a tight corner geometry (curb radius of just 10 ft) and notes a 59% increase in pedestrian crashes at that same corner (CMF = 1.59) when the radius is widened to an excessive 70 ft (and right-turn average speeds went up just 4 mph). On-street parking helps moderate these crash counts, by slowing vehicles (due to sideswipe and open-door concerns) and reducing lines of sight (which also makes drivers more cautious). The authors also give a “general rule of thumb” that a 10% increase in traffic volumes (on either roadway) or pedestrian volumes

is associated with a 5% increase in pedestrian crash counts at that site. Moreover, a left-turn lane (or bay) without a raised pedestrian refuge (a raised median into the crosswalk area, for example) on the major street increased pedestrian crash frequency by 56 percent (!), compared to a no-left-turn bay situation or a raised refuge/median. Their sample size was 299 corners (in OR, VI and WA states), so a larger database could result in more distinctions by median type and left-turn controls on cars and trucks.

1.3.2. Effectiveness of Non-Physical Countermeasures

The introduction of educational programs in schools, particularly those targeted towards younger children, as well as educational campaigns targeted at the broader public, can help to instill more responsible crossing behavior by pedestrians and more care on the part of drivers.

A study of an educational program in Los Angeles elementary schools, conducted by a local hospital network in conjunction with police, used an in-class educational component and an observational component. Students were observed crossing before and after school one week before the education program was administered and one week afterwards. Scores on pedestrian safety knowledge tests, with questions asking, “how do you know a driver has seen you?” or “what should you do if you see a friend going after a ball in the street?” revealed answers that were significantly more conducive to safety after the educational programs were completed. The observational component also noted significant increases in those who looked both ways crossing the street, rising from 10% of children before the program to 41% afterwards. Schools that received the intervention had lower rates of pedestrian injury one year after the program (McLaughlin et al., 2019). There is further evidence that these school-hospital-police partnerships can deliver results by educating children about pedestrian safety, as a similar study on three different educational models (aimed at elementary, middle, and high schoolers) found significant increases in scores on pedestrian safety tests after the programs (Bachman et al., 2015). Table 1.2 summarizes the results of the study on elementary school students.

Other examples of education include billboards and other out-of-home advertising models to increase awareness when crossing the street or at railroad crossings. One such example is Operation Lifesaver on the Minneapolis light rail system, which followed a year of 6 fatal pedestrian crashes there; in the year following the campaign, there were 7 total pedestrian crashes involving the rail system as compared to 14 in the year before the campaign (Conlon, 2017).

Table 1.2 Survey Results of Los Angeles Elementary School Students

Comparison of pedestrian safety examination results before and after educational intervention (with values indicating number of students getting the correct answer)			
Question	Pre (n=1424)	Post (n=1522)	P-Value**
Q1: What sign do you look for to cross the street?	1182 (83)	1280 (84)	0.453
Q2: Who should be with you when crossing the street?	1229 (86)	1365 (90)	0.006
Q3: What is the first thing you do before crossing the street?	423 (30)	688 (45)	<0.001
Q4: Before taking your first step, which way should you look?	1131 (79)	1310 (86)	<0.001
Q5: A police officer's job is to...	1101 (77)	1256 (83)	<0.001
Q6: How do you know if a driver has seen you?	339 (24)	904 (59)	<0.001
Q7: How do you feel when you see a police officer?	987 (69)	1143 (75)	<0.001
Q8: What should you do if you see a friend going after a ball in the street?	836 (59)	1273 (84)	<0.001
Q9: Who should you ask for help?	1224 (86)	1395 (92)	<0.001
Q10: Who is responsible for your safety?	98 (26)*	149 (40)*	<0.001
Results reported as N(%). Obtained from Bachman et al., 2015 *n=672 (Pre) and n=374 (Post) children answered Q10 when it was added to the examination **Chi-square			

1.3.3. Crash Costs

To value the cost-effectiveness of different treatments, one should estimate their costs and benefits. There are a variety of standards and methods for performing benefit-cost analyses for pedestrian countermeasures, and the USDOT offers a more uniform standard, which the Texas DOT uses, albeit in a modified format.

A 2018 USDOT guide uses a KABCO scale for crash severity (Harmon et al, 2018). Crash unit costs are displayed based on the 2001 Consumer Price Index (CPI); the current cost can be obtained

by multiplying the cost by the ratio of the CPI for that year (CPI-U), and then dividing by the 2001 CPI. Table 1.3 shows the breakdown of categories and costs.

The USDOT document (Harmon, et. al, 2018) also includes a breakdown of comprehensive unit costs, including pedestrian and bike accident costs, along with certain types of auto accidents. Pedestrian comprehensive crash costs for non-intersection and intersection collisions are assumed to be \$287,900 and \$158,900, respectively. This document estimates \$242 billion in overall U.S. crash costs. Similarly, it estimates that about 40% of non-fatal injury crashes are not reported to the police, which complicates understanding the crash costs at any geographic level (Harmon et al, 2018).

Table 1.3 Breakdown of Crash Severity Ratings

Category	Severity of Injuries	Crash Unit Costs (2001 CPI)	Mean Person-Unit Costs (2001 CPI)
K	Fatality/Fatalities within 30 Days	\$4,008,900	\$7,119,608
A	Suspected Serious Injury	\$82,600	\$611,932
B	Suspected Minor Injury	\$82,600	\$137,117
C	Possible Injury	\$82,600	\$55,993
O	No Apparent Injury	\$7,400	\$11,539

Note: Table obtained from USDOT report by Harmon et al. (2018)

Texas employs a similar system to the KABCO system, utilizing K, A, and B in severity weighting. In general, higher severities add much more crash costs. Therefore, unweighted crash costs place a high cost on fatal and serious injury crashes, mostly due to short-term benefit-cost analyses being skewed by spikes in crashes. The USDOT report therefore recommends using long-term average predicted or expected crash frequency and applying unweighted crash costs to the estimates (Harmon et al., 2018).

Beyond pricing crash costs, there is also the challenge of pricing specific treatments against the level of benefit in the form of reduction in pedestrian fatalities. Given that there are distinct construction and implementation costs for all 50 states, as well as differing crash estimations among the states, obtaining overall cost-benefit analyses can be difficult.

1.3.4. Obtaining Pedestrian Counts and Usage

Finally, understanding how many pedestrians are in a given space can be particularly challenging, given the relative dearth in permanent pedestrian counting stations. As of April 2020, there were only 86 permanent pedestrian count stations in Texas, with the majority in Houston and Dallas. Other spot counts have been taken, but they only give information over a short period of time in a very specific place, rather than, for example, all crosswalks at an intersection (Texas

Transportation Institute, 2020). However, a small-area estimation method can be used to estimate pedestrian activity within a given census tract. The 2009 National Household Travel Survey (NHTS) and 2010–2012 California Household Travel Survey data were used as the basis for patterns of movement among different transportation modes for each tract, including daily average miles walked. Small area estimates were calculated by summing the surveyed average miles of the gender-age group categories, and then multiplying this by the 2010 population of the given tract. This result is then measured against the direct estimates from the NHTS and California Household Travel Survey data; a correlation can then be calculated between the small-area estimation method and the direct estimates from the NHTS data (Salon, 2016). Also, in relation to how many pedestrians are present, a Norwegian student estimated the cost of a “barrier effect”, as nonmotorized trips shifted to motorized trips. Under the assumption that nonmotorized travel is systematically undervalued from a time perspective, costs were calculated in VMT, per person and per shift from a non-motorized mode to a motorized one (Victoria Transport Policy Institute, 2020). Therefore, there are different approaches to assessing the costs of infrastructure and their effects on the environment for pedestrians. When combined with an understanding of where the highest demand (or potential demand) is present, localized estimation methods combined with cost-benefit analyses can indicate the countermeasures that will have the greatest impact and the locations that will most benefit.

1.4. Conclusions

Given the breadth and depth of the literature on pedestrian-involved crashes and fatalities, this review is meant to include only the most recent, salient information on the subject, and should be understood as a guide to overall trends in pedestrian crash fatalities as well as patterns in risk factors and effective countermeasures.

As indicated by multiple studies of macro trends in the United States and in Texas, pedestrian crash fatalities are on the rise, both in number and as a percentage of total crash fatalities, even as overall traffic fatalities decline (NHTSA, 2019; GHSA, 2018). Key trends beyond the overall increase in pedestrian fatalities include faster rates of increases in fatalities at nighttime and in urban areas as opposed to rural areas, on a per capita basis (IIHS, 2019). Broadly speaking, as light trucks (including standard-sized trucks, crossovers, SUVs, and minivans) have become a larger share of new car sales since 2009 (Ulrich, 2019), pedestrian fatalities have increased, and analyses of accident records show that light-truck-involved pedestrian crash fatalities are about 50% more deadly than those accidents involving passenger cars (GHSA, 2018).

Looking at crash factors, a few categories emerge as contributors to making a pedestrian crash event more likely or deadly. These include pedestrian age, distracted driving (and the idea of “distracted pedestrians,” but this is a highly contested idea), the presence or absence of pedestrian facilities, and vehicle speed. Age affects crossing speed, which can increase the exposure time of pedestrians in a marked or unmarked crosswalk (Avineri, 2012). Beyond increased exposure, older adults are also at an increased risk of injury and death when struck by a car, with the death risk for

a 70-year-old equivalent to the death risk for a 30-year-old who is hit by a vehicle traveling 11.8 mph faster (Tefft, 2013). While overall crash fatality patterns are heterogeneous, distracted driving was found to be a contributing factor in most crash types (Yue, 2019). Regardless of other factors, speed plays an outsized role in the severity of crashes. Between 25 and 40 mph, each increase of 1 mph was associated with a 3% increase in the risk of pedestrian death in any given crash. Death risk at 54 mph was 90% (Tefft, 2013).

Countermeasures can be both non-physical and physical, with benefits and costs weighted based on a variety of factors. NYCDOT has one of the more comprehensive breakdowns of countermeasures taken to reduce pedestrian crash fatalities in their city but broadly focuses on improving pedestrian visibility, reducing speeds, and implementing more bike and pedestrian infrastructure as part of a “road diet” (NYCDOT, 2010). As a result of these measures, New York achieved the steepest decline in pedestrian crash fatalities in the United States for the period 2009–2018 (GHSA, 2018).

Non-physical countermeasures include educational campaigns in schools and advertisements. A study of an educational program conducted by police and hospital staff in Los Angeles elementary schools found higher rates of understanding of pedestrian safety on a written test when compared with scores from a test administered before the lessons (Bachman et al., 2015). Additionally, advertising has been used on billboards and transit stops to educate pedestrians and drivers on the importance of maintaining awareness, leading to a significant decrease in crash fatalities around light rail systems in Minneapolis (Conlon, 2017).

Finally, assessing crash costs and usage are important in determining the salience of potential countermeasures. While the United States does not have a centralized system for assessing crash costs, the KABCO system provides a way to adjust crash costs for inflation and categorize by the scale and scope of the injuries and fatalities in a given crash (USDOT, 2018). While pedestrian count infrastructure in Texas is not yet comprehensive, small-area estimation methods can provide the basis for understanding where the demand for pedestrian facilities is the highest and where the needs for facilities are not being met (Salon, 2016). All countermeasures should be assessed based on the benefits relative to the costs, understanding those costs in terms of both the infrastructure-related expenses and any injuries and fatalities incurred, while assessing benefits in terms of fatalities and injuries prevented, as well as relative demand.

Chapter 2. Assemble and Prepare Relevant Crash Data

This chapter documents the process of assembling multiple data sources for the analysis of pedestrian-vehicle crashes in Texas. The data sources include: 1) Texas Department of Transportation's (TxDOT) Crash Records Information System (CRIS) records, 2) TxDOT Roadway Inventory with historic traffic flow, highway design, and geometric information, 3) detailed CR-3 police records for selected pedestrian crashes used for in-depth analysis, 4) National Household Travel Survey (NHTS) 2017 data to approximate walk distances for pedestrian exposure estimates across Texas, and 5) other relevant spatial and demographic data, including population, job, and rainfall, as well as important location features, such as schools, hospitals, and transit stops. All these datasets are spatially fused at the link and intersection levels using a map-matching mechanism, so that crashes can be counted and compared to vehicle miles traveled (VMT) through each location type to generate appropriate rates.

The sections can be summarized as follows: Section 2.1 describes data sources used for the project, including most prominently CRIS and the TxDOT Roadway Inventory. Section 2.2 explores key findings from looking at data from the census tract and county levels. Section 2.3 outlines challenges encountered, and mitigations developed when importing and beginning to analyze data sources for the project and concludes with an outlook on future project activities. Section 2.4 summarizes the task findings and activities. Finally, several appendices describe data assembly processes and detail the codes used on these huge, complex crash and network files.

2.1. Explaining Data Sources

This section describes the data sources used for the project, including CRIS, road inventory data, or other data sources.

2.1.1. CRIS: Crash Records Information System

A key source of data for this study, the TxDOT CRIS system (TxDOT, 2020a) uses details in police reports generated across all 254 Texas counties and the thousands of municipalities therein. Variables within the database characterize crashes according to time, location, severity, and road conditions. The CRIS data are primarily organized under three files:

- **Crash file:** Includes the variables explaining the crash event, e.g., crash location, date and time, road type, number of vehicles, crash severity, and other site-specific characteristics.
- **Person file:** Includes characteristics of persons involved in the crash (pedestrians, drivers, and occupants), e.g., age, gender, drug and alcohol involvement, injury severity.
- **Vehicle file:** Includes vehicle specific information such as vehicle type, make, model year.

Crash records are not guaranteed to have all variables defined, and many of these data are not provided. Further, not all crashes are reported. For example, if an officer determines that a crash does not incur at least \$1,000 of damages, the incident can be left unreported. Relevant aspects not explicitly captured by CRIS records involving pedestrians include whether each pedestrian is experiencing homelessness or where the pedestrian resides, especially if that pedestrian is lacking identification at the time of the crash. Further, police records are written by officers who are sometimes unable to capture these factors for a variety of reasons. Although these gaps and inconsistencies present challenges to comprehensive crash analyses, CRIS records remain a valuable resource, and offer suitable sample sizes for creating useful prediction models. The CRIS data from 2010 through 2019 comprise the following information categories:

- 5,631,223 crash records
- 9,875,257 roadway vehicles that are explicitly recorded among all crashes
- 4,756,671 crash records that have geographic coordinates, either from GPS latitude/longitude written in the crash record, or geocoded from street names or addresses
- 78,497 crashes that are determined to involve collisions with or avoidance of pedestrians
- 72,243 total pedestrians that are explicitly recorded among all crash records
- 18,265 reported pedestrian-related crashes that occur at intersections
- 5,674 reported pedestrian fatalities

While CRIS data has been obtained for the years 2005 through 2019, the portion included in the database for active analysis is the 2010–2019 period because of the consistent data format used in these years and their relevance, given the data’s recency.

2.1.2. Detailed Crash Records

The publicly available crash records in CRIS do not include certain details about those involved for privacy reasons. Detailed police reports can offer useful information about victims and motorists, including blood-alcohol levels, prior health issues, vehicle movements, and pedestrians’ position and actions prior to the crash. To gather this more detailed information, the research team requested the full police crash reports from TxDOT, selecting 300 pedestrian crashes ($n = 300$), using the following selection criteria:

- Pedestrian crashes that occurred in the last three years (2017, 2018, 2019)
- 70% of the crashes that resulted in fatalities or serious injuries

- 20% of the crashes that could not be geolocated due to missing information in CRIS (latitude, longitude)
- 70% of the crashes that occurred in major hotspot areas of 5 Texas regions (Houston, Dallas/Fort Worth, San Antonio, Austin, and El Paso)

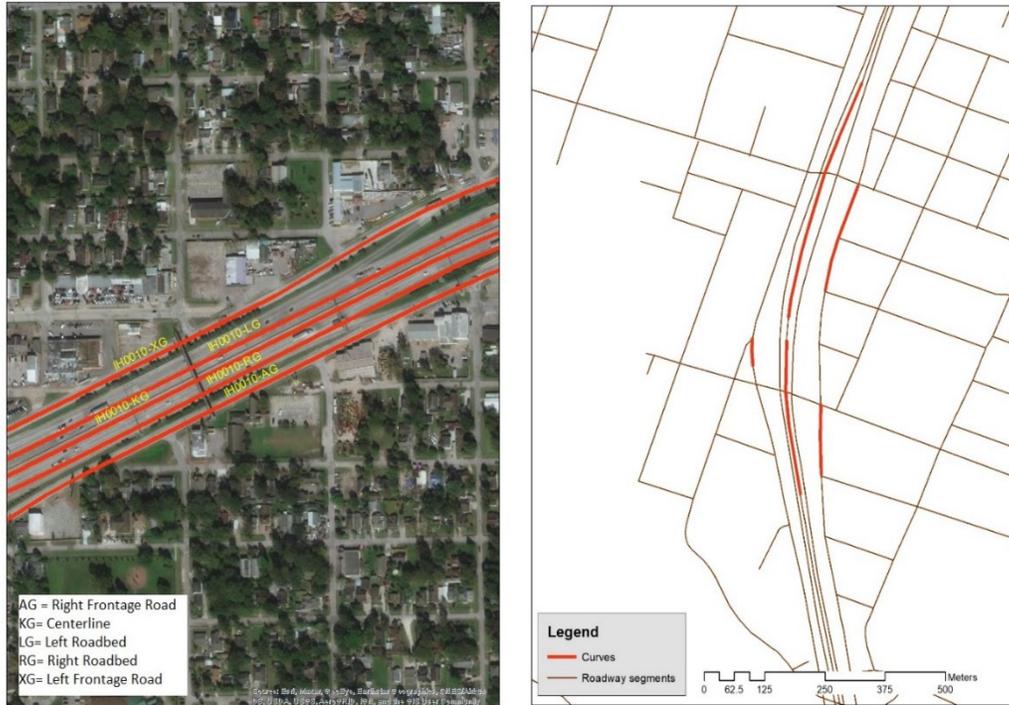
2.1.3. Road Inventory Data

The TxDOT Roadway Inventory database was used to obtain road-specific attributes (TxDOT, 2018). The database is available in GIS shapefile and tabular format. Both on-system (under the jurisdiction of TxDOT) and off-system roads (not under TxDOT jurisdiction) are included in the database. Expressway main lanes and frontage roads are presented as distinct road segments (Figure 2.1). For instance, if a road segment contains left and right frontage roads, then the main lane and frontage roads are usually represented as four unique road segments in the data because each represent a specific direction. The centerline miles show the mileage of a segment, regardless of the number of lanes, while the lane miles include the mileage of all lanes. Accordingly, the database contains a total of 80,455 centerline miles and 196,539 lane miles of highways in Texas (Table 2.1).

Important road attributes include highway design and traffic characteristics such as VMT, average daily traffic (ADT), percentage of truck ADT, shoulder and median types and width, number of lanes, and speed limit. However, the road inventory database contains no geometric information such as curvature length and angle. To map road geometry, the horizontal curves (GEO-HINI) database was spatially matched with the road inventory database.

Table 2.1 Mileage by Road Segments

Road Segments	Centerline Miles	Lane Miles
Mainlanes	72,885	180,669
Left Frontage Road	3,833	8,029
Right Frontage Road	3,737	7,840
Total	80,455	196,539



a) from TxDOT Roadway Inventory

b) Matching Curvature with Road Segments

Figure 2.1 Road Segment Types

2.1.4. Other Data Sources and Walk-Miles Traveled

The CRIS data were spatially matched with land use, population, job, rainfall, and other location features (schools, hospitals, transit stops) to examine the association between pedestrian crash counts and various contributing factors along Texas roads (Figure 2.2). Census tract-level population and job data were obtained from the 2010 US Census and Longitudinal Employer-Household Dynamics (LEHD) dataset respectively. Road segments were matched with the closest census tract centroid using the ArcGIS spatial join routine. Data were normalized by the area of census tracts. Other data sources include annual rainfall data (1981–2010) from the Texas Water Board, school locations from the Texas Education Agency, hospital locations from the Homeland Infrastructure Foundation, and transit stop locations from OpenStreetMap. ArcGIS Spatial Analysis tools were utilized to calculate numbers of transit stops and Euclidean distances from each road segment to the nearest schools and hospitals.

Walk-miles traveled (WMT) for pedestrian exposure measure was estimated using NHTS 2017 Texas add-on data. The full NHTS Texas sample includes information for 42,747 respondents. The dataset is derived from a household travel survey where each respondent provided full details of their travel and activities for a 24-hour travel-day. The origins and destinations of individual trips are included in the dataset, where the trip distance is derived from route geometry returned by Google Maps API. WMT for the travel day was calculated for individual respondents. Next, an

ordinary least-squares regression model was developed for WMT, which includes the variables influencing the WMT such as respondent's age, gender, ethnicity, household income, car ownership, educational attainment, and built environment characteristics such as population density and job density.

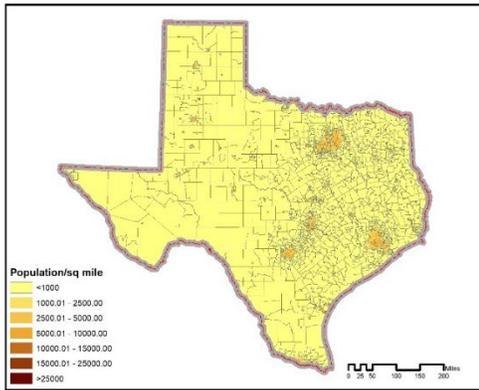
Table 2.2 shows the model estimation results. Age shows non-linear association with WMT. WMT increases with age for younger travelers but the effect of age on WMT diminishes and starts to fall for older adults. Black and Hispanic people have lower WMT compared to White people. Students, people with higher educational attainment, and those without car have higher WMT. Walking tendency also varies depending on the day of week. People tend to walk more on weekends compared to weekdays. Among other variables, higher population density and job density show significant positive association with WMT as more destinations and opportunities are reachable within walking distance with increasing densities.

In the next step, parameters estimated from the model were used to predict WMT at Public Use Microdata Areas (PUMA). PUMAs are contiguous geographic units that contain no fewer than 100,000 people each. Public Use Microdata Sample (PUMS) contains individual demographic and housing information where each record corresponds to a single person. The most detailed geographical unit contained in the PUMS files is the PUMA. PUMS of an individual year for a particular state contains one percent of the state population. PUMS data includes person weight for each observation, which is equal to the number of people that the observation represents. Using the demographic information and person weights in the PUMS data, WMT for each PUMA was predicted based on the estimated model. Finally, PUMA-level WMT estimates were used to calculate county-level WMTs. WMT/capita for each PUMA was assigned to the census tracts contained in that PUMA. Tract-level WMTs were finally scaled up to generate county-level WMT (Figure 2.2).

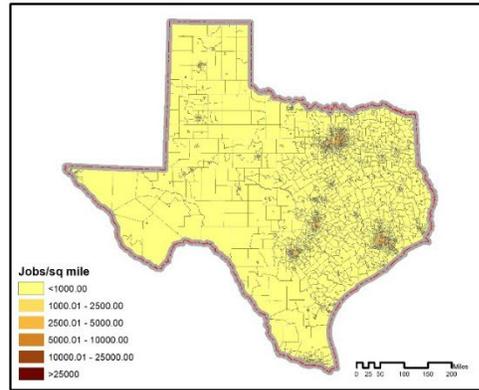
Table 2.2 Ordinary Least-Squares Model Results for Predicting Y = WMT

	Estimates	Std. Error	P-value
(Intercept)	0.0628	2.52E-02	0.013
Age	0.0032	9.04E-04	0.000
Age ²	-3.31E-05	8.84E-06	0.000
Household size	-0.0122	2.73E-03	0.000
Male	0.0096	6.46E-03	0.136
Race			
Black	-0.0416	1.17E-02	0.000
Asian	0.0182	1.39E-02	0.190
Hispanic	-0.0403	8.81E-03	0.000
Other	-0.0311	2.12E-02	0.144
(Reference: White)			
Student	0.0786	1.32E-02	0.000
No car	0.5198	2.47E-02	0.000
Educational attainment			
College degree	-0.0037	9.71E-03	0.701
Bachelor's degree	0.0712	1.01E-02	0.000
Graduate degree	0.1045	1.08E-02	0.000
(Reference: High school or lower)			
Weekends	0.0126	7.18E-03	0.078
Population density	5.09E-06	2.47E-06	0.040
Job density	1.21E-05	1.93E-06	0.000
No of observations	43477		
R ²	0.0212		
F-statistic:	57.77		
P-value	< 2e-16		

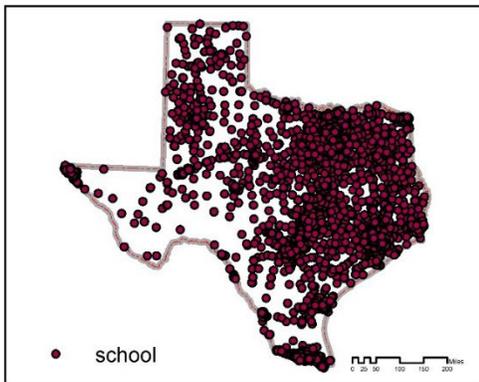
Population Density



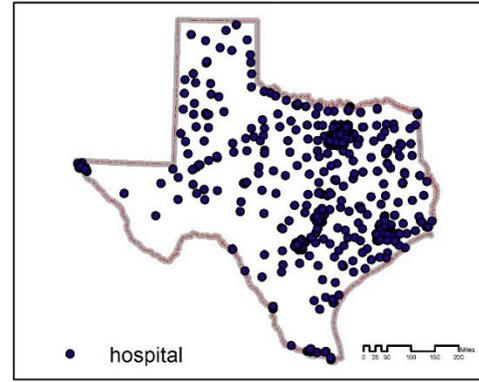
Job Density



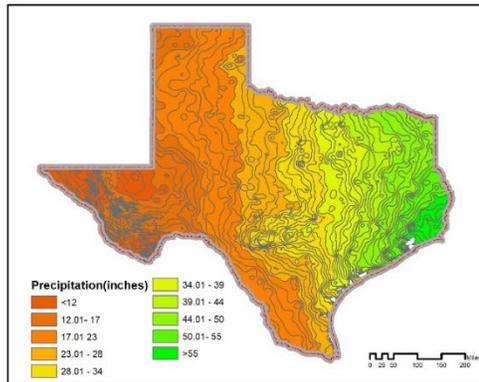
Schools



Hospitals



Average yearly precipitation(1981-2010)



Transit Stops

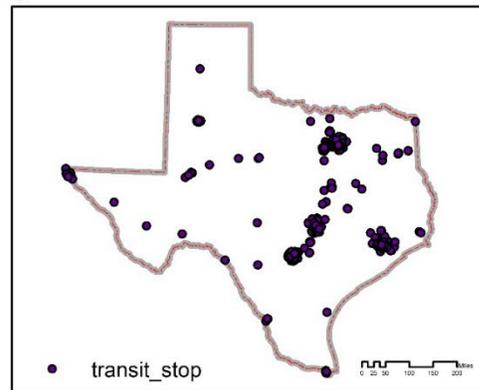


Figure 2.2 Demographic and Spatial Data

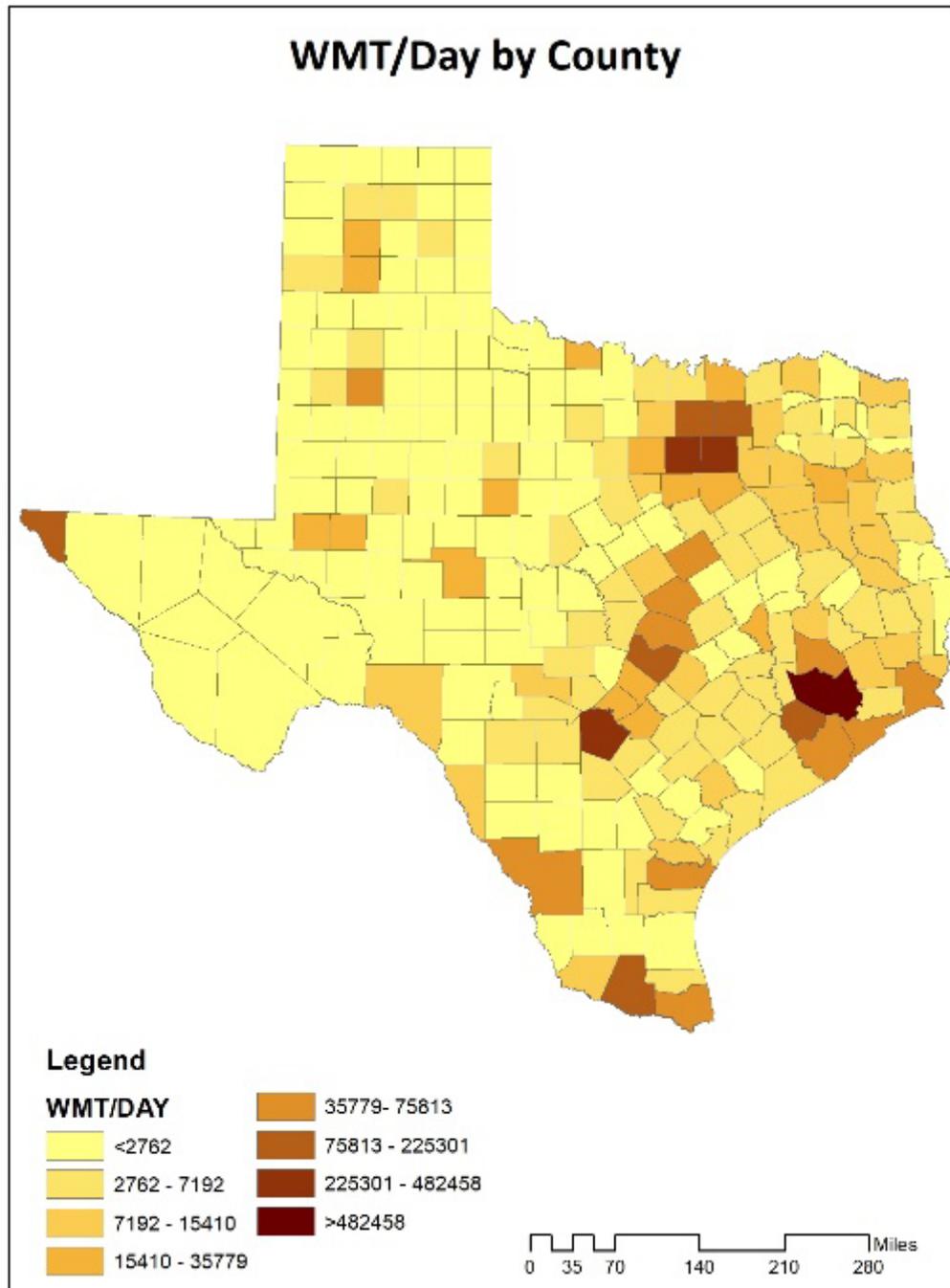


Figure 2.3 WMT by Texas Counties

2.2. Summary Statistics

Over the last ten years (2010–2019), 5.6 million crashes were reported in Texas, resulting in 35,306 traffic fatalities. Just 1.4% of all these reported crashes involved pedestrians, but 16.1% of all Texas’ crash fatalities were pedestrians (5674 pedestrian fatalities), as shown in Table 2.3. The number of pedestrian fatalities has also been rising at a much faster rate (86% over 10 years) than the number of Texas crash fatalities (19%), the statewide population (15%), and Texas VMT (20%). Figure 2.4 illustrates the increase in pedestrian fatalities in Texas.

While Americans are walking somewhat more, their walking distances alone cannot explain this increase. The NHTS data suggest that from 2009 to 2017 WMT per capita rose 13.4% and walking trips per capita rose 6%—from 168.6 walk trips per capita per year to 179.2¹. Yet US pedestrian fatalities per capita rose 46% over that same 9-year period. In 2017, only 10.4% of US person-trips involved walking a block or more (to one’s destination or a bus, for example), yet pedestrian deaths were 16% of all traffic fatalities and thus over-represented, regardless of how one examines the question². Human bodies are simply no match for the weight and density of cars (and buses, trucks, and trains). Many vehicles’ hoods dwarf the humans alongside (and inside) them.

Houston, Dallas, San Antonio, Austin, and Fort Worth are Texas’ 5 largest cities, accounting for 36% of the state’s pedestrian fatalities (Table 2.4) while comprising about 24.3% of Texas’ population. Among these, Austin has the highest share of pedestrian deaths (versus all crash deaths), but 126 miles walked per person per year (versus 112 mi/year/person as the US average and values under 90 mi/year/person in other Texas cities) (Figure 2.5). Dallas has the most pedestrian deaths per capita (40 per 100,000 persons) among Texas cities with populations over 100,000.

As shown in Figure 2.6 and Figure 2.7, pedestrian crashes come with higher levels of injury severity, as compared to other motor-vehicle crashes. For example, 7.12% of pedestrian crashes result in fatalities, versus just 0.47% of all other crashes.

Like many other travelers, pedestrians face elevated injury risk at night. As shown in Table 2.5, 79.28% of Texas pedestrian deaths occurred at nighttime/without daylight. The highest percentage of fatalities occurred from 9 PM to 10 PM (Figure 2.8). There is also higher risk of severe injuries in early morning hours (5 AM to 7 AM). There might be several possible explanations: during these time periods (late night and early morning hours), traffic is lighter than usual, which might cause both pedestrians and drivers to ignore safety rules (drivers might travel at reckless speeds while pedestrians might choose to crossroads abruptly). Moreover, pedestrian activities early in

¹ There was a major change in the method of collecting trip distance in the 2017 NHTS. Trip distance was calculated using the shortest path routes between geocoded origins and destinations in 2017. Previous surveys used self-reported distances.

² A “trip” is defined as “from one address to another,” excluding short trips within the same address, such as going to the mailbox or parking. In 2017, NHTS changed the definition of trip to allow “loop trips” that started and ended at home.

the morning (walking, jogging, physical exercise) and alcohol/drug involvements at night combined with darkness might also contribute to high injury severity during overnight hours.

As shown in Figure 2.9, fatality rate increases with speed limit. The median speed limit in locations of fatal pedestrian crashes is 45 mph, versus a median speed limit of 30 mph for locations of non-fatal crashes. Although the posted speed limit usually influences vehicle speed on roads, a more appropriate indicator would be the actual speed of the vehicle at impact, which is difficult to obtain for many cases.

Among different vehicle types, pickup trucks, sport utility vehicles (SUVs), vans, heavy-duty trucks, and buses significantly increase pedestrian injury severity in pedestrian-motor vehicle crashes. According to CRIS data, the number of light-duty trucks involved in pedestrian deaths is increasing at a fast rate in Texas: during the 2010–2018 period, the number of cars involved in fatal pedestrian crashes increased by 64.7% while the number of SUVs and pickup trucks involved in fatal pedestrian crashes increased by 98.6% and 92.9%, respectively. The growing popularity of SUVs, pickup trucks, and vans partly explains the high injury severity associated with these vehicles. Figure 2.10 indicates, from 2009 to 2016, the share of cars to the total number of light-duty vehicles purchased in the US dropped from 60.5% to 43.8%, while during the same time period, share of SUVs, crossover utility vehicles (CUVs), pickup trucks, and vans increased from 39.4% to 56.2% (EPA, 2017).

Alcohol and/or drugs were involved in 37.6% of pedestrian deaths. As shown in Table 2.6, alcohol was involved in 24.4% of pedestrian deaths—the pedestrians themselves tested as positive for alcohol in 21% of those cases. Of these alcohol-involved fatalities 82% were at night (8 PM to 6 AM), 55% were on weekends (Friday 7 PM to Sunday midnight), and 77.4% were male pedestrians or drivers under the influence. Another 20% of pedestrian deaths involved drugs in some fashion.

Fifty-eight percent of pedestrians killed are 20 to 54 years old. Nearly one-third of the pedestrian fatalities were people 55 or older. As shown in Figure 2.11, the 5-year age group with the most pedestrian fatalities was 55–59. Male fatalities were higher than female: 72% of all the pedestrians killed in traffic crashes were male.

Table 2.3 Number of Crashes Over Time (2010–2019)

Year	Number of Crashes	Total Fatalities	Pedestrian Crashes	Pedestrian Fatalities	Percentage of Total Traffic Fatalities	Population (in millions)	Pedestrian Fatalities per 100,000 people
2010	472,440	3097	6207	374	12.08%	2.51	1.49
2011	456,149	3119	6174	460	14.75%	2.57	1.79
2012	495,890	3461	7418	514	14.85%	2.61	1.97
2013	521,473	3436	7676	511	14.87%	2.65	1.93
2014	555,293	3580	7826	514	14.36%	2.70	1.91
2015	601,170	3625	8267	591	16.30%	2.75	2.15
2016	632,282	3851	8759	715	18.57%	2.79	2.57
2017	620,830	3762	8590	639	16.99%	2.83	2.26
2018	627,760	3700	8521	658	17.78%	2.87	2.29
2019	647,937	3675	9059	698	18.99%	2.90	2.41
Total	5,631,224	35306	78497	5674		n/a	

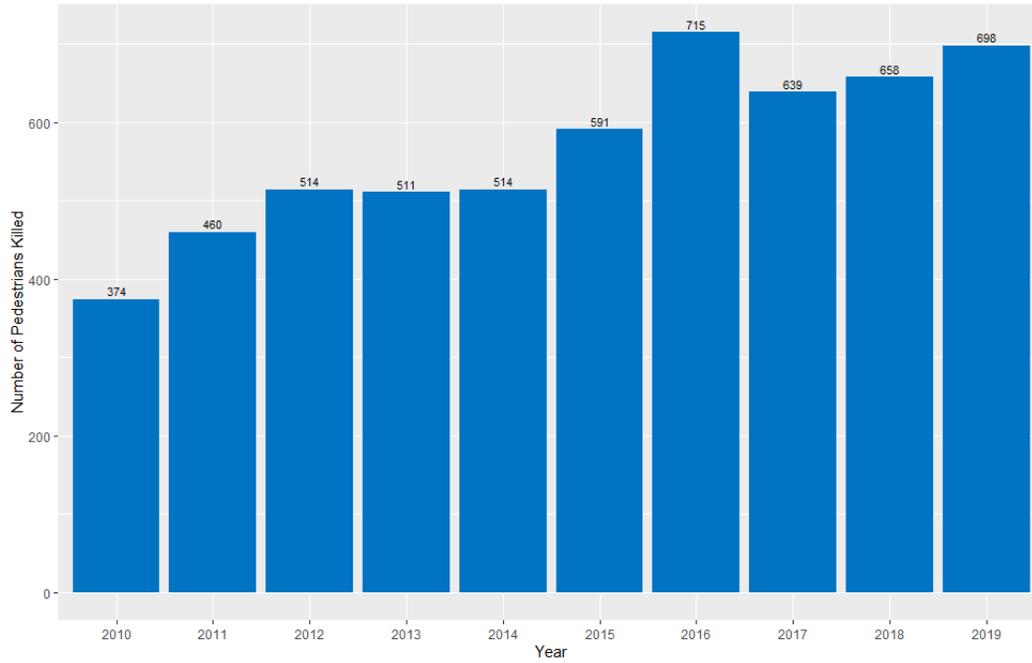
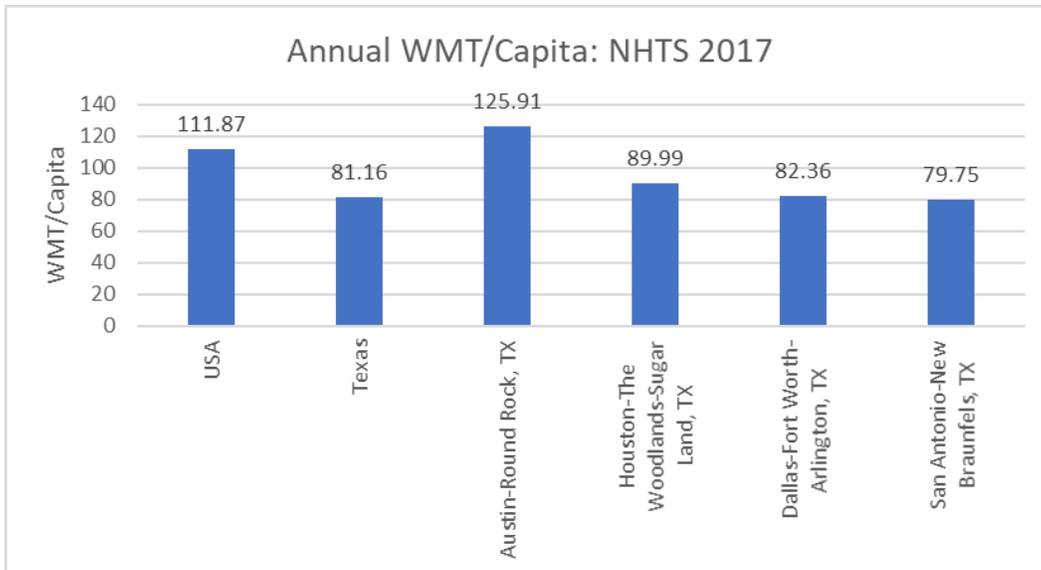


Figure 2.4 Number of Pedestrian Fatalities (2010–2019)

Table 2.4 Pedestrian Deaths in Texas Cities (of Population >100,000)

Name	Population	Pedestrian Deaths	Ped Death per 100,000 Persons
Dallas	1,197,816	478	39.91
San Antonio	1,327,407	474	35.71
Corpus Christi	305,215	106	34.73
Austin	790,390	249	31.50
Beaumont	118,296	37	31.28
Fort Worth	741,206	231	31.17
Houston	2,099,451	635	30.25
El Paso	649,121	192	29.58
McAllen	129,877	37	28.49
Waco	124,805	34	27.24
Lubbock	229,573	62	27.01
Amarillo	190,695	46	24.12
Brownsville	175,023	37	21.14
Killeen	127,921	26	20.33
Laredo	236,091	47	19.91
Arlington	365,438	63	17.24
Garland	226,876	39	17.19
Mesquite	139,824	24	17.16
Midland	111,147	19	17.09
Abilene	117,063	20	17.08
Irving	216,290	36	16.64
Denton	113,383	18	15.88
Wichita Falls	104,553	14	13.39
Grand Prairie	175,396	21	11.97
Carrollton	119,097	13	10.92
Pasadena	149,043	15	10.06
Plano	259,841	16	6.16
McKinney	131,117	8	6.10
Frisco	116,989	0	0.00



Source: NHTS, (2017)

Figure 2.5 Walk-Miles Traveled per Resident per Year

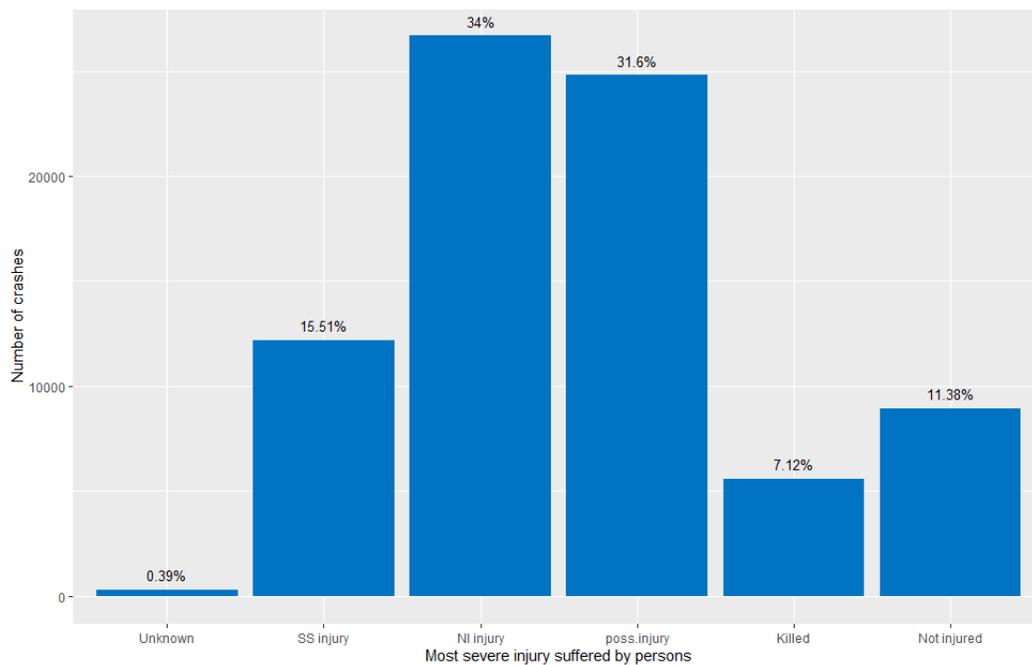


Figure 2.6 Injury Types Suffered in Texas' Pedestrian Crashes³

³ SS Injury= Suspected Serious Injury; NI Injury= Non-Incapacitating Injury; Poss. Injury= Possible Injury

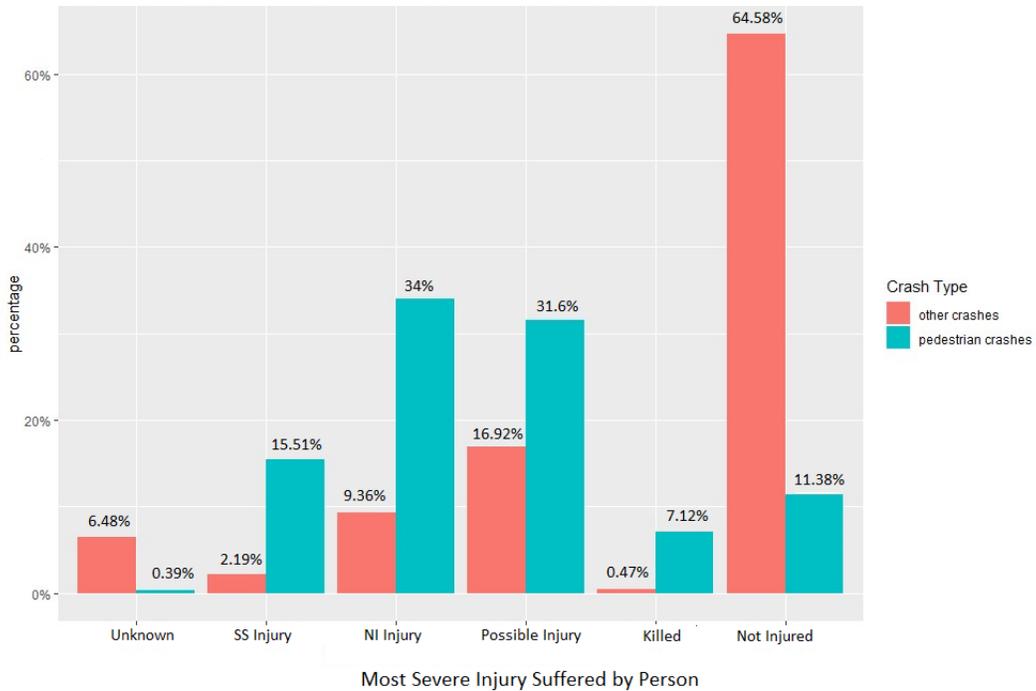


Figure 2.7 Most Severe Injury Suffered by Pedestrians in Texas

Table 2.5 Pedestrian Crashes and Fatalities under Different Daylight Conditions in Texas

	Number of Crashes	% of Total Texas Ped Crashes	Fatalities	% of Total Ped Crash Deaths
Daylight	46,350	59.05%	1167	20.72%
Dark, Not lighted	10,476	13.35%	2250	39.95%
Dark, Lighted	18,701	23.82%	2011	35.71%
Dark, Unknown lighting	737	0.94%	57	1.01%
Dawn	714	0.91%	55	0.98%
Dusk	1148	1.46%	67	1.19%
Other	89	0.11%	7	0.12%
Unknown	282	0.36%	18	0.32%
Total	78,497	100.00%	5632	100.00%

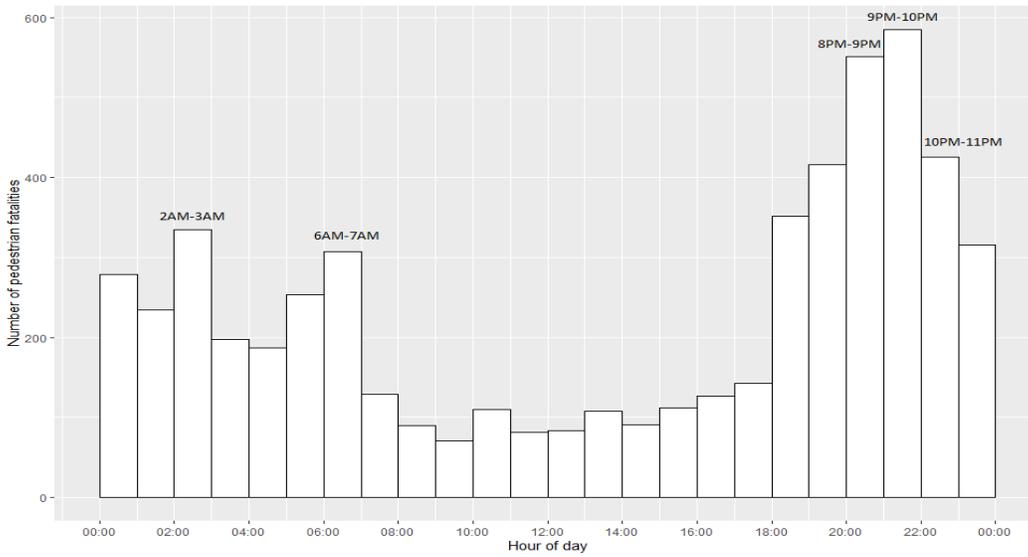


Figure 2.8 Pedestrians Fatalities by Time of Day

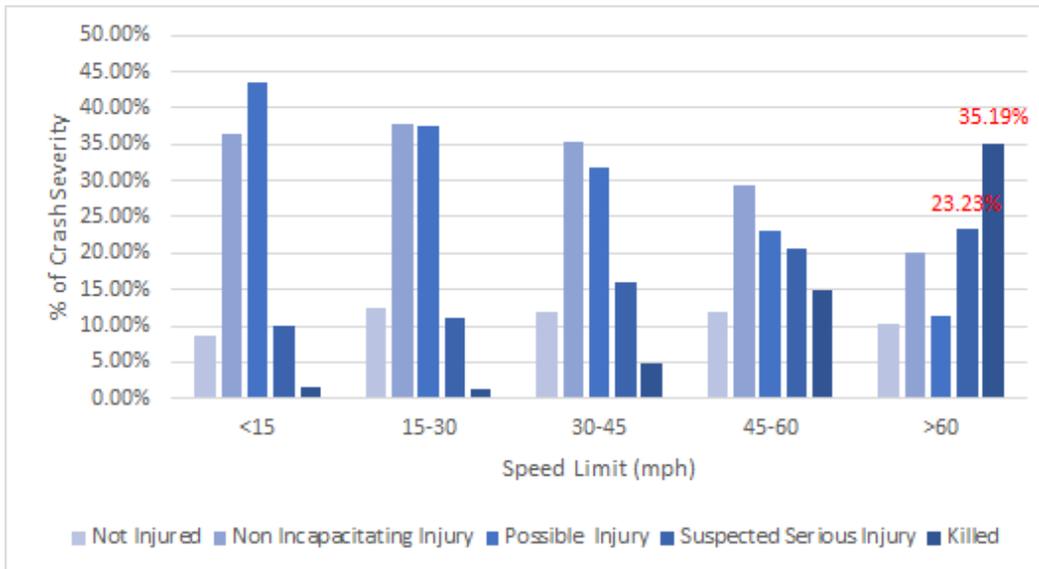
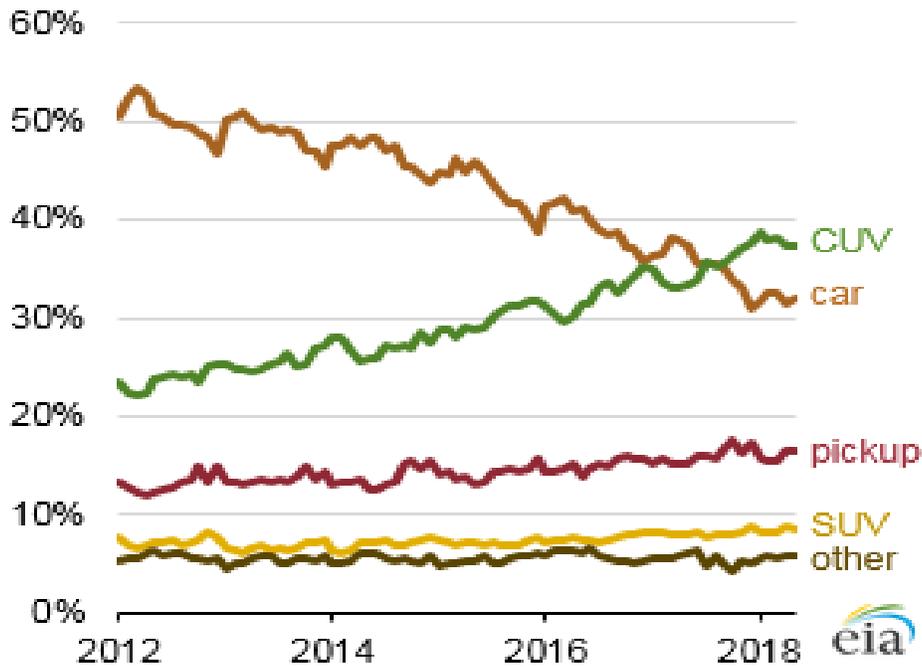


Figure 2.9 Injury Outcomes for Pedestrians on Roadways with Different Speed Limits

Light-duty vehicle sales by class percent



Source: Energy Information Administration

Figure 2.10 Light-Duty Vehicles Sales across US by Body Type, over Time

Table 2.6 Alcohol Involvement in Crashes that Resulted Pedestrian Fatalities

	Driver Positive	Driver Negative	Driver Result NA	Total
Pedestrian Positive	59 (1.04%)	121 (2.13%)	1016 (17.91%)	1196 (21.08%)
Pedestrian Negative	75	201	1303	1579
	(1.32%)	(3.54%)	(22.96%)	(27.83%)
Ped Result NA	114	289	2496	2899
	(2.01%)	(5.09%)	(43.99%)	(51.09%)
Total	249 (4.37%)	623 (10.77%)	5392 (84.86%)	5674 (100%)

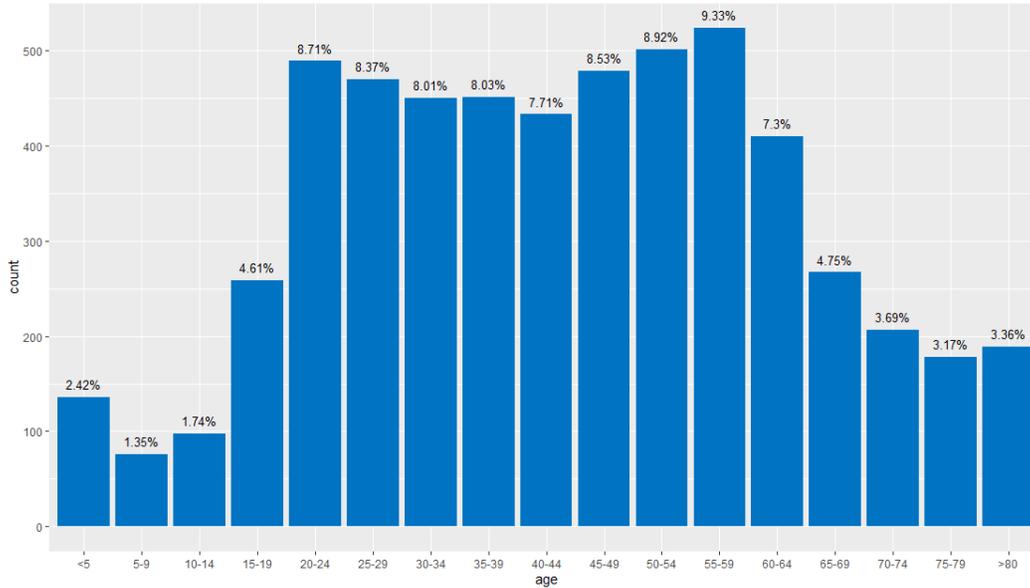


Figure 2.11 Pedestrian Killed in Texas Traffic Crashes by Age Group

2.3. Other Data Processes and Challenges

Three important aspects of using CRIS data for analysis are knowing where a crash is located, determining whether a crash is pedestrian-related, and understanding crash density in a locality. For crash location, many CRIS records contain GPS coordinates that are derived by TxDOT personnel from street address information or GPS coordinates that were provided with the record. However, roughly 15% of all CRIS records over the past decade (2010–2019) contain no geographic information. An initial look indicated a few reasons. For example, street information is often missing or mislabeled (e.g., a street that should be labeled as a “Dr.” is labeled as “Rd.”). Some locations are listed in the CRIS record as a pair of intersecting street names, yet those streets do not actually intersect.

Another challenge that plays into understanding location is in knowing the traffic direction of a divided road or highway on which a crash takes place (e.g., northbound versus southbound lanes). Further, there is sometimes ambiguity concerning whether a crash takes place on freeway mainlines or on the service roads. As researchers become more familiar with the datasets, hints concerning precise crash location continue to be discovered. Meanwhile, most analyses are expected to require less precision.

Along the lines of crash location, understanding whether the crash takes place at an intersection aids analysis. Although CRIS contains the “AT_INTRSCT_FL” and “INTRSCT_RELAT” fields, it is also helpful to consider the crash location with respect to the TxDOT Roadway Inventory. Unfortunately, the Roadway Inventory does not contain intersection information or locations, requiring a derivation of intersections from the roadway line segments themselves. The first attempt at this involved identifying where TxDOT Roadway Inventory geometry crisscrosses

(while filtering out conditions where we do not want to assume the presence of an intersection, such as places that are likely to be grade-separated), and then do a proximity match with crash records. This commonly misclassified crash locations around grade separations and directions of divided highways. A subsequent effort described in Chapter 5 leveraged crowd-sourced intersection locations within OpenStreetMap to map intersections in terms of TxDOT Roadway Inventory (Perrine and Zuniga-Garcia, 2021).

CRIS records do not have a single mechanism for identifying whether a crash is a “pedestrian crash.” Several field values in various portions of a crash record can indicate pedestrian involvement. Another element of uncertainty stems from the possibility of varying consistency among personnel who create the crash reports in terms of how pedestrian involvement is reported. To illustrate the overall complexity, consider that the “harmful event” field can indicate only one value for a given crash, while the actual events could have been more complex. Several hints occur within the CRIS records that can be queried to arrive at a “yes/no” categorization on whether the crash is a “pedestrian crash.” The solution proposed in this chapter is to query for criteria and store those results into a table that can be leveraged for filtering. Several criteria are identified here:

- One or more Units are identified as a pedestrian.
- One or more Persons/Primary Persons are identified as a pedestrian. (This offers an indication, too, of whether there is a pedestrian fatality).
- The “harmful event” of a crash is a pedestrian injury.
- The “other factor” involves a swerve or slowing down because of a pedestrian.

These criteria may or may not be mutually exclusive, and the “yes/no” indication of “pedestrian crash” is determined by the presence of one or more of these criteria.

For determining crash density with respect to TxDOT Roadway Geometry, a process was devised for matching crashes to nearby geometry. However, subsegments in the Roadway Geometry may range from 0.001 miles long to over 50 miles long. This can create high- or low-density measurements that are not accurate or useful. A solution is to resample to a uniform distance. A caveat to remember is that the only CRIS records that can be matched are those that have geographic coordinates—approximately 85% of all records and 74% of the pedestrian-related crash records. While this is a subsample of all records, it still represents a large contingent that is useful for the project’s analysis tasks.

2.4. Conclusions

The dataset constructed in this project provides valuable information about the trend and risk factors associated with pedestrian crashes in Texas. Over the last ten years (2010–2019), traffic fatalities in the state of Texas increased by 19%. By contrast, pedestrian fatalities rose by a

stunning 86%, and their share of deaths went from 12% to 19%. While Americans are walking more (NHTS data suggest that from 2009 to 2017, WMT per capita rose 13% and walking trips per capita rose 6%), their walking distances cannot explain these fatality numbers. Simply put, pedestrians experience dramatically higher risk than those seated inside vehicles (as shown in Table 1.1).

Preliminary data analysis identifies several risk factors. Pedestrians face elevated risk of fatality at night. Seventy-nine percent of Texas pedestrian deaths occurred at nighttime/without daylight. There is also high risk of severe injuries in early morning hours (5 AM–7 AM). Walking and driving under the influence, particularly at night, is another of the major causes of pedestrian fatalities. Alcohol and/or drugs were involved in 37.6% of pedestrian deaths. Both age and gender are pedestrian characteristics having significant correlation with fatality rates. Injury severity increases with most any crash victim's age (inside or outside the vehicle). And male pedestrian deaths are much higher than for females: 72% of all the pedestrians killed in Texas' traffic crashes were male.

Roadway speed limit and vehicle types also influence death rates. Higher speed limits are associated with more severe injuries and deaths. Median speed limit in locations of fatal pedestrian crashes is 45 mph versus 30 mph for non-fatal crashes. Another major concern is the growing popularity of SUVs, pickup trucks, and vans. During the 2010–2018 period, the number of passenger cars involved in fatal pedestrian crashes in Texas increased by 64.7% while the number of SUVs and pickup trucks involved in fatal pedestrian crashes increased by 98.6% and 92.9%, respectively.

Despite the shortcomings noted in the earlier section, the CRIS and TxDOT Roadway Inventory datasets are ready to be used alongside the other datasets described in this chapter for activities that support future project tasks. These include continuing initial analyses and visualizations of the datasets including heat maps and regressions. These efforts lead to the development of crash rate and injury severity prediction models, the building of decision trees to categorize crashes, as well as analyses of countermeasure cost-effectiveness.

Chapter 3. Identify Analysis Methodology and Key Results

This chapter provides a review of the data analysis processes, including models used, methodology, results, and a discussion on the practical significance of results. These various interpretations of the Crash Records Information System (CRIS) data serve to further understanding of trends and patterns of pedestrian crashes and fatalities in the period 2010–2019 and will help to develop specific countermeasures to reduce their frequency and severity. The team conducted analyses such as negative binomial (NB) operations on pedestrian crashes and fatalities and ordered probit (OP) and heteroskedastic ordered probit (HOP) models on a variety of factors, such as roadway geometry, vehicle type, and the presence or absence of roadway lighting. The chapter also includes ordinary least-squares regressions (OLS) on crashes and fatalities per walk-miles traveled (WMT) and per 1 million vehicle-miles traveled (VMT). Additionally, heat maps and cluster analyses show spatial trends among pedestrian crashes and fatalities across Texas. These data analysis techniques are meant to reflect different ways of looking at and thinking about the structure of and trends within the CRIS data.

The rest of the chapter is structured as follows: Section 3.1 presents summary statistics of pedestrian crashes, WMT, and roadway characteristics. Section 3.2 examines the characteristics of individual attributes of the CRIS data, such as crash speed and time of day. Section 3.3 is a brief overview of an ongoing cluster analysis that examines the spatial characteristics of crashes, including heat maps to visualize the spatial distribution of crashes. Analysis methodologies for crash counts in the NB, OP, and fatality count analysis using OP and HOP models are included in Section 3.4, along with an overview of the methodology for the ordinary least-squares regression. Section 3.5 contains zonal statistical regressions utilizing OP and HOP models, as well as ordinary least-squares regressions for crashes and fatalities per WMT and per 1 million VMT. Section 3.6 is a discussion of the results obtained from the methodologies in Sections 3.4 and 3.5.

3.1. Summary Statistics for Pedestrian Crash Variables

Different statistical models help to draw different types of conclusions when analyzing CRIS data. The OLS model, measured at the county level, helps to discern broader trends across variables, particularly socioeconomic variables, and help to understand why pedestrian crashes and fatalities may play out in different types of roadway environments (i.e., urban, suburban or rural). The NB, OP and HOP models provide greater detail through analysis at the link level, allowing for analysis on factors that would otherwise be too noisy at the county level, such as the distance to the nearest hospital or school. Analysis at the segment level also allows for analysis of lane counts and AADT that are not aggregated over potentially hundreds of thousands of lane-miles, allowing the researchers to zero in on potential hotspot segments. OLS models are preferable for describing the rate of change, while the OP and HOP models are preferable for describing the level of variation

to show the strength of the relationship. Overall, the different model estimation approaches used here shed light on specific factors driving the number and severity of pedestrian crashes. Countermeasures can subsequently be developed that target those factors deemed most important or having the strongest relationship to driving up the numbers of pedestrian crashes and fatalities.

3.1.1. Data Details for CRIS and TxDOT Roadway Inventory

A key source of data for this study is TxDOT's CRIS (Texas Department of Transportation, 2020). These records come from police reports generated in all 254 Texas counties and thousands of municipalities therein. Variables within the database characterize crashes according to time, location, severity, and road conditions. Crash records are not guaranteed to have all variables defined, and many of these data are not provided. A relevant aspect not consistently captured by CRIS records involving pedestrians is whether each pedestrian is experiencing homelessness.

Although these characteristics of the database present challenges when performing an analysis on crashes, CRIS remains a valuable resource, and offers suitable sample sizes for creating useful prediction models. Following is a summary of the dataset used for this analysis (encompassing the years 2010 through 2019):

- 5,631,223 crash records are logged
- 9,875,257 roadway vehicles are explicitly recorded among all crashes
- 4,756,671 crash records have geographic coordinates, either from GPS latitude/longitude written in the crash record, or geocoded from street names or addresses
- 78,497 are determined to involve collisions with or avoidances of pedestrians
- 72,243 total pedestrians are explicitly recorded among all crash records
- 5674 pedestrian fatalities are reported

Road-specific attributes were obtained from the TxDOT Roadway Inventory database (Texas Department of Transportation, 2018). The horizontal curves (GEO-HINI) database was spatially matched with the road inventory database to map road geometry. Census-tract-level population and job data were obtained from the 2010 population census and the Longitudinal Employer-Household Dynamics program, respectively. Road segments were matched with the closest census tract centroid using the ArcGIS spatial join routine. All data were normalized by the area of census tracts. Other data sources include annual rainfall data (1981–2010) from the Texas Water Board (the most current available), school locations from the Texas Education Agency, hospital locations from the Homeland Infrastructure Foundation-Level Data catalog, and transit stop locations from OpenStreetMap. Numbers of transit stops and Euclidean distances from each road segment to the nearest schools and hospitals were calculated using ArcGIS Spatial Analysis tools. Table 3.1 provides summary statistics of roadway characteristics from the TxDOT segment dataset.

Table 3.1 Summary Statistics of Variables for Road Segments across Texas

	Mean	Std. Dev	Min	Median	Max
Number of pedestrian crashes	0.080	0.653	0	0	115
Number of fatal pedestrian crashes	0.007	0.102	0	0	10
Segment length (in miles)	0.43	0.81	0.00	0.19	44.24
Number of lanes	2.234	0.784	1	2	14
Median width (in feet)	1.741	11.789	0	0	519
Average shoulder width (in feet)	1.407	3.621	0	0	42
On-system road	0.225	0.417	0	0	1
Indicator of curvature	0.110	0.313	0	0	1
Curve length (in meter)	21.676	125.770	0	0	9630.572
Curve angle (degrees)	3.538	12.954	0	0	331.8
Average daily traffic (ADT) per lane	888.4	2366.0	0.0	165	92090
Percentage of truck ADT	5.960	7.217	0	3.2	95.8
Daily VMT (DVMT)	1035.4	7319.4	0.0	54.4	793941.6
Speed limit (mph)	20.998	28.687	0	0	85
Rural (pop. < 5000)	0.407	0.491	0	0	1
Small urban (pop: 5000–49999)	0.098	0.297	0	0	1
Urbanized (pop: 50000–199999)	0.092	0.288	0	0	1
Large urbanized (pop: 200000+)	0.404	0.491	0	0	1
Population density (per sq. mile)	1671.5	2274.9	0.0	635.8	55239.7
Job density (per sq. mile)	805.0	3285.3	0.0	139.6	130011.1
Average yearly precipitation (1981–2010) (inches)	36.481	11.516	8	37	61
Distance to nearest hospital (miles)	6.822	7.276	0.0018	3.968	98.208
Distance to nearest school (miles)	2.084	3.086	0.01	0.741	53.952
Presence of transit stop within 100-meter buffer	0.006	0.075	0	0	1
Number of transit stops within 100-meter buffer	0.011	0.200	0	0	27

3.1.2. OLS Regression Data Description

The ordinary least-squares regression of CRIS data for 2010–2018 considers a wide variety of demographic, climatological, and roadway factors across the 254 Texas counties to examine their associations with pedestrian crashes and pedestrian fatalities. This model is developed from 78,497 pedestrian crash records in the CRIS system, with county-level covariates on climate, demographics, and geography pulled from a variety of databases, including the US Census Bureau, the 30-year climate normal for Texas from the PRISM Climate Group, and the Texas Association of Counties. Additionally, covariates pulled directly from the CRIS data itself, such as Truck Daily Vehicle Miles Traveled (DVMT) and Annual Vehicle Miles Traveled (AVMT), provide overall DVMT/AVMT with per capita rates for both.

To obtain figures for the homeless population, the study team used point-in-time (PIT) counts (a PIT count is the number of sheltered and unsheltered people experiencing homelessness on a single night). Homeless PIT counts were obtained from Department of Housing and Urban Development (HUD) databases for the areas in which information was available—roughly 100 of the 254 Texas counties. These counts were then divided across the survey area, which often spanned multiple

counties, and then weighted by population, as a county-by-county breakdown was not available for most areas outside of the core urban counties. Climate data were obtained from the US Geological Survey, and the remainder from TxDOT databases updated in 2020. With nearly 30 initial covariates, highly insignificant variables were removed to find the variables of statistical and/or practical significance, defined in this model as variables that had a final p-value below .20; if these variables had high p-values (>.20) in the second run of the model, they were removed from the third and final model even if they had p-values below .20 in the initial model. WMT data were gathered from the 2017 National Household Travel Survey (NHTS) and aggregated up to the county level for the purposes of this model. Table 3.2 provides the summary statistics for all variables used in the ordinary least-squares regression and analysis.

Table 3.2 Summary Statistics for 254 Texas Counties, OLS Regression

Covariate	Mean	Std. Dev	Min	Max	Median
Crashes per 1 million VMT	0.130	0.312	0	4.581	0.0721
Fatalities per 1 million VMT	0.013	0.016	0	0.194	0.0145
Crashes per WMT	0.014	0.02	0	0.203	0.0169
Fatalities per WMT	0.002	0.004	0	0.058	.00134
WMT per Capita (2017)	0.122	0.011	0.11	0.189	0.122
Overall WMT	14,627	58,162	9.85	688,117	2235
Total Crashes	309	1453	0	16,904	19.5
Fatal Crashes	22	90	0	1063	4
Total Daily VMT (DVMT)	3,042,147	9,838,223	51,339	116,251,701	856,479
Centerline Miles	2682	3313	155	35,928	1995
Centerline Miles per Capita	0.185	0.243	0.006	2.182	0.1
Job Density (per sq. mi, 2017)	46.69	175.08	0.03	1879.94	6.0323
Pop Density (per sq. mi, 2017)	124	384	0.22	3086	21.563
Homeless Per 1,000 people	0.357	0.792	0	7.411	0
VMT-weighted Average Speed Limit	59.98	8.21	37.47	77.66	61.15
VMT-weighted Average Lane Count	3.01	0.66	2	5.40	3.07
DVMT per Capita	76	207	8	3008	39
Truck DVMT Per Capita	17	41	1	495	6.931
% Age 17 and Under	24.219	3.822	8.51	35.99	23.963
% Age 65 and Older	17.822	5.234	8.61	35.61	17.215
Median Age (2017)	39	6	27	58	38.2
Growth Rate (2010–2020)	4.376	10.817	-18.6	80.952	2.118
Median Household Income (2017)	51,302	12,196	30,076	102,858	48,542
% of Population in Poverty (2017)	13.76	4.11	13.76	24.60	15.752
Annual Precipitation (in.)	31	12	10	60	29.578
Mean Maximum Temp (°F)	77	3	70	86	77.237
Mean Minimum Temp (°F)	53	5	40	65	52.942

3.2. Key Crash Factors

Minimal cleaning of the data (e.g., standardizing location reporting) was required to perform robust analysis, including generating summary statistics. Additionally, it should be noted that around 48.5% of pedestrian crashes across the United States go unreported, either due to the police not being involved, a failure to disclose hospital or insurance records, or some combination of these factors; this figure has also been anecdotally quoted in Texas as around 50% (Davis, 2015; Oborski, 2020). While many of these unreported crashes ostensibly do not result in injuries, they may still serve to mask potential hotspots where there are more frequent but less severe collisions, such as in residential neighborhoods or parking lots (Reyna, 2020).

3.2.1. Pedestrian Crash and Fatality Trends

In the period 2010–2019, there were 5.6 million reported crashes on Texas roads; of these, 1.4% were pedestrian crashes. In total, there were 35,306 fatalities in the same period, with 5,674 or 16% pedestrian crash fatalities. Pedestrians are therefore disproportionately likely to be killed compared to other road users, excluding cyclists. Furthermore, the per capita rate of pedestrian crash fatalities (per 100,000) has increased in the state from 1.49 in 2010 to 2.41 in 2019, and their percentage of total traffic fatalities has also increased from 12.08% in 2010 to 18.99% in 2019.

The five largest cities in Texas—Houston, Dallas, San Antonio, Austin, and Fort Worth—accounted for 36% of all pedestrian fatalities in Texas within their city limits, while comprising approximately 24.3% of the population. Of Texas cities, Austin led the way in pedestrian fatalities as a proportion of total traffic deaths, with around 33% of traffic fatalities accounted for by pedestrians.

3.2.2. Time of Day

The CRIS data reflect time of day as an important indicator of crash frequency and severity. Perhaps most notably, a roughly inverse relationship is observed between the pedestrian crash frequency and severity. There is some overlap between an elevated risk of fatality and higher numbers of crashes in the hours of 6–10 PM, with the highest frequency of crashes happening in the 6–7 PM hour, and the highest fatality count in the 8–10 PM hours. An overview of the data regarding crash frequency and severity across Texas is featured in Figure 3.1. These patterns in Texas reflect the national trend identified in the literature, which shows an increase in fatalities and crashes at night (NHTSA, 2019; Welch, 2016), although CRIS data is inconsistent when it comes to indicating whether street lighting was present or not. Overall, there are significantly heightened pedestrian fatalities in the nighttime hours over the daytime hours.

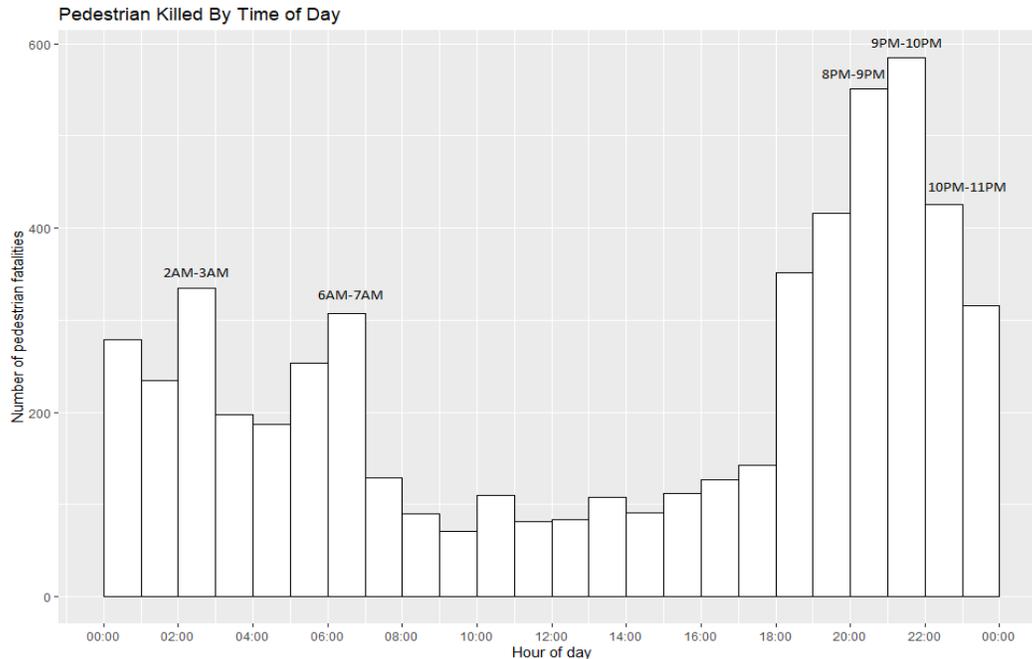


Figure 3.1 Distribution of Pedestrian Fatalities in Texas by Time of Day, 2010–2018

3.2.3. Speed

Speed has more of an impact on crash severity and is less predictive of crash frequency, possibly due to higher posted speed limits on limited access roads on which pedestrian activity is much lower (Tefft, 2013). Generally, the proportion of uninjured pedestrians remains similar across all speed categories, but non-incapacitating injury crashes decline as speed increases, as do crashes where an injury was possible but not confirmed at the time the police report was created.

Deaths increased from near zero on roads with speed limits below 30 MPH to 5% in the 30-to-45-MPH range before climbing significantly to 35% at crashes on roads with speed limits above 60 MPH. The latter category includes, but is not limited to, most limited-access freeways and tollways in Texas, while the under-30-MPH category includes most residential streets and most central business district streets. While this complements the idea that speed is analogous with an increase in fatal crash percentages as outlined in Tefft (2013), these CRIS data are referring to the roadway’s posted speed limit rather than impact speed. Generally, the proportion of uninjured pedestrians remains similar across all speed categories, but non-incapacitating injury crashes decline as speed increases, as do crashes where an injury was possible but not confirmed at the time the police report was created. Nonetheless, the results here support the conclusion put forth in both dc.gov (2018) and Tefft (2013): impact speed increases the likelihood of a pedestrian fatality. Figure 3.2 shows a comprehensive breakdown of the pedestrian injury severity across roadways of given speed limits across Texas.

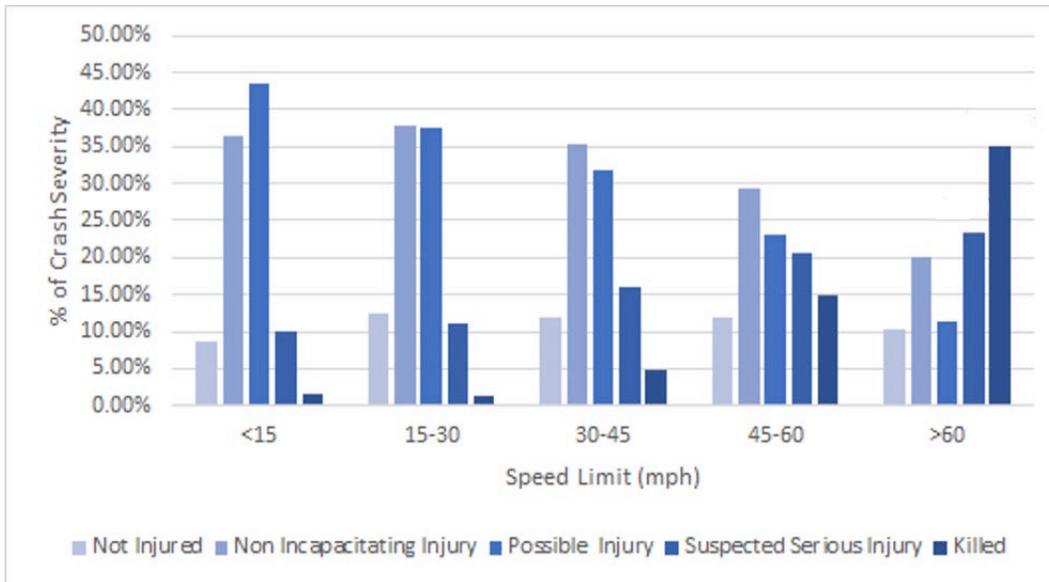


Figure 3.2 Distribution of Injury Severity and Fatalities in Texas by Roadway Speed Limit, 2010–2018

3.3. Heat Maps and Lat-Long Analysis

Heat maps can help determine the spatial patterns of crashes. These patterns reveal the areas that have higher rates of pedestrian crashes and deaths, including intersections, corridors, and even stretches of limited-access highway where individuals routinely attempt to cross. Heat maps from the metropolitan areas of Texas for 2010–2019 total pedestrian crash data, including a few inset maps of downtown areas in Dallas, Houston, San Antonio, and Austin, are provided below in Figures 3.3-3.10. The heat maps section is followed by a short report on missing latitude-longitude data, including the potential reasons why these data were excluded and the temporal distribution of the data.

These heatmaps were created using the online TxDOT CRIS query builder, then displaying the data as a heatmap rather than as discrete crashes. For all maps, pedestrian-involved crashes in the period 2010–2019, resulting in more than \$1,000 in damage and an injury were used as parameters in the query. The scale of all maps, in terms of the concentration of crashes relative to the color on the heatmap can be assumed to be the same, with a red-zone representing multiple injurious crashes over the 10-year period.

The hotspots can be separated into three categories. Those that are located along major commercial corridors, with many driveways and few marked pedestrian crossings, those that are located in high pedestrian and vehicle traffic central business district areas, and those that are located near to major freeways, especially along frontage roads. Major shopping centers invite more pedestrian traffic, and often these hotspots are located at the nexus of one or more of these factors. Additionally, homelessness could play a significant but unproven role in these crashes, as homeless encampments in Texas cities tend to be located in freeway right-of-way (Reyna, 2020; Lee, 2020).

Some of these hotspots, such as in Austin along Ben White Blvd (US-290) and Research Blvd (US-183), or along Webb Chapel Road in Dallas, also coincide with the location of major transit stations or transfer points, with heavier pedestrian traffic along arterial roads. Each of the inset maps below provides a caption describing the locations of several of the top hotspots for each map.

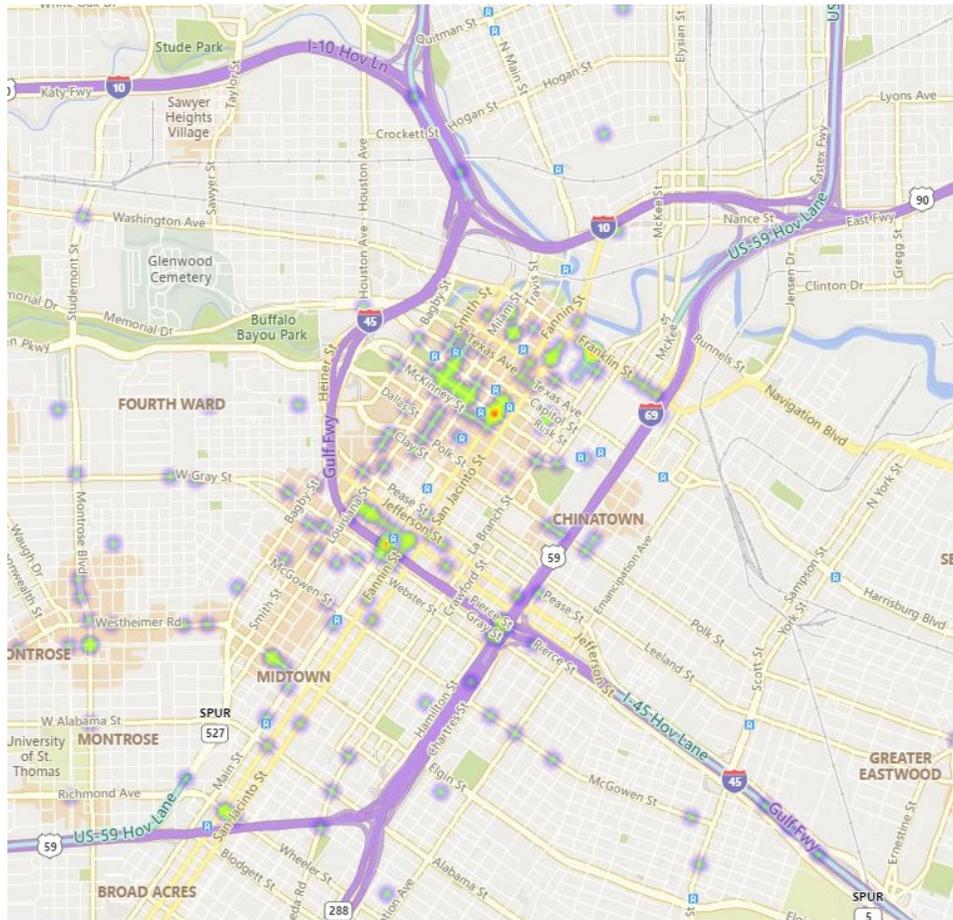


Figure 3.3 Heatmap of Pedestrian Crashes in Downtown Houston

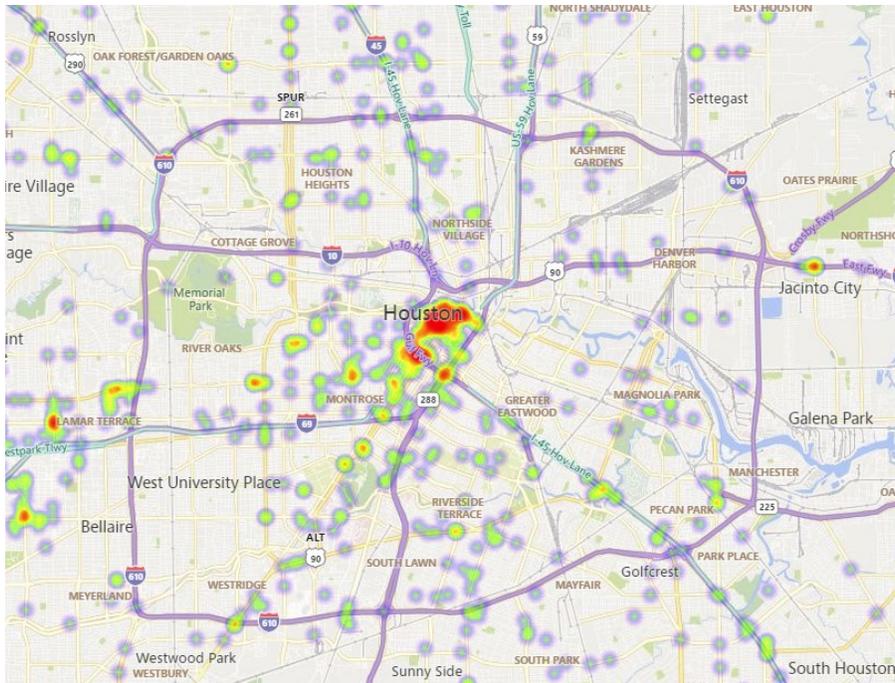


Figure 3.4 Heatmap of Pedestrian Crashes in Central Houston

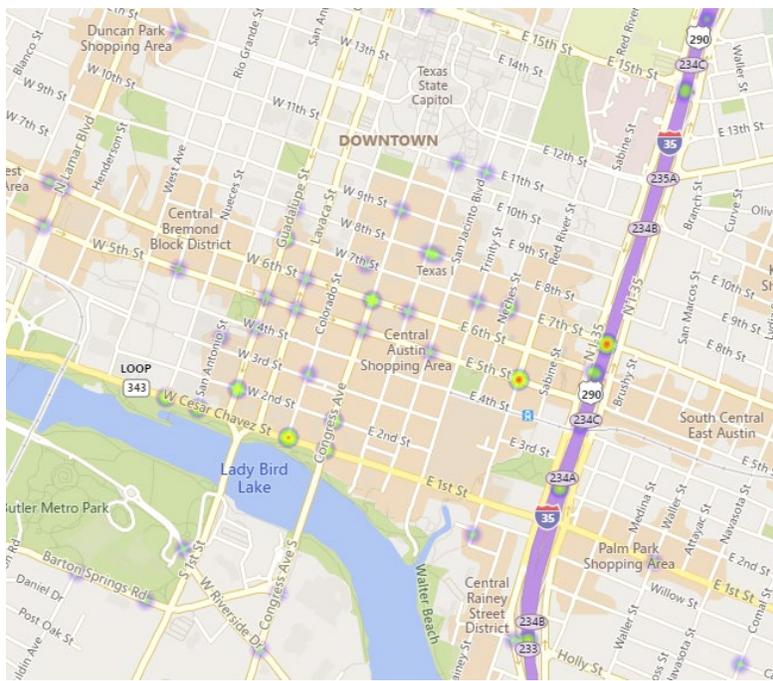


Figure 3.5 Heatmap of Pedestrian Crashes in Downtown Austin

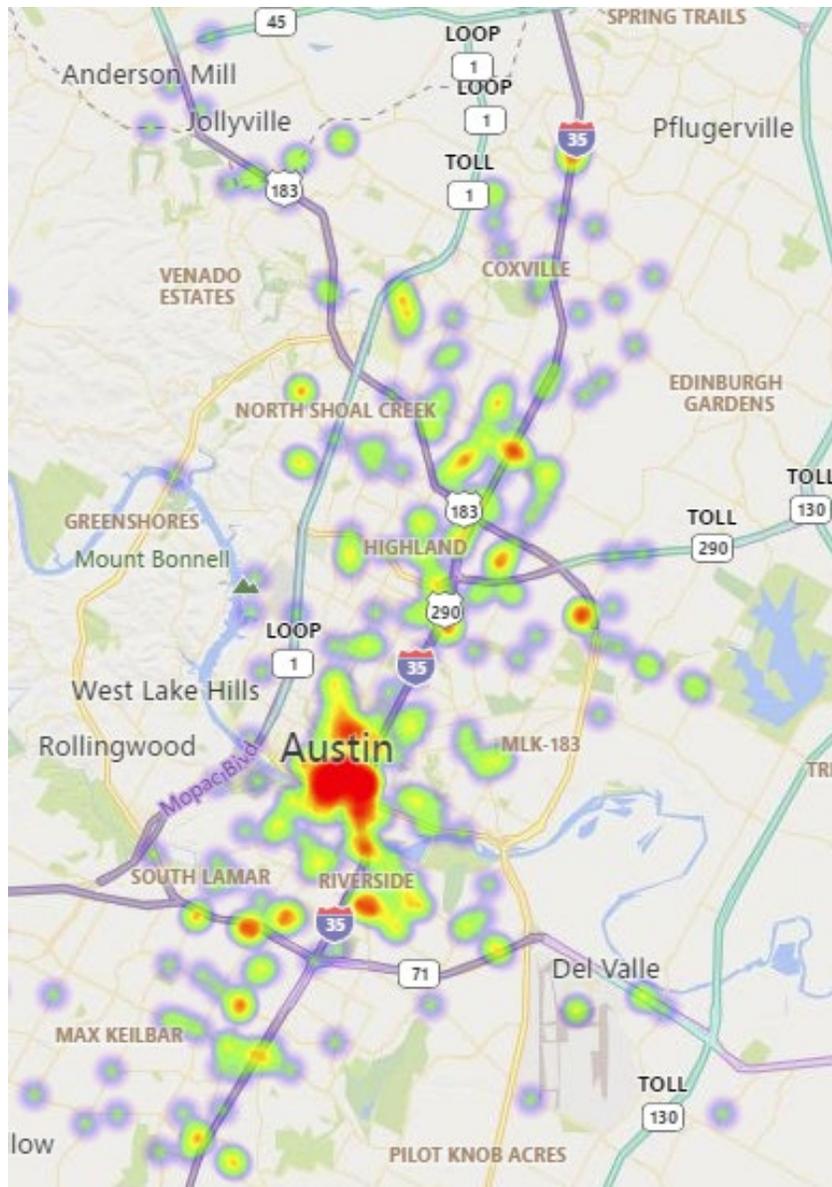


Figure 3.6 Heatmap of Pedestrian Crashes in Greater Austin

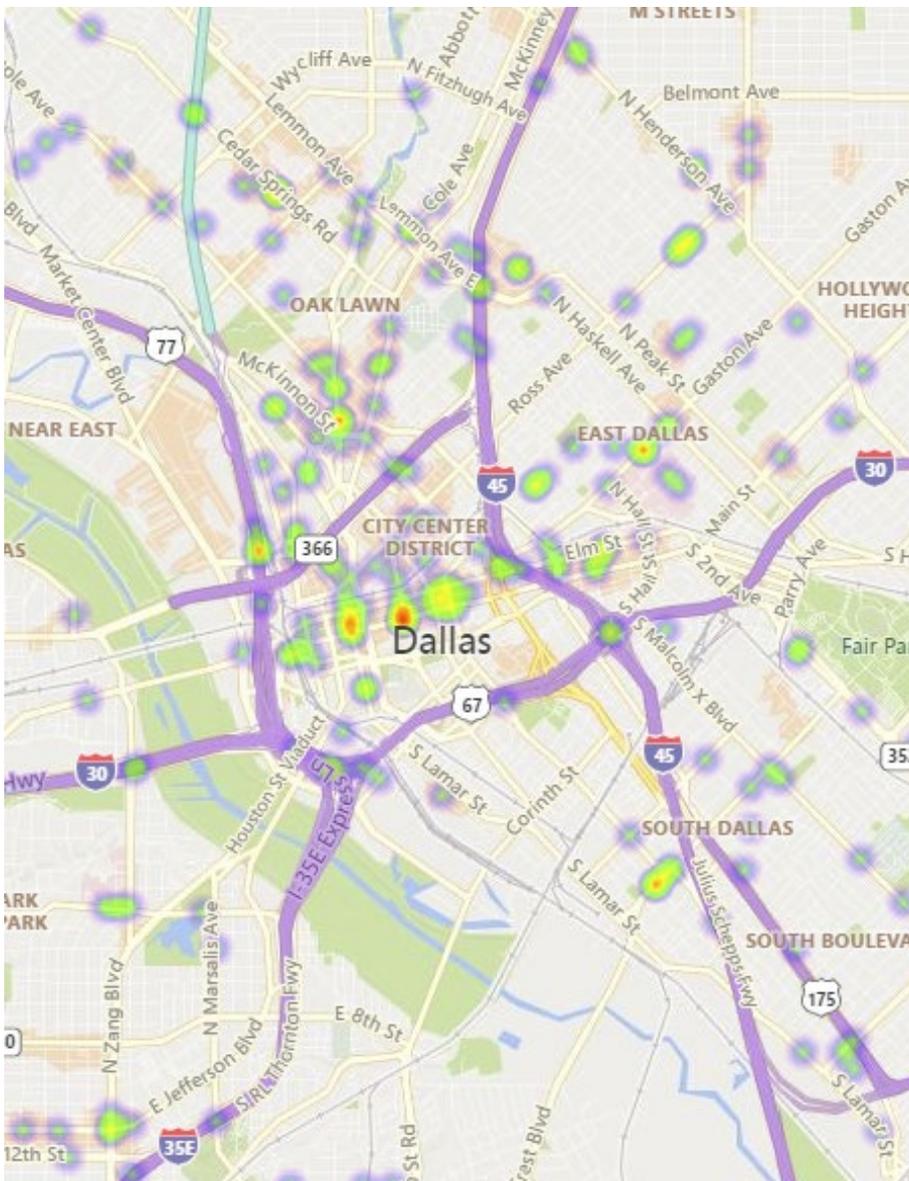


Figure 3.7 Heatmap of Pedestrian Crashes in Downtown Dallas

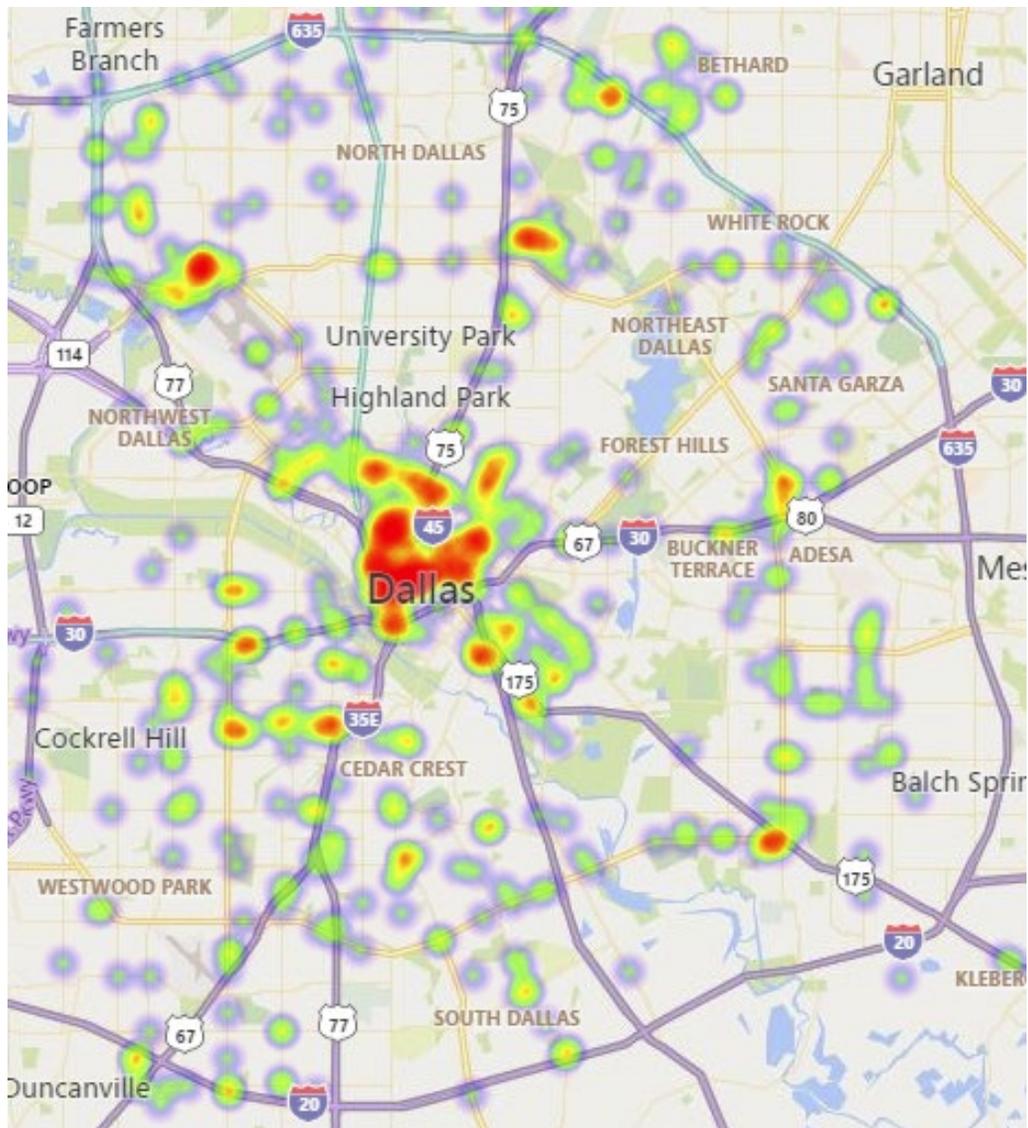


Figure 3.8 Heatmap of Pedestrian Crashes in Central Dallas

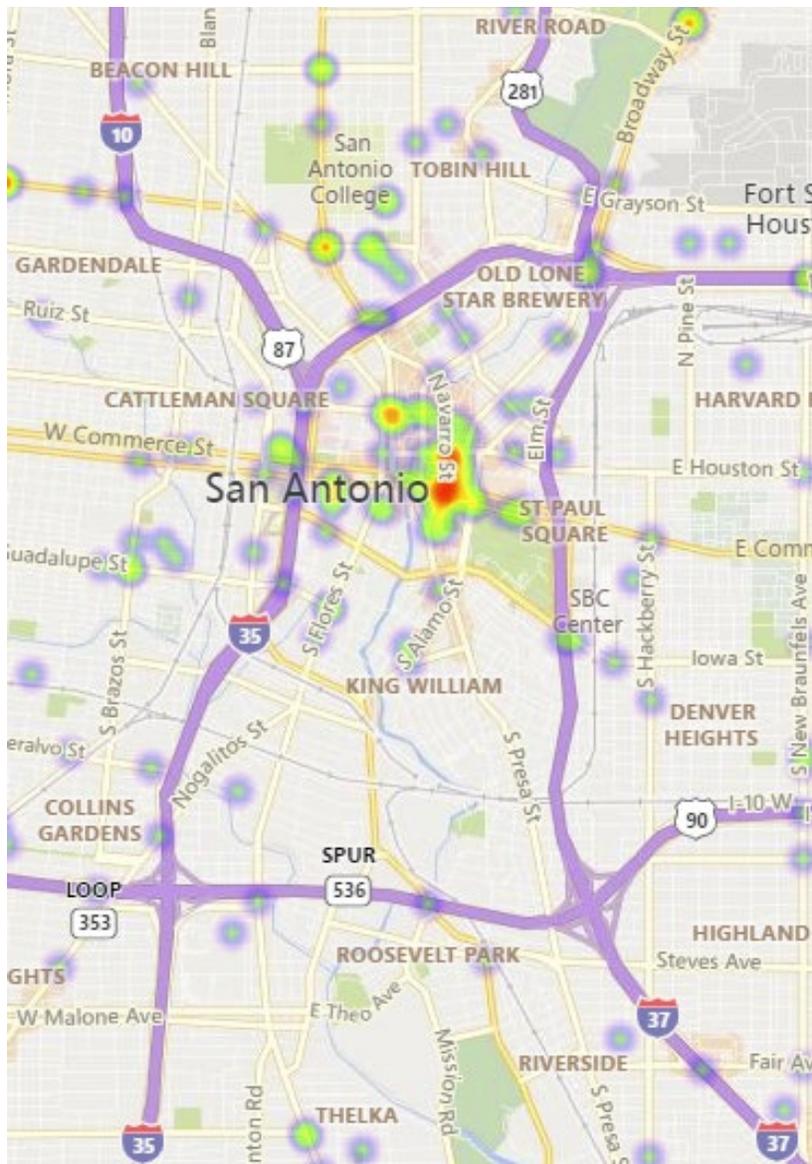


Figure 3.9 Heatmap of Pedestrian Crashes in Downtown San Antonio

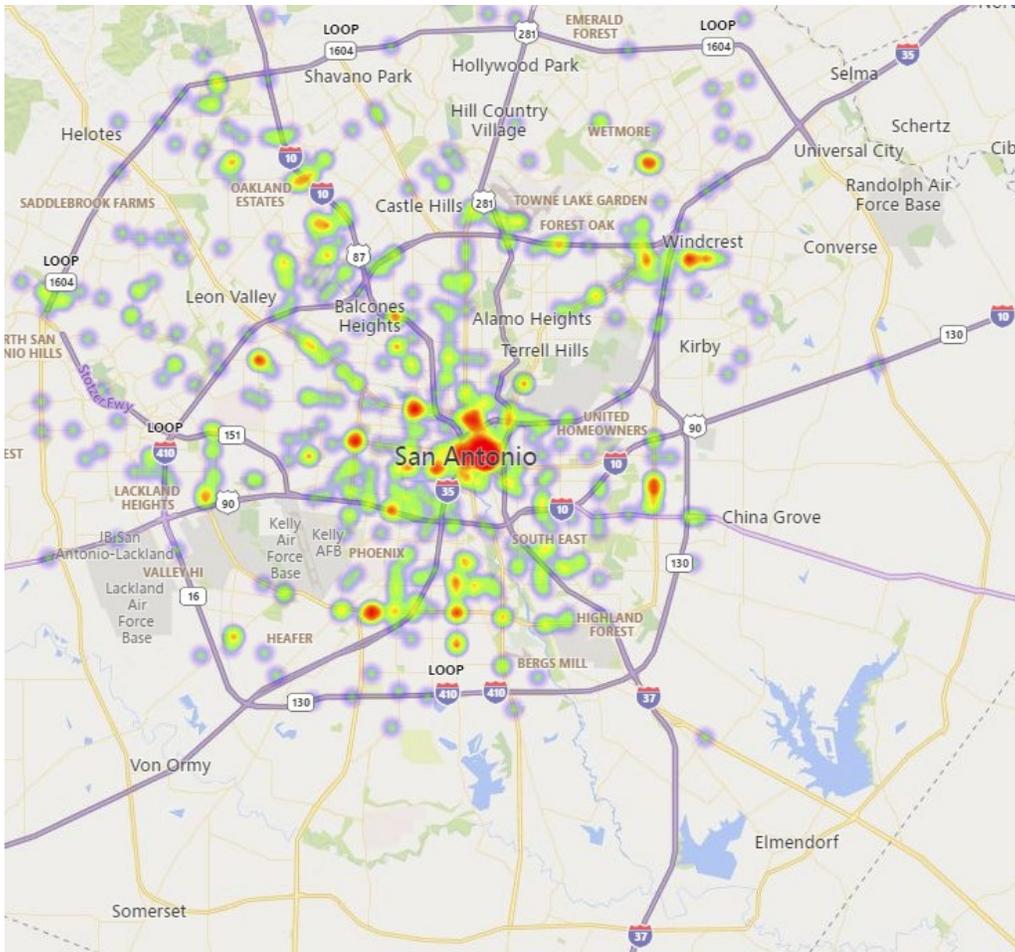


Figure 3.10 Heatmap of Pedestrian Crashes in Greater San Antonio

3.3.1. Investigating the Factors for Missing Geographic Location of Pedestrian Crash Events

About 29% of the records among the 78,497 crash events of CRIS dataset that involved one or more pedestrians do not have geographical location recorded in the database. To investigate this issue a binary logit model has been developed. In this model each crash event has been coded as ‘1’ if it has geographic location recorded in the database and ‘0’ otherwise.

Table 3.3 is showing the estimation results of the final model after trying several model specifications. From the outputs we can conclude that the probability of having geographic location decreases on an average by 26.24% if the crash event occurred outside city limit (Table 3.3). Figure 3.11(C) supports this claim by showing that very high proportion of such events took place outside city limits. Monetary damage of the crash events does not make it more likely to have geographic coordinates in the database. From the model results, the crash events for which the property damage were \$1,000 or more, are on an average 2.7% less likely to have geographic coordinates than a crash event which caused property damage of less than \$1,000.

Table 3.3 Summary statistics of variables used in binary logit model to investigate contributing factors in records lacking geographic location (lat/long) details

Lat-Long Status (1 = Yes/available, 0 = No lat/long info)					
Predictors	Estimate	Odds Ratios	Marginal Effects	z value	p
(Intercept)	0.8921	2.44		5.53	0.000
Outside City Limit [no]	Reference				
Outside City Limit [yes]	-1.898	0.15	-0.2624	-51.17	0.000
\$ Thousand Damage [no]	Reference				
\$ Thousand Damage [yes]	-0.3795	0.68	-0.0279	-10.49	0.000
Private Drive/Property [no]	Reference				
Private Drive/Property [yes]	-8.0325	0.00	-0.7576	-11.34	0.000
TxDOT Reportable Flag [no]	Reference				
TxDOT Reportable Flag [yes]	5.50	244.31	0.7226	66.78	0.000
Crash Severity [killed]	Reference				
Crash Severity [not injured]	0.806	2.24	0.0026	5.11	0.000
Crash Severity [non-incapacit. injury]	-2.807	0.06	-0.0648	-20.70	0.000
Crash Severity [possible injury]	2.9727	0.05	-0.0760	-21.76	0.000
Crash Severity [suspected serious injury]	-3.2401	0.04	-0.0975	-20.43	0.000
Crash Severity [unknown]	0.3797	1.46	0.0014	1.30	0.194
Suspected Serious Injury	0.3796	1.46	0.0272	4.62	0.000
Year [2019]	Reference				
Year [2018]	-0.2016	0.82	-0.0053	-2.52	0.012
Year [2017]	-0.2948	0.74	-0.0081	-3.81	0.000
Year [2016]	-0.0218	0.98	-0.00053	-0.27	0.788
Year [2015]	0.1703	1.19	0.0038	2.03	0.043
Year [2014]	-0.6008	0.55	-0.0192	-8.23	0.000
Year [2013]	-1.0603	0.35	-0.0427	-15.28	0.000
Year [2012]	-1.01398	0.36	-0.0398	-14.35	0.000
Year [2011]	-1.0366	0.35	-0.0412	-14.35	0.000
Year [2010]	-1.2415	0.29	-0.0547	-17.56	0.000
Vehicle Body Style [other]	Reference				
Vehicle Body Style [motorcycle]	-0.3792	0.68	-0.0308	-2.09	0.037
Veh. Body Style [passenger car 4 door]	-0.0875	0.92	-0.0063	-2.66	0.008
Vehicle Body Style [police car/truck]	-0.514	0.60	-0.0440	-2.95	0.003
Crash Speed Limit	0.0056	1.01	0.0004	5.13	0.000
Observations	78497				
McFadden R²	0.713				
AIC	29188.751				

The fact that crash events occurred in a private drive, road or property decreases the probability of the events to have their geographic location recorded in the database on an average by 75.7% (Figure 3.11(B)). The crash events which are categorized as reportable to TxDOT based on certain criteria are highly likely to have their geographic coordinates recorded in the database. TxDOT Reportable Flag indicates whether a crash occurred on a trafficway and resulted in injury or death or \$1,000 damage. Such crash events are 244 times more likely to have geographic coordinates recorded in the crash database. Having a TxDOT reportable flag increases the probability of a crash event to have geographic coordinate by on an average 72.26% than a pedestrian crash event that is not flagged as TxDOT reportable. Figure 3.11(D) supports this claim. There is a general tendency that the higher the level of crash severity the more likely it is to have its geographic

coordinates recorded in the database. Crash events which are labeled as ‘non-incapacitating injury’ for the involved people have on an average 6.5% less probability to have geographic coordinates than a crash event in which someone has been killed.

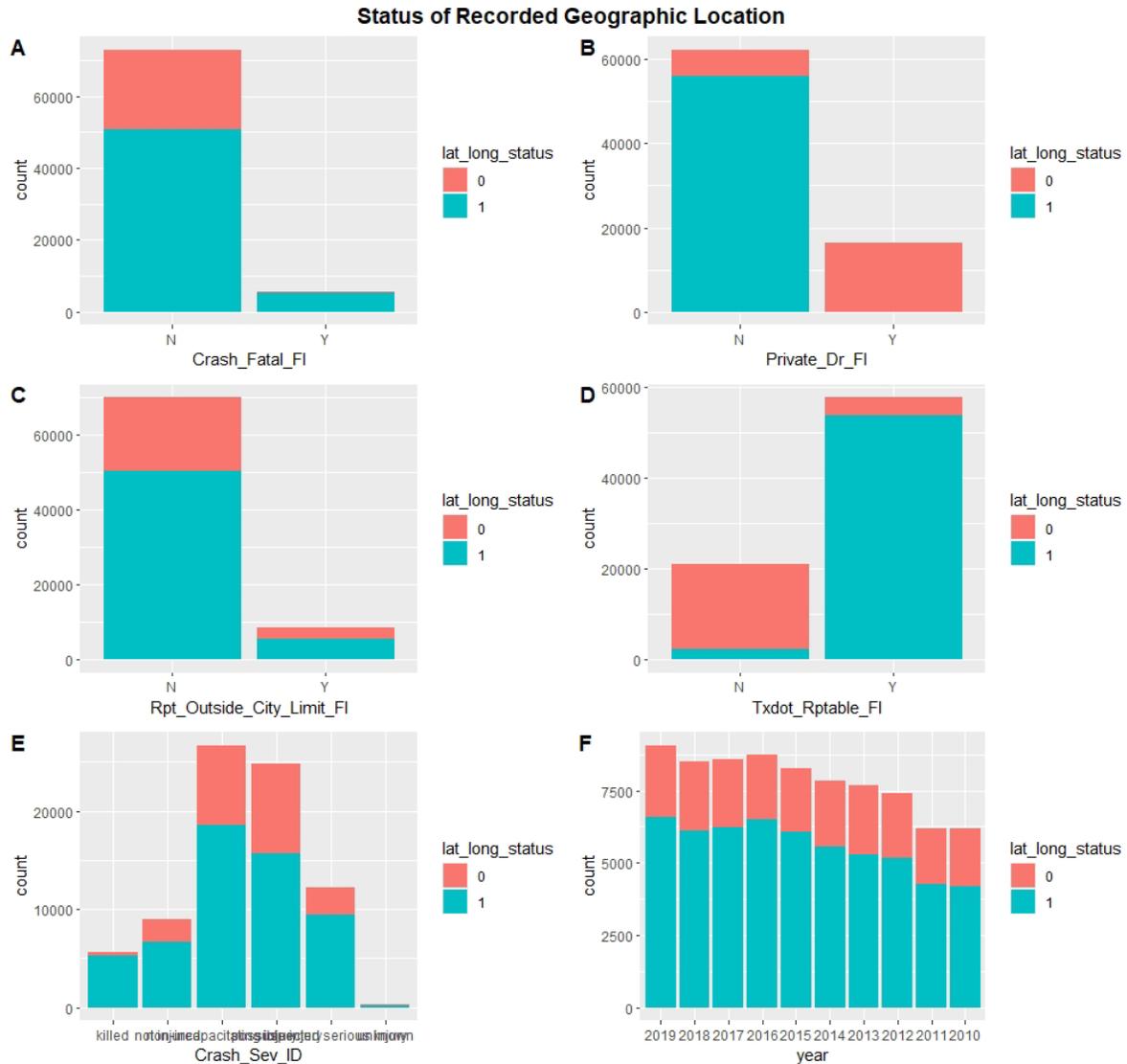


Figure 3.11 Status of Recorded Geographic Location of the Crash Events for Selected Factors

An increase of the number of suspected serious injury in a crash event makes it more likely to have geographic coordinate. Increase of suspected serious injury by 1 person increases the probability of having geographic coordinates by 2.7%. From the model outputs, older crash events are less likely to have geographic coordinate in the database. For example- crash events in year 2010 have on an average 5.5% less probability to have geographic coordinate than crash events in 2019. Figure 3.11(F) shows this trend. If the crash involved motorcycle or police car/truck, then it is less likely to have geographic coordinate. For example- pedestrian crashes involving motorcycle have

on an average 3% less probability to have geographic coordinate recorded. Higher speed limit has very little or no impact on having a crash event geographic location in the database.

In sum, by observing the model outputs it can be concluded that recording of geographic coordinate of a crash event mainly depends on whether it gets the TxDOT reportable flag or not and whether the crash event occurred in a private drive/property or not. Both odds ratios and the marginal effects of the selected contributing factors on geographic location being recorded shows similar results (Table 3.3). ‘TxDOT Reportable Flag’ and ‘Personal Drive/Property’ have the highest marginal effects on the probability of a crash event having geographic location recorded. Figure 3.11 supports the findings from estimated binary logit model. 100% of the pedestrian crash events occurred in a private drive, road or property and about 89.7% of the pedestrian crash event that have not been flagged as TxDOT reportable do not have geographic coordinate recorded in the database.

3.4. Methodologies for Prediction of Crash Counts and Injury Severity

This section discusses the methods used for crash counts and injury severity models. Negative binomial models are used for crash counts, while ordered probit and heteroskedastic ordered probit are used for injury severity.

3.4.1. Analysis of Pedestrian Crash Counts – NB Models

The CRIS data that contained GPS coordinates were spatially matched with the road segments along with land use, population, job, rainfall, and other location features (schools, hospitals, and transit stops) to examine the association between pedestrian crash counts and various contributing factors along Texas roads. A total of 708,738 road segments were included in the analysis. Table 3.1 lists the summary statistics of the roadway segments.

An NB model was used to predict pedestrian crash counts along roadway segments. The expected number of counts $E(Y_i)$ along the i th segment is expressed as follows:

$$E(Y_i) = VMT_i^\alpha \exp(\beta_0 + \sum_K x_{ik}\beta_k + \varepsilon_i) \quad (1)$$

VMT denotes vehicle miles traveled along the i th segment; parameter α shows potential non-linear relationship between crash count and VMT. β_k is the k th covariate, ε_i is the random error that follows gamma distribution $\varepsilon_i \sim \text{gamma}(\gamma, \gamma)$. Y_i represents crash counts with mean $E(Y_i) = \mu_i = VMT_i^\alpha \exp(\beta_0 + \sum_K x_{ik}\beta_k + \varepsilon_i)$ and variance $\text{Var}(Y_i) = \mu_i + \rho\mu_i^2$. Here, ρ is the dispersion parameter, which collapses to a Poisson model when $\rho = 0$.

3.4.2. Analysis of Pedestrian Injury Severity – OP and HOP Models

Injury severity was analyzed at the individual crash level. Both standard OP and HOP models were used to account for the ordinal nature of injury severity. The model specification follows a latent variable framework:

$$y_i^* = \beta X_i + \varepsilon_i \quad (2)$$

y_i^* is the underlying continuous latent variable representing injury severity of the i th pedestrian. X_i is the vector ($k \times 1$) of explanatory variables; β is the vector ($k \times 1$) of unknown parameters to be estimated that is associated with explanatory variables; ε_i is the random error term that is unobserved. In probit, ε_i is assumed to be normally distributed with mean zero and unit variance.

In any given pedestrian crash, we only observe the injury severity y_i as reported by police in crash records. The relationship between the observed discrete variable y_i and the latent variable y_i^* is expressed as follows:

$$y_i = \begin{cases} 0, & \text{if } y_i^* \leq 0 \text{ (Not injured)} \\ 1, & \text{if } 0 < y_i^* \leq \mu_1 \text{ (Possible injury)} \\ 2, & \text{if } \mu_1 < y_i^* \leq \mu_2 \text{ (Non-Incapacitating Injury)} \\ 3, & \text{if } \mu_2 < y_i^* \leq \mu_3 \text{ (Suspected serious injury)} \\ 4, & \text{if } \mu_3 < y_i^* \leq \infty \text{ (Killed)} \end{cases}$$

$\mu_0 = 0$ and μ_j ($j = 1, 2, 3$) are threshold parameters (to be estimated) that align with one of five observed values of injury severity, y_i . In general, the probability of y_i taking on injury severity j on i th pedestrian can be expressed as follows:

$$\Pr(y_i = j | X_i) = \Phi\left(\frac{\mu_j - \beta X_i}{\sigma_i}\right) - \Phi\left(\frac{\mu_{j-1} - \beta X_i}{\sigma_i}\right) \quad (3)$$

Φ is the standard normal cumulative distribution function, and σ_i is variance of the error term. In standard OP models, it is assumed that the variance of error term is constant across all observations. However, the error term can vary across observations: for instance, there can be unobserved heterogeneity in terms of vehicle attributes such as vehicle type, weight, and footprint (Wang and Kockelman, 2005; Chen and Kockelman, 2012; Lemp, Kockelman and Unnikrishnan, 2011) and in terms of pedestrian characteristics (health, weight, and initial response to crashes) (Kim et al., 2010). Failure to account for heteroskedasticity can lead to biased parameter estimates in probit analysis. To overcome this limitation, a HOP was used where variance of the error term is allowed to vary. We follow a flexible specification for the HOP model where σ_i is determined as a function of observed attributes associated with variance as the following equation (Wang and Kockelman, 2005):

$$\sigma_i = \exp(Z_i \gamma) \quad (4)$$

γ is the coefficient for variable Z_i . If γ is not significantly different from zero for all Z_i , then it implies no heteroskedasticity and HOP takes the form of OP. On the other hand, if γ is significantly different from zero, it shows the presence of heteroskedasticity for that particular variable.

The parameters in Equation 3 were estimated by maximizing the log-likelihood function; a sample consisting of n observations would appear as shown in Equation 5:

$$L(\beta, \mu, \gamma) = \sum_{i=1}^n \sum_{j=0}^J I(y_i = j) \ln \left(\Phi \left(\frac{\mu_j - \beta X_i}{\exp(Z_i \gamma)} \right) - \Phi \left(\frac{\mu_{j-1} - \beta X_i}{\exp(Z_i \gamma)} \right) \right) \quad (5)$$

3.4.3. Ordinary Least-Squares Regression

The WMT models utilize a job density variable that controls for the strong collinearity that would be expected for jobs and population. This new “job density residuals” variable takes the residual of the actual versus expected jobs to account for varying job concentrations. For the VMT models, crashes and fatalities per 1 million VMT was chosen as the operational variable to clarify the data, given the high VMT numbers. WMT per capita, similarly, is the operational variable used to partially correct for a modifiable areal unit problem—a challenge that arises when using divisions as large as counties.

3.4.4. Limitations

With the 254 Texas counties as datapoints, there are some limitations to using an ordinary least-squares model, in addition to the geographic issues associated with using county-level data. Given that only county-level, aggregated counts were used, data with a finer resolution was aggregated to the county level, primarily through ArcMap. Recent PIT homelessness count data was recorded for around 100 counties, including all metropolitan statistical areas (MSAs) in Texas. Outside of these areas, it can be assumed that homelessness is at trivially low levels compared to counties within MSAs. Although HUD regulations theoretically require a count in these areas each year, specific methodology is nonbinding (Texas Homeless Network, 2020). Finally, given that around 40 to 50% of pedestrian crashes go unreported in Texas, the CRIS data should be regarded as a dataset that favors severe crashes, and those that occur on public roads (Reyna, 2020; Yang & Diez-Roux, 2012). Crashes that take place on private roads (such as a private parking lot) are often not counted, and crashes that are not reported to the police for any reason are not counted, as CRIS relies primarily on police reports.

3.5. Results from Crash Count and Injury Severity Models

The following tables show the results from 1) NB models for segment wise pedestrian crash counts, 2) OP and HOP models for injury severity and 3) OLS models for county level pedestrian crash counts. Discussion of these results can be found in Section 3.6.

Table 3.4 Estimation Results of NB for All Pedestrian Crashes and Fatal Pedestrian Crashes

	All Ped Crashes			Fatal Ped Crashes			% of Change	
	Coeff	Std. Error	Pr> z	Coeff	Std. Error	Pr> z	All ped crashes	Fatal ped crashes
Ln (DVMT)	0.7390	0.0039	0.000	0.8730	0.0115	0.000		
Highway Design Variables								
Number of lanes	0.0316	0.0060	0.000	0.0459	0.0121	0.000	2.50%	3.60%
Median width	-0.0052	0.0005	0.000	-0.0033	0.0007	0.000	-5.93%	-3.86%
Shoulder width	-0.0187	0.0020	0.000	-0.0164	0.0036	0.000	-6.55%	-5.76%
On-system roads	0.3564	0.0273	0.000	0.8678	0.0617	0.000	42.81%	136.53%
Indicator of curvature	0.0064	0.0281	0.820	-0.0576	0.0524	0.272	0.64%	-3.65%
Curve angle	-0.0047	0.0008	0.000	-0.0028	0.0014	0.044	-5.95%	-2.88%
Speed limit	-0.0093	0.0004	0.000	-0.0024	0.0012	0.037	-23.46%	-6.43%
Traffic Attributes								
ADT per lane	-5.5E-05	2.25E-06	0.000	-3E-05	3.84E-06	0.000	-12.26%	-6.95%
% of truck AADT	0.0054	0.0012	0.000	0.0056	0.0024	0.020	3.95%	4.14%
Land Use Variables								
Population density	0.0001	0.0000	0.000	0.0001	4.89E-06	0.000	35.78%	17.46%
Job density	3.19E-05	7.35E-07	0.000	0.0000	2.07E-06	0.001	11.06%	2.35%
Rural (pop<5000)	-0.6061	0.0321	0.000	-0.6200	0.0746	0.000	-45.45%	-46.20%
Small urban (pop: 5000–49999)	-0.1213	0.0278	0.000	-0.1917	0.0774	0.000	-11.42%	-17.44%
Large urbanized (Pop: 200000+)	0.2074	0.0199	0.000	0.1366	0.0545	0.000	23.05%	14.63%
Ref: Urbanized (pop: 50000–199999)								
Climate and Proximity Factors								
Rainfall	-0.0041	0.0005	0.000	0.0024	0.0014	0.000	-4.63%	2.80%
Distance to the nearest school	-0.2730	0.0083	0.000	-0.0958	0.0137	0.604	-52.45%	-22.92%
Distance to the nearest hospital	-0.0227	0.0021	0.000	0.0022	0.0043	0.000	-15.24%	1.70%
Transit stop indicator	0.6706	0.0484	0.014	0.4290	0.1116	0.339	95.54%	53.46%
Number of transit stops	0.0372	0.0151	0.000	0.0269	0.0281	0.000	0.75%	0.53%
(Intercept)	-7.3860	0.0448	0.000	-11.7900	0.1237	0.000		
No. of observations	708738							
Dispersion Parameter: ρ	2.01			1.39				
McFadden's R2:	0.278			0.335				
LR chi2	89206			17945				
Prob > chi2	0.0000			0.0000				
2x log-likelihood	-231909.99			-35603.96				

Continuous variables show the percentage change for a one-SD increase. Binary variables show the percentage change from 0 to 1. Bolded percentages indicate more practically significant variable.

Table 3.5 Injury Severity Results: OP versus HOP Models

	OP		HOP	
	Estimate	P-value	Estimate	P-value
Vehicle Type				
Pickup trucks	0.0945	0.000	0.1559	0.000
SUV	0.1042	0.000	0.1566	0.000
Heavy-Duty Truck	0.0479	0.029	0.1054	0.001
Van	0.0927	0.000	0.1435	0.000
Bus	0.1883	0.000	0.2665	0.001
Motorcycle	-0.1497	0.011	-0.1452	0.124
Others (ambulance, fire truck, police vehicle, etc.)	0.0159	0.404	0.0262	0.270
(Reference vehicle = Passenger Car)				
Model Year				
After 2016	0.0268	0.200	0.0268	0.315
2011–2015	0.0245	0.045	0.0296	0.056
2005–2010	0.0818	0.000	0.1099	0.000
Unknown	0.0492	0.000	0.0579	0.001
(Reference Data = Before 2005)				
Pedestrian Age	0.0071	0.000	0.0083	0.000
Pedestrian Gender (1=Male)	0.1218	0.000	0.1537	0.000
Driver Age				
Driver Age (<24 years)	0.1550	0.000	0.2139	0.000
Driver Age (>65 years)	0.0357	0.013	0.0493	0.006
Driver Gender (1=Male)	0.1477	0.000	0.1861	0.000
Pedestrian/Driver Intoxicated	1.4382	0.000	2.8614	0.000
Speed Limit (MPH)	0.0171	0.000	0.0215	0.000
Hit-and-Run (1=Yes)	0.1353	0.000	0.1381	0.000
Crash Took Place at Intersection (1=Yes)	-0.1146	0.000	-0.1369	0.000
Road Type				
County Road	0.1097	0.000	0.1560	0.000
Farm-to-Market	0.1247	0.000	0.1597	0.000
Interstate	0.1087	0.000	0.1556	0.000
Non-Trafficway	0.1005	0.000	0.1846	0.000
Other Roads	0.4114	0.000	0.5482	0.000
Tollway/Toll Bridge	-0.4073	0.000	-0.3737	0.011
US State	0.1460	0.000	0.1867	0.000
(Reference type = City Streets)				
Crash Location				
Off Roadway	-0.1564	0.000	-0.0758	0.005
Shoulder	-0.1876	0.000	-0.1338	0.024
Median	-0.4384	0.000	-0.4544	0.000
(Reference location = On Roadway)				
Road Geometry				
Straight Grade	0.1426	0.000	0.2149	0.000
Curved	0.1939	0.000	0.2763	0.000
(Reference = Straight & Level)				
Control Type				
Traffic Sign	0.0224	0.044	0.0423	0.003
Traffic Signal	-0.0786	0.000	-0.0887	0.000
Other (human control, rail gate etc.)	-0.0131	0.556	-0.0034	0.896

	OP		HOP	
	Estimate	P-value	Estimate	P-value
(Reference = No Control)				
Area Population				
<5000	0.2085	0.000	0.2833	0.000
5000–9999	0.1466	0.000	0.1942	0.000
10000–24999	0.1394	0.000	0.2009	0.000
25000–49999	0.1132	0.000	0.1474	0.000
50000–99999	0.1012	0.000	0.1389	0.000
(Reference = 100000+)				
Crash Time				
5AM–7AM	0.3164	0.000	0.3959	0.000
7AM–11AM	0.1837	0.000	0.2190	0.000
4PM–8PM	0.1963	0.000	0.2349	0.000
8PM–11PM	0.2559	0.000	0.3166	0.000
11PM–5AM	0.2863	0.000	0.3799	0.000
(Reference = 11 AM–4PM)				
Lighting Condition				
Dark Lighted	0.1152	0.000	0.1329	0.000
Dark Not Lighted	0.2721	0.000	0.3599	0.000
(Reference = Daylight)				
HOP's Variance Equation				
Intercept			-6.0259	0.000
Pedestrian Age (Years)			0.0008	0.000
Pedestrian Gender (Male)			0.0515	0.000
Crash Speed Limit (MPH)			0.0052	0.000
Pickup Truck Indicator			0.0601	0.000
SUV			0.0337	0.000
Heavy-Duty Truck			0.1458	0.000
Van			0.0277	0.079
Bus			0.1966	0.000
Motorcycle			0.1717	0.000
Other Vehicle Type			-0.0161	0.223
Intersection			-0.0506	0.000
Traffic Sign			-0.0186	0.037
Traffic Signal			-0.0450	0.000
Other Control Type			-0.0258	0.140
Population: <5000 persons			0.0814	0.008
Population: 5000–9999			0.0678	0.004
Population: 10000–24999			0.0535	0.001
Population: 25000–49999			0.0007	0.966
Population: 50000–99999			-0.0666	0.000
Time: 5 AM–7 AM			0.0687	0.000
Time: 7 AM–11 AM			-0.0246	0.024
Time: 4 PM–8 PM			0.0061	0.532
Time: 8 PM–11 PM			0.0217	0.132
Time: 11 PM–5 AM			0.0415	0.006
Dark & Lighted			0.0456	0.000
Dark & Not Lighted			0.0972	0.000
Threshold Parameters				

	OP		HOP	
	Estimate	P-value	Estimate	P-value
μ_0	0	-	0	-
μ_1	1.1813	0.000	1.4569	0.000
μ_2	2.2264	0.000	2.7943	0.000
μ_3	3.1568	0.000	4.1406	0.000
Number of Observations	66,419		66,419	
Model Fit Statistics	OP		HOP	
Log-Likelihood	-88505.78		-87224.93	
Mcfadden's R2:	0.0601		0.0737	
AIC	177111.6		174603.9	
LR Test	X² = 2561.7 (P<0.0001)			

Table 3.6 Marginal Effects of Risk Factors Associated with Pedestrian Injury Severity (HOP Model)

	No Injury	Possible Injury	Non-Incapitating Injury	Suspected Serious Injury	Killed
Car vs. Vehicle Type					
Pickup Truck	-0.0034	-0.0305	-0.0172	0.0277	0.0234
SUV	-0.0084	-0.0299	-0.0080	0.0277	0.0186
Heavy-Duty Truck	0.0190	-0.0256	-0.0470	0.0168	0.0368
Van	-0.0081	-0.0271	-0.0068	0.0251	0.0169
Bus	0.0139	-0.0500	-0.0632	0.0328	0.0665
Motorcycle	0.0541	0.0049	-0.0612	-0.0170	0.0192
Others	-0.0056	-0.0042	0.0065	0.0039	-0.0007
Model Year: 2005/Older Model vs. Newer Model					
After 2016	-0.0044	-0.0085	0.0014	0.0079	0.0035
2011–2015	-0.0049	-0.0094	0.0015	0.0088	0.0039
2005–2010	-0.0170	-0.0348	0.0038	0.0327	0.0154
Unknown	-0.0093	-0.0183	0.0027	0.0172	0.0078
Pedestrian Age (One SD Increase)	-0.0148	-0.0338	0.0000	0.0316	0.0169
Pedestrian Gender (1=Male)	-0.0053	-0.0315	-0.0121	0.0294	0.0196
Driver Age: 25-65 Years vs. Other Age Groups					
Driver Age (<24)	-0.0303	-0.0677	0.0015	0.0635	0.0330
Driver Age (>65)	-0.0079	-0.0156	0.0023	0.0146	0.0066
Driver Gender (1=Male)	-0.0288	-0.0588	0.0063	0.0552	0.0261
Pedestrian/Driver Intoxicated	-0.0497	-0.2467	-0.2673	0.0059	0.5578
Crash Speed Limit (One SD Increase)	-0.0174	-0.0682	-0.0174	0.0634	0.0397
Hit and Run (1=Yes)	-0.0208	-0.0438	0.0037	0.0411	0.0199
Crash Took Place at Intersection	0.0040.	0.0289	0.0117	-0.0271	-0.0175

	No Injury	Possible Injury	Non-Incapacitating Injury	Suspected Serious Injury	Killed
City Street vs. Road Types					
County Road	-0.0226	-0.0495	0.0022	0.0464	0.0234
Farm-to-Market	-0.0231	-0.0507	0.0022	0.0475	0.0240
Interstate	-0.0226	-0.0494	0.0023	0.0463	0.0233
Non-Trafficway	-0.0269	-0.0585	0.0030	0.0549	0.0276
Other Roads	-0.0531	-0.1612	-0.0541	0.1470	0.1214
Tollway/Toll Bridge	0.0881	0.1023	-0.0592	-0.0995	-0.0318
US State	-0.0269	-0.0592	0.0024	0.0555	0.0282
On Roadway vs. Other Location					
Off Roadway	0.0136	0.0236	-0.0061	-0.0221	-0.0090
Shoulder	0.0254	0.0409	-0.0130	-0.0386	-0.0148
Median	0.1143	0.1175	-0.0795	-0.1168	-0.0355
Curvature + Grade + Traffic Control					
Straight Grade	-0.0295	-0.0680	-0.0004	0.0637	0.0341
Curved	-0.0355	-0.0869	-0.0058	0.0813	0.0469
Traffic Sign	-0.0076	-0.0071	0.0079	0.0067	0.0002
Traffic Signal	0.0002	0.0194	0.0120	-0.0183	-0.0132
Other (human control, rail gate etc.)	-0.0046	0.0019	0.0090	-0.0019	-0.0044
Population					
<5000	-0.0101	-0.0509	-0.0265	0.0449	0.0425
5000-9999	-0.0051	-0.0363	-0.0208	0.0323	0.0299
10000-24999	-0.0084	-0.0374	-0.0157	0.0341	0.0274
25000-49999	-0.0133	-0.0275	0.0028	0.0258	0.0122
50000-99999	-0.0268	-0.0221	0.0282	0.0212	-0.0005
Crash Time: 11 AM–4 PM vs. Other Times of Day					
5 AM–7 AM	-0.0210	-0.0701	-0.0245	0.0633	0.0523
7 AM–11 AM	-0.0240	-0.0411	0.0129	0.0384	0.0138
4 PM–8 PM	-0.0208	-0.0438	0.0034	0.0411	0.0201
8 PM–11 PM	-0.0240	-0.0587	-0.0038	0.0548	0.0317
11 PM–5 AM	-0.0253	-0.0694	-0.0123	0.0642	0.0427
Lighting Conditions					
Dark + Lighted	-0.0041	-0.0262	-0.0124	0.0241	0.0185
Dark + Not Lighted	-0.0142	-0.0652	-0.0305	0.0581	0.0518

Table 3.7 OLS Regression – Results for All and Fatal-only Pedestrian Crashes Per 1 M VMT for the 254 Texas Counties

	Y = Crashes Per 1 Million VMT, 2010–2018 Averages						Y = Crashes Per 1 Million VMT, 2010–2018 Averages																	
	<i>Initial Model</i>			<i>Final Model</i>			<i>Initial Model</i>			<i>Final Model</i>														
	<i>Coefficient</i>	<i>Std. Error</i>	<i>P-value</i>	<i>Coefficient</i>	<i>P-value</i>	<i>Std. Coef.</i>	<i>Coefficient</i>	<i>Std. Error</i>	<i>Std. Coef.</i>	<i>Coefficient</i>	<i>P-value</i>	<i>Std. Coef.</i>												
Intercept	-1.593	0.878	0.0709	-1.635			-0.0513	0.0444	0.249	-0.0575														
Lane Miles per Capita	0.0530	0.0941	0.574				0.00513	0.00476	0.282															
Average Speed Limit	5.167E-04	0.00242	0.832				7.506E-06	0.000123	0.951															
Average Lane Count	0.0113	0.0310	0.717				0.000348	0.00157	0.824															
Job Density Residuals	1.025E-04	2.386E-04	0.669				6.696E-06	1.206E-05	0.579															
Homeless Per 1,000	0.0667	0.0238	0.00543	0.0567	0.014	0.144	0.00446	0.00120	2.587E-04	0.00369	0.00173	0.185												
% Age 17 and Under	0.00568	0.00692	0.412				3.502E-04	3.502E-04	0.318															
% Age 65 and Older	0.00520	0.00568	0.360				3.389E-04	2.873E-04	0.240															
Growth Rate	0.00367	0.00195	0.0610				1.991E-04	9.874E-05	0.0451	1.245E-04	0.149	0.085												
Median HH Income	8.291E-06	2.550E-06	0.00132	7.509E-06	0.003	0.293	4.621E-07	1.290E-07	4.168E-04	4.320E-07	9.073E-04	0.334												
% of Pop. in Poverty	0.0187	0.00658	0.00478	0.0210	0.001	2.811E-04	0.00115	3.334E-04	6.391E-04	0.00139	2.910E-05	0.465												
Precipitation	-0.00111	0.00243	0.650				-1.774E-04	1.234E-04	0.153	-1.147E-04	0.145	-0.086												
Mean Max. Temp	-0.0146	0.0147	0.320				-9.294E-04	7.411E-04	0.211															
Mean Min Temp	0.00497	0.00989	0.615				5.271E-04	5.001E-4	0.293															
Truck DVMT	-7.539E-08	3.912E-08	0.255				-2.080E-09	1.979E-09	0.295															
DVMT per Capita	-6.553E-05	7.896E-05	0.407				-4.508E-06	3.996E-06	0.260															
WMT per Capita	12.866	3.307	1.301E-04	8.290	0.001	0.281	0.432	0.167	0.0105	0.234	0.0736	0.157												
<i>n</i>_{obs} = 254	R² = 0.223			Adj. R² = 0.171			R² = 0.182			Adj. R² = 0.166			R² = 0.222			Adj. R² = 0.170			R² = 0.161			Adj. R² = 0.143		

Table 3.8 OLS Regression – Results for All and Fatal-only Pedestrian Crashes per WMT for Texas’ 254 Counties

	Y = Pedestrian Crashes per WMT, 2010–2018 Averages						Y = Fatal Pedestrian Crashes per WMT, 2010–2018 Averages																	
	<i>Initial Model</i>			<i>Final Model</i>			<i>Initial Model</i>			<i>Final Model</i>														
	<i>Coef.</i>	<i>Std. Error</i>	<i>P-value</i>	<i>Coef.</i>	<i>P-value</i>	<i>Std. Coef.</i>	<i>Coef.</i>	<i>Std. Error</i>	<i>P-value</i>	<i>Coef.</i>	<i>P-value</i>	<i>Std. Coef.</i>												
Intercept	-0.0321	0.0389	0.413	-0.0145			0.0227	0.0121	0.063	0.010														
Lane Mi. per Capita	0.00527	0.00417	0.215				0.00155	0.00130	0.234															
Average Speed	-1.920E-04	1.070E-04	0.074	-1.556E-04	0.12	-0.0624	-1.276E-05	3.349E-05	0.703															
Average Lanes	2.780E-04	0.00137	0.84				-2.291E-04	4.287E-04	0.594															
Job Density Residuals	6.312E-06	1.057E-05	0.55				-6.229E-09	3.298E-06	0.985															
Homeless Per 1,000	0.0115	0.00105	6.77E-23	0.0112	0.000	0.433	5.991E-04	3.287E-04	0.069	7.525E-04	0.017	0.143												
% Age 17 and Under	5.260E-04	3.060E-04	0.088	5.111E-04	0.016	0.0955	-3.335E-05	9.568E-05	0.728															
% Age 65 and Older	2.390E-04	2.526E-04	0.344				-7.671E-05	7.851E-05	0.330															
Growth Rate	9.860E-05	8.646E-05	0.255				-7.803E-06	2.701E-05	0.772															
Median HH Income	1.750E-07	1.130E-07	0.123	1.444E-07	0.191	0.0861	3.331E-08	3.532E-08	0.346															
% Pop. in Poverty	3.750E-04	2.916E-04	0.199	5.860E-04	0.023	0.151	1.983E-04	9.105E-05	0.030															
Precipitation	-9.238E-05	1.083E-04	0.394				-5.501E-05	3.374E-05	0.104	-4.184E-05	0.048	-0.118												
Mean Max. Temp	-5.141E-05	6.941E-04	0.937				-2.272E-04	2.202E-04	0.180															
Mean Min. Temp	9.237E-05	4.382E-04	0.833				1.482E-04	1.368E-04	0.280															
Truck DVMT per capita	2.472E-09	1.733E-09	0.255				4.123E-10	5.238E-10	0.454															
DVMT Per Capita	5.783E-05	3.499E-06	2.41E-41	5.754E-05	0.000	0.581	6.008E-07	1.092E-06	9.79E-08	5.823E-06	1.41E-07	0.288												
WMT per capita	0.129	0.147	0.378				-0.0661	0.0457	0.149															
<i>n</i>_{obs} = 254	R² = 0.645			Adj. R² = 0.621			R² = 0.623			Adj. R² = 0.615			R² = 0.168			Adj. R² = 0.112			R² = 0.138			Adj. R² = 0.120		

3.6. Key Results

The key results of the models developed in Section 3.5 are discussed in this section.

3.6.1. NB & OP Models Results Discussion

Table 3.4 (above) shows the parameter estimates of the NB models. Two models were estimated: one for all pedestrian crashes and another for fatal pedestrian crashes. The dispersion parameters, P , for both models are greater than zero, implying that the data are over-dispersed (the variance exceeds the mean of crash counts), and the NB model is preferred over the Poisson regression model.

The association between VMT and pedestrian crash frequencies is positive and non-linear (exponents $\alpha = 0.7390$ for all pedestrian crashes and $\alpha = 0.8730$ for fatal pedestrian crashes), consistent with the expectation that crash frequencies increase with VMT but crash rate effectively falls as VMT of the segment rises. Among highway design variables, on-system roads (state-maintained arterials), median width, shoulder width, and speed limit were found to be practically significant. On-system roads show strong association with fatal crashes: a 42.81% increase of all pedestrian crashes versus 136.53% increase of fatal crashes only. As per CRIS data, two-thirds of all fatal pedestrian crashes in Texas (2010–2019) occurred on on-system roads. Other variables such as shoulder width, median width, and speed limit are negatively associated with pedestrian crashes. Higher speed limit roadways usually have fewer pedestrian activities that might contribute to lower numbers of pedestrian crashes; however, pedestrian crashes on high speed segments are associated with more severe injuries, discussed later in the injury severity analysis.

Surprisingly, ADT per lane is estimated to have negative effects on pedestrian crashes when other variables are controlled (population and job density). Percentage of Truck ADT, however, shows positive association. This might be because the impact of high ADT per lane is captured by population density and job density. Previous studies also found a weak effect of ADT on pedestrian crashes when other variables are controlled (Huang et al., 2017; Pandey and Abdel-Aty, 2009; Zajac and Ivan, 2003).

Population density, job density, and types of urban areas were used as proxies for land use. All of these variables were found to be strong predictors of pedestrian crashes. Pedestrian crashes, including fatal crashes, increase with population and job density, with very high crash rate percentage change (35.78% for population density and 11.06% for job density). This might be partly due to high variance-to-mean ratios for both of these variables; thus, a one-SD change implies a substantial shift. The effect of urbanization should be interpreted with urbanized areas having population 50,000–200,000 as a baseline. Compared to the baseline, large urban areas with populations greater than 200,000 are expected to have 23.05% and 14.63% more pedestrian crashes and fatal pedestrian crashes, respectively. By contrast, small urban areas and rural areas have fewer numbers of crashes. This is consistent with expectations because more dense locations

in large, urbanized areas usually have higher traffic volumes and pedestrian activities, thus increasing the exposure of pedestrian crashes.

Climate, proximity, and transit-related variables—such as rainfall, distance to the closest schools and hospitals, and the number of transit stops—were also included in the model. Among these variables, distance to the closest schools, distance to the closest hospitals, and the presence of transit offer practical significance, although these variables are rarely considered in pedestrian safety literature. Results from the model estimation show that a one-SD decrease in nearest school distance (1 SD = 2.72 miles) is associated with a 52.45% increase in pedestrian crashes and a 22.92% increase in fatal pedestrian crashes. Similarly, hospital distance also shows strong association (except fatal crashes) but is less significant than school distance. Finally, the presence of transit stops along the segments was found to be strongly significant (95.54% increase in pedestrian crashes and 53.46% increase in fatal pedestrian crashes), presumably due to high pedestrian activity near transit stops.

3.6.2. Pedestrian Injury Severity

Both the OP and HOP were estimated using the “`oglmx`” package in R (Carroll, 2017). Results from the likelihood ratio test suggest that heteroskedasticity exists ($\chi^2 = 2561.7$; $P < 0.0001$), and therefore the HOP model was preferred over the OP model (Table 3.3). The coefficients of both models show consistent estimates; however, the main difference is observed in terms of variance components. The HOP model shows significant variance for pedestrian age, gender, speed limit, vehicle type, traffic control type, population of the area, time of day, and lighting condition, suggesting that these variables can affect the spread of latent severity.

Other variables that do not show significant impacts are discarded from the variance equation. The following section discusses details about the impacts of explanatory variables on pedestrian injury severity. Among different vehicle types, pick-up trucks, sport utility vehicles (SUVs), vans, heavy-duty trucks, and buses significantly increase pedestrian injury severity in pedestrian-motor vehicle crashes (Table 3.5). Previous studies also reported similar findings, particularly high injury severity associated with light-duty trucks (SUVs, pickup trucks, and vans) (Lefler & Gabler, 2004; Pour-Rouholamin and Zhou, 2016; Anarkooli et al., 2017; Liu et al., 2019). These vehicles pose higher risks due to heavy mass, higher bumpers, and a more geometrically blunt frontal profile (Lefler & Gabler, 2004). The model also predicts significant variance for vehicle types, suggesting that impacts of unobserved attributes are associated with vehicle types (e.g., shape, stiffness, frontal profile) that increase the range of injury severity prediction. Marginal effects (Table 3.5) show that compared to passenger cars, light-duty trucks (pickup trucks, SUVs, and vans) increase the probability of being killed or seriously injured by 13.9%. According to CRIS data, the number of light-duty vehicles involved in pedestrian deaths is increasing at a fast rate in Texas: during the period 2010–2018, the number of cars involved in fatal pedestrian crashes increased by 64.7% while the number of SUVs and pickup trucks involved in fatal pedestrian crashes increased by 98.6% and 92.9%, respectively. The growing popularity of SUVs, pickup trucks, and vans partly

explains the high injury severity associated with these vehicles. From 2009 to 2016, the share of cars to the total number of light-duty vehicles purchased in the US dropped from 60.5% to 43.8%, while during the same time period, share of SUVs, pickup trucks, and vans increased from 39.4% to 56.2% (EPA, 2017).

Two pedestrian characteristics—age and gender—are found to be significant. Injury severity increases with pedestrians' age, suggesting that older people are vulnerable to more consequential outcomes. An increase of pedestrian age by one SD increases the risk of fatality by 1.69% and serious injury by 3.16%. Male pedestrians are also more likely to sustain severe injury. CRIS data shows that 72.38% of the pedestrians killed in motor-vehicle crashes in Texas from 2010 to 2019 were male. The effect of pedestrian age and gender on injury severity is consistent with the previous findings of Kim et al. (2008), Zhu et al. (2013), and Pour-Rouholamin and Zhu (2016). The model also predicts significant heteroskedasticity for pedestrian gender and age. The unobserved effects of pedestrians on injury severity vary more widely as the age of the pedestrian increases.

Drivers' characteristics also affect pedestrian injury severity. The involvement of younger drivers (aged less than 24) significantly increases the risk of pedestrian injury compared to drivers of the middle-age group (25–64). Male drivers are also more likely to be involved in pedestrian crashes than female drivers. Previous studies also had similar findings regarding male and younger drivers (Kim et al., 2008, Kim et al., 2010; Pour-Rouholamin and Zhu, 2016); however, the effect of older drivers (aged 65 or above) is mixed (Kim et al., 2008; Siddiqui et al., 2006; Mohamed et al., 2013). The results show that drivers aged 65 or above increase injury severity for pedestrians; however, it should be noted that the effect size is small. Wood et al. (2014) found that older drivers (age range 63–80) recognize pedestrians at approximately half the distance required for younger drivers (age range 18–38), which gives less response time to pedestrians.

Among different explanatory variables in the model, intoxication (in drivers and pedestrians) is found to have the strongest effect on pedestrian injury severity. Alcohol- or drug-related crashes are more likely to result in serious injury or deaths for pedestrians. According to CRIS data, alcohol and/or drugs were involved in 37.6% of pedestrian deaths. In most of these cases (33.38% of pedestrian deaths), pedestrians tested positive in alcohol and/or drug screens. Most (88.84%) of alcohol/drug-related pedestrian deaths occurred after dark. Walking under the influence, particularly at night, is one of the major causes of pedestrian fatalities.

Regarding time of day, crashes occurring from 8:00 PM to 5:00 AM showed an increase in the probability of severe pedestrian injuries. Most (79.22%) pedestrian deaths occur at nighttime. This finding is consistent with previous studies (Pour-Rouholamin, 2016; Aziz et al., 2013; Kim et al., 2008). The results also show higher risk of severe injuries in early morning hours (5AM–7AM). There might be several possible explanations: during these time periods (late night and early morning hours), traffic is lighter than usual, which might cause both pedestrians and drivers to ignore safety rules (drivers might travel at reckless speeds while pedestrian might choose to cross

roads abruptly). Moreover, pedestrian activities early in the morning (walking, jogging, or other physical exercise) and alcohol/drug involvements at night (discussed earlier) combined with darkness might also contribute to high injury severity during overnight hours. Lighting conditions also have a separate and significant influence, as a difference in probabilities of severe injuries between lighted roads and unlighted roads is also observed. Roads without streetlights after dark significantly increase the risk of pedestrian fatalities.

Roads with higher speed limits also lead to more severe pedestrian injuries. Table 3.4 shows the change in predicted probabilities by injury severity levels due to a one-SD increase of speed limit. The positive association between speed limit and injury severity is consistent with previous studies (e.g., Halem et al., 2015; Chen and Fan, 2019). Although the posted speed limit usually influences vehicle speed on roads, a more appropriate indicator would be the actual speed of the vehicle at impact, which is difficult to obtain for many cases. Speed limit increases the variance and outcome uncertainty: the unobserved effect varies more widely as the speed limit increases. Hit-and-run crashes increase injury severity levels. Almost 20% (19.4%) of pedestrian deaths are hit-and-run cases. Fleeing drivers increase the risk of pedestrian fatality because this often causes a delay in emergency service arrival and there is also the possibility that a pedestrian might get hit again by another vehicle after the first impact.

With regard to roadway characteristics, compared to city streets, the risk of severe pedestrian injury increases if a crash takes place on interstate, US, and state highways; county roads; and other types of roads not classified. Generally, city streets have lower speed limits and incorporate traffic controls, which reduces pedestrian crash severity. Analyzing CRIS data, we find that interstate highways account for 5.5% of pedestrian crashes but 20.6% of pedestrian fatalities in Texas. This percentage becomes higher when restricted to major urban areas. For instance, IH-35 alone accounts for 28.2% of pedestrian deaths in Austin over the last ten years. Higher speeds, poor lighting conditions, pedestrians entering to the highways, and lack of countermeasures might contribute to the severity of crashes on highways. Road geometry also affects crash severity: pedestrian-involved crashes on curved roads are more likely to result in more severe injuries than those on straight roads. The marginal effect shows that curved roads increase the probability of fatal crashes by 4.7% and serious injury by 8.1%. The location of the crash affects the type of injury. Crashes that occur at an intersection are associated with less severe injuries. Most pedestrian fatalities (89.16%) occur at non-intersection locations. The probability of severe injury decreases when the crash takes place on non-roadway sites (e.g., parking lots, driveways), shoulders, and medians, as compared to on roadways. Vehicle impact speed is usually lower in these locations, decreasing the likelihood of severe injury. The presence of traffic controls, such as traffic signals, reduces the probability of fatal and severe injuries. Pedestrians and drivers are better informed of each other's right of way and expected movements when there are traffic signals or traffic signs. As seen in studies on traffic calming in urban areas, drivers are usually more cautious and drive at lower speeds in the presence of traffic controls, as compared to places where there are no such controls (Ewing, 1999).

3.6.3. OLS Regression Model Results Discussion

Tables 3.7 and 3.8 contain a column of standardized coefficient values, which can help in comparing the relative predicted impacts of each explanatory variable, implying the practical significance of a variable. Since the final model specification includes variables that are statistically significant at the 90% confidence level, this approach can show the relative efficacy of certain predictor variables. This standardized coefficient is the estimate of how much change in crashes or fatalities per 1 million VMT or crashes or fatalities per WMT will result from an increase of one standard deviation (one-SD) in the explanatory variable, with all else constant.

For crashes per 1 million VMT (Table 3.7), the strongest relationships are between median household income and per capita WMT, for which there are positive relationships, with a practically significant relationship for rates of homelessness as well. Literature has shown that higher-income persons tend to walk longer distances (Yang & Diez-Roux, 2012), although the county level is at a far more aggregate level than the NHTS data from which the WMT figures are sourced, which is primarily at the census tract level. Thus, higher rates of WMT would, in this case, point to higher rates of pedestrian crashes per 1 million VMT. For fatalities per 1 million VMT, the picture is a bit clearer in terms of practical relationships. Median household income and the percentage of population in poverty both display stronger, positive relationships, pointing to more urban counties where both median income and the population in poverty tend to be higher in Texas. This may be due to a larger wealth gap within urban areas as opposed to suburban counties, which are more uniform in income; lower-income people also tend to walk for longer durations (and less distance), which may also increase exposure time among those who cannot own a car due to the financial burden (Yang & Diez-Roux, 2012). A weaker but still statistically significant relationship exists between growth rate and fatalities; in the VMT crash model, growth rates initially displayed a low p-value, but had a much higher p-value in the final model, suggesting that growth rates play only a minor role in rates of pedestrian crashes and fatalities. Exurban Texas counties, such as Hays, Kaufman, and Montgomery, would be areas that could shine more light on this through tract-level analysis.

The crashes per WMT model (Table 3.8) also shows a strong, positive relationship with homelessness rates and poverty, but has a weaker relationship with household income, as well as the curious addition of a positive relationship with the percentage of the population under the age of 17. Tract-level analysis would be helpful here, as this could further be broken down among school-age children to show where the strongest relationships lie. Studies in Los Angeles schools show that there are risks for children walking to school (Bachman et al., 2015), which can be ameliorated through pedestrian safety educational programs and improved pedestrian infrastructure (DiMaggio et al., 2015). The weak, negative relationship with average speed limit would also point to urban counties having higher rates of crashes per WMT, as the lane-miles of rural roads is more limited to trunk highways that have higher speed limits than many urban and suburban roads, particularly residential streets.

Fatalities per WMT results are less conclusive. There continues to be a positive, practical relationship with homelessness rates, as well as daily VMT per capita, suggesting that counties with higher VMT per capita experience higher rates of fatalities. Fatality rates in rural counties would seem to reinforce this, as pedestrian crashes there tend to be less frequent but more fatal (Hall et al., 2004). Notably absent from the final model for either WMT model is WMT per capita, which has a far higher p-value in the final model for both crashes and fatalities per WMT. This provides some evidence against the ‘safety in numbers’ idea behind pedestrian safety, particularly in terms of crash rates. Higher WMT rates do not necessarily move crash and fatality rates among pedestrians in either direction, at least at the county level. Tract-level analysis may also be useful for examining this issue in-depth, particularly in areas of exceptionally high foot traffic, such as university campuses, central business districts, and entertainment districts.

3.7. Conclusions

A wide variety of data analysis methods can examine the characteristics of a litany of impacts on pedestrian crashes and fatalities. The NB, OP, and HOP models can reveal the effects of lighting, vehicle type, size, environmental variables, and crash time. Furthermore, heat maps and cluster analysis can show the spatial patterns of crashes, helping practitioners prioritize sites for countermeasures and showing, from a practical perspective, which types of roadway designs and factors may lead to higher rates of pedestrian crashes and fatalities.

Findings from the NB model indicate the practical significance of micro-level variables in predicting pedestrian crashes. Proximity to schools, hospitals, and transit lines are associated with higher crash frequencies, although these variables are rarely included in pedestrian crash frequency models. Total crash rates and fatal crash counts rise with number of lanes and population and job densities, while greater median and shoulder widths provide some protection. Higher speed limits are associated with lower crash frequencies but increase the likelihood of more severe injuries, as shown by the HOP model.

Results from the HOP model identified several risk factors at the pedestrian, driver, roadway, and vehicle levels that significantly affect pedestrian injury severity. Crashes occurring at night (8 PM–5 AM), without overhead lighting, involving intoxicated pedestrians or drivers, and light-duty trucks (SUVs, pickup trucks, CUVs, and vans) are associated with more severe injuries. In contrast, being a younger and female pedestrian, on a straight segment off the state (and interstate) highway systems, in the presence of a traffic control device (stop sign or signal) lowers the likelihood of pedestrian injury. The involvement of vehicles from more recent model years was not found to lower pedestrian injury; instead, increasing SUV and CUV purchases in recent years further raises concerns about pedestrian safety. Findings from this study underscore the importance of enhanced vehicle safety features for pedestrians, campaigns against driving and walking while intoxicated, improved roadway design, enforcement of safety countermeasures near schools and bus stops, and installation of additional traffic controls and streetlights where there are more pedestrian activities.

The results of the ordinary least-squares regression point to a practical, positive relationship between crash rates and both household income and homelessness, with a weaker positive relationship to the growth rate. Income, poverty, and homelessness rates all have significant impacts in crash and fatality rates per WMT models, as does percentage of young people; a greater proportion of individuals under 17 within a county corresponded with increased crash rates. However, there was no strong relationship between crash rates and WMT per capita in either model. In practical terms, future census tract-level models should closely examine income as a predictor of pedestrian injury and death and clearly define urban, suburban, and exurban tracts to reveal the individual patterns within these areas.

The rise of pedestrian crashes and fatalities across the United States is a worrying trend (NHTSA, 2019), and one for which there is no one specific answer. Policymakers may consider faster-acting countermeasures to lower speeds and educate drivers and pedestrians alike on safe driving and walking behaviors, such as those described in Tefft (2013) and Bachman et al. (2015). They may then turn to design investments that have been shown to reduce the risk for pedestrians, such as the path widening and path segregation discussed in Carroll et al. (2019), as well as improved lighting and signage (Welch, 2016; DiMaggio et al., 2015). In this way, policymakers and DOT leaders can address the issue on both ends, creating a more welcoming environment for pedestrians while simultaneously working to curb the factors that lead to greater pedestrian injury severity. An analysis of the CRIS data shows that nighttime and higher speed are two major drivers of pedestrian crashes and fatalities, mirroring national trends; these may be two good frameworks around which to build effective countermeasures.

Further work on understanding how homelessness plays into the bigger picture of pedestrian crashes and fatalities is important to advance the understanding of pedestrian crash trends, given the limited existing research conducted and data collected by governments across Texas and the United States. While a stronger relationship than many other variables was found between the prevalence of homelessness and rates of pedestrian crashes in this model, only anecdotal evidence—provided by DOT officials, law enforcement, and city transportation and safety staff in conversations regarding this issue—supports the idea that homelessness is a major contributor to pedestrian crash counts (Oborski, 2020; Lee, 2020). This model was derived by piecing together HUD PIT count data with TxDOT data; independent data on pedestrian crashes collected by cities would be a good first step towards better understanding the nature of the interactions between homelessness and pedestrian crashes and fatalities. For example, the City of Austin only started collecting data on homelessness and pedestrian crashes in 2019, so a comprehensive dataset on pedestrian crashes involving suspected homeless individuals remains a distant goal.

Chapter 4. Develop Decision Trees to Classify Crashes across Contributing Factors

Decision trees are a relatively transparent form of machine learning (ML) or artificial intelligence (AI) for data analysis. They break a dataset or observational records into smaller and smaller subsets, to illustrate likely outcomes from different values of input variables (or branching “decisions”/directions that are based on explanatory variable values). This chapter uses pedestrian-crash data from TxDOT’s Crash Records Information System (CRIS) over the years 2010 through 2019 to develop a series of simple-to-use decision trees to classify various human factors (both pedestrian and driver), vehicle factors, roadway variables, and environmental factors associated with pedestrian crashes. These simple trees visually highlight influential factors to assist in understanding cause-and-effect relationships across a potentially wide set of contributing factors.

Although decision trees are intuitive, they do not have the same predictive accuracy as other statistical modeling approaches (like ordered probit models for severity and negative binomial models for crash count prediction) or ML methods. By aggregating many decision trees, using methods like bagging, random forest, and boosting, tree-based ML methods can substantially improve their predictive performance. Therefore, an analysis using tree-based methods is also developed to help improve the models. Methods such as random forests, gradient boosting, and Bayesian additive regression trees are applied and compared using pedestrian crash data. For estimates based on ordered probit and negative binomial models, please see a paper by Rahman et al. (2020) using these same data and predictors.

This chapter is organized into the following sections: Section 4.1 introduces the use of decision trees to develop pedestrian severity models and presents findings based on the 66,245 crash-involved pedestrians in CRIS records over the 10-year period. Section 4.2 describes tree-based methods for analyzing both pedestrian crash severity and crash count models across Texas using more than 700,000 roadway segments from TxDOT roadway inventory. Sections 4.3 and 4.4 conclude the chapter by summarizing the main findings.

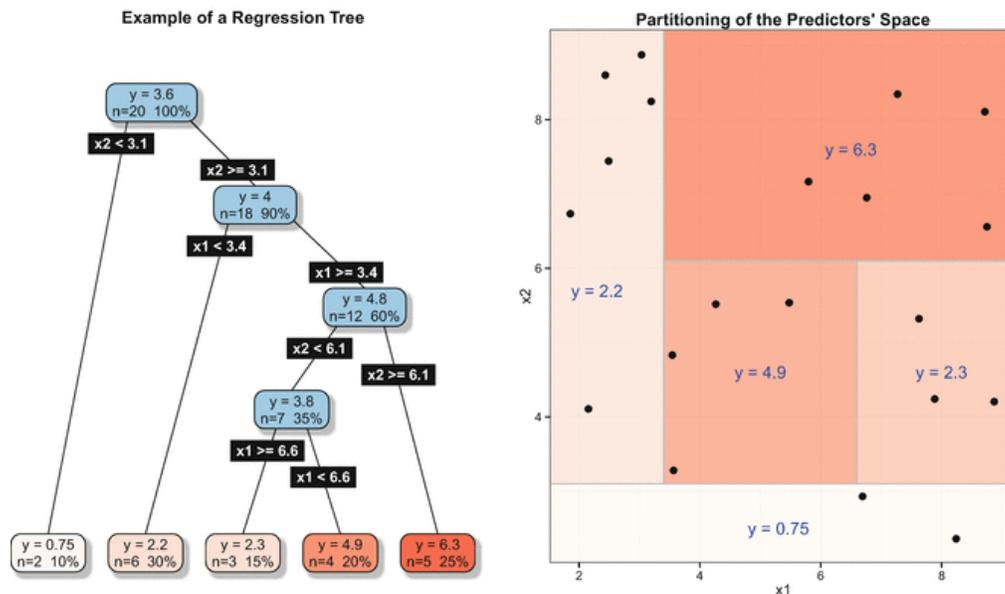
4.1. Decision Trees

The method used to estimate best-fit decision trees is known as recursive partition (Therneau and Atkinson, 1997), which is a modification of the Classification and Regression Trees (CART) method (Breiman et al., 1984). Decision trees are a type of supervised learning algorithm that can be used in both regression and classification problems (where the response variable is cardinal [like crash cost or crash count], ordinal [like severity], or categorical [like vehicle type]). Relevant terms are as follows:

- *Root node* represents the entire population or sample. It is divided into two or more homogeneous sets.

- *Splitting* is a process of dividing a node into two or more sub-nodes.
- When a sub-node splits into further sub-nodes, it is called a *decision node*.
- Nodes that do not split are referred to as *terminal nodes* or *leaves*.
- *Pruning* is the process of removing sub-nodes of a decision node. The opposite of pruning—adding sub-nodes to a decision node—is splitting.
- A sub-section of an entire tree is called a *branch*.
- A node divided into sub-nodes is called a *parent node* of the sub-nodes, whereas the sub-nodes are called the *child* of the parent node.

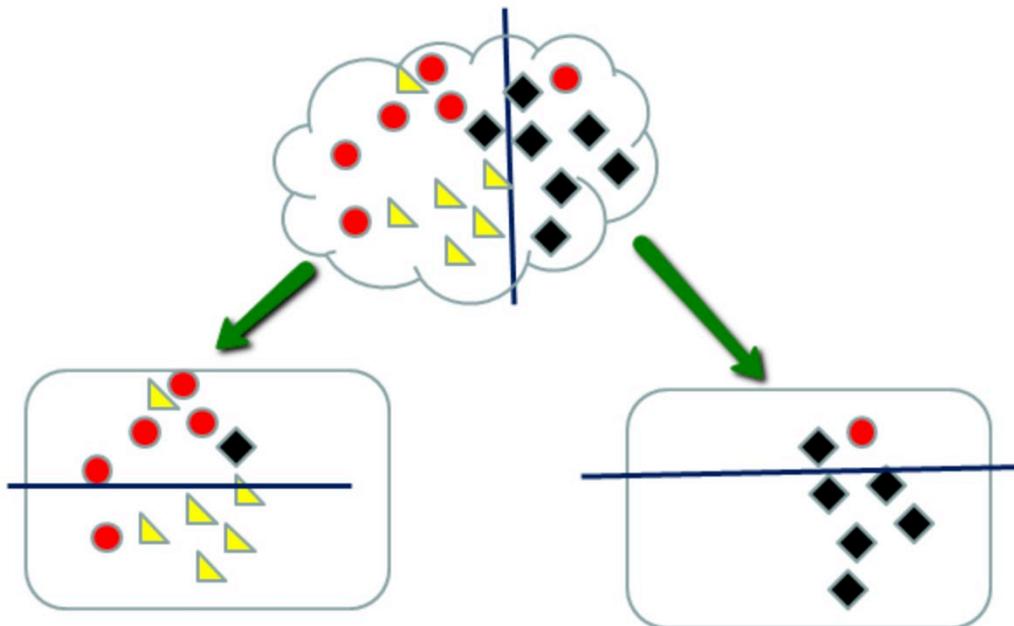
Regression trees are used to predict quantitative responses (like crash count or crash cost), while classification trees predict categorical or qualitative responses (like vehicle type). Both methods use a recursive binary splitting algorithm to develop the tree. The recursive partition is used for regression trees to minimize the Residual Sum of Squares (RSS), which is the sum of squared differences in predicted versus actual/observed values. The process can be summarized as follows: first, the data are divided into two branches based on different criteria using the input variables. The method creates a partition of two spaces, and the resulting two-prediction RSS is estimated. The method finds the variable split criteria that minimize the (current) RSS. Finally, a pruning process is applied, introducing tuning parameters that balance the tree's depth (number of branches) and overall goodness of fit (RSS). Figure 4.1 illustrates the results of such a method, using a five-branch (maximum depth) tree based on two variables (x_1 and x_2) to predict y values, with samples represented as dots within the predictor's spaces.



Source: Torgo, 2017

Figure 4.1 Example of a Regression Tree

The recursive partition method also uses a two-stage process to build trees for classification of categorical response (like vehicle type). As with a regression tree, the single best variable to split the data into two groups is sought, the data are separated, and this process is applied again and again, to each sub-group, recursively, as illustrated in Figure 4.2 until no improvements can be made. The splitting criteria are based on a measure of impurity or diversity within nodes (rather than RSS, since the Y variable is not cardinal or ordinal in nature). The most popular impurity measure is the Gini index, calculated by subtracting the sum of the squared probabilities of each class from one (the Gini index is zero if all the nodes are perfectly classified). In the case of categorical responses with m categories, the split criteria are tested using $2m-1$ comparisons, while for ordinal responses, only $m-1$ comparisons are tested. For example, for an ordinal response with three categories (small < medium < large), two comparisons are tested: {small} vs. {medium, large}, {small, medium} vs. {large}). If these categories were treated as unordered, the comparisons would include {small, large} vs. {medium} as well. The second stage of the procedure consists of using cross-validation to trim back the full tree (pruning).

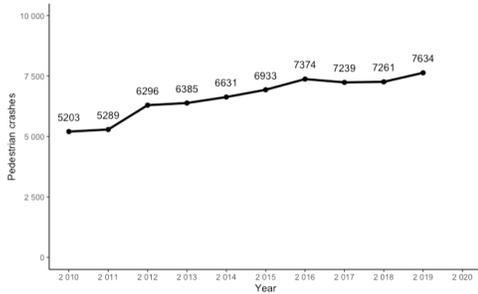


Source: Le, (2018)

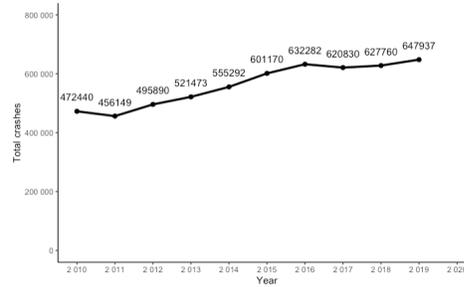
Figure 4.2 Example of a Classification Tree

4.1.1. Data and Method Description

In this chapter, the recursive partition method was used to develop decision trees to classify pedestrian crash severity levels. The data used consists of CRIS records from 2010 to 2019. Additional vehicle information (such as weight and dimensions) was provided by the Insurance Institute for Highway Safety (IIHS) using the reported vehicle identification number (VIN) for each vehicle involved in pedestrian crashes.



a) Annual pedestrian crashes



b) Total crashes in Texas

Note: results reported by police.

Figure 4.3 Texas' Annual Crash Counts

During the period of study, 66,245 crashes involved a pedestrian. Figure 4.3 shows a distribution of the annual count of pedestrian crashes compared to the total annual count of crashes in Texas. On average, pedestrian crashes represent 1.2% of the total crashes. Recent years show a significant increase in the number of both pedestrian and total crashes. For example, the total number of crashes in Texas increased 37% over 10 years, while pedestrian crashes increased 47% in the same period. The pedestrian severity level is divided into five categories:

- Not injured
- Possible injury
- Non-incapacitating injury
- Incapacitating injury
- Killed

Figure 4.4 shows the distribution of the data into pedestrian severity levels. Most crashes (69.1%) correspond to possible and non-incapacitating injuries; incapacitating injuries correspond to 15.8%. Of the pedestrians involved in crashes, 90.9% survived, 7.7% were killed, and the injury status of the remaining 1.5% is unknown (that percentage is omitted in the modeling section).

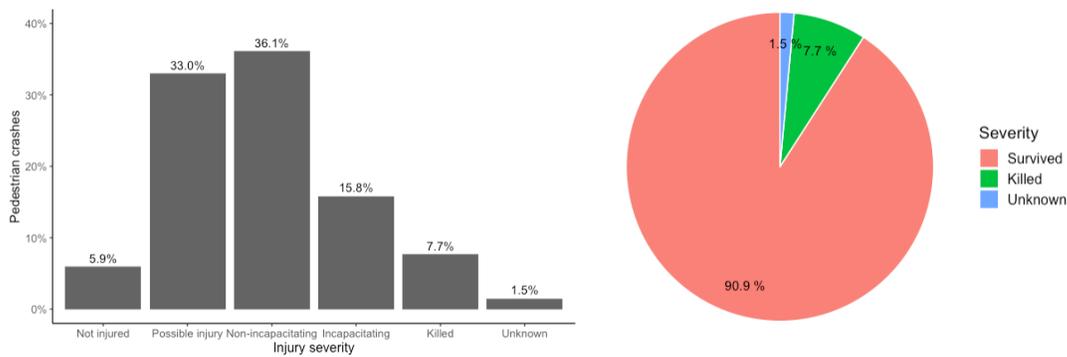


Figure 4.4 Pedestrian Injury Severity Distribution

After processing the dataset, invalid and extreme values are removed from the analysis. The data is divided into a training set, consisting of 80% of the sample, and a testing set (20% of the sample) to be used for the accuracy estimates. The models are estimated using R software and the “rpart” library (Therneau and Atkinson, 1997) for the recursive partition estimates.

The decision trees are divided into four categories:

- Highway design characteristics
- Pedestrian and driver characteristics
- Environmental, temporal, and locational conditions
- Vehicle characteristics

A final decision tree is developed using all the variables in the previous four categories to analyze their interaction. Table 4.1 describes the variables’ uses in each of the decision trees.

Table 4.1 Variable Description

Variable name	Variable type	Description
<i>Highway design characteristics</i>		
Road_Type	Categorical	Interstate, US state, farm to market, county road, non trafficway, city streets, others
Speed (limit)	Continuous	Range: 10 to 80 mph
Geometry	Categorical	Curved, straight grade, straight & level
Control	Categorical	Traffic signal, traffic sign, other, no control
Located	Categorical	Intersection, roadway segment
<i>Pedestrian and driver characteristics</i>		
Ped_Age	Continuous	Range: 0 to 100 years
Dr_Age	Continuous	Range: 16 to 100 years
Ped_Gender	Categorical	Male, female
Dr_Gender	Categorical	Male, female
Ped_Intoxicated	Categorical	Intoxicated (equal to 1), non-intoxicated (0)
Dr_Intoxicated	Categorical	Intoxicated (equal to 1), non-intoxicated (0)
<i>Environmental, temporal, and locational conditions</i>		
Time	Categorical	Night, morning, afternoon
Location	Categorical	Median, off roadway, shoulder, on roadway
Lighting	Categorical	Dark lighted, dark not lighted, daylight
Weather_Fog	Categorical	Yes, no
Weather_Rain	Categorical	Yes, no

Variable name	Variable type	Description
<i>Vehicle characteristics</i>		
Vehicle_Type	Categorical	Car, SUV/CUV, pickup, van
Curb_Weight	Continuous	Range: 2000 to 8600 pounds
Height	Continuous	Range: 40 to 100 inches
Model_Year	Continuous	Range: 1980 to 2020

4.1.2. Data and Method Description

A decision tree was estimated for each category, and graphical descriptions are presented in Figures 4.5 to 4.9. The decision trees are color-coded based on the pedestrian injury level classification within each node. This classification was made based on the level with the highest amount of data within each node. The nodes show the distribution of the data across the different injury severity levels represented as four decimal numbers for (1) possible injury, (2) non-incapacitating injury, (3) incapacitating injury, and (4) killed, respectively. They also show the percentage of data within the node as a single percentage number located at the bottom of the node. The splitting criteria are shown under the description of the node, and the decision is explained in the root node: the right node indicates that the data does not follow the splitting criteria (“no”), and the left node indicates that the data follows the criteria (“yes”). This logic continues until the terminal nodes are reached.

In addition to the decision trees, Figure 4.10 shows the variable importance for each of the models. The variable importance is estimated by calculating the relative influence of each variable and how much the impurity was improved as a result of introducing it into the model. Finally, as a measure of goodness of fit, Table 4.2 shows the model accuracy estimate using the test dataset.

The decision tree for highway design characteristics (Figure 4.5) shows that the speed and roadway type criteria are determinants in the classification of the injury severity levels. The variable importance indicates that these two features account for 98.3% of the importance metric. Similar results were obtained by (Rahman, M. and Kockelman, 2020) and presented in Chapter 3 for this project. The authors developed ordinal probit (OP) and heteroskedastic ordinal probit (HOP) models for pedestrian injury severity using the same dataset and found that more severe pedestrian injuries occur on roads with higher speed limits. Furthermore, the authors found that, compared to City Streets, the risk of severe pedestrian injury is higher if a crash takes place on interstates, US and state highways, or country roads.

The decision tree allows for observation of the interaction between speed limits and other variables such as roadway type. For example, 11% of the crashes were located on city streets, county roads, non-trafficways, and other roadways (refer to Table 1) with speed limits between 48 and 58 mph. Among these crashes, 38% resulted in a pedestrian being killed. Similarly, 89% of the crashes happened in areas with speed limits lower than 48 mph. Among these, 40% of the crashes had non-

incapacitating injuries. Furthermore, variables such as geometry (curved, straight grade, straight & level), and location (intersection, roadway segment) were not highly influential for the classification of the injury severity level.

In terms of driver and pedestrian characteristics, pedestrian intoxication is the most relevant variable for the classification of injury severity levels. The variable importance metric ranks this variable at an importance level of 89.9%. Similar results were found by (Rahman and Kockelman, 2020). However, the authors did not separate intoxication according to affected party (whether pedestrian and driver). Instead, they used a single variable for these characteristics. Their results showed that intoxication had the strongest effect on pedestrian injury severity. The decision tree (Figure 4.6) shows that 3% of the crashes involved intoxicated pedestrians, and 94% of those individuals were killed. The results suggest that young male and female pedestrians have less likelihood of suffering severe injury. The results do not indicate the effect of driver age; however, literature shows that, generally, older drivers tend to be more experienced and cause less severe injuries than young drivers.

Variables representing environmental, temporal, and locational conditions are also used for the development of a decision tree. Results shown in Figure 4.7 indicate that an interaction between lighting conditions and time of day significantly affect the severity of the crashes. Specifically, 61% of the pedestrian crashes occurred during daylight. Among those crashes, 42% were non-incapacitating and 3% resulted in a fatality. In contrast, 39% of the crashes occurred in dark and not lighted conditions, where 34% were non-incapacitating and 16% were fatal crashes. The variable importance suggests that weather conditions (rain and fog) and crash location (median, off roadway, shoulder, on roadway) did not greatly influence the severity of the crashes.

Vehicle characteristics such as model year, weight and dimensions, and vehicle body style also influence the pedestrian crash severity levels. Figure 4.8 shows that cars, SUVs, and vans have a lower severity than pickups. The resulting tree suggests that 11% of pickup crashes were fatal while only 8% of the crashes involving the other vehicle types were fatal. Newer vehicles (2006 and later) represent 9% of fatalities while older vehicles represent 7% of fatalities. However, in all the cases, the percentages are similar, and further analysis is required to understand the true effect of the model year. The variable importance for the decision tree of vehicle characteristics shows that the variables have similar weight in defining the model. Vehicle model year and type have importance in a range of 34% to 35%. Compared to the other models, where only one or two variables defined the importance, the vehicle characteristics show a more uniform distribution of the variable importance. Also, this model uses less data than the other estimated models because it does not include trucks, motorcycles, and other special vehicles, as their dimension and weight information was not available.

A final decision tree is modeled using all the variables described previously to understand the interaction across the different categories analyzed. The results of this model, shown in Figure 4.9, indicate that the most significant variable is pedestrian intoxication. Interestingly, this variable has

an importance of 60.8%, the highest among all the other variables included. The speed limit is also a relevant variable in determining the severity of the crashes. Crashes in locations with high-speed limits tend to be more severe, as expected. Other relevant variables include lighting conditions. As described previously, crashes in dark conditions cause more severe pedestrian injuries. Time of day, roadway type, and pedestrian age are also relevant for determining the severity of the crash.

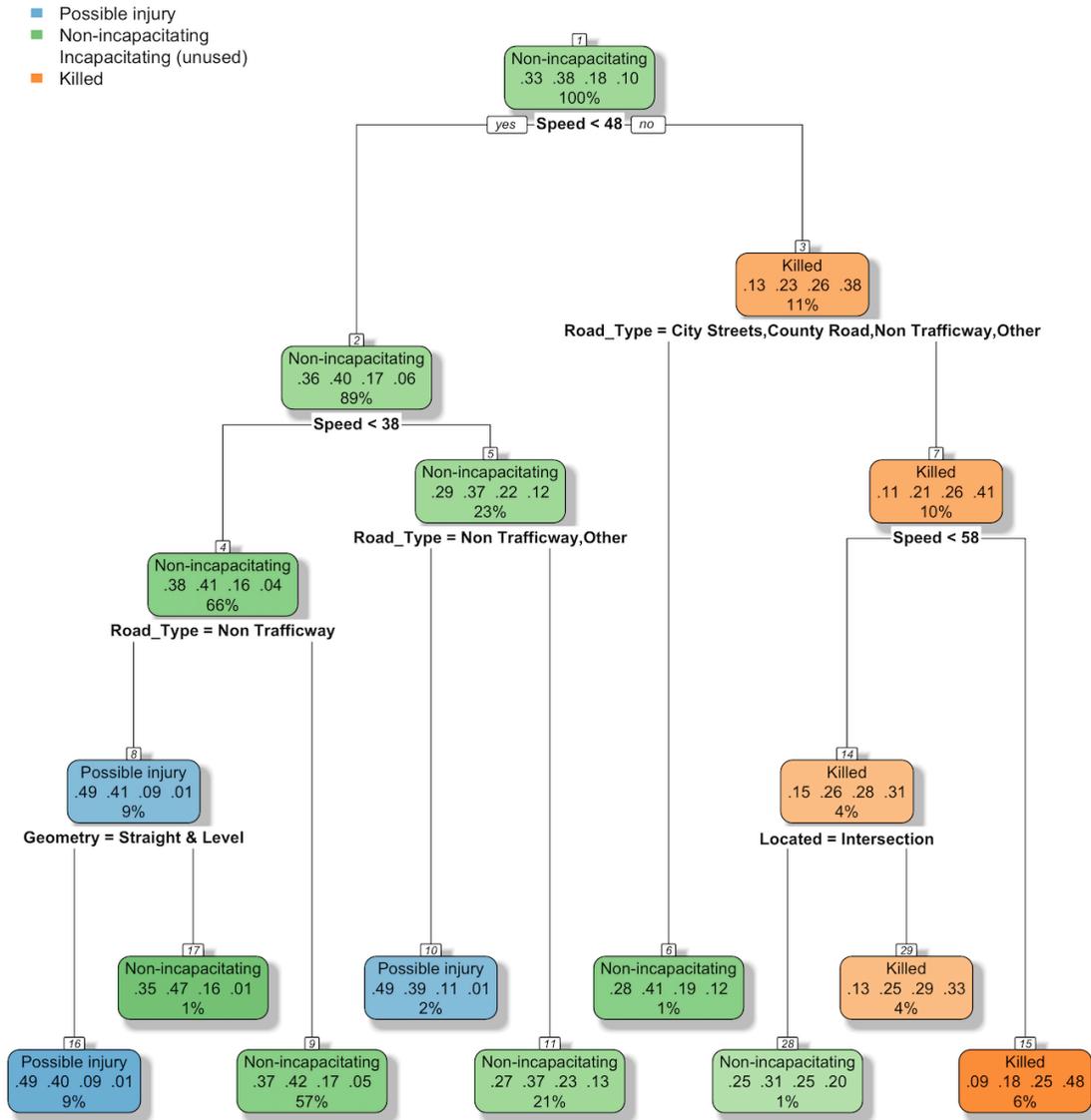


Figure 4.5 Decision Tree for Highway Design Characteristics

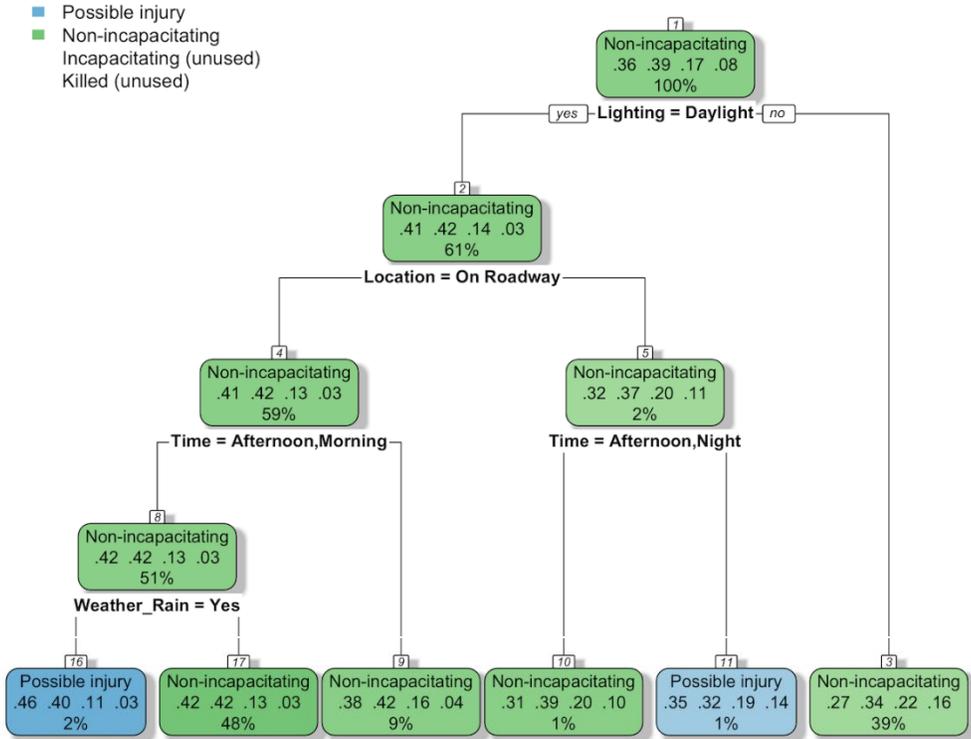


Figure 4.7 Decision Tree for Environmental and Other Conditions

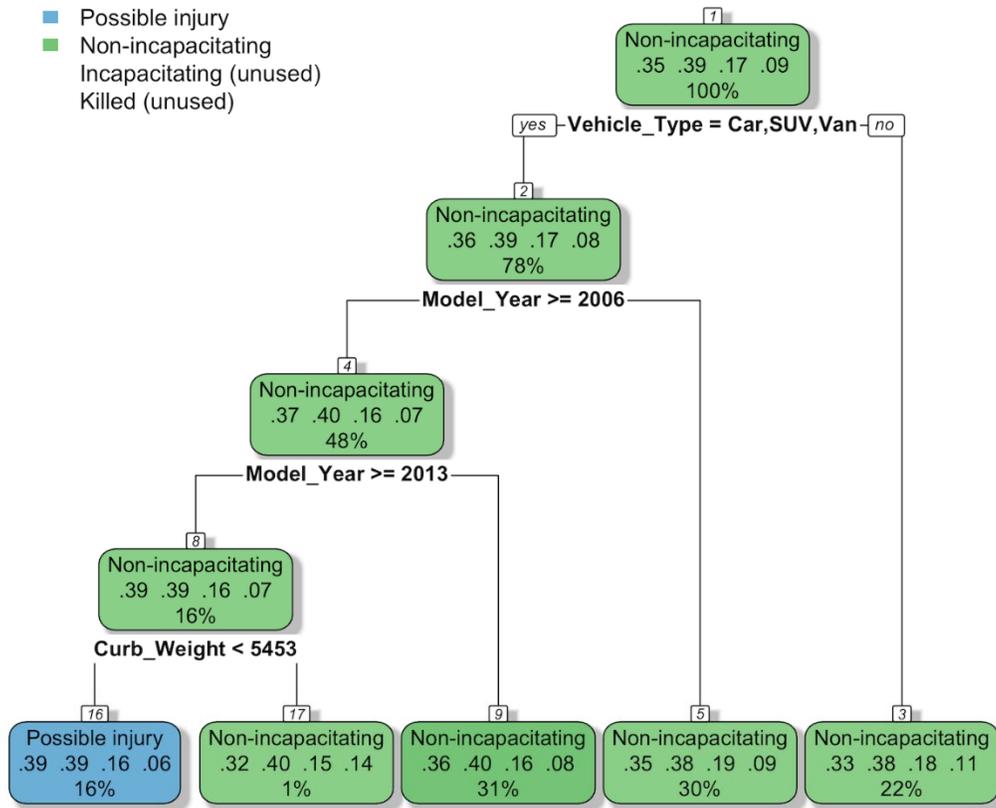


Figure 4.8 Decision Tree for Vehicle Characteristics

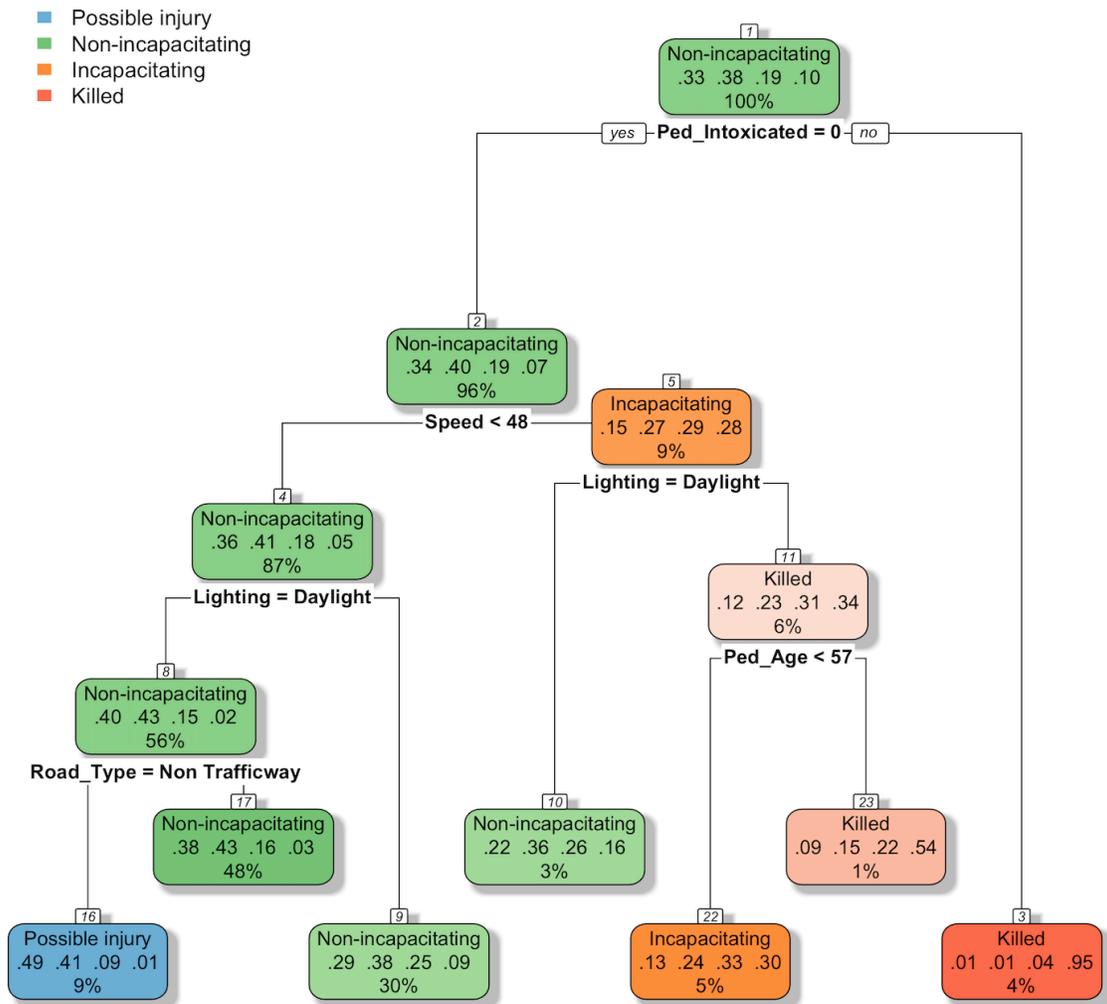
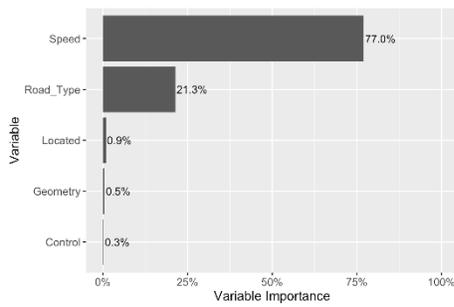
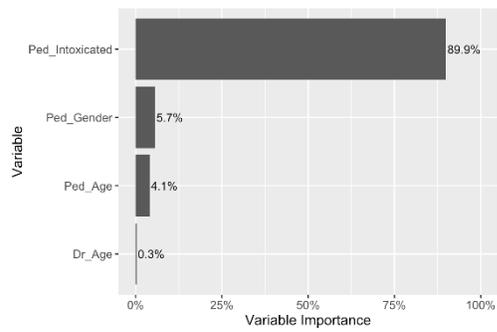


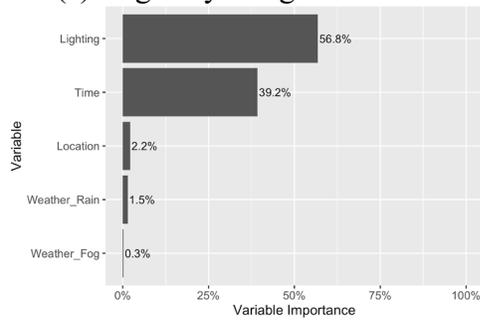
Figure 4.9 Model with All Variables



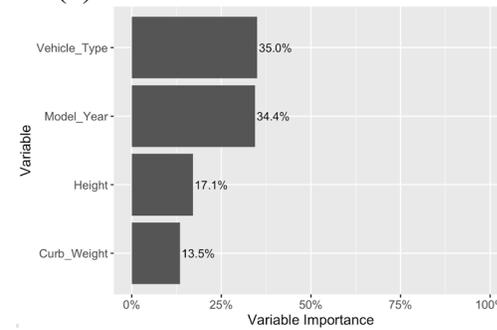
(a) Highway design characteristics



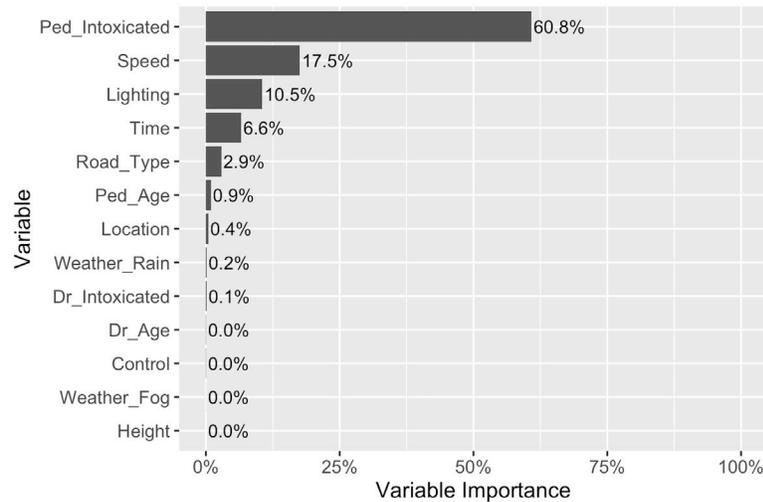
(b) Pedestrian and driver characteristics



(c) Environmental, temporal and locational conditions



(d) Vehicle characteristics



(e) Model with all variables

Figure 4.10 Variable Importance

The model accuracy is estimated using the test dataset, representing a random sample of 20% of the total sample size. As shown in Table 4.2, results indicate a model accuracy between 38% and 43% for all the decision trees. This accuracy level is expected in decision trees. Furthermore, as mentioned before, the main disadvantage of the decision tree is the lack of robustness in the results and the high variance.

Table 4.2 Model Accuracy

Description	Sample Size	Decision Tree Accuracy
Highway design characteristics	47,349	41.5%
Pedestrian and driver characteristics	47,163	41.6%
Environmental and other conditions	58,747	39.1%
Vehicle characteristics	42,591	38.9%
Model with all variables	33,079	43.2%

4.1.3. Summary

Five decision trees were developed to model different characteristics influencing pedestrian crash severity levels. The decision trees provide an interpretable and visual analysis of the main variables affecting the severity of the crashes. However, decision trees tend to have high variance and low accuracy. To address this limitation, the research team also calibrated and evaluated tree-based ML models, which can help reduce the variance and increase model accuracy significantly while taking advantage of the flexibility provided by the decision trees. The following section presents the results using different tree-based models.

4.2. Tree-based Methods

In this section, tree-based methods are applied and compared using pedestrian crash data. An extended version of this section can be found at:

https://www.cae.utexas.edu/prof/kockelman/public_html/TRB22MLpedcrashes.pdf

Tree-based models use a series of if-then rules to generate predictions from one or more decision trees. Various methods combining a set of tree models, i.e., ensemble methods, have attracted much attention and have been widely used for supervised learning tasks. These include random forests (Breiman, 2001; Liaw and Wiener, 2002), gradient boosting (Chen and Guestrin, 2016; Ke et al., 2017), and Bayesian additive regression trees (Chipman et al., 2012; He et al., 2018), each of which uses different techniques to fit a linear combination of trees. This section investigates the performance of different tree-based ML models in predicting pedestrian crash occurrence and injury severity in Texas. The following subsections briefly introduce the investigated tree-based models.

4.2.1. Background

Four different tree-based methods will be described in this section: random forest, extreme gradient boosting, efficient gradient boosting decision tree, and accelerated Bayesian additive regression trees.

4.2.1.1. Random Forest (RF)

An RF model comprises decision trees constructed by splitting each node using the best among a subset of predictors randomly chosen at that node with a different bootstrap sample of the data (Breiman, 2001). Running an RF algorithm can be described as follows (Liaw and Wiener, 2002): (1) draw n_{tree} bootstrap samples from the original data; (2) for each bootstrap sample, grow an unpruned tree using the following procedure: at each node, randomly sample m_{try} of the predictors and choose the best split from among those variables; and, (3) predict new data by aggregating the prediction of the n_{tree} trees, i.e., majority votes for classification, average for regression. With the two layers of randomness, i.e., random feature selection and bootstrap/bagging, RF is powerful at handling complex and non-linear relationships. RF can also be trained quickly since the trees do not rely on each other and thus can be trained in parallel. However, RF is known to be less accurate for regression problems as it tends to overfit.

4.2.1.2. Extreme Gradient Boosting (XGBoost)

XGBoost is a scalable ML system for gradient tree boosting, which gives state-of-the-art results on a wide range of problems (Chen and Guestrin, 2016). Boosting is an ensemble tree method that builds consecutive small trees with each tree focused on correcting the net error from the previous trees. For example, the first tree is split on the most predictive feature, and then the weights are updated to ensure that the subsequent tree splits on whichever feature allows it to correctly classify the data points that were misclassified in the initial tree. The next tree will then focus on correctly classifying errors from that tree, and so on. The final prediction is a weighted sum of all individual predictions. Gradient boosting is the most popular extension of boosting and uses the gradient descent algorithm for optimization.

4.2.1.3. Efficient Gradient Boosting Decision Tree (LightGBM)

LightGBM is another popular gradient boosting tree (GBT) model. Compared with XGBoost, LightGBM incorporates gradient-based one-side sampling (GOSS) to improve computational efficiency (Ke et al., 2017). The basic assumption behind GOSS is that those samples with larger gradients, i.e., under-trained instances, will contribute more to the information gain. Therefore, to retain the accuracy of information gain estimation, GOSS keeps all the instances with large gradients (e.g., larger than a pre-defined threshold or among the top percentiles) and only randomly drops those instances with small gradients. It was shown that LightGBM could lead to a more accurate gain estimation than uniformly random sampling, with the same target sampling rate, especially when the value of information gain has a large range.

4.2.1.4. Accelerated Bayesian Additive Regression Trees (XBART)

XBART is a variant of the Bayesian additive regression tree (BART) model with improved computational efficiency (He et al., 2018). Conceptually, BART is a Bayesian nonparametric approach that fits a parameter-rich model using a strongly influential prior distribution (Chipman

et al., 2012). BART is similar to GBT models, i.e., XGBoost and LightGBM, in that they all sum the contribution of sequential weak learners. However, BART weakens the individual trees using a prior, instead of multiplying each sequential tree by a small constant, i.e., the learning rate, as in GBT models. Additionally, BART performs the iterative fitting by using the back-fitting Monte Carlo Markov Chain (MCMC) algorithm rather than using gradient descent algorithms. The Bayesian perspective yields a number of practical advantages of BART, including the robustness to hyperparameter settings, more accurate predictions, and the inherent Bayesian measure of uncertainties. On the other side, the incorporation of the MCMC algorithm also imposes severe computational demands, especially in the application of high-dimensional large datasets. XBART improves the computational efficiency by adopting the novel stochastic hill-climbing algorithms, which follow the Gibbs update framework in BART but replace the Metropolis-Hasting updates of each tree with a novel grown-from-root back-fitting strategy (He et al., 2018). XBART is shown to yield very similar results to BART, but with much higher computational efficiency (He et al., 2018).

4.2.2. Model Fit Evaluation

Several metrics are used to evaluate and compare the performance of the models described above. The choice of evaluation metrics differs in crash occurrence and severity prediction models because crash occurrence models are regression models where the response variable is the total or fatal pedestrian crash occurrence, while crash severity models are classification models where the response variable is the severity of the crash as represented by categorical values. The following two sections discuss model evaluation metrics for crash occurrence and severity prediction models, respectively.

4.2.2.1. Crash Occurrence Prediction Model

The crash occurrence prediction models are formulated as regression problems, where the independent variables are the characteristics of roadway segments, and the response variable is the total or crash pedestrian crash occurrence on that roadway segment. R-square and root mean square error (RMSE) are adopted in this project as evaluation metrics for regression models. R-square represents the proportion of variance in the response variable that has been explained by the independent variables in the model. It provides an indication of goodness of fit and therefore a measure of how well unseen samples are likely to be predicted by the model, through the proportion of explained variance. The RMSE is a measure of how spread out these residuals are. It evaluates how concentrated the model-predicted response values are around the true response values. Lower RMSE scores indicate better model performance.

4.2.2.2. Crash Severity Prediction Model

The output of classification models can be class labels or the probability for the class membership. The discrete outputs of class labels form a confusion matrix, on which accuracy, precision, recall,

and F1 score are constructed (Tharwat, 2018). Accuracy is measured as the percentage of time for which the predicted labels for samples exactly match the corresponding true labels. Accuracy sometimes is sensitive to imbalanced dataset. Therefore, the geometric mean (GM) is also used as an evaluation metric. The precision measures the proportion of positive identifications that are actually correct, while the recall measures the proportion of actual positives identified correctly. The F1 score is a function of precision and recall, typically used in unbalanced samples (a large number of actual negatives) and seeks to provide a balance between precision and recall.

The area under the curve (AUC)—specifically under the Receiver Operating Characteristic (ROC) curve—is a widely used measure of performance of classification (Hand and Till, 2001). The ROC is a two-dimensional graph that plots the false positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis according to the discrete classification outputs. The continuous probability output can be discretized with a threshold to plot ROC. The value of AUC is between 0.0 and 1, and a higher AUC suggests better classification. AUC is equivalent to the probability that a classifier ranks a randomly chosen positive sample higher than a randomly chosen negative sample (Fawcett, 2006). The margin of a classifier was proposed by (Breiman, 2001). It measures the extent to which the classifier votes an instance to the correct class rather than to an incorrect class. The margin is in $[0,1]$, and the larger the margin is, the more confidence there is in the classification.

4.2.3. Hyperparameter Tuning

The aim of hyperparameter optimization is to find the hyperparameters of a given ML algorithm that return the best performance as measured by the specified evaluation metric. The optimization of hyperparameters (θ) can be represented in equation form as:

$$\theta^* = \operatorname{argmin}_{\theta \in \Theta} f(\mathcal{M}, \theta) \quad (1)$$

where, \mathcal{M} is the ML model; $f(x)$ represents an objective function to minimize, such as RMSE for regression models or F1 score for classification models, evaluated on the validation set; θ^* is the set of hyperparameters that yields the lowest value of the score; and θ can take on any value in the domain Θ . Bayesian hyperparameter optimization methods build a probability model of the objective function, i.e., $P(f(\mathcal{M}, \theta)|\theta)$, by tracking the past evaluation results and using them to select the most promising hyperparameters to evaluate in the true objective function (Klein et al., 2017). Specifically, the process of Bayesian hyperparameter tuning can be described as follows: (1) build a surrogate probability model of the objective function; (2) find the hyperparameters that perform best on the surrogate; (3) apply these hyperparameters to the true objective function; (4) update the surrogate model incorporating the new results; and (5) repeat steps 2–4 until the maximum number of iterations or specified time is reached.

4.2.4. Sensitivity Analysis

ML models excel at capturing complex relationships between input independent and output response variables. However, they can be less intuitive in explaining how and why such relationships are captured. Several sensitivity analysis methods were developed to mitigate the interpretability deficiency, aiming to unveil the cause-and-effect relationship between the input and output variables. Sensitivity analysis is a simple yet powerful way to understand an ML model by examining what impact each feature has on the model's prediction. The feature value was changed to calculate feature sensitivity, while all the other features stay constant, and the output of the model was recorded. If the model's outcome has been altered drastically by changing the feature value, it means that this feature significantly impacts the prediction.

Specifically, given a test set X , the process of evaluating the sensitivity of feature X_i can be described as follows: (1) train the baseline model on X and denote the prediction vector as y ; (2) create a new set X^* where a transformation was applied, such as reshuffling or dropping, over feature X_i ; (3) perform prediction on X^* and denote the prediction vector as y^* ; (4) measure the change in the outcome using the percentage change in the prediction mean, i.e., $\frac{\bar{y}^* - \bar{y}}{\bar{y}} \times 100\%$. In pedestrian crash occurrence prediction, the transformation is defined as in Li and Kockelman (2020): an increase of one standard deviation for continuous input variables and binary (0 to 1) change for binary input variables. Specifically, for each input variable, one standard deviation or binary change is applied to each data point. The modified variables are passed to the model to calculate the prediction, i.e., permuted prediction. Then, the difference between the mean of original prediction and permuted prediction is calculated to represent the contribution of that feature. In injury severity prediction, the probability of each class was obtained for every single data point. The same imputation and computation approach was used to analyze the marginal effects, as in pedestrian crash occurrence prediction, except that each class's probability is used instead of class values.

4.3. Results

Results of pedestrian crash occurrence and severity prediction models are discussed in this section.

4.3.1. Pedestrian Crash Occurrence Prediction

Four tree-based ensemble ML models were developed to predict pedestrian crash occurrence: RF, XGBoost, LightGBM, and XBART. For each model configuration, two models were trained—one for total pedestrian crashes and another for fatal pedestrian crashes. The optimal hyperparameters for each model were obtained using Bayesian optimization.

4.3.1.1. Model Performance Evaluation

Table 4.3 summarizes the model performance measured by R-square and RMSE on the testing set to predict total and fatal pedestrian crash occurrence. For the total pedestrian crash occurrence prediction model, LightGBM achieves the best performance in terms of both R-square and RMSE, while RF yields the best performance for fatalities. For fatalities, R-square values are lower than the values for the total pedestrian crashes, which can be related to the significantly low number of fatal pedestrian crashes. The computation times, including training and testing after the optimal hyperparameters are also obtained. LightGBM is the most computationally efficient model due to the efficient GOSS optimization algorithm. XBART is the most computationally expensive model, which can be explained by the expensive MCMC connection between the trees.

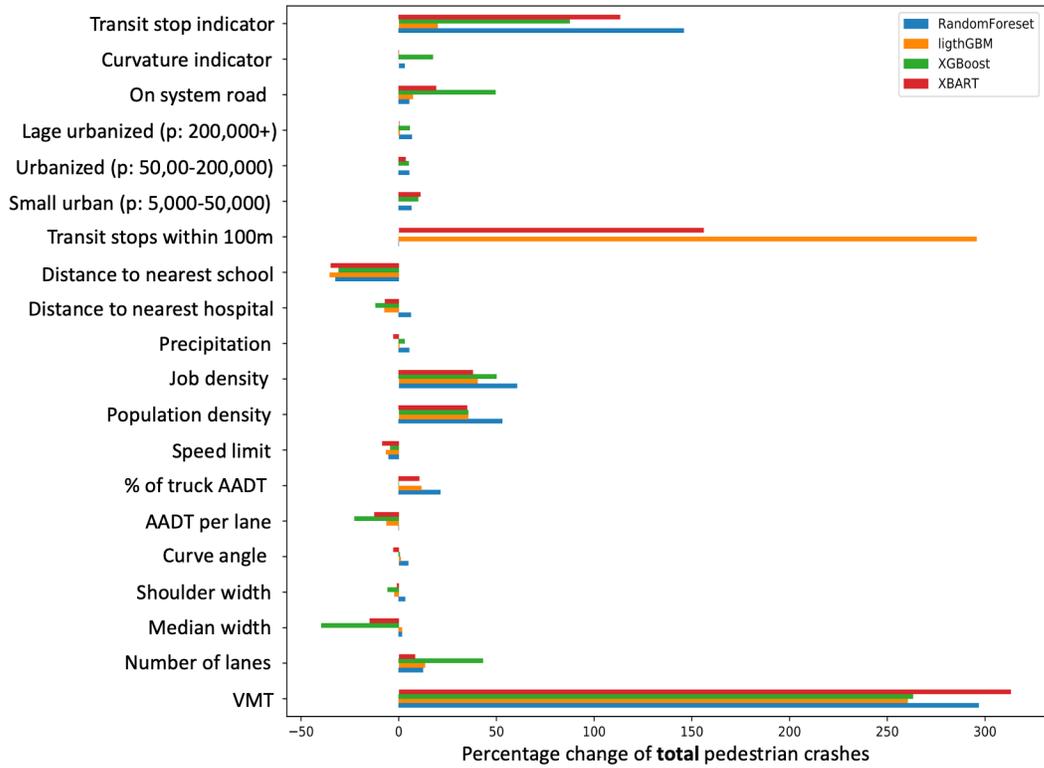
Table 4.3 Comparison of Model Performance and Computation Time

Model	Total pedestrian crash occurrence			Fatal pedestrian crash occurrence		
	R-square	RMSE	Time [s]	R-square	RMSE	Time [s]
RF	0.359	0.242	216	0.148	0.008	278
XGBoost	0.318	0.258	126	0.070	0.009	133
LightGBM	0.363	0.241	43	0.133	0.009	25
XBART	0.351	0.245	354	-0.001	0.010	5110

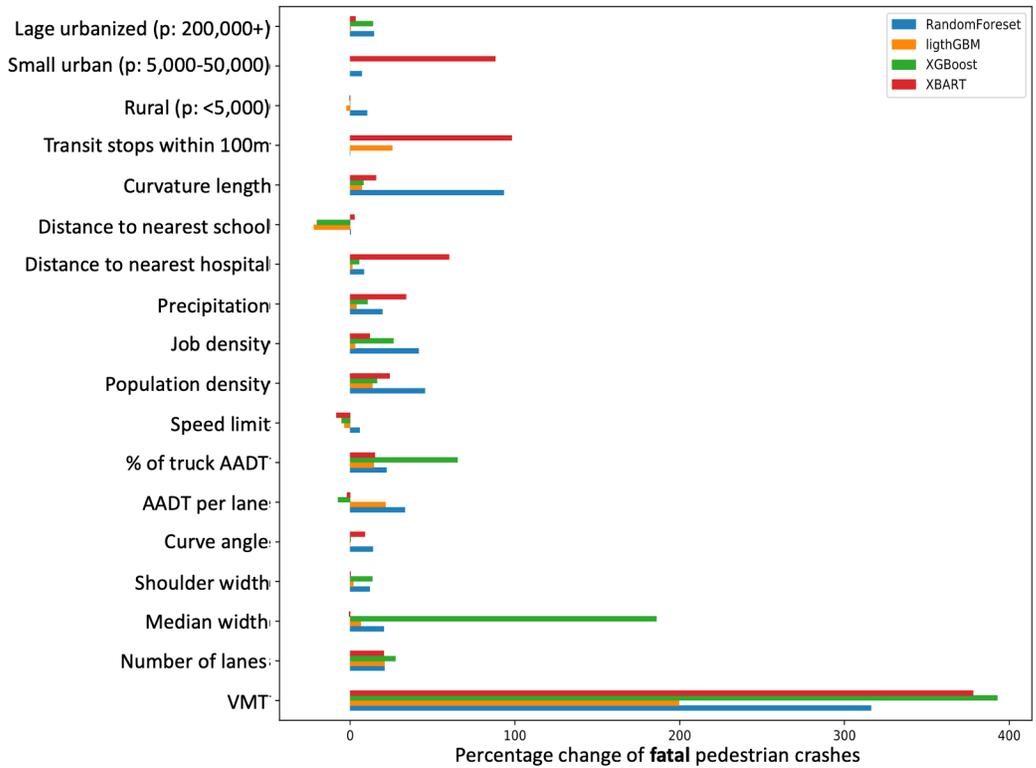
4.3.1.2. Feature Sensitivity Analysis

The practical importance of input variables can be estimated using the proposed sensitivity analysis approach. The value of continuous features is increased by one standard deviation, and binary changes are made on binary features for each data point in the dataset. Then the percentage change in the mean of the model prediction is estimated. The estimated feature importance for total and fatal pedestrian crash occurrence are shown in Figure 4.10. The y-axis shows the name of the input variables. The x-axis represents the percentage change in the mean of model prediction, i.e., total or fatal pedestrian crash occurrence, after applying the proposed transformation on the corresponding input features. Different colors represent different ML models: blue for RF, orange for LightGBM, green for XGBoost, and red for XBART.

As shown in Figure 4.11, the VMT figure has the most significant impact on total and fatal pedestrian crash occurrence. One standard deviation increase on VMT can lead to around a 270% and more than 300% increase in the total number and number of fatal pedestrian crash occurrences per roadway segment, respectively. However, one standard deviation increase of VMT (7,319) on all roadway segments is not practical, considering the capacity limit of segments. Therefore, the process was repeated considering a double VMT on each segment. The results indicate that the total and fatal pedestrian occurrence will increase by 50%, which is still a significant impact. These results, consistent with literature findings (Nashad et al., 2016), represent the higher crash risk faced by pedestrians with increasing VMT, which is consistent with the expectation that crash frequencies increase with the increase in pedestrian exposure to motorized vehicles.



(a) Total pedestrian crash occurrence prediction



(b) Fatal pedestrian crash count prediction

Figure 4.11 Sensitivity Analysis for Crash Occurrence Predictions

The number of transit stops within a buffer of 100 meters is a relevant variable in predicting total and fatal pedestrian crashes in roadway segments, according to the results of the LightGBM and XBART models. This variable is an indirect measure of pedestrian exposure as pedestrian activity surrounding transit stops is high. Similarly, variables for the distance to nearest school and distance to nearest hospital offer a practical significance. The number of pedestrian crashes increases in areas near schools and decreases as distance from schools increases, consistent with literature findings (Warsh et al., 2009). Interestingly, the hospital proximity is particularly significant for fatal crashes, where the frequency increases as the distance to the hospital increases, possibly related to the response time of emergency services. Although relevant, these variables are rarely considered in pedestrian safety literature (Rahman, M. and Kockelman, 2020).

Highway design variables such as on-system roads (or state-maintained arterials), number of lanes, curve angle, curvature indicator, and curvature length have a significant positive impact on pedestrian crash frequencies. One standard deviation increment on the number of lanes can lead to more than a 25% increment in total or fatal crash counts. On-system roads are found to be strongly correlated to the number of total crashes. The speed limit is found to be negatively correlated to the number of crashes. This can be related to the reduced exposure of pedestrians to high-speed roadway segments. However, high speed limits lead to more severe injuries, as discussed in the following section. Variables such as median and shoulder widths show diverse variations across the different models, limiting the conclusions for these variables.

Land use characteristics are described by variables such as population, job density, and types of urban areas. These metrics are directly related to pedestrian exposure. For example, dense urban areas with high job density usually have higher traffic volumes and pedestrian activity. Changes in one standard deviation led to a positive, significant increment in pedestrian crash frequencies, as expected. The number of pedestrian occurrences increases by approximately 50% for total counts and 30% for fatal counts when the population and job density are increased by one standard deviation. Large urbanized, urbanized, and small urban locations have more conservative increments of 10% for the total pedestrian crashes. However, for the fatal count model, the effect differs significantly across models, possibly related to the low count number within the different categories.

The four different models come to a similar conclusion about the significance of some features, such as distance to the nearest school, job density, population density, and VMT. However, the results diverge on other features, such as the number of transit stops within a 100-meter buffer. XBART and LightGBM consider the number of transit stops a very important feature in predicting the total pedestrian crash occurrence. One standard deviation increases on the transit stop variable can lead to an increase of 150% and 300% of total pedestrian crash occurrence, respectively. However, results from LightGBM and XGBoost show that the number of transits stops has little impact on the total pedestrian crash occurrence. This observation indicates that different ML models interpret the significance of the input features differently. It might make more sense to look only at the model that yields the best performance, i.e., prediction accuracy. Noticeably, the

discrepancies in the results from different models are even more obvious in fatal pedestrian occurrence prediction as compared with total pedestrian occurrence. This again stresses the importance of choosing the best performing model when comparing the evaluation metrics and then analyzing the feature importance using the chosen optimal model.

4.3.2. Pedestrian Crash Injury Severity Prediction

To estimate the crash injury severity models, first, the complete dataset was randomly split into training and testing datasets to predict crash injury severity. Then, four models (RF, XGBoost, LightGBM, and XBART) were fitted on the training dataset and tested on the testing dataset. Before model fitting, parameters for each model were tuned with the Bayesian optimization method to obtain optimal parameters. The simulation was repeated ten times.

4.3.2.1. Model Performance Evaluation

Table 4.4 summarizes model performance on model running time, accuracy, precision, recall, F1 score, and GM. In general, RF, XGBoost, and LightGBM behave similarly in terms of these classification metrics, but LightGBM runs much faster than the other two GBT models. Even with higher precision, XBART shows lower recall and F1 and it is heavily time-consuming.

Table 4.4 Summary of Model Performance on Injury Severity Prediction

Models	Time(s)	Accuracy	Precision	Recall	F1 score	GM
RF	36.55	0.42	0.49	0.33	0.34	0.53
XGBoost	75.17	0.42	0.43	0.33	0.34	0.53
LightGBM	18.59	0.42	0.45	0.34	0.34	0.53
XBART	1447.56	0.42	0.53	0.31	0.32	0.51

The crash injury data is imbalanced with 7%, 33%, 36%, 17%, and 7% of class 0 (not injured), 1 (possibly injury), 2 (non-incapacitating), 3 (incapacitating) and 4 (killed), respectively. The GM metric is less sensitive to an imbalanced dataset (Tharwat, 2018). Results show that RF, XGBoost, and LightGBM achieve the same GM value of 0.53, which is higher than XBART with a 0.51. This indicates that XBART may be more sensitive to imbalanced data.

The model margins show the confidence of a classifier making a correct classification. A positive margin value means the classifier voted for the right classification (Breiman, 2001), and a negative margin value indicates the classifier voted incorrectly. Figure 4.12 shows ten-time repeated results plotted in a bar graph; model margins are affected by the injury class. For example, all four models achieve higher margins on class 2 and class 4 (around 0.2), while the margin values for class 0 and class 3 are negative (-0.4 and -0.6), indicating a high discrepancy between true class and predicted class. This discrepancy is not necessarily related to data imbalance as class 3 makes up a greater proportion of the data than class 4, which has a higher margin. Also, XBART shows a much greater

capacity for correctly classifying class 2, while for other classes XBART is limited to a weak classification capacity.



Figure 4.12 Margins of Classification Models on Different Classes

The ROC curve is one of the most important metrics to visualize the performance of multiclass classification models. It quantifies the extent to which the model is able to distinguish between classes (Narkhede, 2018). The AUC is the quantitative measure of ROC. The point (0,1) in the ROC curve represents the perfect classifier, meaning no false-positive error happens (Fawcett, 2006). The ROC curve of one simulated result is presented in Figure 4.13 to analyze how those models behave in the different classes. Based on the results, RF, XGBoost, and LightGBM are capable of classifying on class 4 while achieving an AUC around 0.9. However, XBART shows less ability to vote for the right classification on class 4 (AUC = 0.72). For other injury classes, RF, XGBoost, and LightGBM obtain an AUC value ranging from 0.59 to 0.67, which exceeds that of XBART.

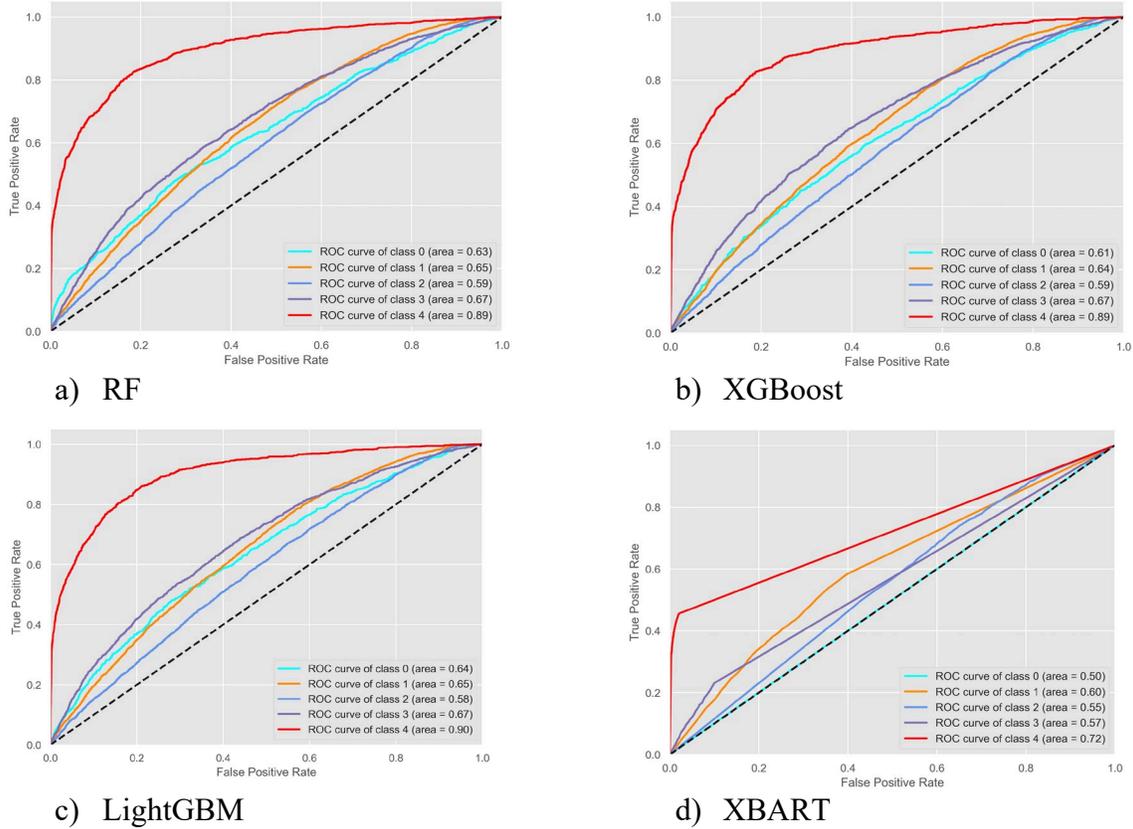


Figure 4.13 ROC Curve of Different Classification Models

4.3.2.2. Marginal Effects

The marginal effects of each variable are analyzed by data imputation to identify the most important factors that contribute to fatal pedestrian injury. First, the data is randomly split for training and testing data. Then, the hyperparameters are optimized on the training data, and models are trained on training data with optimal hyperparameters. Finally, the model is used for prediction on the testing dataset before and after one variable is imputed, and the probability difference between two times of prediction is defined as the marginal effect for that imputed variable. This process is repeated ten times, and the results are summarized in Figure 4.14.

The variables are classified into ten categories. In terms of driver and pedestrian characteristics, driver age seems to have both positive and negative marginal effects. This can be related to the impact of driver age observed in some literature, with the involvement of younger drivers increasing the risk of high severity as compared to the presence of middle-aged drivers (Kim et al., 2010; Pour-Rouholamin and Zhou, 2016). One study found that drivers aged 65 and older also increase the risk for pedestrian injury severity (Mohamed et al., 2013). However, some researchers have found that this is not always the case, as older drivers may also be more experienced (Wood et al., 2014). Furthermore, a high value for pedestrian age has a high likelihood of increasing injury severity, which might be due to the greater physical vulnerability of older people.

Several human violation variables were tested to analyze the effect of pedestrian and driver intoxication. Figure 4.14b shows that the intoxicated pedestrian variable is the most important factor that contributes to fatal pedestrian injury among those variables. The probability of a pedestrian fatality shows significant positive changes after data imputation, indicating that pedestrian intoxication greatly increases the risks of pedestrian death in a crash. When pedestrian intoxication is imputed, the probability of pedestrian death increases by 15% on average in RF, XGBoost, and LightGBM models, and XBART shows a probability change as great as 30%. Driver intoxication also leads to an increase in pedestrian death probability change. Intoxication is more likely to cause pedestrian fatality in a crash, and this result is also supported by the CRIS dataset, where intoxication is involved in 38% of fatal crashes. The simulation result is consistent with the previous report that intoxication has the strongest effect on pedestrian death (Rahman, M. and Kockelman, 2020). Hit-and-run accidents are also related to a high pedestrian severity level. In the CRIS data, 19% of the pedestrian deaths are hit-and-run cases, highlighting the relevance of this variable. The high severity is largely due to the time delay incurred when the driver leaves the crash location, which delays emergency services and prompt attention to the pedestrian.

Speed limit contributes significantly to the probability of pedestrian death, and the marginal effect ranges from 2% to 6% among these classification models. As expected, roads with higher speed limits led to higher pedestrian injuries, consistent with previous studies (Chen and Fan, 2019). However, crash frequency is reduced as speed limit increases, as found in the analysis of the previous section. Approximately 21% of the crashes were located at intersections. The results indicate that those crashes have a lesser risk of pedestrian fatalities than mid-segment crashes, likely due to reduced speed at these locations. Traffic control, including traffic signs and traffic signals, can help to reduce the probability of a crash and thus pedestrian death. Results suggest that traffic control is predicted to reduce pedestrian death probability on average by 3% in LightGBM, 2% in XGBoost, 1% in XBART, and 0.5% in RF. Also, data imputation on traffic signals decreases the pedestrian death probability by as much as 2% on average in LightGBM.

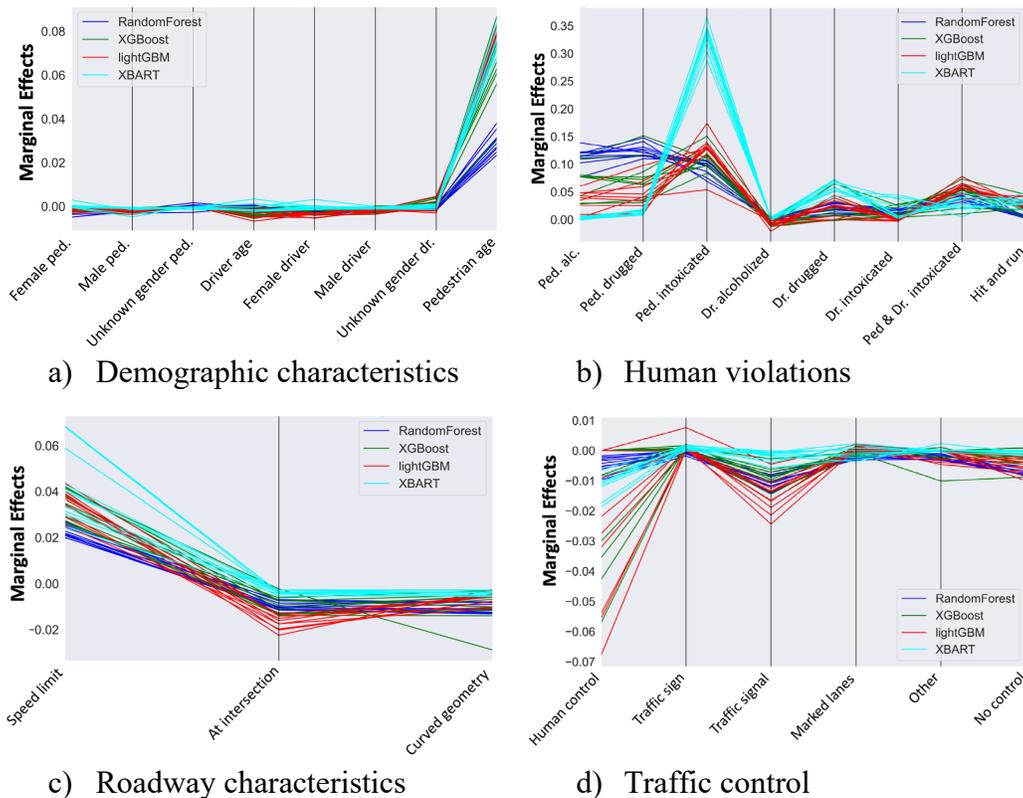
Roadway functional classification is also an important factor. Different road types (such as country roads, city streets, and interstate roads) also play distinct roles in fatal pedestrian injury. For example, city streets and non-trafficways help reduce the marginal effect in RF, XGBoost, and LightGBM models, while the interstate seems to increase the marginal impact on all models. In Texas, interstate highways account for 6% of pedestrian crashes but 21% of pedestrian fatalities. This outcome is likely related to the speed of the crash. As analyzed previously, high-speed roadway segments tend to have fewer pedestrian crashes, but the severity is higher due to the speed of impact. The crash location analysis seems to indicate that crashes occurring on the roadway shoulder have a higher risk of causing pedestrian fatalities compared to crashes on the roadway and in the median area.

The area type is also an important factor for injury severity. The results suggest that rural and small urban areas present a higher risk for pedestrian fatalities. Factors such as distance to hospitals and speed limits can influence this finding. Rural and small urban areas tend to be less dense, and the

emergency response time is higher than in urban areas. However, results from the previous section indicate that pedestrian activity is lower in these areas compared to large urban areas.

In terms of vehicle types, research in the field indicates that high injury severity is associated with light-duty vehicles, such as SUV/CUV, pickup trucks, and vans, due to the heavy mass involved in the collision (Anarkooli et al., 2017; Liu et al., 2019; Pour-Rouholamin and Zhou, 2016; Rahman, M. and Kockelman, 2020). However, this study shows that trucks involved in a crash are more likely to cause the death of pedestrians, but the effects of vans and SUV/CUVs are not significant. Buses also have a significant effect, but it is important to mention that the number of crashes involving buses is low compared to other vehicle body types.

Environmental factors such as crash time and lighting condition strongly affect the pedestrian injury severity (Aziz et al., 2013; Pour-Rouholamin and Zhou, 2016). The time of day is found to influence its marginal effect. Specifically, the period after 8 PM and before 7 AM (under dark conditions) has a positive effect on pedestrian deaths. Approximately 80% of pedestrian deaths occurred at this time. In contrast, in the daytime, the probability of pedestrian death is reduced in all models. Similarly, the brighter daytime conditions significantly help to lower the likelihood of pedestrian death in all models, but not the other light types. This finding highlights the importance of streetlight improvements to reduce pedestrian crashes.



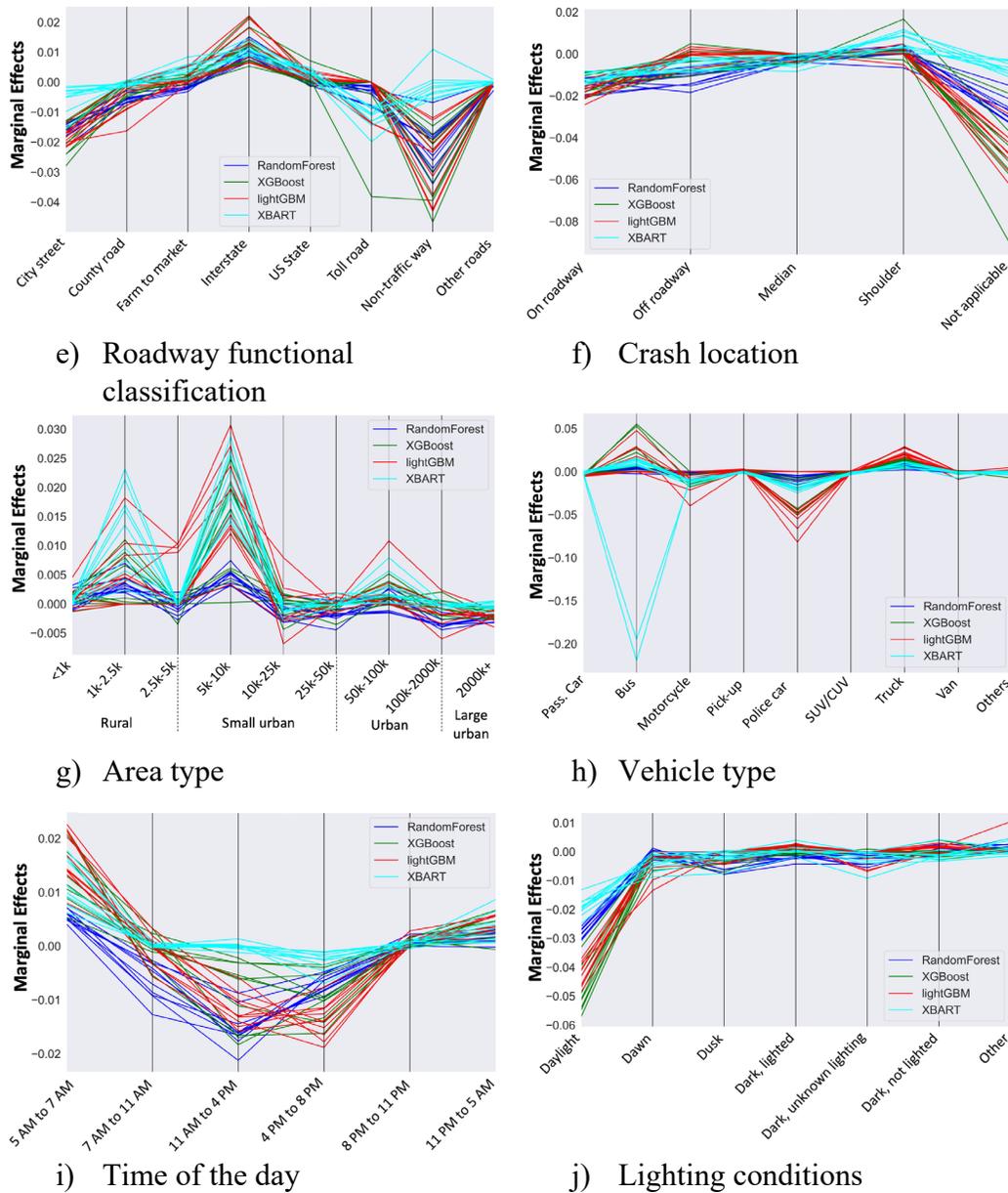


Figure 4.14 Parallel Coordinate Plots of Marginal Effects for Fatal Crashes

4.3.3. Summary

In addition to the decision tree analysis, tree-based ML methods were applied to provide robust analysis and improve the accuracy of the estimates. Tree-based ML models are popular methods for making predictions. Ensemble tree models implementing bagging or boosting approaches usually outperform traditional, statistically based prediction models due to the informative and deliverable prediction. Methods such as RF, gradient boosting (Light GBM and XGBoost), and BART were applied and compared. This modeling included pedestrian severity classification and pedestrian crash counts (total counts and fatal counts). The pedestrian severity models developed

using tree-based models offered an improvement in the precision metric, with values ranging between 43% and 53%.

4.4. Conclusions

This chapter develops decision trees to classify pedestrian crash severity using a recursive partition algorithm. Furthermore, different tree-based models are developed to describe pedestrian crash counts and severity levels. Although the two approaches yield similar findings, decision trees provide intuitive and easy-to-explain results while tree-based models attain greater accuracy. The graphical representation of the decision tree models facilitates understanding. However, decision trees present a low accuracy and high variance. Tree-based models using ML algorithms provide more robust results. But, due to the lack of transparency, these methods are more difficult to interpret.

The main results suggest that pedestrian characteristics and highway design characteristics are the most significant variables influencing pedestrian crash severity. The most relevant variable to determine severity levels was pedestrian intoxication, with almost all the crashes involving an intoxicated pedestrian resulting in fatalities (although these represent only 3% of the sample). In terms of highway design, the speed limit significantly influenced the crash severity, with findings indicating that high-speed roadways increase the risk of pedestrian fatalities. Results also indicate that pedestrian crash frequencies are lower in these locations. Lighting conditions were also relevant, with more fatal crashes occurring in dark conditions. Other factors such as vehicle type, crash location, and traffic control type were also analyzed in the models. In terms of pedestrian crash counts, VMT was the most significant variable correlated to pedestrian crash frequencies. Other factors, such as the number of transit stops, the distance to the nearest school, and the distance to the nearest hospital, offer practical significance; more crashes occur near transit stops and schools. Results also indicate that highway design variables such as on-system roads (or state-maintained arterials), number of lanes, curve angle, curvature indicator, and curvature length have a significant positive impact on pedestrian crash frequencies.

This study also showed a comparison across the tree-based models analyzed. In the pedestrian crash occurrence prediction, the principal results showed that all four models perform similarly, with close root mean square error (RMSE) and R-square for total crash occurrence. Still, LightGBM exceeds the other three models in terms of computational efficiency. For fatal crash occurrence, LightGBM and RF have comparable performance. However, XGBoost and XBART showed significantly lower goodness of fit values. Also, XBART is more sensitive to imbalanced data than are the other models. In the injury severity prediction, RF, XGBoost, and LightGBM achieved similar goodness of fit performance, evaluated by the metrics' accuracy, precision, recall, F1, and geometric mean. XBART obtained a higher precision value, but the other metrics were lower, with a significantly high computational time.

Findings from this study underscore the importance of campaigns against driving and walking while intoxicated, installation of streetlights in pedestrian-active areas, improved roadway design, and enforcement of safety countermeasures in areas where pedestrians are more vulnerable (such as near bus stops and schools). It also highlights the importance of detailed police reports to develop analyses of this type that can be used to improve pedestrian safety.

Chapter 5. Evaluate Value and Cost-Effectiveness of Treatments

This chapter provides an overview of the benefit-cost ratios (BCRs) of certain suggested treatments as applied to corridors and intersections throughout Texas. Since pedestrian crashes have been increasing in the state over the last 10 years (TxDOT, 2021), it is important to understand, at a micro level, where these crashes are occurring, and which specific corridors are experiencing disproportionate rates of pedestrian crashes and fatalities.

Using a novel clustering methodology, this chapter describes several of the highest-crash segments and intersections in the state, along with benefit-cost calculation methods. This chapter is arranged as follows: an overview of cost and benefit calculation models, along with definitions of terms used throughout this chapter. An overview of the corridor ranking and creation methodology is provided next, which creates a list of ranked corridors by total comprehensive costs for pedestrian crashes and fatalities. A comprehensive breakdown of estimated costs and crash modification factor (CMF) estimates follows. Lastly, the research team presents an overview of the 10 highest-crash corridors as provided by the clustering model, with BCRs for selected treatments.

5.1. Segment Ranking and Creation Methodology

To perform an effective BCR analysis, the research team needed to generate a corridor-based representation of the worst segments and intersections. The solution was to create groupings of combinations of intersections and 0.1-mile segments along individual high-ranking roadways, using the KABCO scores. This score was developed by the National Safety Council (NSC) and is frequently used by law enforcement for classifying injuries: K – Fatal, A – Incapacitating injury, B – Non-incapacitating injury, C – Possible injury, and O – No injury. (See Section 1.3.3 for an introduction).

5.1.1. Dataset Preparation

To prepare for building up representative corridors, the entire 2018 version of the TxDOT Roadway Inventory was resampled at 0.1-mile increments, producing about 3.4 million segments available for analysis. This resampling is available online (Perrine and Zuniga-Garcia, 2021). Along with this, crash records from the Texas Crash Records Inventory System (CRIS) were matched with each 0.1-mile incremental segment according to these criteria:

- The crash record has geographic coordinates. (About 83% of all crash records satisfy this criterion.)

- The record is not marked to have occurred at an intersection. (About 72% of all records meet this criterion.)
- The record is classified as a pedestrian-related crash. (About 1.4% of all crash records are determined to be predominantly pedestrian-related as explained in Section 2.3.)
- The record’s geographic coordinates sit within 329 feet (100 meters) of a TxDOT Roadway Inventory segment.
- Once a crash is matched with the nearest segment, it is not eligible to be matched to any other segments.

The end result was that 41,131 crash records (0.7% of all of CRIS within the 2010–2019 analysis period) were matched to roadway segments for the purpose of finding and ranking crash-prone corridors for BCR analysis.

In a related effort, intersections were derived from OpenStreetMap and matched to corresponding locations within the TxDOT Roadway Inventory; a dataset is available online (Perrine and Zuniga-Garcia, 2021). (The TxDOT Roadway Inventory on its own does not explicitly represent intersections.) Similar to the process used on segments, CRIS crash records were matched to these intersections according to these criteria:

- The crash record has geographic coordinates, is marked to occur at an intersection, and is classified as a pedestrian-related crash.
- The record’s geographic coordinates sit within 329 feet (100 meters) of a derived OpenStreetMap intersection.
- Once a crash is matched with the nearest intersection, it is not eligible to be matched to any other intersection.

This produced 16,502 crash records available for the purpose of finding and ranking crash-prone corridors.

5.1.2. Finding Corridors

These 0.1-mile segments, OpenStreetMap-derived intersections, and selected crash records were then used as inputs to generate analysis corridors. The algorithm for performing this followed a “greedy” pattern of picking up the worst intersections first and building off of them:

- Pick the next worst intersection in terms of the number of pedestrian-related crashes that are matched with it.
- For each cross street, “walk” down each direction of eligible 0.1-mile segments from the starting intersection until:

- o 3 successive segments and intersections that coincide with them each have fewer than 5 pedestrian-related crashes, or
- o The end of the street is reached.
- Record all segments and intersections traversed as a new corridor. Include the cumulative KABCO score for all of the pedestrian crashes therein and make those segments ineligible for inclusion in future corridors.
- Loop again until no more intersection/cross street combinations remain.
- Rank all the corridors according to decreasing KABCO score.

At the completion of this algorithm, 7,945 corridors were discovered. However, corridors beyond the 500 highest scoring are considered insignificant for this study, as most are composed of a small handful of intersections. In looking at just the top 100, 1,274 intersections and 1,116 0.1-mile segments comprise the 100 corridors, encompassing 4,295 crash records.

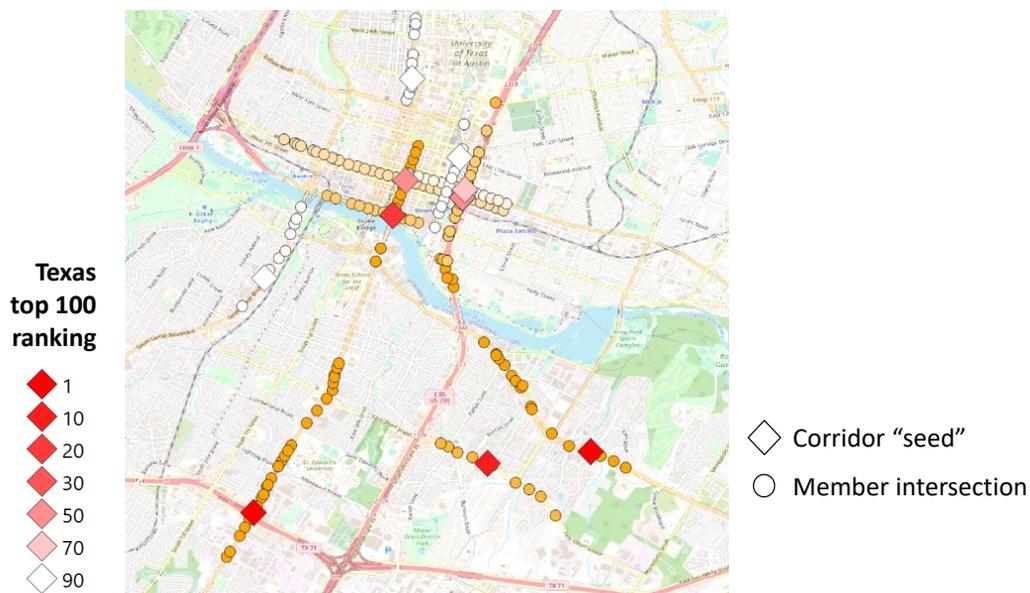


Figure 5.1 Corridors Among the Texas “top 100” That Exist within the Central Austin Area

5.2. Crash Count Estimation Model

CRIS contains records from police crash reports across Texas’ 254 counties and 268,597 square miles (Texas Department of Transportation, 2020). Crash variables include time and location, persons and vehicles involved, injury severities, and road conditions. Many crashes are never reported to police or are not flagged for CRIS inclusion. These are typically property-damage-only or no-injury crashes, but drivers and pedestrians will leave the scene for other reasons as well. And

police who deem a reported crash to be worth less than the \$1,000 minimum crash cost threshold (for recording purposes) often do not record the crash formally.

5.2.1. Negative Binomial Count Model

A negative binomial (NB) count model was used for pedestrian crash counts. The expected number of pedestrian crash counts $E(Y_i)$ along the i th intersection or mid-block segment is expressed as follows:

$$E(Y_i) = \exp(\beta_0 + \sum_k x_{ik}\beta_k + \varepsilon_i) \quad (1)$$

where β_k is the k^{th} covariate, ε_i is a random error term that follows a Gamma distribution $\varepsilon_i \sim \text{Gamma}(\gamma, \gamma)$, Y_i represents the average daily e-scooter trip count with mean $E(Y_i) = \mu_i$ and variance $\text{Var}(Y_i) = \mu_i + \rho\mu_i^2$, and ρ is the dispersion parameter ($\rho = 0$ for a Poisson model).

5.2.2. Model Results

The results from the NB model are summarized in Table 5.1 (Texas) and Table 5.2 (Austin). The dispersion parameter (ρ) of the two models is greater than one, indicating that the data is over-dispersed, and an NB model is preferred over a Poisson model.

The Texas model (Table 5.1) shows a positive correlation between the walk-miles traveled (WMT) and the number of pedestrian crashes across intersections and mid-block-segment models, likely due to increased exposure levels. However, previous research also found that the relationship between crash exposure and crash rates is non-linear. It has rates falling off dramatically as walk levels rise, presumably due to drivers expecting more pedestrians in high-WMT zones and safer pedestrian environments that encourage walking (Wang and Kockelman, 2013).

Table 5.1 Estimation Results of NB for Pedestrian Crashes, Texas

	Intersections			Mid-block Segments		
	Coeff.	Std. Error	P-value	Coeff.	Std. Error	P-value
(Intercept)	-8.694	0.216	0.000	-8.035	0.098	0.000
WMT per pop dens. (log)	0.335	0.013	0.000	0.305	0.007	0.000
Signalized intersection (ind.)	1.426	0.032	0.000			
Number of approaches Intersections crossed	0.398	0.019	0.000	0.093	0.002	0.000
DVM (log) [major]	0.195	0.008	0.000	0.522	0.006	0.000
Speed limit (mph) [major]	-0.020	0.002	0.000	-0.013	0.001	0.000
Number of lanes [major]	0.132	0.012	0.000	0.217	0.010	0.000
Lane width (ft) [major]	0.033	0.004	0.000	0.041	0.003	0.000
Median width (ft) [major]	-0.006	0.001	0.000	-0.014	0.001	0.000
One-way road (ind.) [major]	0.095	0.052	0.068	-0.906	0.048	0.000
DVM (log) [minor]	0.136	0.008	0.000			
Speed limit (mph) [minor]	-0.021	0.002	0.000			
Number of lanes [minor]	-0.004	0.018	0.842			
Lane width (ft) [minor]	0.040	0.005	0.000			
Median width (ft) [minor]	-0.027	0.005	0.000			
One-way road (ind.) [minor]	-0.211	0.063	0.000			
AADT per lane [major]	1.76E-05	4.53E-06	0.000	-7.67E-05	4.19E-06	0.000
Truck percentage [major]	0.020	0.003	0.000	0.003	0.002	0.100
Arterial (ind.) [major]	0.444	0.037	0.000	0.198	0.028	0.000
On system roadway (ind.)	-0.230	0.036	0.000	0.209	0.028	0.000
Rural (ind.)	-0.107	0.087	0.218	-0.339	0.041	0.000
Small urban (ind.)	-0.108	0.055	0.050	0.049	0.034	0.154
Large urbanized (ind.)	0.171	0.037	0.000	0.170	0.025	0.000
Distance to nearest hospital (mi)	-0.023	0.006	0.000	-0.009	0.003	0.002
Transit stops (ind.)	0.525	0.047	0.000	0.526	0.033	0.000
Number of stops	0.042	0.008	0.000	0.049	0.004	0.000
City of Austin (ind.)	0.327	0.047	0.000	-0.392	0.042	0.000
No. of observations	699,954			574,910		
Dispersion Parameter (ρ):	0.393			0.575		
McFadden's R2:	0.483			0.543		
Likelihood ratio test (χ²)	32,515			62,980		
Prob > χ²	0.000			0.000		
2 x log-likelihood	-86,105			-161,539		

Table 5.2 Estimation Results of NB for Pedestrian Crashes, City of Austin

	Intersections			Mid-block Segments		
	Coeff.	Std. Error	P-value	Coeff.	Std. Error	P-value
(Intercept)	0.360	0.061	0.000	-5.098	0.473	0.000
WMT per pop. dens. (log)	1.671	0.103	0.000	0.068	0.043	0.114
Signalized intersection (ind.)	0.123	0.067	0.067			
Number of approaches Intersections crossed				0.001	0.012	0.904
DVM (log) [major]	0.166	0.028	0.000	0.245	0.024	0.000
Speed limit (mph) [major]	-0.007	0.006	0.248	-0.036	0.004	0.000
Number of lanes [major]	0.375	0.046	0.000	0.411	0.033	0.000
Lane width (ft) [major]	0.030	0.011	0.009	0.081	0.010	0.000
Median width (ft) [major]	0.001	0.002	0.726	-0.017	0.004	0.000
One-way road (ind.) [major]	0.159	0.175	0.735	-1.437	0.160	0.000
DVM (log) [minor]	0.145	0.029	0.000			
Speed limit (mph) [minor]	-0.012	0.009	0.177			
Number of lanes [minor]	0.057	0.063	0.359			
Lane width (ft) [minor]	0.036	0.013	0.006			
Median width (ft) [minor]	-0.049	0.020	0.013			
One-way road (ind.) [minor]	-0.458	0.182	0.012			
AADT per lane [major]	4.45E-05	1.17E-05	0.000	2.02E-05	9.70E-06	0.038
Truck percentage [major]	-0.019	0.037	0.610	-0.049	0.029	0.084
Arterial (ind.) [major]	0.229	0.141	0.105	-0.167	0.085	0.049
On system roadway (ind.)	-0.231	0.131	0.077	-0.010	0.108	0.923
Distance to nearest hospital (mi)	0.089	0.046	0.050	0.035	0.031	0.252
Transit stops (ind.)	0.378	0.116	0.001	0.647	0.091	0.000
Number of stops	0.028	0.016	0.070	0.013	0.012	0.275
Population density (sq mi)	2.11E-05	7.41E-06	0.005	5.30E-05	6.30E-06	0.000
Employment density (sq mi)	-1.59E-06	1.61E-06	0.324	3.33E-05	6.82E-06	0.000
Median income (\$10k)	-0.099	0.015	0.000	-0.119	0.011	0.000
CBD (ind.)	0.738	0.182	0.000	1.453	0.157	0.000
No. of observations	19,194			41,107		
Dispersion Parameter (ρ):	0.821			0.430		
McFadden's R2:	0.616			0.370		
Likelihood ratio test (χ2)	2,808			2,718		
Prob > χ2	0.000			0.000		
2 x log-likelihood	-5,618			-10,567		

The Austin-specific model in Table 5.2 shows an intersection model that is less sensitive to WMT and signalized intersections than the Texas model but still shows a significant value. It is likely that pedestrian crashes in non-signalized intersections are more frequent in this area compared to the state of Texas. The number of approaches also shows a positive coefficient. In terms of mid-block segments, the number of intersections crossed is not significant, possibly due to the size of the segments. In this case, the segments are 0.1-mile long, and the number of intersections crossed is significantly lower than the case where 1-mile segments were used in the Texas model.

5.3. Treatments by Category

These treatments were taken from a variety of sources across the internet and highway safety manuals, including the Crash Modification Factor (CMF) Clearinghouse, and the report by UNC Highway Safety Research Center by Bushell et al. (2013). The highlighted cells are based on estimates for these treatments, as there is either too small of a sample size to obtain a high or low CMF, or there is no applicable data available for that treatment. In this case, estimates were made based on similar treatments or other studies that reported results but did not provide a CMF. For some treatments, enough data is available to give an average CMF, but not enough to provide a high or low number.

The treatment list is broken up into seven categories; these are arranged based on general purpose of the treatment, as well as which roadway users are primarily affected. Traffic calming is included as it has a special role in determining speed. Following are the seven categories:

- Basic roadway treatments
- Roadway treatments – traffic calming
- Pedestrian-specific infrastructure
- Street furniture
- New sidewalks
- Education
- Homelessness-centric treatments, direct outreach to pedestrians

Treatments in Table 5.3 primarily involve adding treatments to the roadway that do not affect the material roadway conditions for drivers. For the most part, these treatments include enhanced signage, attention-getting measures such as the rectangular red flashing beacon and pedestrian-hybrid beacons. These treatments can typically be applied at the corridor level along corridors experiencing high rates of crashes.

Treatments in Table 5.4 primarily relate to traffic calming measures, which reduce vehicle speed. As speed is a major factor in pedestrian crashes that involve fatalities (Tefft, 2013; Bernhardt & Kockelman, 2021), these treatments seek to limit the impact of speed and narrow the roadway so that pedestrians have less exposure time when crossing. While the cost of implementing these treatments varies widely, from roadway reconfiguration to simple signage, their impacts can be significant when implemented in high pedestrian traffic areas. Examples of methods to reduce speed include speed table, center rumble strips, and zigzag pavement marking (Boodlal et al., 2015), and 1 or a combination of 2 or more methods can be used to reduce the vehicle's speed.

The addition of infrastructure, as Table 5.5, that tends to pedestrian needs ranges from signage to barriers and signal improvements. A few of these treatments can limit pedestrian contact with vehicles altogether, such as pedestrian bridges, but these are typically very high cost. Additionally, traffic signals can help provide a controlled crossing at an intersection where a treatment such as a pedestrian-hybrid beacon would not be appropriate, and some of these treatments also have crossover safety improvements with drivers. Treatments such as signal re-timings, leading intervals, and scramble intervals can increase driver delays, but lead to positive outcomes in pedestrian safety.

Street furniture, as Table 5.6, is another potential option that can help both with traffic calming and provide additional services to pedestrians. As the studies on the crash reduction effects are limited, these estimates are provided by Bushell et al. (2013). The presence of street furniture can also communicate to drivers that they are entering a crowded area, or one with high pedestrian activity, and in response drivers are more likely to reduce their speed, improving pedestrian safety outcomes (Bushell et al., 2013). For instance, although Table 3.4 implied that the existence of transit stops may increase pedestrian crashes due to high pedestrian activity near that transit stop, installing street furniture (e.g., bus shelter) can contribute to improving pedestrian's safety and can help to get the driver's attention.

New sidewalks, as Table 5.7, are among the most basic treatments available to improve pedestrian safety in areas where they currently do not exist. While grade separation can be costly, even providing a basic sidewalk can lead to reductions in crashes by 75% or more (CMF Clearinghouse, 2021).

The efficacy and cost-effectiveness of education programs, as Table 5.8, is disputed (Arellano, 2021; Bachman et al., 2015), but they can be an option when implemented alongside roadway safety improvements. Specifically, increased traffic law enforcement alongside treatments, such as prohibiting right turns on red or lowering the speed limit through certain corridors, may be an effective countermeasure to bring lasting improvements in pedestrian safety.

Table 5.3 Basic Roadway Treatments

	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Avg. CMF	Hi/Lo CMF
Basic curb and gutter ⁴	108	\$20	\$21	\$1.05/\$120	Linear foot	0.89	
“Daylighting” left turns & crossing locations ⁴	2	\$300	\$300	\$50/\$250	Each	0.75	0.52/1.49
Gateway signage (see examples) ⁴	6	\$15,350	\$22,750	\$5,000/\$64,330	Sign + structure (each)	0.83	0.68/0.98
Reduced curb radii ⁴	12		\$32,500	\$15,000/\$40,000	Per corner	0.81	
Pedestrian-hybrid beacons ⁴	9	\$51,460	\$57,560	\$21,440/\$128,660	Each	0.71	0.63/0.84
Prohibition of left turns ⁴	6		\$800		Per sign	0.28	0.23/0.36
Prohibition of right turn on red ⁴	4		\$800		Per sign	0.77	0.70/0.97
Crosswalk (hi-vis; see citation for specs) ⁴	4	\$3,070	\$2,540	\$600/\$5,710	Each	0.63	
Raised crosswalk ⁴	6		\$18,995	\$7,110/\$30,080	Each	0.64	0.55/0.7
Flashing beacon ⁵	25	\$5,170	\$10,010	\$360/\$59,100	Each	0.85	
Rectangular red flashing beacon ⁵	4	\$14,160	\$22,250	\$4,520/\$52,310	Each	0.53	
Raised median (controlled) ⁶	9	\$22,500		\$15,000/\$30,000	100 ft.	0.6	0.33/0.75
Raised center medians (uncontrolled) ⁴	30	\$6	\$7.26	\$1.86/\$44	Square foot	0.93	0.61/1.94
Freeway fencing (both sides) ⁴			\$25	\$1/\$100	Linear foot	0.63	0.10/0.87
Advanced stop/yield sign ⁷		\$520	\$570	\$100/\$1150	Each	0.75	
Install crosswalk sign ⁷	23	\$520	\$570	\$100/\$1150	Each	0.91	0.86/0.95
Narrow roadway from 4 lanes to 3 lanes ⁸			\$20,000	\$12,500/\$50,000	Per mile	0.71	

⁴ CMF Clearinghouse, 2021

⁵ Bushell et al., 2013

⁶ FHWA, 2018b

⁷ CMF Clearinghouse, 2021

⁸ FHWA, 2018b

Table 5.4 Traffic Calming Treatments

Treatment	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Average CMF	Hi/Lo CMF
Speed humps ⁷	14	\$2,130	\$2,640	\$690/\$6860	Each	0.64	0.73/0.55
Speed limit reductions - 15% decrease ⁷			\$135		Each (sign)	0.89	0.83/0.95
Speed limit reductions - 10% decrease ⁷			\$135		Each (sign)	0.79	0.68/0.9
Speed limit reductions - 5% decrease ⁷			\$135		Each (sign)	0.705	0.56/0.85
Chicanes ⁷	9	\$8,050	\$9,960	\$2140/\$25,730	Each	0.69	0.64/0.75
Diverters ⁷	6	\$22,790	\$26,040	\$10,000/\$51,460	Each	0.69	0.64/0.75
Curb extensions (bulb-outs) ⁷		\$10,150	\$13,000	\$1070/\$41,170	Each	0.75	0.51/1.07
Traffic circle ⁷	14	\$27,190	\$85,370	\$5,000/\$523,080	Each	0.75	0.51/1.07
Road diet ⁹	10		\$40,000	\$25,000/\$100,000	Per mile	0.71	
Hardened left turns ¹⁰	20	\$2,500	\$2,500.00	\$2000/\$3000	Each	0.65	

Table 5.5 Pedestrian-specific Treatments

Treatment	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Average CMF	Hi/Lo CMF
Streetlight ¹¹	17	\$3,600	\$4,880	\$310/\$13,900	Each	0.44	0.19/0.69
In-pavement lighting (flashing crosswalks) ¹¹	4	\$18,250	\$17,260	\$6,480/\$40,000	Complete system	0.71	
Pedestrian leading intervals ¹¹	4	\$1,750	\$1,750	\$0/\$3500		0.85	0.71/1.48
Crosswalk signage (for road users) ¹²		\$30	\$30	\$25/\$35	Square foot	0.84	0.75/0.88
Bollards (at crossing points) ¹¹	42	\$650	\$730	\$62/\$4,130	Each	0.93	

⁹ Fitzpatrick et al., 2014

¹⁰ <https://www.autoblog.com/2020/04/12/iihs-left-turn-pedestrian/>

¹¹ CMF Clearinghouse, 2021

¹² <http://www.trafficsign.us/signcost.html>

Treatment	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Average CMF	Hi/Lo CMF
Curb ramps (to crossings) ¹¹	74	\$740	\$810	\$89/\$3,600	Each	0.95	
Pedestrian refuge islands ¹¹	15	\$9.80	\$10	\$2.28/\$26	Square foot	0.44	0.25/0.76
Fence (general purpose) ¹¹	7	\$120.00	\$130	\$17/\$370	Linear foot	0.63	0.10/0.87
Pedestrian overpass (wooden) ¹³	8	\$122,610.00	\$124,670	\$91,010/\$165,710	Each	0.63	0.10/0.87
Pedestrian overpass (steel) ^{11, 13}	5	\$191,400	\$206,290	\$41,580/\$653,840	Each	0.63	0.10/0.87
Pedestrian underpasses ^{11, 13}		\$120			Square foot	0.63	0.10/0.87
Sidewalk railings ¹¹	33	\$95	\$100	\$7.20/\$690	Linear foot	0.83	0.52/1.18
Access management improvements (esp. at commercial centers) ¹⁴	3	\$4,000	\$4,000	\$3000/\$5000	Per driveway removed	0.5	
Full street closure (one city block) ¹¹				\$500/\$120,000		0.05	
Partial street closure (depends on treatment) ¹¹		\$37,500		\$10,290/\$41,170		0.71	
Ped detection - detector (actuate) ¹⁵	14	\$180	\$390	\$68/\$1330	Each	0.55	
Ped detection - push button ¹⁵	34	\$230	\$350	\$61/\$2510	Each	0.83	
Audible pedestrian signal ¹⁵	4	\$810	\$800	\$550/\$990	Each	0.72	
Increase crossing time ¹⁵	10	\$1,750			Per re-timing	0.49	
Countdown timers ¹⁵	18	\$600	\$740	\$190/\$1930	Each	0.48	0.3/0.75
Pedestrian signal (complete) ¹⁵	70	\$2,680	\$3,260	\$850/\$13,410	Each	0.6	0.45/0.85
Traffic signal (new) ¹⁶	25		\$90,000	\$80,000/\$100,000	Each	0.44	0.5/1.48
Dedicated pedestrian interval ¹⁵	4		\$1,750	\$0/\$3500	Per re-timing	0.41	0.16/0.49
Speed trailers ¹⁵	6	\$9,480	\$9,510	\$7000/\$12,410	Each	0.95	0.93/0.95

¹³ Fitzpatrick et al., 2014

¹⁴ <https://mobility.tamu.edu/mip/strategies-pdfs/system-modification/technical-summary/Access-Management-4-Pg.pdf>

¹⁵ CMF Clearinghouse, 2021

¹⁶ https://ftp.txdot.gov/pub/txdot-info/pio/casbrochures/pub_signals.pdf

Table 5.6 Street Furniture Treatments

Treatment	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Average CMF	Hi/Lo CMF
Street trees ¹⁵	7	\$460	\$430	\$54/\$940	Each	0.82	
Bench ¹⁵	17	\$1,660	\$1,550	\$220/\$5750	Each	0.82	
Bus shelter ¹⁵	4	\$11,490	\$11,560	\$5,230/\$41,850	Each	0.82	
Trash/recycling receptacle ¹⁵	13	\$1,330	\$1,420	\$310/\$3,220	Each	0.82	

Table 5.7 New Sidewalk Treatments

Treatment	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Average CMF	Hi/Lo CMF
Widen paved shoulder ^{15, 17}	4	\$5.81	\$5.56	\$2.96/\$7.65	Square foot	0.72	0.54/1.01
Asphalt sidewalk ¹⁵	11	\$16.00	\$35.00	\$6.02/\$150	Linear foot	0.26	
Concrete sidewalk ¹⁸	164	\$27	\$32	\$2.09/\$410	Linear foot	0.26	
Concrete sidewalk w/curb ¹⁸	7	\$170	\$150	\$23/\$230	Linear foot	0.26	
Multi-use trail - paved ^{18, 19}	42	\$261,000	\$481,140	\$64,470/\$4,228,520	Mile	0.14	
Multi-use trail - unpaved ^{18, 19}	7	\$83,870	\$121,390	\$29,520/\$412,720	Mile	0.14	

¹⁷ Fitzpatrick, et al., 2014

¹⁸ CMF Clearinghouse, 2021

¹⁹ Fitzpatrick, et al., 2014

Table 5.8 Education Treatments

Treatment	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Average CMF	Hi/Lo CMF
“Be safe, be seen” ²⁰	1			\$18,000	Campaign implementation	0.93	
Primary school training from local police department ²¹	1			\$18,000	Campaign implementation	0.9	
Out-of-home advertising campaigns ²²	1			\$18,000	Campaign implementation	0.93	
Anti-distracted driving campaign ²²	1			\$18,000	Campaign implementation	0.93	
Increased traffic law enforcement ²³	1			\$18,000	Campaign implementation	0.77	0.60/1.28
Safe Routes to School - educational programs ²⁴	5		\$10,298		Curriculum implementation	0.93	

Table 5.9 Homelessness-centric Treatments, Direct Outreach to Pedestrians

Treatment	N =	Cost (median)	Cost (average)	Cost (min/max)	Cost Unit (i.e., linear foot)	Average CMF	Hi/Lo CMF
Hi-vis vests ²⁵	40	\$12	\$10	\$4/\$50	Each	0.85	
Tiny housing to decrease freeway camps ²⁶	5	\$60,000	\$45,000	\$7500/\$150,000	Each	0.9	
Lights for pedestrians ²⁷	10	\$20	\$40	\$7.50/\$60	Lights + implementation	0.79	
Flags for pedestrian crossings ²⁸	3	50 cents (unit)	500 (total)	\$50/\$18,000	Total program cost	0.9	

²⁰ Arellano, 2021

²¹ Bachman et al., 2015

²² Cantulupo, 2021

²³ FHWA, 2018b

²⁴ Muennig et al., 2014

²⁵ <https://www.homedepot.com/b/Safety-Equipment-Safety-Vests/N-5yc1vZc29h>

²⁶ Nowacki, 2021

²⁷ Madsen et al., 2013

²⁸ Davis, 2014

In Table 5.9, since persons experiencing homelessness are often present in freeway right-of-way, as well as along arterial roads, they are important to address when addressing pedestrian safety (Bernhardt & Kockelman, 2021). While these treatments have also had mixed success, and “treatments” such as housing or connecting those experiencing homelessness to institutional resources is often preferable (Arellano, 2021), these treatments should be examined in the context of addressing the structural issues that create higher rates of persons experiencing homelessness.

5.4. Cost and Benefit Calculation Models

A benefit-cost ratio (BCR) is simply a measurement of a specific countermeasure’s benefits, measured in the anticipated reduction in crash costs multiplied by the treatment’s CMF, divided by the sum of countermeasure implementation plus the costs from the preceding 10 years at the affected part of the intersection.

Although Texas already uses Safety Improvement Index (SII) for identifying, ranking, and selecting eligible projects by comparing its benefit and cost, it is outdated (developed in 1974), assumes too high annual inflation rate (8%), and requires too many variables (e.g., projected ADT for all future years). Thus, in this project, a rather straightforward approach of relying on benefits coming from crash reduction, delay cost, and construction cost is proposed to estimate a treatment’s benefit-cost ratio.

The BCRs constructed from the BCA and CMFs have a few underlying assumptions. Crash costs were derived using TxDOT’s most recent Highway Safety Improvement Program manual from 2020 (TxDOT, 2020a). In the KABCO system, when adjusted for inflation to 2021, the average comprehensive cost (which includes quality-of-life costs and lost productivity) of a non-incapacitating injurious crash (B) was approximately \$500,000 and the average incapacitating crash injury (A) and average fatal crash cost (K) is around \$3,500,000. Estimated future crash costs for each individual intersection in the corridor were determined through an NB model using CR-3 Texas Peace Officer’s Reports through the CRIS database and determining the severity of the injury and location of the crash. Specific countermeasure recommendations are based on the problem areas within each intersection, defined as one or more instance of a particular event within the same part of the intersection in the next 10-year period. For benefits measuring, CMFs were taken primarily from the CMF Clearinghouse and the UNC Highway Safety Research Center’s 2013 report on CMFs (Bushell et al., 2013). Highway Safety Improvement Program (HSIP) in Texas also provides some CMFs (Texas Department of Transportation, 2021), but HSIP emphasizes physical features including barriers, lighting, signage, and pavement markings. Unfortunately, some benefits are difficult to measure, such as an ensuing mode shift from private cars to walking or other active modes because of improved pedestrian safety and comfort and the resultant improvements in greenhouse gas emissions and air quality. These benefits are left out of the calculations for the purposes of this chapter but should be considered when evaluating BCRs for pedestrian safety. Ranges for CMFs, as calculated, can be found starting at Table 5.3 in this chapter.

The United States Department of Transportation (2021) assumes a cost of \$15 per vehicle delay hour. A cost of \$14.14 per vehicle delay hour was used in this study. Delay costs use relatively conservative delay assumptions across all vehicles in the affected travel lanes, such as the average delay for the entire leading interval or lack of right turn on red (1 seconds and 10 seconds for a cycle, respectively), as well as a 2-second average delay for pedestrian-hybrid beacons. Calculations for each BCR can be found in Section 5.5 of this chapter, as well as a comprehensive list of treatments considered with costs and CMFs.

5.4.1. BCR Calculation Example

Below is an example of a calculated BCR for an intersection in Austin that is one of the 50 most crash-prone in Texas. These BCR calculations are written out in full, showing each step and mirroring the calculation logic used for the large-scale treatment recommendations seen in Section 5.5. Additional detailed calculations can be found in later sections, when looking at Texas' 10 corridors with the highest pedestrian crash rates, as well as in Appendix A.

Congress Avenue and Cesar Chavez Street – Austin

Treatment: Pedestrian Leading Interval (1 second for all crossings)

Cost of installation: \$1,750, CMF: 0.85²⁹

Delay costs = (ADT on Cesar Chavez (28,625) + ADT on Congress (15,785)) × 365 days in the year = 16,202,350 vehicles (162,023,500 in the 10-year period)

Assuming that each vehicle is delayed by 1 second, there will be 162,023,500 seconds of delay overall and, when divided by 3,600 seconds in an hour, 45,007 hours of delay.

Assuming of a cost of \$14.14 per vehicle delay-hour, the total cost of delay is as follows:

$\$14.14 \times 45,007 \text{ delay hours} = \$636,939 + \$1,750 = \$638,689$ in total costs

Benefits:

Costs of 7 non-incapacitating injuries and 1 incapacitating injury in the period 2010–2019 = $(\$500,000 \times 7) + \$3,500,000 = \$7,000,000$ in crash costs

CMF = 0.85

$\$7,000,000 \times 0.15$ (the CRF, or crash reduction factor) = \$1,050,000 in benefits

$\text{BCR} = \$1,050,000 / \$638,689 = 1.65$

$\text{BCR (without delay costs)} = \$1,050,000 / \$1,750 = 600$

²⁹ CMF Clearinghouse, 2021

5.5. Recommendations for Texas' Highest-Crash Corridors

Based on the ranking methodologies described in Section 5.2, following is a comprehensive ranking of the 10 corridors with the highest pedestrian crash frequency and crash costs across Texas, along with recommended treatments. Detailed calculations can be found in Appendix A.

5.5.1. I-35 Southbound Frontage Road – Martin Luther King Jr. Boulevard to Holly Street – Austin

6th Street and SB I-35 Frontage Road - Pedestrian Leading Interval (1 second average delay for all crossings)

Cost of installation: \$1,750, CMF: 0.85³⁰

Delay Costs = ADT on I-35 SB Frontage Road (30,614) + ADT on 6th Street (11,695) × 365 days
= 15,442,420 vehicles × 1 second × 10 years = 154,424,200 seconds of delay over 10 years

154,424,200 vehicle delay-seconds / 3600 seconds in an hour = 42,896 vehicle delay-hours

42,896 × \$14.14 (discounted average over 10 years) = \$606,544

Total Costs = \$606,544 + \$1,750 = \$608,294 in total costs

Benefits:

5 non-incapacitating injuries, 2 incapacitating injuries 2010–2019

(\$500,000 × 5) + (\$3,500,000 × 2) = \$9,500,000

CMF = 0.85; 0.15 (CRF) × \$9,500,000 = \$1,425,000 benefits

BCR = \$1,425,000 / \$608,294 = 2.34

BCR, without delay costs = \$1,425,000 / \$1,750 = 814.29

5.5.2. Tomball Parkway (SH-249) – Sam Houston Tollway (SL-8) to Breen Road – Houston

Tomball Parkway (SH-249) - Speed Limit Reductions - 10% decrease

Cost of installation: \$135 per speed limit sign × 10, CMF: 0.79³⁰

Speed limit from 35 mph to 30 mph results in travel time increasing from 838 seconds to 977 seconds for 43,000-foot segment = 139 seconds lost per vehicle

³⁰ CMF Clearinghouse, 2021

Delay Costs = ADT on SH-249 (27,040) × 139-second delay × 365 days = 1,371,874,400 seconds delay

1,371,874,400 seconds delay / 3,600 seconds in an hour = 381,076 hours delay

381,076 hours × \$14.14 per vehicle delay-hour = \$5,388,417

\$5,388,417 + \$1,350 = \$5,389,767 in total costs

Benefits:

9 non-incapacitating injuries, 11 incapacitating injuries for 2010–2019

$(\$500,000 \times 9) + (\$3,500,000 \times 11) = \$43,000,000$

CMF = 0.79; 0.21 (CRF) × \$43,000,000 = \$9,030,000

BCR = \$9,030,000 / \$5,389,767 = 1.67

BCR, without delay costs = \$9,030,000 / \$1,350 = 6,689

5.5.3. Westheimer Road – Fondren Road to Chimney Rock Road – Houston

Westheimer Road - Speed Limit Reductions - 10% decrease

Cost of installation: \$135 per sign × 10, CMF: 0.79³¹

Speed limit from 35 mph to 30 mph results in travel time from 292 s to 340 s for 15,000 ft segment.

48 second loss per vehicle

Delay Costs = ADT on Westheimer Road (15,211) × 48 second delay × 365 days = 266,496,720 seconds delay

266,496,720 seconds delay / 3600 seconds in an hour = 74,026 hours delay

74,026 hours × \$14.14 per vehicle delay-hour = \$1,046,727

\$ 1,046,727 + \$1,350 = \$1,048,077 in total costs

Benefits:

10 non-incapacitating injuries, 2 incapacitating injuries for 2010–2019

$(\$500,000 \times 10) + (\$3,500,000 \times 2) = \$12,000,000$

³¹ CMF Clearinghouse, 2021

$$\text{CMF} = 0.79 \rightarrow \$ 2,520,000$$

$$\text{BCR} = \$2,520,000 / \$1,048,077 = 2.40$$

$$\text{BCR, without delay costs} = \$2,520,000 / \$1,350 = 1,867$$

5.5.4. Congress Avenue – 12th Street to Barton Springs Road – Austin Congress & Cesar Chavez: Prohibit Right on Red

Cost of installation: \$800, CMF: 0.77³²

Delay costs = ADT on Congress (11,157) × 365 days = 4,072,305 vehicles delayed in a year

40,723,050 vehicles delayed in 10 years / 3 lanes = 13,574,350 vehicles likely to be impacted in one lane (prohibition of right turn only applies to right lane)

Each vehicle is delayed by an average of 10 seconds. 13,574,350 × 10 seconds = 135,743,500 seconds of delay

135,743,500 vehicle delay-seconds / 3600 seconds in an hour = 37,706 hours

37,706 hours × \$14.14 per vehicle delay-hour = \$533,162

\$533,162 + \$800 = \$533,962 in total costs

Benefits:

4 non-incapacitating injuries + 3 incapacitating injuries for 2010–2019

$(\$500,000 \times 4) + (\$3,500,000 \times 3) = \$2,325,000$

CMF = 0.77 → \$2,875,000

BCR = \$2,875,000 / \$533,962 = 5.38

BCR, without delay costs = \$2,875,000 / \$800 = 3,594

Congress Avenue and 6th Street: Pedestrian Leading Interval

Cost of Installation: \$1,750, CMF: 0.85³²

Delay Costs = ADT on Congress (11,157) + ADT on 6th (7,706) × 365 days in the year = 6,884,995 × 10 years = 68,849,950 vehicle delay-seconds

³² CMF Clearinghouse, 2021

68,849,950 / 3600 hours = 19,124 vehicle delay-hours

19,124 vehicle delay-hours × \$14.14/hour = \$270,413 in delay costs

\$270,413 + \$1,750 = \$272,163 in total costs

Benefits:

3 non-incapacitating injuries, 1 incapacitating injury for 2010–2019

$(\$500,000 \times 3) + (\$3,500,000 \times 1) = \$5,000,000$

CMF = 0.85 → \$750,000

BCR = \$750,000 / \$272,163 = 2.76

BCR, without delays = \$750,000 / \$1,750 = 428

Congress Avenue & 6th Street: Prohibit right-turn on red

Cost of installation: \$800, CMF: 0.77³³

ADT on 6th (7,706) × 365 days in the year × 10 years = 28,126,900 vehicles over 10 years

1 of 4 lanes is impacted = 7,031,725 vehicles delayed

7,031,725 vehicles × 10 seconds average delay = 70,317,250 seconds of delay

70,317,250 vehicle delay-seconds / 3600 seconds in an hour = 19,533 hours

19,533 hours × \$14.14/hour = \$276,197

Total Costs = \$276,197 + \$800 = \$276,997

Benefits:

3 non-incapacitating injury, 1 incapacitating injury for 2010–2019

$(\$500,000 \times 3) + (\$3,500,000 \times 1) = \$5,000,000$

CMF = 0.77 → \$1,150,000

BCR = \$1,150,000 / \$276,997 = 4.15

BCR, without delays = \$1,150,000 / \$800 = 1,437

³³ CMF Clearinghouse, 2021

5.5.5. Lamar Boulevard – Masterson Pass to Payton Gin Road – Austin

Rundberg Ln. and Lamar Boulevard – Pedestrian Refuge Islands

Cost of installation (average): \$9.80/square foot, CMF: 0.44³⁴

4 refuge islands, 2 @ 108 square feet (crossing Rundberg), 2 @ 67 square feet (crossing Lamar)

$(108 \times 2) + (67 \times 2) = 350$ square feet total, $\times \$9.80 = \$3,430$ in total installation costs

Benefits:

1 non-incapacitating injuries, 1 incapacitating injury for 2010–2019

$(\$500,000 \times 1) + (\$3,500,000 \times 1) = \$2,185,000$

CMF = 0.44 \rightarrow \$2,240,000

BCR = $\$2,240,000 / \$3,430 = 653$

Payton Gin Road and Lamar Boulevard - Refuge Island & Streetlight – Northern Crosswalk

Cost of installation, refuge island (average): \$9.80/square foot, Cost of installation (streetlight): \$4,880, CMF (for both treatments): 0.44³⁴

1 refuge island @ 131 square feet (can use the entire width of the turn lane if acceptable; a left-turn lane is not necessary as this is a three-way intersection)

$131 \text{ square feet} \times \$9.80 = \$1,283$

Total cost of installation = $\$4,880 + \$1,283 = \$6,163$

Benefits:

6 non-incapacitating injuries, 2 incapacitating injuries for 2010–2019

$(6 \times \$500,000 + 2 \times \$3,500,000) = \$10,000,000$

CMF = 0.44 \rightarrow \$5,600,000

BCR = $\$5,600,000 / \$6,163 = 908$

³⁴ CMF Clearinghouse, 2021

5.5.6. Congress Avenue – Woodward Street to St. Elmo Road - Austin

Congress Avenue - Speed Limit Reductions - 10% decrease

Cost of installation: \$135 per sign × 10, CMF: 0.79³⁵

Speed limit from 25 mph to 20 mph results in travel time from 129 s to 165 s for 4,800 ft segment.

36 second loss per vehicle

Delay Costs = ADT on Congress Avenue (12,437) × 36 second delay × 365 days = 163,422,180 seconds delay

163,422,180 seconds delay / 3600 seconds in an hour = 45,395 hours delay

45,395 hours × \$14.14 per vehicle delay-hour = \$641,886

\$ 641,886 + \$1,350 = \$643,236 in total costs

Benefits:

3 non-incapacitating injuries, 3 incapacitating injuries for 2010–2019

$(\$500,000 \times 3) + (\$3,500,000 \times 3) = \$12,000,000$

CMF = 0.79 → \$ 2,520,000

BCR = \$2,520,000 / \$643,236 = 3.92

BCR, without delay costs = \$2,520,000 / \$1,350 = 1,867

5.5.7. E. Riverside Drive – Pleasant Valley Road to Faro Drive – Austin

E. Riverside Drive and Pleasant Valley Road - Speed Limit Reductions - 10% decrease

Cost of installation: \$135 per sign × 5, CMF: 0.79³⁵

Speed limit from 35 mph to 30 mph results in travel time from 49 s to 56.8 s for 2,500 ft segment.

7.8 second loss per vehicle

Delay Costs = ADT on Pleasant Valley Road (18,134) × 7.8 second delay × 365 days = 51,627,498 seconds delay

51,627,498 seconds delay / 3600 seconds in an hour = 14,340 hours delay

14,340 hours × \$14.14 per vehicle delay-hour = \$202,767

³⁵ CMF Clearinghouse, 2021

$\$202,767 + \$675 = \$203,442$ in total costs

Benefits:

1.29 non-incapacitating injuries, 0.48 incapacitating injuries predicted for 2020-2029

$(\$500,000 \times 1.29) + (\$3,500,000 \times 0.48) = \$2,325,000$

CMF = 0.79 → \$488,250

BCR = $\$488,250 / \$203,442 = 2.40$

BCR, without delay costs = $\$488,250 / \$675 = 723$

E. Riverside Drive and Wickersham Lane - Pedestrian Leading Interval (1 second average delay for all crossings)

Cost of installation: \$1,750, CMF: 0.85³⁶

Delay Costs = ADT on E. Riverside Drive (11,871) + ADT on Wickersham Lane (5,599) × 365 days = 6,376,550 vehicles × 10 years = 63,765,500 seconds of delay over 10 years

63,765,500 vehicle delay-seconds / 3600 seconds in an hour = 17,712 vehicle delay-hours

41,199 × \$14.14 (discounted average over 10 years) = \$ 250,448

Total Costs = \$ 250,448 + \$1,750 = \$ 252,198 in total costs

Benefits:

3 non-incapacitating injuries, 1 incapacitating injury predicted for 2020-2029

$(\$500,000 \times 3) + (\$3,500,000 \times 1) = \$ 5,000,000$

CMF = 0.85 → \$750,000

BCR = $\$750,000 / \$ 252,198 = 2.97$

BCR, without delay costs = $\$750,000 / \$1,750 = 429$

³⁶ CMF Clearinghouse, 2021

Zarzamora Street – Cincinnati Street to Delgado Street – San Antonio

Zarzamora Street and Culebra Road – Pedestrian Leading Interval (1 second average delay for all crossings)

Cost of installation: \$1,750, CMF: 0.85³⁷

Delay Costs = ADT on Zarzamora Street (5,352) + ADT on Culebra Road (3,636) × 365 days = 3,280,620 vehicles × 10 years = 32,806,200 seconds of delay over 10 years

32,806,200 vehicle delay-seconds / 3600 seconds in an hour = 9,112 vehicle delay-hours

9,112 × \$14.14 (discounted average over 10 years) = \$128,843

Total Costs = \$128,843 + \$1,750 = \$130,593 in total costs

Benefits:

3 non-incapacitating injuries + 3 incapacitating injuries for 2010–2019

(\$500,000 × 3) + (\$3,500,000 × 3) = \$12,000,000

CMF = 0.77 → \$2,760,000

BCR = \$2,760,000 / \$130,593 = 21.13

BCR, without delay costs = \$2,760,000 / \$1,750 = 1,577

5.5.8. Fannin Street – Commerce Street to Jefferson Street – Houston

Fannin Street and Walker Street – Pedestrian Leading Interval (1 second average delay for all crossings)

Cost of installation: \$1,750, CMF: 0.85³⁷

Delay Costs = ADT on Fannin Street (12,542) + ADT on Walker Street (1,019) × 365 days = 4,949,765 vehicles × 10 years = 49,497,650 seconds of delay over 10 years

49,497,650 vehicle delay-seconds / 3600 seconds in an hour = 13,749 vehicle delay-hours

13,749 × \$14.14 (discounted average over 10 years) = \$194,415

Total Costs = \$194,415 + \$1,750 = \$196,165 in total costs

Benefits:

1 non-incapacitating injuries + 2 incapacitating injury for 2010–2019

³⁷ CMF Clearinghouse, 2021

$$(\$500,000 \times 1.64) + (\$3,500,000 \times 0.60) = \$7,500,000$$

$$\text{CMF} = 0.77 \rightarrow \$1,725,000$$

$$\text{BCR} = \$1,725,000 / \$196,165 = 8.79$$

$$\text{BCR, without delay costs} = \$1,725,000 / \$1,750 = 986$$

Fannin Street and Congress Street – Pedestrian Leading Interval (1 second average delay for all crossings)

Cost of installation: \$1,750, CMF: 0.85³⁸

Delay Costs = ADT on Fannin Street (12,542) + ADT on Congress Street (9,681) × 365 days = 8,111,395 vehicles × 10 years = 81,113,950 seconds of delay over 10 years

81,113,950 vehicle delay-seconds / 3600 seconds in an hour = 22,531 vehicle delay-hours

22,531 × \$14.14 (discounted average over 10 years) = \$318,588

Total Costs = \$318,588 + \$1,750 = \$320,338 in total costs

Benefits:

1 non-incapacitating injuries + 2 incapacitating injury for 2010–2019

$$(\$500,000 \times 1.64) + (\$3,500,000 \times 0.60) = \$7,500,000$$

$$\text{CMF} = 0.77 \rightarrow \$1,725,000$$

$$\text{BCR} = \$1,725,000 / \$320,338 = 5.38$$

$$\text{BCR, without delay costs} = \$1,725,000 / \$1,750 = 986$$

5.5.9. Milam Street – McGowan Street to Alabama Street – Houston

Milam Street – McGowan Street to Alabama Street – Road Diet

Cost of installation: \$4,000 × 0.6 mi, CMF: 0.71³⁹

Delay costs = ADT on Milam Street (14,530) × 365 days = 5,303,450 vehicles delayed in a year

From 4 lanes to 3 lanes results in 25% decreased capacity.

³⁸ CMF Clearinghouse, 2021

³⁹ Fitzpatrick et al., 2014

Current travel time through corridor = 4 min

Assuming 25% longer travel time after road diet, 5 min required.

Each vehicle is delayed by an average of 60 seconds. $5,303,450 \times 60$ seconds = 318,207,000 seconds of delay

$318,207,000$ vehicle delay-seconds / 3600 seconds in an hour = 88,390 hours

$88,390$ hours \times \$14.14 per vehicle delay-hour = \$ 1,249,834

Total cost of installation = \$2,400 + \$1,249,834 = \$1,252,234

Benefits:

5 non-incapacitating injuries + 4 incapacitating injuries for 2010–2019

$(\$500,000 \times 5) + (\$3,500,000 \times 4) = \$16,500,000$

CMF = 0.77 \rightarrow \$3,795,000

BCR = $\$3,795,000 / \$1,252,234 = 3.03$

BCR, without delay costs = $\$3,795,000 / \$2,400 = 1581$

5.6. Recommendations for Pedestrian Safety in Texas

Below is a comprehensive BCR analysis of applying select treatments to the Texas segments and intersections with the highest pedestrian crash risks to reduce such crashes. While Section 5.5 was focused on site-specific treatments, this section aims to apply four design treatments to all segments in Texas and deliver the top 100 hotspots' average BCR. The treatments applied in Section 5.6 includes pedestrian leading interval, speed hump, raised median (signalized / unsignalized), and 10% speed limit reduction. These treatments generally do not have site-specific restrictions and can be applied to any place in Texas. Four design treatments are applied, and the average of the top-100 highest BCR results out of the top-10,000 deadliest hotspots in Texas for all four treatments are presented below. The results suggest that the average BCR ranges from 0.82 to 19.20, implying that meaningful crash reductions can be expected with these treatments.

5.6.1. Pedestrian Leading Interval

Cost of installation: $\$1,750 \times 699,594$, CMF: 0.85⁴⁰

Delay cost = Assume 1 second of delay per vehicle, using AADT of each intersection.

⁴⁰ CMF Clearinghouse, 2021

Total cost = cost of installation + delay cost

Benefit = \$500,000 × estimated non-incapacitating injuries & \$3.5M × estimated incapacitating injuries predicted for 2020-2030, for each intersection. Assume CMF = 0.85.

Top 100 hotspots' average BCR = 0.82

5.6.2. Speed Hump

Cost of installation: \$2,640 × 699,594, CMF: 0.64⁴¹

Delay cost = Assume 3 seconds of delay per vehicle, using AADT of each segment.

Total cost = cost of installation + delay cost

Benefit = \$500,000 × estimated non-incapacitating injuries & \$3.5M × estimated incapacitating injuries predicted for 2020-2030, for each segment. Assume CMF = 0.64.

Top 100 hotspots' average BCR = 4.72

5.6.3. Raised Median

Cost of installation: \$22,500 (signalized) & \$18,000 (unsignalized)⁴²

CMF: 0.6 (signalized), 0.94 (unsignalized)

Delay cost = # of vehicles affected is estimated using AADT of each segment.

Total cost = cost of installation + delay cost

Benefit = \$500,000 × estimated non-incapacitating injuries & \$3.5M × estimated incapacitating injuries predicted for 2020-2030, for each segment. Assume CMF = 0.6 / 0.94.

Top 100 hotspots' average BCR = 11.19

5.6.4. Speed Limit 10% Reduction

Cost of installation: \$675⁴¹

Delay cost = 2500-foot length and calculate the delay per vehicle using AADT of each segment.

CMF: 0.79

Total cost = cost of installation + delay cost

⁴¹ CMF Clearinghouse, 2021

⁴² FHWA, 2018b

Benefit = \$500,000 × estimated non-incapacitating injuries & \$3.5M × estimated incapacitating injuries predicted for 2020-2030, for each segment/intersection. Assume CMF = 0.79

Top 100 hotspots' average BCR = 19.20

5.7. Policy Implementations for Vehicles Sold in the US

The United States National Highway Traffic Safety Administration (US NHTSA) has a “5-Star Safety Ratings” program to measure crash safety. It conducts frontal, side, and rollover tests on vehicles, with a total of four distinct tests. All test scenarios evaluate the injuries of dummies in the driver’s seat and passenger’s seat. They are concerned with collisions involving moving and fixed barriers, poles, and rollover events. The NHTSA rating standard does not consider collisions involving pedestrians (NHTSA, 2021).

The Insurance Institute for Highway Safety’s (IIHS) testing is more elaborate than that of NHTSA. Its six tests examine frontal collisions with moderate overlap (40% of the total width of the vehicle), driver-side overlap (25%), and passenger-side overlap (25%) along with side impact, roof strength for rollovers, and head restraints or seats. IIHS also assesses headlight systems and child seat attachment hardware. Several of their tests involve barriers and measure vehicle and driver impact. The IIHS does have three vehicle-to-pedestrian tests. These evaluate vehicles’ speed reduction capabilities for vehicles with autobrake when the vehicle is travelling perpendicular to an adult pedestrian, perpendicular to a child pedestrian, and when parallel to an adult pedestrian. However, it does not measure pedestrian injury (IIHS-HLDI, 2021).

The European Union provides the European New Car Assessment Programme (Euro NCAP) to evaluate the safety of vehicles sold in Europe. Euro NCAP issues star ratings based on five criteria: frontal impact test results (using barriers), a side impact test, pole test, child protection protocol, pedestrian protection tests, and electronic stability control. The pedestrian protection tests look at how vehicles’ bumpers and hoods (leading edge and top area) protect pedestrians’ lower legs at 40 kph (24.85 mph) for both children and adult pedestrians (ECMAT, 2021).

In the US, only the IIHS includes a test involving pedestrians, and this test evaluates the quality of vehicles’ autobrake rather than its ability to protect pedestrians in the event of a collision (IIHS-HLDI, 2021). In the US, OEMs need not protect pedestrians in the case of an impact. Like the US, the EU assesses autobrake systems. EMCAT examines performance when a pedestrian crosses in front of the vehicle, is walking in the same direction, crosses a road in which the vehicle is turning, and when it is reversing. There are conditions with adults and children and in low light. However, Euro NCAP runs these tests only when the vehicle performs well in the 40 kph pedestrian impact tests. The Euro NCAP measures injuries to a pedestrian’s head, upper leg, and lower leg for both adults and children using head form and leg form impactors. Ratings are given based on damages to dummies, hoods, and bumpers. Vehicles with energy-absorbing structures, deformation

clearance, and deployable protection systems for pedestrians are encouraged for better Euro NCAP ratings.

American vehicle brands specifically score very poorly in pedestrian safety according to Euro NCAP. For example, Chevrolet vehicles like the Malibu, Trax, and Captiva were rated less than 70% as of 2013. They received good ratings on other factors, such as adult occupant and child occupant safety and safety assist. Jeep has models whose Euro NCAP crash ratings have fallen over time. The Jeep Renegade and Cherokee showed lower ratings over six years. These vehicles score poorly in all categories of adult occupant, child occupant, safety assist, and pedestrian safety. Most notably, they rated in the 50s for vulnerable road user safety. Still, these vehicles received mixed ratings from the IIHS, ranging from ‘good’ to ‘poor’ for crashworthiness and crash avoidance across scenarios (front, side) (IIHS-HLDI, 2020a, 2020b).

In order to improve the pedestrian safety, the focus should be moved from testing the vehicles’ quality of autobrake to the actual ability to protect pedestrians in the event of a collision. The 40 kph speed limit suggested by Euro NCAP’s vehicle test should be increased to a higher value that may result in severe pedestrian injuries. OEMs’ responsibilities and obligations should incorporate the duty to protect pedestrians in case of an impact, which are currently not included.

5.8. Conclusions

This chapter provides algorithms to identify the Texas corridors, and intersections within those corridors, with the highest pedestrian crash risks, a list of treatments to improve pedestrian safety by decreasing the number of crashes, and the BCRs for application of those treatments to the corridors and intersections. Historical crash counts from 2010–2019 are used to assess the impact of each treatment on reducing potential threats to the pedestrians. Additionally, a negative binomial model to predict crash count estimates is suggested. Recommendations for Texas to apply the four different treatments are made with a BCR analysis

The BCR analyses show that the analyzed treatments—including prohibited right turns on red, speed limit reductions, pedestrian leading intervals, road diets, pedestrian refuge islands, and streetlights—have BCRs ranging from 1.67 to 5.38 when the delay costs are added, and from 428 to 6,689 when delay costs are ignored. Since a BCR greater than 1 indicates that the treatments are cost effective, implementation of the suggested treatments should lead to a cost savings while also increasing the safety conditions for pedestrians on otherwise high-risk corridors in Texas. A comprehensive BCR analysis of applying four design treatments (pedestrian leading interval, speed hump, raised median, and 10% speed limit reduction) to all segments in Texas suggests that the average BCR ranges from 0.82 to 19.20, implying that under most circumstances, meaningful crash reductions can be expected with these treatments. These treatments generally do not have site-specific restrictions and can be applied to any place in Texas.

A comparison of policies for promoting pedestrian safety in the US to European regulations was conducted, which found that a transition from testing vehicles' autobrake to testing the actual ability to protect pedestrians (e.g., the ability to stop properly before crashing with pedestrians) is needed. The speed limit of 40 kph (24.85 mph) used in the vehicle test should be increased, and the OEM's responsibilities to protect pedestrians should be clearly defined.

The strategies used in this chapter have been concisely formulated into the 35-page guidebook *Developing Countermeasures to Decrease Pedestrian Deaths* that is targeted at practitioners, managers, and other who have a vested interest in pedestrian safety. It is Product P1 of this project. The guidebook's methodology was piloted through a workshop presented to an audience of relevant professionals and enthusiasts at state and local levels of government. The video recording of the workshop is Product P2 of this project. Researchers recommend further work to find opportunities to more closely tie the guidebook into existing TxDOT road safety guidance and BCR methodologies, such as those documented within the *Highway Safety Improvement Program Manual* (TxDOT, 2015).

References

- Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of road crash frequency with spatial models. *Transportation Research Record*, 2061(1), 55-63.
- Anarkooli, A. J., Hosseinpour, M., & Kardar, A. (2017). Investigation of factors affecting the injury severity of single-vehicle rollover crashes: a random-effects generalized ordered probit model. *Accident Analysis & Prevention*, 106, 399-410.
- Arellano, Miguel. (2021). E-mail conversation regarding TxDOT Homelessness Efforts. January 3, 2021.
- Avineri, E., Shinar, D., & Susilo, Y. O. (2012). Pedestrians' behaviour in cross walks: The effects of fear of falling and age. *Accident Analysis & Prevention*, 44(1), 30-34.
- Aziz, H. A., Ukkusuri, S. V., & Hasan, S. (2013). Exploring the determinants of pedestrian-vehicle crash severity in New York City. *Accident Analysis & Prevention*, 50, 1298-1309
- Bachman, S. L., Arbogast, H., Ruiz, P., Farag, M., Demeter, N. E., Upperman, J. S., & Burke, R. V. (2015). A school-hospital partnership increases knowledge of pedestrian and motor vehicle safety. *Journal of community health*, 40(6), 1057-1064.
- Bernhardt, M., Kockelman, K., (2021). An analysis of pedestrian crash trends and contributing factors in Texas. *Journal of Transport & Health* 22, 101090.
<https://doi.org/10.1016/j.jth.2021.101090>
- Breiman, L., (2001). Random forests. *Mach. Learn.* 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., (1984). *Classification and regression trees*. CRC press.
- Boodlal, L., Donnell, E. T., & Porter, R. J. (2015). *Factors Influencing Operating Speeds and Safety on Rural and Suburban Roads*. Federal Highway Administration, McLean, VA, USA. Report FHWAHRT-15-030.
- Bureau of Transportation Statistics, US DOT, 2019. *National Transportation Statistics: U.S. Passenger-Miles*. URL: <https://www.bts.gov/content/us-passenger-miles> (accessed Dec. 2021)
- Bureau of Transportation Statistics, US DOT, 2020. *National Transportation Statistics: Injured Persons by Transportation Mode*. URL: <https://www.bts.gov/content/injured-persons-transportation-mode> (accessed Dec. 2021)
- Bushell, M. A., Poole, B. W., Zegeer, C. V., & Rodriguez, D. A. (2013). *Costs for pedestrian and bicyclist infrastructure improvements*. University of North Carolina Highway Safety Research Center, University of North Carolina, Chapel Hill, 45.
- Cantulupo, J. (2021). Email correspondence by Kara Kockelman with Texas Dept of Transportation Austin District Engineer, Joseph Cantulupo regarding pedestrian deaths.
- Carroll, N. (2017). *oglmx: Estimation of Ordered Generalized Linear Models*. R package version, 2(0.3).

- Carroll, P., Caulfield, B., & Ahern, A. (2019). Modelling the potential benefits of increased active travel. *Transport Policy*, 79, 82-92.
- Chen, T. D., & Kockelman, K. M. (2012). Roles of vehicle footprint, height, and weight in crash outcomes: application of a heteroscedastic ordered probit model. *Transportation Research Record*, 2280(1), 89-99.
- Chen, T., Guestrin, C., (2016). XGBoost: A scalable tree boosting system, in: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785--794. <https://doi.org/10.1145/2939672.2939785>
- Chen, Z., & Fan, W. D. (2019). A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. *International Journal of Transportation Science and Technology*, 8(1), 43-52.
- Chipman, H.A., George, E.I., McCulloch, R.E., (2012). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* <https://doi.org/10.1214/09-AOAS285>
- Crash Modification Factors (CMFs) Clearinghouse. (2021). How to Develop and Use CMFs. Accessed on April 27, 2021 at: http://www.cmfclearinghouse.org/cmf_data.cfm
- Davis, A. (2014). U.S. City Thinks Having Pedestrians Carry Flags Will Keep Them Safe. *Wired*, Conde Nast. Accessed at: <https://www.wired.com/2014/06/bridgeport-pedestrian-flags/>
- Davis, M. (2015, July). National telephone survey of reported and unreported motor vehicle crashes. (Findings Report. Report No. DOT HS 812 183). Washington, DC: National Highway Traffic Safety Administration. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812183>
- Dc.gov (2020). Vision Zero Action Plan | Ddot. Accessed May 25, 2020, <https://ddot.dc.gov/page/vision-zero-action-plan>.
- DfT (2010) Relationship between Speed and Risk of Fatal Injury: Pedestrians and Car Occupants. UK's Department for Transport Road Safety Web Publication No. 16. Available at: https://nacto.org/docs/usdg/relationship_between_speed_risk_fatal_injury_pedestrians_and_car_occupants_richards.pdf.
- DiMaggio, C., Brady, J., & Li, G. (2015). Association of the Safe Routes to School program with school-age pedestrian and bicyclist injury risk in Texas. *Injury epidemiology*, 2(1), 1-8.
- European Commission Mobility and Transport. (2021). Vehicle Safety. Accessed on November, 2021 at: https://ec.europa.eu/transport/road_safety/specialist/knowledge/safetyratings/safety_ratings_in_use/vehicle_safety_en.
- Ewing, R. (1999). *Traffic Calming: State of the Practice*, ITE/FHWA, August 1999 (No. FHWA-RD-99-135). United States. Federal Highway Administration.
- Fawcett, T., (2006). An introduction to ROC analysis. *Pattern Recognit. Lett.* <https://doi.org/10.1016/j.patrec.2005.10.010>

- Federal Highway Administration (FHWA, 2018). Summary of travel trends: 2017 National Household Travel Survey (Report FHWA-PL-18-019). Federal Highway Administration, US Department of Transportation, Washington, DC.
- Federal Highway Administration (FHWA, 2018). Toolbox of Pedestrian Countermeasures and Their Potential Effectiveness (Report FHWA-SA-018-41). Federal Highway Administration, US Department of Transportation, Washington, DC.
- Fitzpatrick, K., Iragavarapu, V., Brewer, M., Lord, D., Hudson, J., Avelar, R., & Robertson, J. (2014). Characteristics of Texas Pedestrian Crashes and Evaluation of Driver Yielding at Pedestrian Treatments. Texas A&M Transportation Institute, Texas A&M, 50.
- Fitzpatrick, K. et al. (2022) Crash Modification Factor for Corner Radius, Right-Turn Speed, and Prediction of Pedestrian Crashes at Signalized Intersections. FHWA Report HRT-21-105. At: <https://www.fhwa.dot.gov/publications/research/safety/21105/21105.pdf>.
- Goodkind et al. (2019) Fine-scale damage estimates of particulate matter air pollution reveal opportunities for location-specific mitigation of emissions. PNAS 116 (18): 8775-8780. <https://www.pnas.org/content/116/18/8775>
- Governors Highway Safety Association (2019). Pedestrian traffic fatalities by state: 2018 preliminary data. Accessed on April 18, 2020 at: https://www.ghsa.org/sites/default/files/2019-02/FINAL_Pedestrians19.pdf.
- Governors Highway Safety Association (2020). Pedestrian Traffic Fatalities by State: 2019 Preliminary Data. Governors Highway Safety Association. Retrieved from <https://www.ghsa.org/sites/default/files/2020-02/GHSA-Pedestrian-Spotlight-FINAL-rev2.pdf>
- Haleem, K., Alluri, P., & Gan, A. (2015). Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accident Analysis & Prevention*, 81, 14-23
- Hall, J.W., Brogan, J.D., and Kondreddi, M. (2004). Pedestrian Safety on Rural Highways. Federal Highway Administration Report #FHWA-SA-04-008, http://www.pedbikeinfo.org/cms/downloads/Ped_Safety_RuralHighways.pdf.
- Han, Y., Yang, J., Mizuno, K., & Matsui, Y. (2012). Effects of vehicle impact velocity, vehicle front-end shapes on pedestrian injury risk. *Traffic injury prevention*, 13(5), 507-518.
- Hand, D.J., Till, R.J., (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Mach. Learn.* <https://doi.org/10.1023/A:1010920819831>
- He, J., Yalov, S., & Hahn, P. R. (2019). XBART: Accelerated Bayesian additive regression trees. In *The 22nd International Conference on Artificial Intelligence and Statistics*, 1130-1138.
- Hu, W., & Cicchino, J. B. (2018). An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *Journal of safety research*, 67, 37-44.
- Huang, H., Zhou, H., Wang, J., Chang, F., & Ma, M. (2017). A multivariate spatial model of crash frequency by transportation modes for urban intersections. *Analytic Methods in Accident Research*, 14, 10-21.

- IIHS-HLDI (2020a). 2020 Jeep Grand Cherokee. Accessed March 2021: <https://www.iihs.org/ratings/vehicle/jeep/grand-cherokee-4-door-suv/2020>.
- IIHS-HLDI (2020b). 2020 Jeep Renegade. Accessed March 2021: <https://www.iihs.org/ratings/vehicle/jeep/renegade-4-door-suv/2020>.
- IIHS-HLDI (2021) About our tests. Accessed March 2021: <https://www.iihs.org/ratings/about-our-tests>.
- Islam, S., Hossain, A. B., & Barnett, T. E. (2016). Comprehensive injury severity analysis of SUV and pickup truck rollover crashes: Alabama case study. *Transportation Research Record*, 2601(1), 1-9.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., (2017). LightGBM: A highly efficient gradient boosting decision tree, in: *Advances in Neural Information Processing Systems*.
- Kim, J. K., Ulfarsson, G. F., Shankar, V. N., & Kim, S. (2008). Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention*, 40(5), 1695-1702.
- Kim, J. K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis & Prevention*, 42(6), 1751-1758.
- Klein, A., Falkner, S., Bartels, S., Hennig, P., Hutter, F., (2017). Fast Bayesian optimization of machine learning hyperparameters on large datasets, in: *Artificial Intelligence and Statistics*. pp. 528--536.
- Kockelman, K., Bottom, J., Kweon, Y. J., Ma, J., & Wang, X. (2006). Safety impacts and other implications of raised speed limits on high-speed roads (Vol. 90). Washington, DC, USA: Transportation Research Board.
- Le, J., (2018). Decision Trees in R [WWW Document]. DataCamp Community. URL <https://www.datacamp.com/community/tutorials/decision-trees-R> (accessed 5.26.21).
- Lee, C., & Abdel-Aty, M. (2005). Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. *Accident Analysis & Prevention*, 37(4), 775-786.
- Lee, Shaun (Heart of Texas Region MHMR, 2020). E-mail conversation regarding the state of homelessness in Texas and PIT count methodologies. Accessed July 15, 2020.
- Lefler, D. E., & Gabler, H. C. (2004). The fatality and injury risk of light truck impacts with pedestrians in the United States. *Accident Analysis & Prevention*, 36(2), 295-304.
- Lemp, J. D., Kockelman, K. M., & Unnikrishnan, A. (2011). Analysis of large truck crash severity using heteroskedastic ordered probit models. *Accident Analysis & Prevention*, 43(1), 370-380.
- Li, G., Wang, F., Otte, D., Cai, Z., & Simms, C. (2018). Have pedestrian subsystem tests improved passenger car front shape? *Accident Analysis & Prevention*, 115, 143-150.
- Li, W., Kockelman, K.M., (2020). How does machine learning compare to conventional econometrics for transport data sets? A test of ML vs MLE. *Transp. Res. Rec.*
- Liaw, A., Wiener, M., (2002). Classification and Regression by randomForest. *R News* 2, 18–22.

- Liu, J., Hainen, A., Li, X., Nie, Q., & Nambisan, S. (2019). Pedestrian injury severity in motor vehicle crashes: an integrated spatio-temporal modeling approach. *Accident Analysis & Prevention*, 132, 105272.
- Liu, X. J., Yang, J. K., & Lövsund, P. (2002). A study of influences of vehicle speed and front structure on pedestrian impact responses using mathematical models. *Traffic Injury Prevention*, 3(1), 31-42.
- Lobo, A., Ferreira, S., Iglesias, I., & Couto, A. (2020). Daily and Latent Lagged Effects of Rainfall on Pedestrian–Vehicle Collisions. *Weather, climate, and society*, 12(2), 279-291.
- Luoma, J., & Peltola, H. (2013). Does facing traffic improve pedestrian safety?. *Accident Analysis & Prevention*, 50, 1207-1210.
- Ma, J., Kockelman, K. M., & Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention*, 40(3), 964-975.
- Madsen, J., Andersen, T., & Lahrman, H. (2013). Safety effects of permanent running lights for bicycles: A controlled experiment. *Accident Analysis and Prevention* 50: 820-9.
- Martensen, H., Focant, N., & Diependaele, K. (2016). Let's talk about the weather–Interpretation of short term changes in road accident outcomes. *Transportation research procedia*, 14, 96-104.
- Martin, Peter (2022) Lead engineer, Office of Crashworthiness Standards, US NHTSA (National Highway Transportation Safety Administration). Phone interview by Kara Kockelman, January 28.
- Massachusetts Department of Transportation (2020). Risk Factors for Older Pedestrian Injuries and Fatalities in MA. Accessed on April 18, 2020 at: https://www.mass.gov/files/documents/2019/10/02/RiskFactorsOlderPedestrian_August_2019.pdf.
- McAndrews, C., Pollack, K.M., Berrigan, D., Dannenberg, A.L., Christopher, E.J., (2017). Understanding and Improving Arterial Roads to Support Public Health and Transportation Goals. *Am J Public Health* 107, 1278–1282. <https://doi.org/10.2105/AJPH.2017.303898>
- Mohamed, M. G., Saunier, N., Miranda-Moreno, L. F., & Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. *Safety Science*, 54, 27-37.
- Muennig, P., Epstein, M., Guohua, L., & DiMaggio, C. (2014). The Cost-Effectiveness of New York City’s Safe Routes to School Program. *American Journal of Public Health* 104(7): 1294-1299.
- NHTSA (2021) Ratings. Accessed March 2021: <https://www.nhtsa.gov/ratings>.
- Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M.A., (2016). Joint Modeling of Pedestrian and Bicycle Crashes: Copula-Based Approach. *Transp. Res. Rec.* 2601, 119–127. <https://doi.org/10.3141/2601-14>

- National Association of City Transportation Officials. (2020). Vehicle Stopping Distance and Time. Accessed on June 22, 2020 at: https://nacto.org/references/a-hrefdocsusdgvvehicle_stopping_distance_and_time_upenn/
- National Highway Traffic Safety Administration. (2019). 2018 Fatal Motor Vehicle Crashes Overview. Accessed on April 19, 2020 at: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812826>.
- National Oceanic & Atmospheric Administration (NOAA, 2020). Climate Prediction Center - GIS Data (Shapefile and Raster). Accessed July 15, 2020 at: https://www.cpc.ncep.noaa.gov/products/GIS/GIS_DATA/.
- New York City Department of Transportation (NYCDOT, 2020). NYC DOT - Pedestrian Safety Report. Accessed on April 18, 2020 at: <https://www1.nyc.gov/html/dot/html/pedestrians/pedsafetyreport.shtml>.
- Nie, B., & Zhou, Q. (2016). Can new passenger cars reduce pedestrian lower extremity injury? A review of geometrical changes of front-end design before and after regulatory efforts. *Traffic Injury Prevention*, 17(7), 712-719.
- Nowacki, Lauren (2022). How Much Does a Tiny Home Cost? Rocket Homes, Rocket Homes. Accessed at: <https://www.rockethomes.com/blog/home-buying/tiny-house#:~:text=The%20average%20cost%20of%20a,frills%20you%20want%20to%20include>.
- Oborski, Patrick (Detective – Austin Police Department). E-mail conversation regarding Austin Police Department and Pedestrian Crashes and Fatalities. Accessed July 15, 2020.
- Oregon State University. (2020). PRISM Climate Group, 1981-2010 Normals. Accessed on October 27, 2021 at: <https://prism.oregonstate.edu/normals/>.
- Pai, C. W., Chen, P. L., Ma, S. T., Wu, S. H., Linkov, V., & Ma, H. P. (2019). Walking against or with traffic? Evaluating pedestrian fatalities and head injuries in Taiwan. *BMC public health*, 19(1), 1-11.
- Pande, A., & Abdel-Aty, M. (2009). A novel approach for analyzing severe crash patterns on multilane highways. *Accident Analysis & Prevention*, 41(5), 985-994.
- Perrine K. A. and Zuniga-Garcia, N. (2021) Repository peds-midblocks-intersections. Accessed Nov. 2021 online at <https://github.com/ut-ctr-nmc/peds-midblocks-intersections>
- Pour-Rouholamin, M., & Zhou, H. (2016). Investigating the risk factors associated with pedestrian injury severity in Illinois. *Journal of Safety Research*, 57, 9-17.
- Rahman, M. and Kockelman, K.M., (2020). Predicting Pedestrian Crash Occurrences and Injury Severity in Texas Presented at the 100th Annual Meeting of the Transportation Research Board. Under review for publication in *Traffic Injury and Prevention*.
- Ralph, K., & Girardeau, I. (2020). Distracted by “distracted pedestrians”? *Transportation research interdisciplinary perspectives*, 5, 100118.
- Reyna, Sean (Communications – Austin Police Department, 2020). E-mail conversation regarding Austin Police Department and Pedestrian Crashes & Fatalities. Accessed July 15, 2020.

- Ruikar, M. (2013). National statistics of road traffic accidents in India. *Journal of Orthopedics, Traumatology and Rehabilitation*, 6(1), 1.
- Schneider, R. J., Diogenes, M. C., Arnold, L. S., Attaset, V., Griswold, J., & Ragland, D. R. (2010). Association between roadway intersection characteristics and pedestrian crash risk in Alameda County, California. *Transportation Research Record*, 2198(1), 41-51.
- Schneider, R. J., Qin, X., Shaon, M. R. R., Sanatizadeh, A., He, Z., Wkyhuis, P., ... & Bill, A. (2017). Evaluation of Driver Yielding to Pedestrians at Uncontrolled Crosswalks. University of Wisconsin-Madison Traffic Operations and Safety (TOPS) Laboratory: Madison, WI, USA.
- Schultz, G. G., Eggett, D. L., & Berrett, J. J. (2019). Pedestrian Walking Speeds at Signalized Intersections in Utah (No. UT-19.06). Utah. Dept. of Transportation.
- Siddiqui, N. A., Chu, X., & Guttenplan, M. (2006). Crossing locations, light conditions, and pedestrian injury severity. *Transportation Research Record*, 1982(1), 141-149.
- Simmons, S. M., Caird, J. K., Ta, A., Sterzer, F., & Hagel, B. E. (2020). Plight of the distracted pedestrian: a research synthesis and meta-analysis of mobile phone use on crossing behaviour. *Injury Prevention*, 26(2), 170-176.
- Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M. N., Peden, M., & Ewing, R. (2015). Pedestrian safety and the built environment: a review of the risk factors. *Journal of Planning Literature*, 30(4), 377-392.
- Sullivan, J. M., Adachi, G., Mefford, M. L., & Flannagan, M. J. (2004). High-beam headlamp usage on unlighted rural roadways. *Lighting Research & Technology*, 36(1), 59-65.
- Tefft, B. C. (2013). Impact speed and a pedestrian's risk of severe injury or death. *Accident Analysis & Prevention*, 50, 871-878.
- Texas Association of Counties (TAC, 2020). County Information System. Accessed on July 15, 2020 at: <https://imis.county.org/iMIS/CountyInformationProgram/QueriesCIP.aspx>
- Texas Department of Transportation (2015). Highway Safety Improvement Program Manual. Accessed Dec. 2021: <http://onlinemanuals.txdot.gov/txdotmanuals/hsi/hsi.pdf>
- Texas Department of Transportation (2021). Highway Safety Improvement Program Guidelines. Accessed on February 9, 2022 at: <https://www.txdot.gov/inside-txdot/forms-publications/publications/highway-safety.html>
- Texas Department of Transportation (TxDOT, 2018). Roadway Inventory. Accessed July, 2020 at: <https://www.txdot.gov/inside-txdot/division/transportation-planning/roadway-inventory.html>
- Texas Department of Transportation (TxDOT, 2020a). Crash Data Analysis and Statistics. Accessed on July, 2020 at: <https://www.txdot.gov/government/enforcement/crash-statistics.html>
- Texas Department of Transportation (TxDOT, 2020b). Texas Motor Vehicle Crash Statistics: 2020. Accessed on December, 2021 at: <https://www.txdot.gov/inside-txdot/forms-publications/drivers-vehicles/publications/annual-summary.html>

- Texas Department of Transportation (TxDOT, 2021). CRIS Query. Retrieved April 27, 2021, from <https://cris.dot.state.tx.us/public/Query/app/welcome>
- Texas Homeless Network (2017). Point in Time (PIT) Count and HIC Reports. Accessed on May 31, 2017 at: <https://www.thn.org/texas-balance-state-continuum-care/data/pit-count-and-hic/>.
- Tharwat, A., (2018). Classification assessment methods. *Appl. Comput. Inform.* <https://doi.org/10.1016/j.aci.2018.08.003>
- Therneau, T.M., Atkinson, E.J., (1997). An introduction to recursive partitioning using the RPART routines. Technical report Mayo Foundation.
- Torgo, L., (2017). Regression Trees, in: Sammut, C., Webb, G.I. (Eds.), *Encyclopedia of Machine Learning and Data Mining*. Springer US, Boston, MA, pp. 1080–1083. https://doi.org/10.1007/978-1-4899-7687-1_717
- U.S. Environmental Protection Agency (2017). Light-Duty Automotive Technology, Carbon Dioxide Emissions, and Fuel Economy Trends: 1975 Through 2017. Appendix D. U.S. EPA: Washington DC. <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100TGDW.pdf>
- US Energy Information Administration (EIA). (2020). Crossover Utility Vehicles Overtake Cars as the Most Popular Light-Duty Vehicle Type - Today in Energy - U.S. Energy Information Administration (EIA). Accessed on June 22, 2020 at: <https://www.eia.gov/todayinenergy/detail.php?id=36674>.
- US Environmental Protection Agency (EPA, 2020). Highlights of the Automotive Trends Report [Data and Tools]. Accessed on May 4, 2020 at: <https://www.epa.gov/automotive-trends/highlights-automotive-trends-report>
- USDOT. (2018). Summary of Travel Trends 2017 National Household Travel Survey (NHTS). Accessed on July 2018 at: <https://doi.org/10.2172/885762>.
- USDOT. (2021). Summary of Travel Trends 2017 National Household Travel Survey (NHTS). Accessed on July 2018 at: <https://doi.org/10.2172/885762>.
- Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. *Safety Science*, 50(4), 1141-1151.
- United States Census Bureau. (2020). 2017 ACS 1-Year Estimates. Accessed on July 15, 2020 at: <https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2017/1-year.html>.
- Wang, X., & Kockelman, K. M. (2005). Use of heteroscedastic ordered logit model to study severity of occupant injury: distinguishing effects of vehicle weight and type. *Transportation Research Record*, 1908(1), 195-204.
- Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention*, 60, 71-84.
- Warsh, J., Rothman, L., Slater, M., Steverango, C., Howard, A., (2009). Are school zones effective? An examination of motor vehicle versus child pedestrian crashes near schools.

- Inj. Prev. J. Int. Soc. Child Adolesc. Inj. Prev. 15, 226–229.
<https://doi.org/10.1136/ip.2008.020446>
- Welch, E. A. (2016). Identifying factors explaining pedestrian crash severity: A study of Austin, Texas (Doctoral dissertation, The University of Texas at Austin).
- WHO (2018) Global Status Report on Road Safety 2018. World Health Organization, France. Available via: <https://www.who.int/publications/i/item/9789241565684>.
- Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, 41(1), 137-145.
- Wood, J. M., Lacherez, P., & Tyrrell, R. A. (2014). Seeing pedestrians at night: effect of driver age and visual abilities. *Ophthalmic and Physiological Optics*, 34(4), 452-458.
- Wood, J. M., Tyrrell, R. A., & Carberry, T. P. (2005). Limitations in drivers' ability to recognize pedestrians at night. *Human factors*, 47(3), 644-653.
- Xu, J., Kockelman, K. M., & Wang, Y. (2014). Modeling crash and fatality counts along mainlanes and 1 frontage roads across Texas: 2 the roles of design, the built environment, and weather 3. In 93rd Annual Meeting of the Transportation Research.
- Yang, Y., & Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American journal of preventive medicine*, 43(1), 11-19.
- Yue, L., Abdel-Aty, M., Wu, Y., Zheng, O., & Yuan, J. (2020). In-depth approach for identifying crash causation patterns and its implications for pedestrian crash prevention. *Journal of safety research*, 73, 119-132.
- Zajac, S. S., & Ivan, J. N. (2003). Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accident Analysis & Prevention*, 35(3), 369-379.
- Zegeer, C. V., & Bushell, M. (2012). Pedestrian crash trends and potential countermeasures from around the world. *Accident Analysis & Prevention*, 44(1), 3-11.
- Zhao, B., Zuniga-Garcia, N., Xing, L., Kockelman, K., (2021). Predicting Pedestrian Crash Occurrence and Injury Severity in Texas Using Tree-Based Machine Learning Models. Under review for publication in *Transportation Research Record*.
- Zhu, M., Zhao, S., Coben, J. H., & Smith, G. S. (2013). Why more male pedestrians die in vehicle-pedestrian collisions than female pedestrians: a decompositional analysis. *Injury Prevention*, 19(4), 227-231.

Appendix A. Additional Examples of BCR Calculations

S. 1st Street at William Cannon Boulevard – Austin

Refuge Islands

$$\text{CMF} = 0.44^{43}$$

Cost of installation: \$9.80 per square foot

4 refuge islands, two @ 36 sq feet (crossing S. 1st), two @ 64 square feet (crossing William Cannon) = 100 sq ft.

$$100 \text{ sq. ft} \times \$9.80 = \$980$$

Benefits

3 non-incapacitating injuries

$$3 \times \$500,000 = \$1,500,000$$

$$\text{CMF} = 0.44 \rightarrow \$840,000$$

$$\text{BCR} = \$840,000 / \$980 = 857$$

Oltorf Street and Pleasant Valley Road – Austin

Refuge Islands

$$\text{CMF}: 0.44^{43}$$

Cost of installation: \$9.80 per square foot

3 refuge islands @ 40 square feet

$$120 \text{ sq. ft.} \times 9.80 = \$1,176$$

Benefits

Three non-incapacitating injuries

⁴³ CMF Clearinghouse, 2021

$$3 \times \$500,000 = \$1,500,000$$

$$\text{CMF} = 0.44 \rightarrow \$840,000$$

$$\text{BCR} = 840,000 / 1,176 = 714$$

Pedestrian Hybrid Beacon – west of western crosswalk approx. 800 feet

$$\text{CMF} = 0.71^{44}$$

Cost of installation (avg): \$57,000

Delay Costs

ADT on Oltorf: 18,894

$$18,894 \times 365 = 6,896,310 \text{ vehicles delayed} \times 10 \text{ years} = 68,963,100 \text{ vehicles}$$

Average delay: 2 sec

$$68,963,100 \times 2 \text{ seconds delay} = 137,926,200 \text{ seconds delay}$$

$$137,926,200 / 3600 \text{ seconds in an hour} = 38,313 \text{ hours of delay}$$

$$38,313 \times \$14.14/\text{hour} = \$541,743$$

$$\$57,460 + \$541,743 = \$599,203$$

Benefits

10 years – 2 non-incapacitating injuries, 1 incapacitating injury

$$(\$500,000 \times 2) + \$3,500,000 = \$4,500,000$$

$$\text{CMF} = 0.71 \rightarrow \$1,305,000$$

$$\text{BCR} = \$1,305,000 / \$599,203 = 2.18$$

$$\text{BCR, without delay} = \$1,305,000 / \$57,460 = 22.7$$

William Cannon Drive and Bluff Springs Road – Austin

⁴⁴ CMF Clearinghouse, 2021

Reduction in Speed Limit, 10% (2500 ft. of William Cannon between I-35 and Elm Creek Drive)

$$\text{CMF} = 0.79^{45}$$

Installation Cost: \$135/sign

$$\text{Est. signs needed} = 5 \times \$135 = \$675$$

Delay Costs:

Travel time increased by 10%

45 mph is 66 fps

40 mph is approximately 59 fps

Travel time over the 2,500 ft segment is about 38 seconds at 66 fps

Travel time over the 2,500-foot segment is about 42.6 seconds at 59 fps

4.6 second loss per vehicle

ADT on William Cannon: $34,131 \times 365 \text{ days} \times 10 \text{ years} \times 4.6 \text{ second delay} = 573,059,490 \text{ seconds delay}$

$573,059,490 \text{ seconds delay} / 3600 \text{ seconds in an hour} = 159,183 \text{ hours delay}$

$159,183 \text{ hours delay} \times \$14.14 \text{ per vehicle-hour} = \$2,250,850$

$\$2,250,850 + \$675 = \$2,251,525$

Leading Interval

$$\text{CMF} = 0.85^{45}$$

Installation cost: \$1750

Delay Costs

ADT on William Cannon: 34,131

ADT on Bluff Springs: 14,352

AADT: $(34,131 + 14,352) \times 365 = 17,696,295 \times 10 \text{ years} = 176,692,950$

⁴⁵ CMF Clearinghouse, 2021

176,692,950 vehicles delayed x 1 second delay = 176,692,950 seconds delayed

176,692,950 / 3600 seconds in an hour = 49,081 hours delayed

49,081 hours x \$14.14 (discounted cost of delay over 10 years) = \$694,005

\$694,005 + \$1,750 = \$695,755

Benefits

8 non-incapacitating injuries, 1 incapacitating injuries, 1 fatality

$(8 \times 500,000) + (2 \times 3.5m) = \$11,000,000$

CMF = 0.85 x \$11,000,000 = \$1,650,000 or more in benefits

BCR = \$1,650,000 / \$695,755 = 2.37

BCR, without delay = \$1,650,000 / \$1,750 = 942.85

Manor Road and Susquehanna Ln. – Austin

Leading Interval

CMF = 0.85⁴⁶

Installation cost: \$1750

Delay Costs:

ADT on Manor: 16,815

ADT on Susquehanna (unknown; comparable streets in the area have around 3,000 ADT)

AADT $(16,815 + 3,000) \times 365 = 7,232,475 \times 10 = 72,324,750$

72,324,750 vehicles delayed x 1 seconds delay = 72,324,750 seconds delayed

180,811,875 / 3600 seconds in the hour = 20,065 hours delayed

20,065 x \$14.14 (discounted cost of delay over 10 years) = \$283,722

\$283,722 + \$1,750 = \$285,472

Benefits:

⁴⁶ CMF Clearinghouse, 2021

8 non-incapacitating injuries, 2 incapacitating injuries

$$(8 \times \$500,000) + (2 \times \$3,500,000) = \$11,000,000$$

$$\text{CMF} = 0.85 \times \$11,000,000 = \$1,650,000$$

$$\text{BCR} = \$1,650,000 / \$285,472 = 5.78$$

$$\text{BCR, without delays} = \$1,650,000 / \$1,750 = 942.85$$

Pedestrian Refuge Island – southern and northern crossings (of Manor Road)

$$\text{CMF} = 0.44^{47}$$

Installation cost: \$9.80/sq. ft.

2 @ 80 square feet = 160 square feet

$$\$9.80 \times 160 = \$1568$$

6 non-incapacitating injuries, 1 incapacitating injury (difference from above: the two others were not caused by left-turners and leading intervals could have improved visibility, theoretically)

$$(6 \times \$500,000) + \$3,500,000 = \$6,500,000$$

$$\text{CMF} = 0.44 = \$3,640,000 \text{ in benefits}$$

$$\text{BCR} = \$3,640,000 / \$1568 = 2080$$

6th Street and Lamar Boulevard – Austin

Leading Interval

$$\text{CMF} = 0.95 \text{ CMF}^{47}$$

Installation costs: \$1750

Delay Costs

ADT on Lamar: 33,757

ADT on W. 6th: 17,223

⁴⁷ CMF Clearinghouse, 2021

AADT: $(33,757 + 17,223) \times 365 = 18,607,700 \times 10 = 186,077,000$ vehicles delayed over 10 years

$186,077,000$ vehicles delayed $\times 1$ seconds delay = $186,077,000$ seconds of delay

$186,077,000 / 3600$ seconds in an hour = $51,688$ hours of delay

$51,688$ hours $\times \$14.14$ (discounted cost of delay over 10 years) = $\$730,869$

$\$730,869 + \$1,750 = \$732,619$

Benefits

6 non-incapacitating injuries, 1 incapacitating injuries

$(\$500,000 \times 6) + \$3,500,000 = \$6,500,000$

$CMF = 0.85 \times \$6,500,000 = \$975,000$

$BCR = \$975,000 / \$732,619 = 1.33$

BCR, without delays: $\$975,000 / \$1750 = 557.14$

Appendix B. CRIS Online Hot Spot Analysis Tool's Intersection Evaluations

All images in this appendix have been obtained from Google Maps.

Hopkins Street, San Marcos

Description: Major entertainment corridor in downtown San Marcos, popular with Texas State University students. Controlled crosswalks at either end of the block but a lack of pedestrian space in front of the bars, no mid-block crossings.

Recommendations: Eliminate parking in front of bars (and replace with curb extensions), create dedicated pedestrian intervals on adjacent blocks to ensure safe access to and from the area as well as barriers on any new curb extensions to prevent unauthorized mid-block crossings.

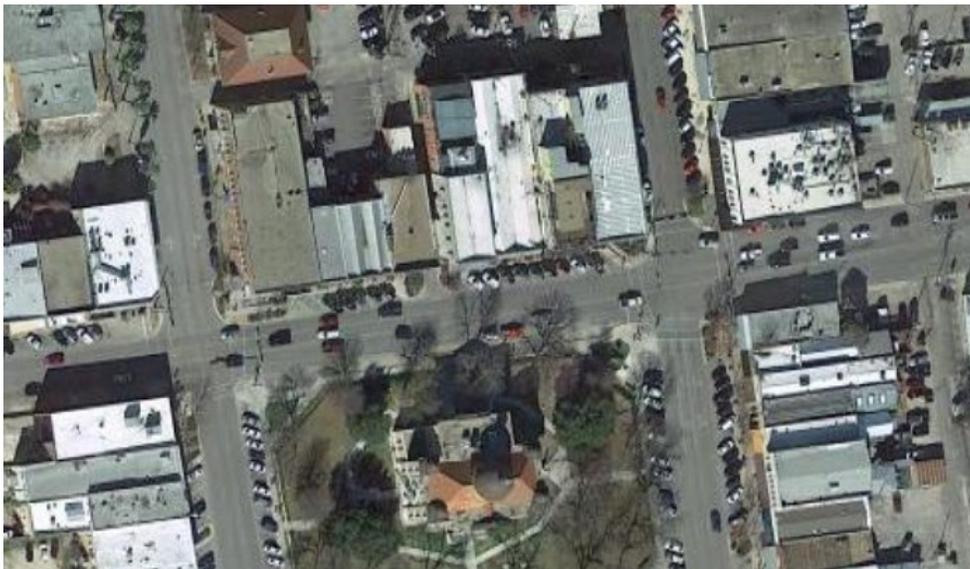


Figure B. 1 Hopkins Street, San Marcos

Commerce Street & N. St. Mary's Street, San Antonio

Description: This intersection sits in the middle of downtown San Antonio, a very high pedestrian-traffic area for both businesspeople as well as tourists. Both intersecting streets are one-way.

Recommendations: Dedicated pedestrian interval to give pedestrians more visibility, longer pedestrian signal times to handle the volume of pedestrians close to tourist sites and downtown businesses, “watch for pedestrians” signs at left turns (as these are two one-way streets, left turns are made close to the curb) AND/OR no left/right on red



Figure B.2 Commerce Street & N. St. Mary's Street, San Antonio

Zarzamora Street & Culebra Road, San Antonio

Description: This intersection is in a dense urban neighborhood and is abutted by public transit stops and commercial development with numerous driveways.

Recommendations: Given the width of the street and volume of pedestrian traffic, refuge islands and a leading pedestrian interval would be appropriate for this intersection. The ability for VIA buses to utilize their own light controls may also help exiting passengers.

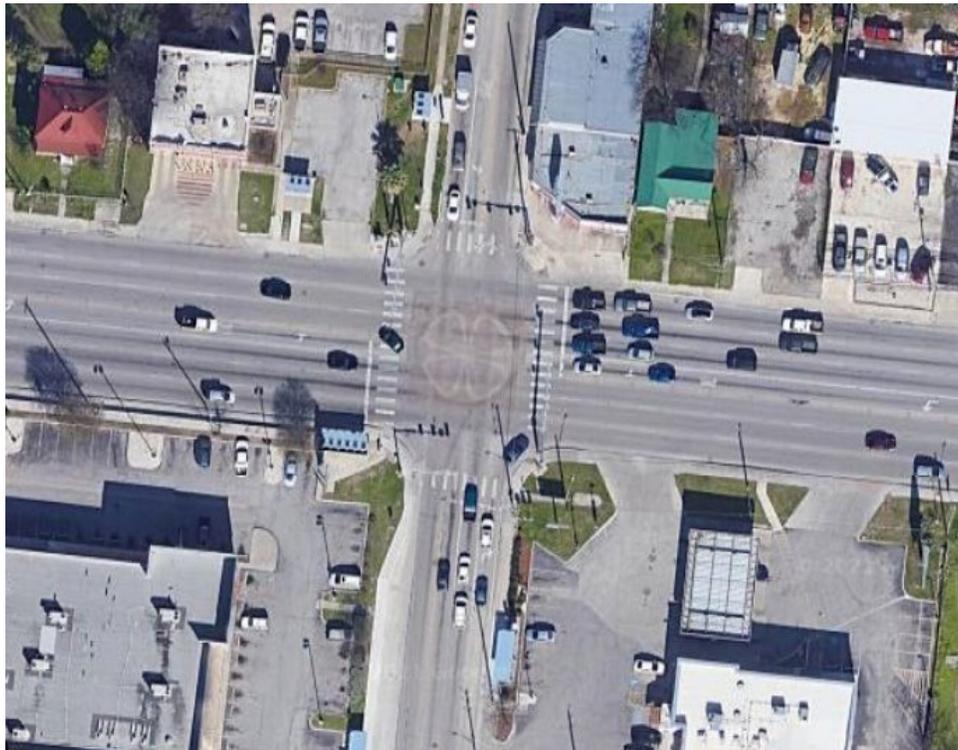


Figure B.3 Zarzamora Street & Culebra Road, San Antonio

Paisano Drive & Oregon Street, El Paso

Description: This intersection lies in a commercial corridor near the Mexican border, as well as a regional bus terminal, and has higher levels of pedestrian activity. Families (including those with young children) will come from Mexico for the day to shop in this area.

Recommendations: Leading or dedicated pedestrian intervals will help to move larger amounts of pedestrians across the intersection without conflict from vehicles. The median can also be extended to create a refuge island and decrease the left turn radius (and speed).



Figure B.4 Paisano Drive & Oregon Street, El Paso

Zarzamora Street & Guadalupe Street, San Antonio

Description: This intersection lies in a dense urban residential neighborhood with multiple schools, commercial buildings, and public transit stations.

Recommendations: As this intersection is located in a school zone, a dedicated pedestrian cycle (in which only pedestrians are allowed to negotiate the intersection) during before- and after-school hours would be helpful, along with a crossing guard presence.

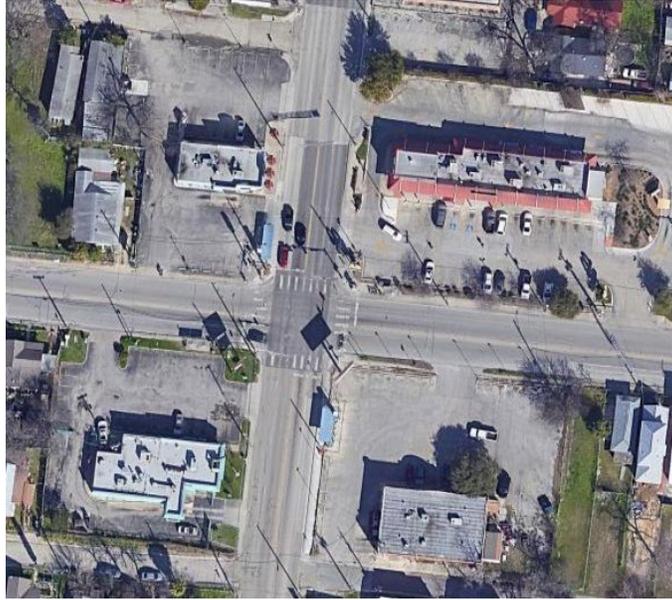


Figure B.5 Zarzamora Street & Guadalupe Street, San Antonio

Voss Road/Hillcroft Avenue & Westheimer Road, Houston

Description: This is a very typical intersection for Houston: two wide arterial roads surrounding by commercial driveways and high-density apartment complexes.

Recommendation: Pedestrian refuge islands would be particularly helpful on Westheimer Road (the horizontal street), along with dedicated left-turn cycles that preclude the possibility of pedestrians legally crossing while cars are making left turns. Elimination of the channelized right turn present at the NW corner would also prioritize pedestrian visibility and decrease right-turn traffic speeds.



Figure B.6 Voss Road/Hillcroft Avenue & Westheimer Road, Houston

Callaghan Road & Ingram Road, San Antonio

Description: This intersection is bounded by commercial centers, a church, and the entrance to a single-family residential neighborhood. A transit station abuts two of the corners.

Recommendations: A center median with a refuge island would help reduce the amount of roadway pedestrians need to cross at once while decreasing the left-turn radius, slowing speeds, and discouraging vehicles from cutting corners. Additionally, access management improvements could be made to the NE corner, as there are multiple driveways close to the corner.



Figure B.7 Callaghan Road & Ingram Road, San Antonio

Westheimer Road & Sage Road, Houston

Description: This intersection is located in a very busy shopping corridor in Uptown Houston. Despite the ornamentation of the intersection, this is a very wide crossing, especially on the Sage Road axis, which lacks refuge islands.

Recommendations: Given the heavy volume of traffic here, a leading interval would be useful to increase pedestrian visibility. Given that there is a dedicated left-turn cycle, refuge islands would also help to split up the crossing of Westheimer Road, especially for slower walkers.



Figure B.8 Westheimer Road & Sage Road, Houston

Riverside Drive & Lancaster Avenue, Ft. Worth

Description: This section lies in between downtown Ft. Worth and a large area of homeless shelters along Lancaster avenue Two sides of the intersection come from limited-access highways.

Recommendation: A refuge island, coupled with dedicated intervals for pedestrians and left-turners, will help to minimize conflict between pedestrian and vehicle traffic. Additionally, lighting at the refuge islands, lowering speed limits along Lancaster avenue, and clearly posting the speed limits could be helpful, as the speed limit is not signed along Lancaster avenue after the freeway off-ramp.



Figure B.9 Riverside Dr. & Lancaster Avenue, Ft. Worth

Congress Avenue & Cesar Chavez Street, Austin

Description: This intersection lies at the heart of downtown Austin, between the central business district and the South Congress Avenue bridge, a popular tourist attraction.

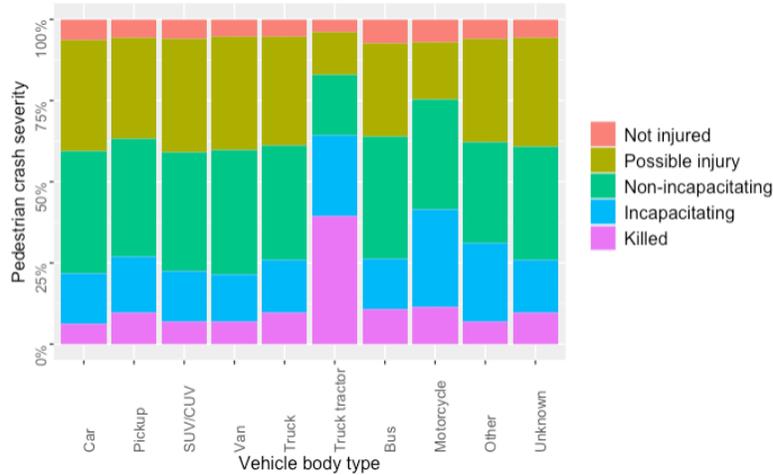
Recommendations: Improving lighting conditions along the channelized right turn, as well as requiring right-turning vehicles to stop for pedestrians, rather than yield, would help to slow down speeds at this crossing. Additionally, given the high volume of pedestrian traffic, a dedicated pedestrian interval would help to keep pedestrian and vehicle traffic separate, especially during the frequent special events that downtown Austin sees.



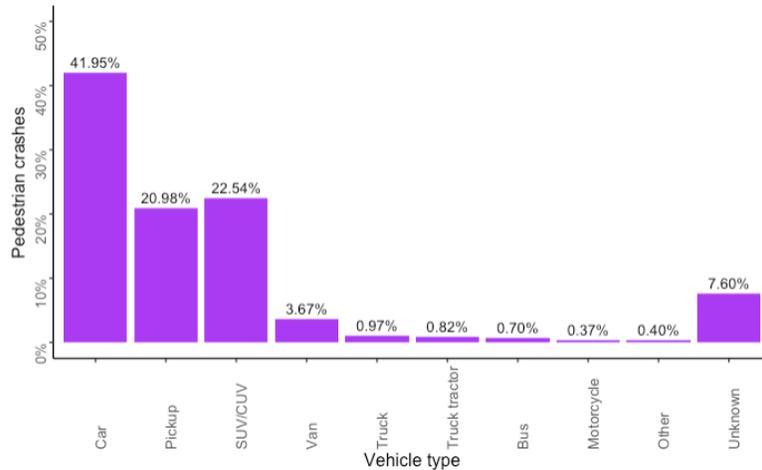
Figure B.10 Congress Avenue & Cesar Chavez Street, Austin

Appendix C. Vehicle Make/Model Obscurity

A preliminary hypothesis was considered postulating that a vehicle's shape or type affects pedestrian crash severity. CRIS data from the 2010–2019 analysis period reveals that pickup vehicles are associated with more fatal pedestrian-related crashes than passenger cars, even though cars are represented in approximately twice as many crashes as pickup trucks, as shown in Figure C.1.



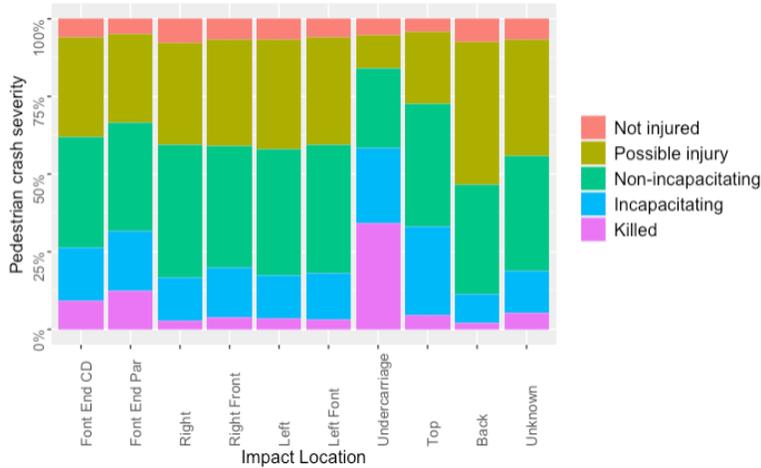
a) Pedestrian crash severity by vehicle body type



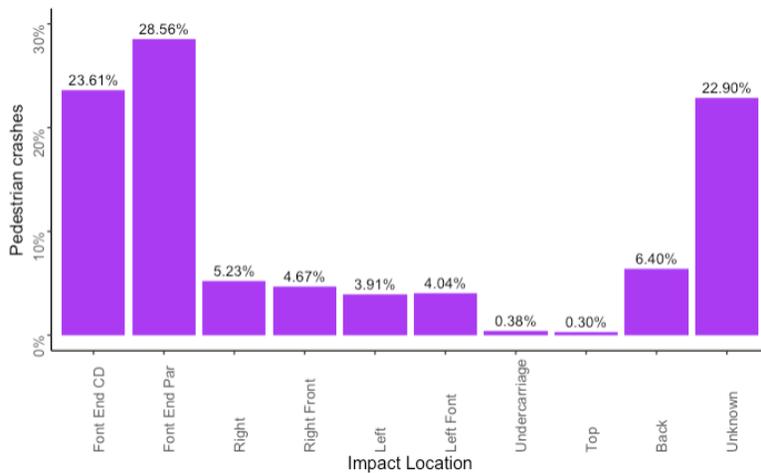
b) Percentage of pedestrian crashes by vehicle type

Figure C.1. Analysis of vehicle type relative to pedestrian crashes

The analysis of the impact location, shown in Figure C.2, suggests that this factor also affects the severity of the pedestrian crash. Impacts in the front areas are significantly more frequently observed than other locations. More than half of all crashes have a front impact location. The undercarriage is the most dangerous location, with nearly 30% of these crashes ending in a fatality, although the number of crashes with this type of impact is low (less than 1%).



a) Pedestrian crash severity by impact location



b) Percentage of pedestrian crashes by impact location

Figure C.2. Analysis of impact location relative to pedestrian crashes

Percentage Obscured

Head injuries are the dominant cause of fatality, followed by chest injuries (Ruikar, 2011). The effect of the vehicle front-end shape suggests that different vehicle types have different effects on the impact to the head and chest, as shown in Figure C.3. In efforts to characterize vehicle type's effect on pedestrian crash severity, a simple measurement of "vehicle obscurity" was devised. This measurement is expected to vary between different vehicle types, covering the area of the vehicle that would most likely collide with the head and chest of a pedestrian. As such, through regression analysis, it can be a possible predictor for severity. For these preliminary efforts, vehicle obscurity was obtained for about 60% of all vehicle makes/models involved in Texas pedestrian crashes, accounting for 66% of the total US VMT.

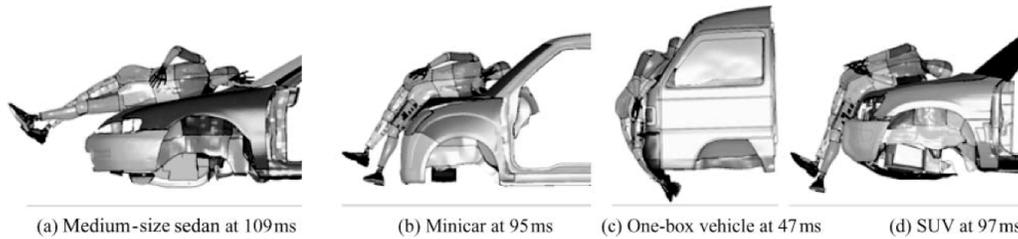


Figure C.3. Head dynamic kinematics in collisions with different vehicle types (Han et al., 2012)

The vehicle obscurity “%obscured” was estimated as the percentage of a vehicle image occupied by black pixels using a scaled profile picture of each vehicle and a selection of a 7-foot-square area, as exemplified in Figure C.4. On average, the vehicles analyzed have 49.8% obscurity with a standard deviation of 5.8%. Table C.1 shows descriptive statistics of the estimates.

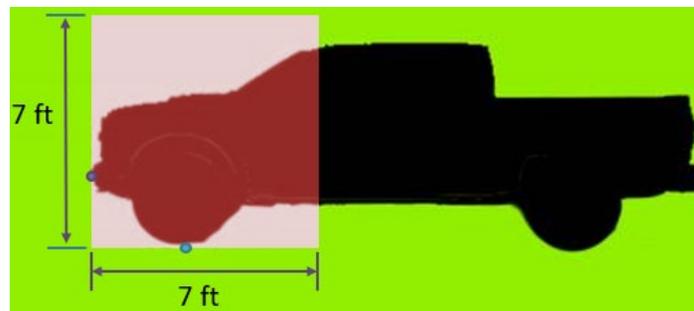
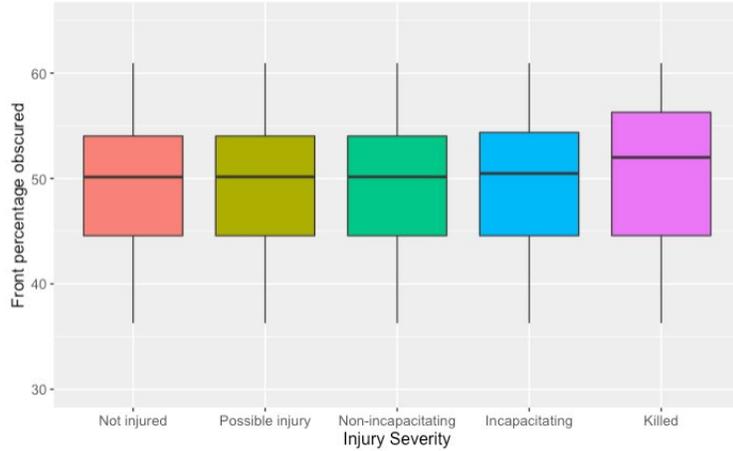


Figure C.4 Vehicle obscurity measure example for a pickup truck

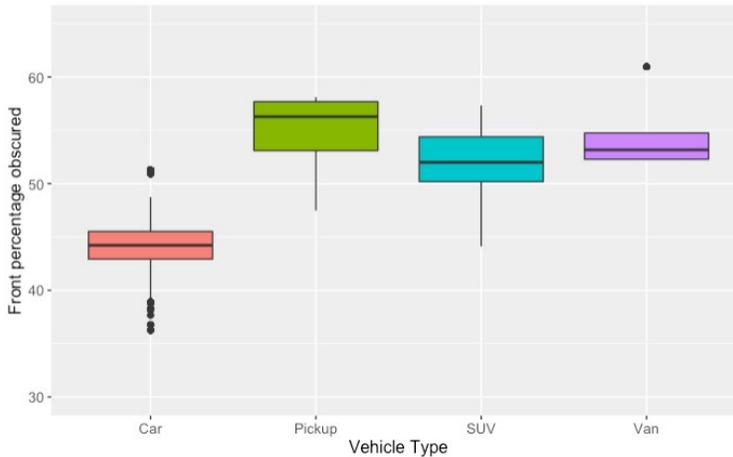
Table C.1. Percentage obscure descriptive statistics

Description	Estimates
Minimum	36.7%
Median	44.6%
Mean	49.8%
Standard Deviation	5.8%
Maximum	61.0%

Figure C.5 provides a description of the distribution of the percent obscured by injury severity and vehicle type. The results suggest that, on average, a higher front percent obscured is related to an increase in fatal accidents. The distribution by vehicle type suggests that pickup vehicles have the highest values, followed by vans, SUVs, and passenger vehicles.



a) Distribution of percentage obscured by injury severity



b) Distribution of percentage obscured by vehicle type

Figure C.5. Analysis of front percentage obscurity relative to crash severity and vehicle type

Most notably, vehicle obscurity for pickup trucks and SUV is higher than for sedans. For example, a Ford F250 pickup truck has a vehicle obscurity of 57.1%, while a Honda Accord sedan has a vehicle obscurity of 42.9% (refer to Figure C.6). Figure C.7 shows a summary of vehicle makes/models along with vehicle representation on Texas roadways in terms of VMT, the vehicle obscurity measure, and percentage of pedestrian-related crashes that result in fatalities and severe injuries.



a) Ford F250
%Obscured = 57.1%



b) Honda Accord
%Obscured = 42.9%



c) Acura Legend
%Obscured = 37.7%

Figure C.6. Examples of specific vehicles' obscurity estimates

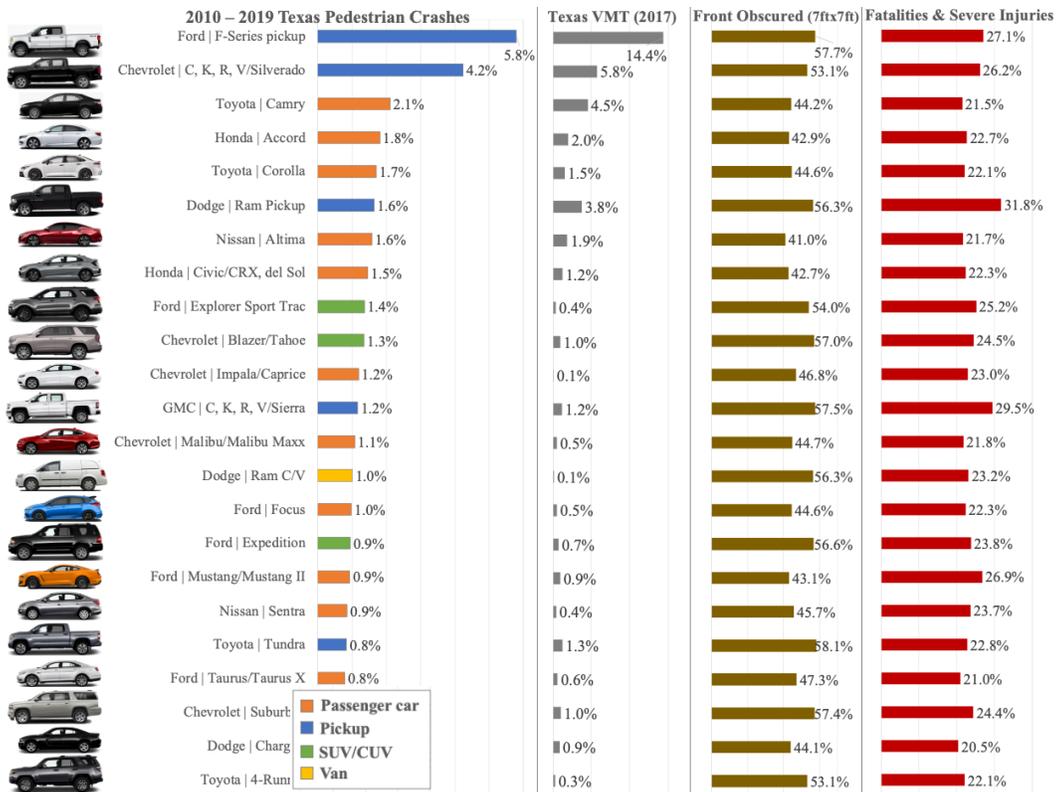


Figure C.7. Vehicle make/model characteristics pertaining to fatalities and severe injuries

Modeling Results

It is possible to investigate the use of vehicle obscurity and other factors such as impact type and location in the prediction of pedestrian crashes' severe injury or fatality likelihood. Injury severity models were constructed using ordered probit (OP) and heteroskedastic ordered probit (HOP) to understand the effects of different variables. Results are shown in Table C.2 along with a marginal effect analysis in Table C.3. A summary of the main results is as follows:

- A high front percentage obscured generally increases pedestrian severity.
- Newer car models generally lead to more severe injuries.
- An increase of 19.5% (3.36 standard deviations) in the front percentage obscured increases the probability of a pedestrian being killed or seriously injured by 0.43%.
- The front and top impact locations increase pedestrian severity compared to impacts on the back end (5 to 11%).
- Left front impacts increase severity compared to the right front.
- Undercarriage damage area collisions are significantly more severe.

- Compared to back-end collisions, undercarriage damage leads to a 36% increase in the probability of being severely injured or killed.
- Pedestrian and driver intoxication is the most significant variable. The probability of pedestrian fatality in these crashes is 69% greater.
- The severity increases with the pedestrian's age (vulnerability) and decreases with the driver's age (experience).
- Male pedestrians experience more severe injuries compared to women.
- Male drivers tend to cause crashes with higher severity of injuries.
- Hit-and-runs result in higher severity of injuries.
- Compared to city streets, crashes on interstate, US state, and farm-to-market highways are more severe.
- Interstate crashes increase the probability of being killed or seriously injured by 14.3%.
- County road and non-traffic ways tend to have crashes with lesser injury severity.
- Irregular geometry and speed increase the likelihood of severe pedestrian injuries.
- Crashes at intersections are more likely to be less severe.
- Areas with traffic control are less likely to be locations of severe crashes.
- Urban areas tend to have less severe pedestrian crashes.
- The probability of crashes in the early morning hours leading to death or serious injury is 5.5% greater compared to night hours.

Summary and Future Work

An exploratory creation of OP and HOP models shows promise for predicting injury severity using both vehicle obscurity and model year. Both of these demonstrated a significantly positive relationship with increased injury severity; in addition, undercarriage crash damage was closely associated with severe injuries.

This exploration was not pursued further within this project, which prioritizes reducing pedestrian crash likelihood through roadway treatments. However, further work, especially in formalizing the model and validating results, could significantly aid in policy creation around and general awareness of how vehicle design impacts pedestrian crash outcomes.

Table C.2. Injury severity models

	OP			HOP		
	Estimate	p-value		Estimate	p-value	
Vehicle Information						
Front percentage obscured	0.0055	0.0000	*	0.0071	0.0000	*
Model year	0.0004	0.0000	*	0.0005	0.0000	*
Impact Location						
Front end concentrated & distributed	0.2456	0.0000	*	0.2733	0.0000	*
Front end left/right damage partial	0.3325	0.0000	*	0.3793	0.0000	*
Right back/side & top/side angular	-0.0886	0.0243	*	-0.0548	0.2319	
Right front quarter angular	0.0451	0.2547		0.0768	0.0966	
Left back/side & top/side angular	-0.0436	0.3025		0.0006	0.9911	
Left front quarter angular	0.1301	0.0016	*	0.1791	0.0002	*
Undercarriage damage	0.9731	0.0000	*	1.2897	0.0000	*
Top damage	0.3443	0.0024	*	0.3794	0.0028	*
Unknown	0.0202	0.5459		0.0325	0.3937	
(Reference type = Back end)						
Pedestrian & Driver Characteristics						
Pedestrian age	0.0040	0.0000	*	0.0043	0.0000	*
Pedestrian gender (Male = 1)	0.0328	0.0142	*	0.0355	0.0298	*
Driver age	-0.0013	0.0004	*	-0.0015	0.0009	*
Driver gender (Male = 1)	0.0436	0.0012	*	0.0548	0.0010	*
Pedestrian/driver intoxicated (Yes = 1)	1.7354	0.0000	*	3.7812	0.0000	*
Hit-and-run (Yes = 1)	0.0987	0.0006	*	0.0732	0.0532	
Roadway Type						
Interstate	0.3418	0.0000	*	0.4591	0.0000	*
U.S. State	0.1906	0.0000	*	0.2361	0.0000	*
Farm to market	0.1528	0.0000	*	0.1749	0.0000	*
County road	-0.0904	0.0283	*	-0.1342	0.0199	*
Non-traffic way	-0.1663	0.0000	*	-0.2059	0.0000	*
Other (e.g., tollway, toll bridges)	0.0945	0.3703		0.1374	0.3089	
(Reference type = City streets)						
Road Geometry & Speed						
Straight grade	0.1648	0.0000	*	0.2128	0.0000	*
Curved	0.2199	0.0000	*	0.2755	0.0000	*
(Reference type = Straight & level)						
Intersection (Yes = 1)	-0.3418	0.0000	*	-0.2514	0.0000	*
Speed limit (mph)	0.0103	0.0000	*	0.0110	0.0000	*
Control Type						
Traffic sign	-0.1099	0.0000	*	-0.1353	0.0000	*
Traffic signal	-0.1917	0.0000	*	-0.2174	0.0000	*
Other (e.g., human control, rail gate)	-0.0444	0.2062		-0.0479	0.2364	
(Reference type = No control)						
Area Population						
Rural	0.0826	0.0031	*	0.1278	0.0013	*
Small town (500 - 100k)	0.0854	0.0000	*	0.1061	0.0000	*
Large town (100k - 250k)	0.0690	0.0003	*	0.0878	0.0001	*

	OP			HOP		
(Reference type = Urban > 250k)						
Crash Time						
Early morning (12 AM to 6 AM)	0.0928	0.0007	*	0.1673	0.0000	*
Morning (6 AM to 12 PM)	-0.0368	0.0892		-0.0539	0.0432	*
Afternoon (12 PM to 6 PM)	-0.0469	0.0280	*	-0.0555	0.0324	*
Reference (Night from 6 PM to 12 AM)						
Lighting Condition						
Dark lighted	0.1199	0.0000	*	0.1303	0.0000	*
Dark not lighted	0.3092	0.0000	*	3.75E-01	0.0000	*
(Reference type = Daylight)						
Threshold Parameters						
μ_1	1.3429	0.0000	*	1.6294	0.0023	*
μ_2	2.4319	0.0000	*	2.9940	0.0023	*
μ_3	3.3754	0.0000	*	0.4088	0.0023	*
Model Fit Statistics						
Log-Likelihood	-34,920			-34,495		
McFadden's R2	0.0944			0.1055		
Akaike information criterion (AIC)	69,922			69,138		
χ^2 Likelihood ratio test (LRT)	849.40	0.0000	*			
Sample size	31,772			31,772		

*Note: conditions to reject the null hypothesis with a 95% confidence level.

Table C.3. Marginal Effects (HOP)

		No Injury	Possible Injury	Non-Incapacitating Injury	Suspected Serious Injury	Killed
Vehicle Information						
Front percentage obscured	*	-0.05%	-0.15%	0.01%	0.13%	0.00%
Model year	*	0.00%	-0.01%	0.00%	0.01%	0.06%
Impact Location						
Front end concentrated & distributed		-1.04%	-5.49%	-1.71%	4.76%	3.48%
Front end left/right damage partial		-1.60%	-7.65%	-1.98%	6.61%	4.62%
Right back/side & top/side angular		-0.12%	1.17%	1.06%	-1.13%	-0.97%
Right front quarter angular		-0.53%	-1.61%	0.12%	1.37%	0.66%
Left back/side & top/side angular		-0.12%	-0.02%	0.27%	-0.01%	-0.12%
Left front quarter angular		-1.10%	-3.72%	-0.07%	3.18%	1.71%
Undercarriage damage		-3.27%	-18.55%	-14.38%	11.43%	24.77%
Top damage		-3.09%	-9.37%	3.18%	7.65%	1.64%
Unknown		-0.13%	-0.67%	-0.17%	0.59%	0.38%
(Reference type = Back end)						
Pedestrian & Driver Characteristics						
Pedestrian age	*	-0.02%	-0.09%	-0.02%	0.08%	0.05%
Pedestrian gender (Male = 1)		0.46%	-0.72%	-1.60%	0.77%	1.08%
Driver age	*	0.01%	0.03%	0.00%	-0.03%	-0.01%
Driver gender (Male = 1)		-0.26%	-1.15%	-0.19%	1.00%	0.60%
Pedestrian/driver intoxicated (Yes = 1)		-3.83%	-27.73%	-32.41%	-5.09%	69.06%
Hit-and-run (Yes = 1)		0.96%	-1.42%	-3.15%	1.23%	2.38%
Roadway Type						
Interstate		-0.87%	-7.97%	-5.45%	6.29%	8.00%
U.S. State		-0.33%	-4.53%	-2.88%	3.91%	3.83%
Farm to market		-0.61%	-3.47%	-1.26%	3.01%	2.32%
County road		0.96%	2.84%	-0.29%	-2.39%	-1.10%
Non-traffic way		0.85%	4.47%	0.85%	-3.92%	-2.25%
Other (e.g., tollway, toll bridges)		-0.04%	-2.64%	-2.02%	2.29%	2.40%
(Reference type = City streets)						
Road Geometry & Speed						
Straight grade		-1.99%	-7.14%	-0.53%	6.14%	3.53%
Curved		-2.39%	-9.11%	-1.22%	7.86%	4.87%
(Reference type = Straight & level)						
Intersection (Yes = 1)		1.27%	5.41%	0.53%	-4.68%	-2.54%
Speed limit (mph)	*	-0.05%	-0.23%	-0.04%	0.20%	0.12%
Control Type						
Traffic sign		0.91%	2.85%	-0.18%	-2.42%	-1.17%
Traffic signal		1.42%	4.63%	-0.22%	-3.93%	-1.90%
Other (e.g., human control, rail gate)		0.12%	1.02%	0.41%	-0.91%	-0.63%
(Reference type = No control)						
Area Population						
Rural		1.83%	-2.46%	-5.65%	1.83%	4.45%
Small town (500 - 100k)		0.73%	-2.07%	-3.18%	1.87%	2.66%

		No Injury	Possible Injury	Non-Incapacitating Injury	Suspected Serious Injury	Killed
Large town (100k - 250k)		-0.62%	-1.85%	0.18%	1.56%	0.73%
(Reference type = Urban > 250k)						
Crash Time						
Early morning (12 AM to 6 AM)		-0.23%	-3.24%	-2.03%	2.82%	2.68%
Morning (6 AM to 12 PM)		0.26%	1.14%	0.18%	-0.99%	-0.58%
Afternoon (12 PM to 6 PM)		0.21%	1.17%	0.30%	-1.02%	-0.66%
Reference (Night from 6 PM to 12 AM)						
Lighting Condition						
Dark lighted		-0.20%	-2.61%	-1.49%	2.32%	1.98%
Dark not lighted		-1.29%	-7.18%	-2.91%	6.12%	5.26%
(Reference type = Daylight)						

* Note: continuous variables include the effect of one standard deviation change.

Appendix D. CR-3 Training Results

CR-3 forms are instrumental in facilitating the collection and filing of information for each reported crash in Texas. TxDOT needs accurate CRIS data for good analysis, appropriate funding, and public information campaigns. For each reported crash, the peace officer or other qualified crash reporter that visits the crash site to report it fills out a CR-3 form. While the CR-3 form requires the entry of narrative text to describe the crash, as well as a sketch depicting an overhead view of vehicle positions and movements that contribute to a crash, most of the CR-3 involves entering distinct pieces of information into a series of fields, including a number of fields that request numeric codes that signify categorical, predefined answers.

Because it is a high priority for TxDOT to store most of the contents of CR-3 fields in a database for future analysis, it is important to ensure a high degree of accuracy. Much of this accuracy comes from a consistent understanding among all who are qualified to fill out CR-3 forms. For example, for the “At Intersection” field to be trustworthy in queries, all personnel should have a common understanding of what exactly an “intersection-related crash” is, including where to delineate an intersection on roadways. To address the most common mistakes seen in reporting, TxDOT offers a free 2-hour interactive training class called “On the Road to Zero: How to Best Complete a Texas CR-3” that covers a number of key factors in consistently filling out a CR-3.

Researchers have found that this training is instrumental for them in properly interpreting fields found within the CRIS database. Not only do researchers better understand the purpose of various fields, they also have realistic ideas on where inconsistencies caused by variations in understanding or common mistakes may occur within crash records. Researchers also learn the data entry and cleaning process that TxDOT utilizes when CR-3 submissions are processed, and how the CRIS “interpreted fields” (or those that are derived from narrative, drawings, and other information) are provided.

These are key takeaways researchers have noted from attending the training:

Ground Rules

- By law, officers must report within 10 days to TxDOT. But many agencies have a close-of-business-day or next-day requirement. Most reports are done offsite. TxDOT also provides an online app for CRIS reports (called “CRASH”), to make reporting and editing more accessible. Every report is reviewed by TxDOT staff.
- When a CR-3 form is coded and analyzed by TxDOT, it is subjected to a series of business rules to check for correctness and consistency. If flags arise, the officer is informed of these and is requested to submit a supplemental report to correct those flags. When such a request is issued, the report has the “Open ETL” status.

- Fatalities must have occurred within 30 days of a crash. A 31-days-post-crash death cannot be labeled as a fatal crash.
- Non-collision events are reportable.
- For fields that allow for numeric, categorical entry (e.g., contributing unit factor), the option to enter “98 Other” is provided to cover situations where no other option is appropriate, and explanation must be put into the narrative. Officers are strongly encouraged to only use “98 Other” as a last resort, as it greatly reduces the usefulness of the crash record in the database.
- There are about 33,000 crashes that are reported each year in Texas (leaving the possibility of many others not being reported or recorded). Average number of reports per eligible officer per year is 1. This underscores the challenge for all officers to fill out the CR-3 form properly when many are reporting very few crashes per year.

Locations

- In cases where the crash reporter does not know the date and time (or even exact location) of a crash, the discovery time and location should be reported.
- Highway number is preferred over the street name.
- To distinguish between a frontage road and mainline, the “street description” field should be used, as well as “roadway part.”
- A crash is only to be regarded as “at intersection” if the crash occurs inside the “box” of an intersection. If the crash occurs in a separate right-turn yield section of roadway near an intersection, or in a roundabout, this is not to be regarded as “at intersection”
- If the crash even partly occurs on a public roadway, the “private drive” box should not be checked. A crash is to be regarded as “on a private drive” if the crash originated on a private drive (including a driveway or parking lot).
- A crash originating in a publicly owned parking lot, such as at a state university or government building, should be marked as occurring “on a private drive.” The main distinction is that a public right-of-way is not involved in the crash.
- Geographic coordinates (e.g., GPS lat/lon) are not often reported by officers. The officer provides distances to mile markers, or nearby intersections or addresses. There are 33,000 officers in Texas, and many don’t have internet access. As a result, they can’t locate the crash sites with GPS unless they use their own phone, which then opens them up for FOIA requests.

- Crash locations are often pinned to the centerline, because roads' alignments are altered, roads are widened, etc., which would require "re-pinning" the crash, a time-consuming task that is not undertaken. This also reduces the risk that crash locations would be lost as GIS layers are updated year by year. This project's researchers devised a process for improving this that is documented in Appendix E ("Crash Direction" tech doc), but the process has not yet been implemented.

Units

- All vehicles involved in a crash, including parked and hit-and-run vehicles, are to be recorded as "units."
- It is usual but not necessary that Unit #1 is what the officer suspects to be the "at fault" vehicle. However, Unit #1 must be a vehicle in motion, never parked. Unit #1 also cannot be a train.
- People in a parked vehicle should not be listed as a driver (even if they're in the driver's seat), since they're not driving at the time.
- Everyone's address must be provided. Registered owner's address can be retrieved via the license plate in the Texas Law Enforcement Telecommunications System (TLETS).
- VIN numbers preceding 1980 are shorter than the VIN numbers seen today.
- There are a number of subtleties around proper vehicle classification:
 - A truck (which must be marked as either a medium-duty truck or heavy-duty truck, MDT or HDT respectively) is to be reported as a different vehicle type than a pickup truck (<8500 GVWR, a light-duty truck or LDT).
 - A FedEx delivery vehicle is a MDT, but it should be listed as a "box truck."
 - Any school transport should be coded as a "yellow school bus."
 - If an emergency response vehicle is involved in a crash, the CR-3 field having to do with "emergency response vehicle" should only be checked if the emergency response vehicle was actually en route to an emergency.
- Items involved in a collision that don't have an owner should not be listed as units. But, they should be described in the narrative. For example, a tree in a right-of-way or an alligator that crawled from a nearby swamp don't have owners. Correspondingly, items that do have owners (such as a tree that is a prominent fixture in a property owner's front yard or a pet alligator) should be listed if there is consequential monetary damage to the owner.

- An “automation” field now exists in the CR-3 form. This is appropriate for vehicles that have some degree of driving automation enabled at the time of the crash.

What to Report

- The training offers a number of uncommon use cases in order to exemplify correct reporting. In some instances CR-3 reporting is not necessary, e.g., when a crash occurs off of a roadway. If a crash occurs off the roadway with a train, and the train hits a parked vehicle, it is not reported in a CR-3, or it is classified as non-reportable by the agency at a later time.
- Tractors with trailers should not list the trailer as a separate unit (so the commercial driver doesn’t have his/her record suggest two crashed units by him/her). It is to be regarded as an added towed unit that is part of the main (power) unit.
- If a vehicle that is parked on top of something else is hit (e.g., a flatbed trailer), that vehicle is considered property, not a separate driving unit.

Any records that can’t be properly coded are put into an “unprocessed queue” for review by TxDOT. The officer should submit a supplement, and the report will be kept for 10 years. (An example that would cause this: a parked car is hit and mentioned in the narrative, but not entered as a separate unit.)

Appendix E. Datasets and Technical Documentation

Technical documentation outlining data processing aspects of the project, including database creation, data importing, algorithms, data table creation, etc., are provided as Markdown documents hosted in the GitHub “ped-crash-techvol” repository. This is titled “Texas Ped Crash Tech Volume Pack.” This is also where datasets produced through the course of the project that are to be publicly shared are referenced, including the “Texas Top 100” most crash-prone corridors, as well as the Python code that created them. The following is a reprint of the repository’s “readme” file. All contents are accessible at <https://github.com/ut-ctr-nmc/ped-crash-techvol>

Introduction

This repository contains technical documentation and source code that were used to analyze data from TxDOT’s Crash Records Information System (CRIS) and other sources to determine causes of pedestrian-related crashes, and to assist in determining the best roadway treatments for mitigating the most severe pedestrian injuries and fatalities. While this was a project using Texas data, the processes and results may be applicable to other locations.

Another repository “peds-midblocks-intersections” (<https://github.com/ut-ctr-nmc/peds-midblocks-intersections>) had been created that contains the results of the methods described for finding intersections and 0.1-mile resampled roadway segments from the TxDOT Roadway Inventory.

This documentation was written by Kenneth Perrine, Research Associate at Center for Transportation Research at The University of Texas at Austin. Licensed under the MIT License.

Contents

Database Preparation

- Database Functions: Outlines database tables, queries and access
- Importing Major Data Files: Importing CRIS Share and TxDOT Roadway Inventory into the database
- Crash Statistics for 2010–2019: Summary queries for crash data
- Other Lookup Tables: Preparing for queries around vehicle make/model

Initial Crash Matching and Analysis

- Crash Stats Segments Breakdown: Queries for summarizing crash-prone areas of TxDOT Roadway Inventory data

- Clustering: A first attempt at grouping clusters of crashes around intersections for hotspot analysis

Roadway Inventory

- Uniform Segments: First round of resampling TxDOT Roadway Inventory to 1-mile segments, plus crash-matching
- 0.1-mile Uniform Segments: Second round of resampling 0.1-mile segments plus crash-matching
- Intersection: Strategies for mapping intersections to TxDOT Roadway Inventory, including the use of OpenStreetMap
- Multi-Year Intersections: Processing multi-year AADT estimates from TxDOT Roadway Inventory for intersections

Subsequent Analysis

- BCR Corridors: Documents the final “Top 100 worst corridors” ranking strategy used in the project
- Analysis that Includes Sidewalks: Further statistics on Roadway Inventory plus use of sidewalk data

Supporting Activities

- GitHub Preparations: Instructions for preparing the “peds-midblocks-intersections” (<https://github.com/ut-ctr-nmc/peds-midblocks-intersections>) dataset
- VIN Testing: Additional analysis that uses VIN numbers as recorded in CRIS
- Crash Direction: Documents future work that would be needed to more positively position crashes relative to roadway geometry