8-1-2020

Comprehensive Safety Analysis of Vulnerable Road User Involved Motor Vehicle Crashes

Farah Jazi Al-Mahameed
University of Wisconsin-Milwaukee

COMPREHENSIVE SAFETY ANALYSIS OF VULNERABLE ROAD USER INVOLVED MOTOR VEHICLE CRASHES

by

Farah J. Al-Mahameed

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Engineering

at

The University of Wisconsin-Milwaukee

August 2020

ABSTRACT

COMPREHENSIVE SAFETY ANALYSIS OF VULNERABLE ROAD USER INVOLVED MOTOR VEHICLE CRASHES

by

Farah J. Al-Mahameed

The University of Wisconsin-Milwaukee, 2020
Under the Supervision of Professor Xiao Qin

This dissertation explores, identifies, and evaluates a multitude of factors significantly affecting motor vehicle crashes involving pedestrians and bicyclists, commonly defined as vulnerable road users (VRUs). The methodologies are guided by the concept of safe behavior of different parties that are primary responsible for a crash, either a pedestrian, a bicyclist or a driver, pertaining to roadway design, traffic conditions, land use and built environment variables; and the findings are beneficial for recommending targeted and effective safety interventions.

The topic is motivated by the fact that human factors contribute to over ninety percent of the crashes, especially the ones involving VRUs. Studying the effect of road users' behavior, their responses to the dynamics of traveling environment, and compliance rate to traffic rules is instrumental to precisely measure and evaluate how each of the investigated variables changes the crash risk. To achieve this goal, an extensive database is established based on data collected from sources such as the linework from topologically integrated geographic encoding and referencing, Google maps, motor vehicle accident reports, Wisconsin Information System for Local Roads, and Smart Location Dataset from Environmental Protection Agency. The crosscutting datasets represent various aspects of motorist and non-motorists travel decisions and behaviors, as well as their safety status. With this comprehensive database, intrinsic relationships between pedestrian-vehicle crashes and a broad range of socioeconomic and demographic factors, land use and built

environment, crime rate and traffic violations, road design, traffic control, and pedestrian-oriented design features are identified, analyzed, and evaluated.

The comprehensive safety analysis begins with the structural equation model (SEM) that is employed to discover possible underlying factor structure connecting exogenous variables and crashes involving pedestrians. Informed by the SEM output, the analysis continues with the development of crash count models and responsible party choice models to respectively address factors relating to roles in a crash by pedestrians and drivers. As a result, factors contributing to crashes where a pedestrian is responsible, a driver is responsible, or both parties are responsible can be specified, categorized, and quantified. Moreover, targeted and appropriate safety countermeasures can be designed, recommended, and prioritized by engineers, planners, or enforcement agencies to jointly create a pedestrian-friendly environment.

The second aspect of the analysis is to specify the crash party at-fault, which provides evidence about whether pedestrians, bicyclists or drivers are more likely to be involved in severe crashes and to identify the contributing factors that affect the fault of a specific road user group. An extensive investigation of the available information regarding the crash (i.e., issued citations, actions/circumstances that may have played a role in the crash occurrence, and crash scenario completed by the police officer) are considered. The goal is to recognize and measure the factors affecting a specific party at-fault. This provides information that is vital for proactive crisis management: to decrease and to prevent future crashes. As a part of the result, a guideline is proposed to assign the party at-fault through crash data fields and narratives. Statistical methods such as the extreme gradient boosting (XGboost) decision tree and the multinomial logit (MNL) model are used. Appealing conclusions have been found and suggestions are made for law enforcement, education, and roadway management to enhance the safety countermeasures.

The third aspect is to evaluate the enhancements of crash report form for its effectiveness of reporting VRU involved motor vehicle crashes. One of the State of Wisconsin projects aiming to develop crash report forms was to redesign the old MV4000 crash report form into the new DT4000 crash report form. The modification was applied from January 1, 2017, statewide. The reason behind this switch is to resolve some matters with the old MV4000 crash report form, including insufficient reporting in roadway-related data fields, lack of data fields describing driver distraction, intersection type, no specification of the exact traffic barrier, insufficient information regarding safety equipment usage by motorists and non-motorists, unclear information about the crash location, and inadequate evidence concerning non-motorists actions, circumstances and condition prior to the crash. Hence, the new DT4000 crash form modified some existing data fields incorporated new crash elements and more detailed attributes. The modified and new data fields, their associated attribute values have been thoroughly studied and the effectiveness of improved data collection in terms of a better understanding of factors associated with and contributing to VRU crashes has been comprehensively evaluated. The evaluation has confirmed that the DT4000 crash form provided more specific, details, and useful about the crash circumstances.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

Acronyms, Abbreviations	Explanation
AADP/AADB	Pedestrian/Bicycle Volume
AADT	Average Annual Daily Traffic
AI	Artificial Intelligence
BAC	Blood Alcohol Concentration
CHAID	Chi-Square Automatic Interaction Detector
CART	Classification And Regression Trees
CA	Correspondence Analysis
CCTV	Closed-Circuit Television
CFA	Confirmatory Factor Analysis
CP	Complexity Parameter
CPM	Crash Prediction Model
DMV	Department of Motor vehicle
DMTV	Daily Vehicle Miles Traveled
DOT	Department of Transportation
DT	Decision Tree
DTM	Document Term Matrix
DUI	Driving Under Influence
EFA	Exploratory Factor Analysis
GEV	Generalized Extreme Value
FA	Factor analysis
FARS	Fatality Analysis Reporting System
GBDT	Gradient Boosted Decision Tree
GLM	Generalized Linear Model
HSIS	Highway Safety Information System
LSA	Latent Semantic Analysis
MCA	Multiple Correspondence Analysis
MDA	Mean Decrease Accuracy
MDG	Mean Decrease Gini
ML	Mixed Logit
MLR	Multinomial Logistic Regression
MMUCC	Model Minimum Uniform Crash Criteria
MNL	Multinomial Logit
MV4000	Wisconsin Motor Vehicle Accident Report
MVPLN	Multivariate Poisson-Lognormal Models
MVA	Multivariate analysis
NB	Negative Binomial
NHTSA	National Highway Traffic Safety Administration
NLP	Natural Language Processing
OL	Ordered Logit
OOB	Out-of-Bag
OL/OP	Ordered Logit/Probit
PA	Path Analysis

PBCAT	Pedestrian and Bicycle Crash Analysis
PCA	Principal Component Analysis
PLN	Poisson-lognormal
PPO	Partial Proportional Odds
QUEST	Quick, Unbiased and Efficient Statistical Trees
RF	Random Forests
ROW	Right-of-Way
SEM	Structural Equation Model
SVM	Support Vector Machine
SVD	Singular Value Decomposition
SLD/EPA	Smart Location Dataset from The Environmental Protection Agency
SRM	Structural Regression Model
TAZ	Traffic Analysis Zone
TDM	Term Document Matrix
TCD	Traffic Control Device
TF-IDF	Term-Frequency-Inverse Document Frequency
TIGER/LINE	Topologically Integrated Geographic Encoding and Referencing
UA	Univariate Analysis
VKT	Vehicle Kilometers Traveled
VMT	Vehicle Miles Traveled
VRU	Vulnerable Road User
WISLR	Wisconsin Information System for Local Roads
XGBoost	Extreme Gradient Boosting
ZINB	Zero-inflated Negative Binomial
ZIP	Zero-Inflated Poisson

GREEK SYMBOLS

ξ	n x 1 Column Vector of Latent Exogenous Variables
η	m x 1 Column Vector of Latent Endogenous Variables
δ	q x 1 Column Vector of Measurement Error Terms for Observed Variables x
ε	p x 1 Column Vector of Measurement Error Terms for Observed Variables y
Γ	The Matrix (m x n) of Regression Effects for Exogenous Latent Variables to Endogenous Latent Variables
β	The Coefficient Matrix (m x m) of Direct Effects Between Endogenous Latent Variables
ς	Vector of Error Terms

SUBSCRIPTS AND SUPERSCRIPTS

x	Vector of Observed Exogenous Variables
y	Vector of Observed Exogenous Variables
τ_i	Degree of Misspecification of the Model

ACKNOWLEDGMENTS

With God all things are possible. First of all, I thank with humility my Lord and Almighty for his grace and blessing hands, in helping me through everything.

This journey would not have been possible without the support of my supervisor (Prof. Xiao Qin), professors and mentors (Assoc. Prof. Robert Schneider, Asst. Prof. Jie Yu, Asst. Prof. Yin Wang, Assoc. Prof. Wilkistar Otieno), husband (Dr. Ahmad Abbas), son (Amir), family and friends.

With great pleasure and heartfelt gratitude, I express my thanks to my main research supervisor, Professor Xiao Qin, the Director of the Safe and Smart (S^2) Traffic Laboratory at the University of Wisconsin-Milwaukee, for his constant intellectual support for more than four years. He is the busiest and the most available person at the same time. Prof. Qin, your clarity of thinking is something I will always aim for. Prof. Robert Schneider, I owe you a debt of gratitude for your time and careful attention to details, your vision always succeeded to enrich my knowledge. Prof. Yu, you are my role model as one of the few female faculty members, thank you for showing me what it means to be dedicated in your own unique way. Prof. Wang, thank you for your willingness to help me out, for offering an open ear to all of my questions. Prof. Otieno, thank you for modeling great teaching and for furthering my thinking in Statistics. I thank you all for your untiring support and guidance throughout my journey and for giving your time to attend and give your feedback through my proposal and dissertation defense.

To my husband Ahmad, thank you for supporting me as we go through the Ph.D. journey together. My son Amir, thank you for adding joy to my days and teaching me how to be more patient. To my family, father Jazi, mother Asma, father-in-law Ibrahim, mother-in-law Samira, brothers Maen and Mahmoud, sisters Halla and Marah, thank you all for encouraging me in all of my pursuits and inspiring me to follow my dreams. I am especially grateful to my parents, who supported me emotionally. I always knew that you believed in me and wanted the best for me. Thank you for teaching me that my job in life was to learn, to be happy, and to know and understand myself; only then could I know and understand others. To my friends in the united states and others who are scattered around the world, thank you for listening, well-wishes/prayers, phone calls, texts, and supporting me through this entire process.

I am thankful to the University of Wisconsin-Milwaukee for the Chancellor's Award and the teaching assistantships, which made this dissertation possible and provided me with both a teaching experience in the United States and great collogues.

Farah

Chapter 1 : Introduction

1.1 Pedestrian and Bicyclist Safety in Numbers

Walking and cycling are unique modes of transportation that provide a wide range of benefits for individuals and the public, such as increasing physical and psychological health, reducing environmental pollution, and maintaining low-stress levels. Cities around the world are becoming more aware of the added value that walking and biking bring to their communities in terms of healthier, safer, and more sustainable communities. This awareness has been informing the policymaking of city authorities that help grow this group of road users. According to USA Today, pedestrians are the largest single traveler group; moreover, the bicyclist's group is growing for different activities including work, commuting, and leisure during the past decade. In fact, the percentage of people biking to work has increased by 60% over the past decade, while the percentage of people walking to their jobs has remained the same. With the growth of the walking and biking population, non-motorists safety becomes a pressing issue. Pedestrians and bicyclists are commonly referred to as vulnerable road users (VRUs) since they cannot be protected by a seat belt or an airbag like a motor-vehicle driver, hence they are more likely to be severely injured when involved in a crash with a motor vehicle. In 2016, a total of 37,461 people lost their lives in car accidents, among which pedestrian and bicyclists' fatalities accounted for 16 percent and 2 percent, respectively (R. Retting and Consulting 2017).

Although a large number of cities across the United States have been trying to provide VRUs with a safer road environment. Statistics show that crashes related to such road users remain a major concern. In 2009, 4,092 fatalities (which accounted for 12 percent of all traffic fatalities) and 59,000 injuries (which made up 3 percent of all the people injured in traffic crashes) were recorded among pedestrians. On average, a pedestrian was killed every two hours and injured every nine minutes in traffic crashes. While in 2016, 5,900 traffic crashes occurred in the United States resulted in 6,080 deaths among pedestrians, showing that some of these crashes involved one or more pedestrian fatalities. A pedestrian was killed every 1.5 hours in traffic crashes. Fatalities among bicyclists accounted for 2.2 percent of all traffic fatalities in 2016 ("Fatality Analysis Reporting System (FARS) | NHTSA" 2016). Although VRUs reported crashes to represent low numbers, they make up most of the fatalities among all crashes. In Wisconsin, a pedestrian was injured/killed every 7.1 hours, and a bicyclist was injured/killed every 10.2 hours. According to the year 2016 crash statistics, 1,252 vehicle crashes involving pedestrians, resulted in 1,181 injuries and 49 fatalities among pedestrians. Whereas, 918 vehicle-crashes involved bicyclists, resulting in 849 injuries and 11 fatalities among bicyclists in the State of Wisconsin (R. Retting and Consulting 2017).

Walking and cycling are active transportation modes that are believed to help our communities be less car dependence and increase the overall physical activity levels (Lindsay, Macmillan, and Woodward 2011). Those sustainable modes of transportation play a vital role in reducing vehicle crashes and traffic congestion. This research is motivated by reducing the increasing traffic causalities and injuries concerning to VRUs, pedestrians in particular because of the high fatality count, and by contributing to a safe walking environment through extensive data collection and rigorous safety analysis and evaluation.

Speaking of pedestrian safety, pedestrians are more subject to certain risks and more vulnerable than occupants in a vehicle: i) unlike drivers, pedestrians lack physical protection which makes them unprotected from the vehicle's mass and speed (Wegman, Aarts, and Bastiaans 2006), ii) the size of a pedestrian is very small compared to the size of a vehicle, hence, it is not surprising that drivers overlook pedestrians in the roadway (Langham and Moberly 2003), iii) pedestrians vary in their knowledge level on traffic rules, thus, drivers may fail to predict their behavior which depends on their level of compliance with the roadway rules drivers (Zhou et al. 2013), iv) the dedicated space for both roadway users is different, for example a driver can only use specific lanes, whereas a pedestrian may use any chance to cross whether at marked/unmarked locations or any other convenient location (Hill 1984), v) pedestrians are more prone to misjudging cars' velocities or distances (Hills 1980) and usually commit failure to perceive vehicles due to obstructions or failing to maintain a high level of attention (Schofer et al. 1995; DiMaggio and Durkin 2002), and vi) in order for drivers to spot pedestrians, a wide horizontal angle of vision is required to locate pedestrians that are usually standing on the side of the road (Nugter et al. 2017;Shahar et al. 2010). Accordingly, research of factors affecting the behavioral errors or faults committed by both road users to support the development of targeted countermeasures designed for drivers, with focus also on pedestrians, is essential. Several pedestrians, roadway, driver, vehicle, temporal, and environmental-related factors are considered in this research.

1.1.1 Pedestrian-Bicyclist Issues

Commuters can be discouraged to use active transportation modes considering those modes are less safe than transporting by vehicles (R. J. Schneider 2013; Buehler and Dill 2016). If this situation continues, pedestrian/bicyclist facilities will be underutilized and investment decisions on non-motorists road users will become more difficult when competing with other demands. Numerous studies have been conducted to address the urgent need of accommodating safe travels for non-motorists, and despite the substantial progress made in the area of crash data analysis and evaluation, pedestrian crashes in the United States continue to increase at a steady pace. More studies are needed to fill the knowledge gaps, create new analytical frameworks, and generate new strategies to inform decision-making for all types of safety stakeholders: legislators and policymakers, engineers, planners, law enforcement, and safety education communities.

Several main gaps in pedestrian crash model development have been identified. Generally, conventional crash prediction models are too simplistic to handle complex relationships among variables related to pedestrian crashes. Past research has shown that crashes involving pedestrians are strongly attributable to various factors such as land use, built environment, roadway, traffic, weather, and human factors. Furthermore, factors contribute to crashes in different manners, i.e., directly (the direct relationship between variables), indirectly (a relationship that is mediated via intervening variables), or collectively (direct and indirect). Researchers have developed statistical regression models (i.e., multivariate regression, multinomial logistic regression (MLR), and negative binomial (NB) models) to explain and quantify the relationships between explanatory variables and crashes. However, conventional crash prediction models are limited in establishing and identifying relationships beyond direct effects. The underlying relationships among these variables are very complicated, especially in the presence of omitted variables and/or confounding variables. Likely, such models may produce conflicting or biased estimates that confuse the users and decision-makers. Moreover, roadway design and traffic variables are often correlated, creating difficulty in precisely measuring the contribution of individual variables to crashes.

Furthermore, in most pedestrian crash analysis, there is no distinction between a primary responsible party and a non-primary responsible party. Nor the corresponding contributing factors have been analyzed separately. Not every factor contributes equally to a crash. Some factors are more relevant to crashes where the driver is the primary responsible party than crashes where the pedestrian is the primary responsible party or vice versa; while other factors may affect both types of crashes without discrimination. It is necessary to include all variables in a single model while still addressing, comparing, and measuring their respective roles towards different parties involved. Due to strikingly different consequences of parties involving in a pedestrian crash, investigating the association between primary responsible parties in a crash and their covariates is the top priority of this dissertation.

However, determining the primary responsible party in a crash can be tricky as such information is not available in a crash report. Although few studies have explicitly researched this subject, a pedestrian and bicycle crash analysis study (R. Schneider, Stefanich, and Corsi 2015) in which the primary responsible party in crashes was identified for the State of Wisconsin offers emerges to be a crucial reference. In the study, the authors stated that citations and police narratives provided a reliable source to assign primary responsibility. Their analysis provided details on the primary responsible party of a crash occurrence for fatal pedestrian/bicycle crashes, and for severe pedestrian/bicycle crashes. The authors also highlighted the crash types that attributed crash responsibility to the pedestrian as well as describing the faults committed. Only 31% of the issued citations were relevant to faults and errors committed and believed to cause the crash, suggesting underreporting might be an issue. It is also noted that citations in fatal crashes were not captured in the database, this might occur since details of the crash are still not ready by the time the police officer issues the citation because of the absence of the involved parties. Another key reference is the pedestrian laws that provide a legal basis for making the decision of a responsible party.

Lastly, the absence of strong measures describing pedestrian exposure (e.g., walking trips, and bicycle kilometers traveled) at specific locations can be a challenge for accurately modeling the number of crashes. The common practice to handle the exposure issue is to use proxy variables such as built environment, land use, and socio-demographic variables. Whilst these variables represent the level of pedestrian activities, many of them do not always translate into reliable exposure measures. In this study, a comprehensive dataset integrating both site and area-level variables were assembled with the aim to retrieve accurate information regarding pedestrian activities from intersection and corridor features. Variables such as the presence of a pedestrian signal, the presence of curb extensions, and intersection signage and lighting are considered. The site-level related analysis is anticipated to be useful for correcting design and operational deficiencies but offer limited help for the planning of pedestrian networks. Like a highway network, the pedestrian-friendly facilities need to be consistent, continuous and connected. Therefore, the corridor level analysis may provide a promising connection for developing models to identify elements that stretch beyond a specific location. Furthermore, in the context of spatial continuity, appropriate treatments can be effectively identified and implemented based on the corridor analysis, in support of a proactive instead of a reactive planning process.

1.1.2 At-Fault Party

The importance of determining the fault status of each of the parties emerged from the transportation professional's need to understand the injury severity of traffic crashes where the driver is at-fault or not-at-fault. This knowledge may be used to educate at-fault drivers and at-fault pedestrians about the possible risk produced to other not-at-fault pedestrians and drivers. Besides, comparing the injury severity of the at-fault party with injury severity of the not-at-fault party allows the identification of the major factors affecting both parties. In this study, fault investigation included four different outputs: pedestrian at-fault, driver at-fault, both parties at-fault, and none of the parties are at fault.

This study carries an investigation of whether pedestrians or drivers were more likely to be involved in severe crashes, identified the contributing factors that affect the fault of a specific road user group, using 2017 to 2018 Wisconsin crash data provided by the WisTransPortal system. To fulfill a gap in previous research, broad considerations of multiple sources of information regarding with the crash (e.g., driver behavior, pedestrian violating traffic law, citations issued at the crash scene) are researched. Identifying and quantifying the influential factors on crash occurrence can benefit incident management to build proactive crisis management plans and therefore, lessen the impact of future accidents. While studying the effect of the responsible party on the severity of a crash, researchers have often encountered the problem of the lack of a reliable methodology to determine crash responsibility. Researchers considered different responsibility analysis methods, trying to overcome this problem and to allow for a good prediction of the reason behind severe crashes.

Over the last 20 years, developments have led to the use of several methods that may be categorized into two groups. According to the first group, trained police officers were asked to investigate the responsibility through the narrative description of the pre-prepared crash report. The second group developed crash fault scoring guidelines to allow researchers to evaluate the crash responsible party based on several explanatory variables. A value of the total score was used to consider a road user to be at or not at fault. Both methods imply responsibility evaluation without taking into consideration vulnerable road user (VRU) laws, citations issued at the scene, observations, and descriptions provided by witnesses at the crash scene. The goal of the developed guideline is to show a summary of studies utilizing these two methods to highlight the slight differences, provide a new guideline to enhance both methods, and apply the new guideline in a comparative analysis for crash severity prediction using the extreme gradient boosting (XGboost) decision tree technique and the classical discrete choice model; the multinomial logit (MNL) model. XGboost, a data mining technique, is known to inherit the advantages of artificial intelligence (AI) approaches and statistical models. Interesting conclusions have been founded by comparing the variables affecting crash severity levels triggered by road user's faults. Suggestions were also formulated from different perspectives such as law enforcement, education, and roadway management to enhance crash injury severity.

In this part of the study, the aim is bridging the gap through analyzing the crash injury severity caused by drivers and pedestrians separately. The special contributing factors leading a specific road user to be at-fault of the crash would be identified while the importance of each contributing factor on both, the pedestrian and the driver faulty would be examined and discussed as well.

One of the study objectives focused on identifying the primary responsible party in a crash by reviewing citations and the police narrative details in crash reports. The analysis focused on highlighting the primary responsible party in fatal and severe injury VRU crashes. For example, among 80 fatal crashes involving pedestrians, the analysis showed that the driver is the primary responsible party in 53% of crashes, in 29% of crashes it was the pedestrians' responsibility, in 7.5% of crashes it was both parties' responsibility, and 11% of crashes did not suggest any responsible party. The authors recommended improvements in police crash reporting practices to record more details of the crash (e.g., existing crash reports shows alcohol involvement in a crash instead of showing the intoxicated party).

1.2 DT4000 Crash Form

One of the safety initiatives to improve traffic safety is to reform/revise the crash reporting form (MV4000) into (DT4000). Some issues with the MV4000 crash form were regarding specific engineering fields such as poor reporting of roadway curvature, no data field indicating driver distraction, no specification of the exact traffic barrier, safety equipment used by the individual (motorist and non-motorists), imprecise location of the non-motorists at the time of the crash. The change in the DT4000 crash form involved proposing new crash elements and more detailed attributes. It is clear that the DT4000 crash form captures more details about the crash circumstances.

An understanding of the newly added data fields in the new driver crash report form (DT4000), and how they are differently identified from the Motor Vehicle (MV4000) accident report form, in addition to determining if the new attributes add significant value to the previously used data fields, is important to assess the benefit of the redesign of the crash report and is important to make the transportation system safer.

1.3 Dissertation Structure

The organization of this dissertation is as follows: succeeding in this chapter, Chapter 2 provides a salient review of relevant research conducted previously, including current issues in safety research of VRUs, especially pedestrians, the findings, and limitations. Chapter 3 attempts to summarize methodologies used in pedestrian and bicyclist-vehicle crash analysis. Chapter 4 analyzes a large number of variables involving VRUs, especially pedestrians, based on a comprehensive dataset of socio-demographic, temporal, and environment-related variables, traffic and road characteristics, and human characteristics at the area wide-level using the SEM technique. Chapter 5 provides the comparison of crash patterns based on the primary responsible party for a crash. Finally, Chapter 6 provides an extensive search through the newly proposed and reorganized crash data fields of the DT4000 crash form.

1.4 References

Buehler, Ralph, and Jennifer Dill. 2016. “Bikeway Networks: A Review of Effects on Cycling.” *Transport Reviews* 36 (1): 9–27. https://doi.org/10.1080/01441647.2015.1069908.

DiMaggio, Charles, and Maureen Durkin. 2002. “Child Pedestrian Injury in an Urban Setting: Descriptive Epidemiology.” *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine* 9 (1): 54–62.

“Fatality Analysis Reporting System (FARS) | NHTSA.” 2016. 2016. https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars.

Hill, Michael R. 1984. “Walking, Crossing Streets and Choosing Pedestrian Routes: A Survey of Recent Insights from the Social/Behavioral Sciences,” 79.

Hills, B. L. 1980. “Vision, Visibility, and Perception in Driving.” *Perception* 9 (2): 183–216. https://doi.org/10.1068/p090183.

Langham, Martin P., and Nicholas J. Moberly. 2003. “Pedestrian Conspicuity Research: A Review.” *Ergonomics* 46 (4): 345–63. https://doi.org/10.1080/0014013021000039574.

Lindsay, Graeme, Alexandra Macmillan, and Alistair Woodward. 2011. “Moving Urban Trips from Cars to Bicycles: Impact on Health and Emissions.” *Australian and New Zealand Journal of Public Health* 35 (1): 54–60. https://doi.org/10.1111/j.1753-6405.2010.00621.x.

Nugter, A., Faculty of Social & Behavioral Sciences, Psychology (Master), Applied Cognitive Psychology, and s1249371. 2017. “The Effect of Driving Experience on Hazard Perception in Relation to Visual Attention.” Master thesis. 2017. https://openaccess.leidenuniv.nl/handle/1887/48219.

Retting, Richard, and Sam Schwartz Consulting. 2017. “Governors Highway Safety Association,” 38.

Schneider, Robert J. 2013. “Theory of Routine Mode Choice Decisions: An Operational Framework to Increase Sustainable Transportation.” *Transport Policy* 25 (January): 128–37. https://doi.org/10.1016/j.tranpol.2012.10.007.

Schneider, Robert, Joseph Stefanich, and Larry Corsi. 2015. “Wisconsin Pedestrian and Bicycle Crash Analysis: 2011-2013,” 251.

Schofer, J. L., K. K. Christoffel, M. Donovan, J. V. Lavigne, R. R. Tanz, and K. E. Wills. 1995. “Child Pedestrian Injury Taxonomy Based on Visibility and Action.” *Accident; Analysis and Prevention* 27 (3): 317–33.

Shahar, Amit, Concetta F. Alberti, David Clarke, and David Crundall. 2010. “Hazard Perception as a Function of Target Location and the Field of View.” *Accident Analysis & Prevention* 42 (6): 1577–84. https://doi.org/10.1016/j.aap.2010.03.016.

Wegman, Fred, L.T Aarts, and René Bastiaans. 2006. *Advancing Sustainable Safety: National Road Safety Outlook for 2005-2020*. Leidschendam: SWOV.

Zhou, Zhu-Ping, Ying-Shun Liu, Wei Wang, and Yong Zhang. 2013. “Multinomial Logit Model of Pedestrian Crossing Behaviors at Signalized Intersections.” Research article. Discrete Dynamics in Nature and Society. 2013. https://doi.org/10.1155/2013/172726.

Chapter 2 : Literature Review

This chapter is provided to highlight issues in the literature conducted throughout the years for developing VRU crash prediction models. This section will discuss several exposure measures and risk factors associated with crashes involving VRUs. Considering the broad field of crash modeling, it briefly discusses two parts of the methodology (1) traditional crash prediction models (CPMs), (2) structural equation model (SEM). In the end, a reader, whether a policymaker, an engineer, or a planner will be more familiar with the gaps and challenges in crash modeling that are associated with the previous research. Lastly, this chapter serves as a base for the understanding of the current research framework.

2.1 Exposure

"Pedestrian exposure is defined as the exposure risk of pedestrians to collisions with motor vehicles (R. Schneider, Stefanich, and Corsi 2015)". The definition is not prescriptive as exposure can be measured from different perspectives. One of them is the geographic scope and spatial unit of the study. The geographic scale has always been a crucial component of safety research. To select the suitable aggregation level for their safety problems, researchers needed to understand the effect of spatial aggregation on their analysis outcome. A wide range of exposure measures and corresponding risk factors have been used by researchers, such as site-level or micro-level (Elvik 2009), and area wide-level or macro-level (Wier et al. 2009). For instance, the site-level analysis is conducted at points (i.e., mid-blocks, road segments, and intersection street crossings), whereas, area wide-level analysis takes place at corridors, zones (i.e., TAZ, Census tract, and Census block group), and regions (i.e., city, county, and state). Regardless of the scale, VRU volume data is rarely available and difficult to measure and have always been one of the limitations facing researchers. A summary of VRU proxy exposure measures used throughout prior studies at the site-level and areawide-level is as follows.

2.1.1 Site-Level VRUs Related Exposure Measures

Due to Radford and Ragland 2004, pedestrian exposure measured by pedestrian volume is the pedestrian's rate of contact with potentially harmful vehicular traffic. Pedestrian exposure is expressed in pedestrians per hour (Dong et al. 2019; Sivasankaran and Balasubramanian 2020)(Raford and Ragland 2004). Mainly, pedestrian volume data is derived from pedestrian travel demand models (Raford, Street, and Ragland 2006).

Studies used proxy variables of VRU exposure due to its measurement difficulty, and the fact that the actual observation of the activity is not available. Such studies introduced proxy variables for pedestrian/bicycle volumes, intending to study the exposure effects of this group of road users by including several proxy variables. For example, some researchers used "population" and "percentage of people walking/biking to work" as a proxy of pedestrian/bicycle exposure due to lack of AADT data. Others used vehicle kilometers traveled (VKT), vehicle miles traveled (VMT), and the number of VRUs walking and biking to work. Other exposure measures that are exclusive for pedestrians are i) several trips, ii) traveled distance, iii) several pedestrians, and iv) time spent traveling. Sometimes, roadway characteristics such as (higher functional classes) might provide a general conclusion that the area of focus will attract more pedestrian and traffic exposure. **Error! Reference source not found.** provides a review of proxy exposure measures used at the site-level.

Table 2-1: A Review of Proxy Exposure Measures Used At the Site-Level.

Author(s), (year)	Exposure Measures
J. Lee et al. (J. Lee et al. 2015); Y. Zhang et al. (J. Lee et al. 2015; Y. Zhang et al. 2015); Bu et al. (J. Lee et al. 2015; Y. Zhang et al. 2015)	Population
Cameron (Cameron 1982); Tobey (Tobey 1983); Papadimitriou et al. (Papadimitriou, Yannis, and Golias 2012)	Vehicles volume encountered while crossing
Cameron (Cameron 1982); Tobey (Tobey 1983)	The product of Pedestrian and vehicle volumes
(Strauss, Miranda-Moreno, and Morency 2013) a (Strauss, Miranda-Moreno, and Morency 2013); Strauss et al. b (Strauss, Miranda-Moreno, and Morency 2014)	Million cyclists/pedestrians per unit of time
Lindsey et al. (Lindsey, n.d.); (Hankey and Lindsey 2016)(Hankey and Lindsey 2016)	Bicyclist volumes
Raford and Ragland (Raford and Ragland 2004)	Average annual pedestrian volume
Papadimitriou et al. (Papadimitriou, et al. 2012)	The product of vehicle volume and pedestrian crossing time
Amoh-Gyimah et al. (Amoh-Gyimah, Saberi, and Sarci 2016)	Roadway characteristics (i.e., higher functional classes), VKT,

	VMT, and number of pedestrian and bicyclists walk and bike to work
Hankey and Lindsey (Hankey and Lindsey 2016)	Pedestrian volumes
Liggett et al. (Liggett et al. 2016)	Average Number of Riders
Qin and Ivan (Qin and Ivan 2001); Schneider et al. (R. J. Schneider et al. 2012); Schneider et al. (R. J. Schneider, Grembek, and Braughton 2013); Radwan et al. (Radwan et al., 2016); Molino et al. (Molino et al. 2009); Radwan et al. (Radwan et al., 2016); Bu et al. (Bu et al. 2007); Greene-Roesel et al. (Greene-Roesel, Diogenes, and Ragland 2007)	Number of VRUs, Time spent traveling, Number of trips, Traveled distance

2.1.2 Area-Level VRUs Related Exposure Measures

Area-level studies on the VRU crash used a different set of exposure measures. VMT, daily vehicle miles traveled (DMTV), vehicular traffic volume, employment data, parking signs density, traffic signal density, and population density have been used as both proxy bicycle and proxy pedestrian exposure measures. Explicitly, bike lane density and number of bicycle commuters are commonly used bicycle exposure measures.

Table 2-2 provides a review of research analysis using exposure variables conducted at the areawide-level. According to Turner and colleagues (2017), the areawide-level includes several area scales, such as networks, neighborhoods, systems, regional, city, and state.

Table 2-2: Exposure Measures at the Area-Level

Author(s), (year)	Exposure Measures
Chu (Chu 2009); Siddiqui et al. (Siddiqui, Abdel-Aty, and Choi 2012a); National Complete Streets Coalition 2014; Schneider et al. (Schneider et al. 2015); Alluri (Alluri et al. 2015); Retting and Rothenberg (Retting and Rothenberg 2015); X. Wang et al. (X. Wang et al. 2016); Cai et al. ((Siddiqui, Abdel-Aty, and Choi 2012a; J. Lee, Abdel-Aty, and Jiang 2015; Cai et al. 2017); Saha et al. (Siddiqui, Abdel-Aty, and Choi 2012a; J. Lee, Abdel-Aty, and Jiang 2015; Cai et al. 2017); Alliance for Biking and Walking 2016; Cai et al. (Cai et al. 2017); (J. Lee, Abdel-Aty, and Jiang 2015)((J. Lee, Abdel-Aty, and Jiang 2015)	Population Density
(J. Lee, Abdel-Aty, and Jiang 2015) (Rasmussen, Rousseau, and Lyons 2013); Lyons et al. (Lyons et al. 2013); Siddiqui et al. (Siddiqui, Abdel-Aty, and Choi 2012a); Schneider et al. (Schneider et al. 2015); J. Lee et al. (J. Lee, Abdel-Aty, and Jiang 2015); Cai et al. ((Siddiqui, Abdel-Aty, and Choi 2012a; J. Lee, Abdel-Aty, and Jiang 2015; Cai et al. 2017)	VMT
Siddiqui et al. (Siddiqui, Abdel-Aty, and Choi 2012a); (J. Lee, Abdel-Aty, and Jiang 2015) (J. Lee, Abdel-	Employment density

Aty, and Jiang 2015); Cai et al. ((Siddiqui, Abdel-Aty, and Choi 2012a; J. Lee, Abdel-Aty, and Jiang 2015; Cai et al. 2017)	
(Siddiqui, Abdel-Aty, and Choi 2012a; J. Lee, Abdel-Aty, and Jiang 2015; Cai et al. 2017) (Rasmussen, Rousseau, and Lyons 2013); Lyons et al. (Lyons et al. 2013); Blaizot et al. (Blaizot et al. 2013)	Number of trips per mode
Saha et al. (Saha et al. 2018); Cai et al. (Cai et al. 2017); Schneider et al. (Schneider et al. 2015)	Daily vehicle miles traveled (DMTV)
Schneider et al. (Schneider et al. 2015); (J. Lee, Abdel-Aty, and Jiang 2015) (J. Lee, Abdel-Aty, and Jiang 2015); NACTO 2016; Alliance for Biking and Walking 2016; Saha et al. (J. Lee, Abdel-Aty, and Jiang 2015)	Number of bike commuters
Jacobsen (Jacobsen 2003); Chu (Chu 2009)	A portion of walking/biking to work trips
Chen (Chen 2015); J. Lee (Guler and Grembek 2016); Jacobsen (Jacobsen 2003); Blaizot et al. (Blaizot et al. 2013); National Complete Streets Coalition 2014; Schneider et al. (Schneider et al. 2015); Chen (Chen (2015); Alluri et al. (Alluri et al. 2015); J. Lee et al. (Guler and Grembek 2016); Alliance for Biking and Walking 2016; Salon (Salon 2016)	Number of total trips; Kilometers walked/biked; Distance traveled, time spent traveling; Number/percent of walk commuters; Traffic signals density, bike lane density, and parking signs density

2.1.3 Summary

As clearly shown in **Error! Reference source not found.** and **Table 2-2**, there is a variety of proxy exposure measures, and that is regarding the fact that there is no uniform definition of exposure. For example, Greene-Roesel et al. (2007) expressed that “there is no single best definition of pedestrian exposure”, and this applies to bicycle exposure as well.

Researchers have used exposure measures at the site-level such as population density, AADT, pedestrian/bicycle volume (AADP/AADB), VMT, as well as walking and bicycling distance which may be used as site-level proxy exposure measures for future research. At the area-level, commonly used proxy exposure measures are found to be population density, VMT, employment density, number of pedestrians/bicyclists, and number of trips per mode. VMT and population density are the most popular proxy exposure measures for VRUs. However, it is noted that studies that included VMT as a proxy exposure measure, showed that the performance did not show a level of efficiency as compared to studies conducted at a specific facility (site-level).

2.2 Risk Factors

This section briefly discusses various factors that showed an influence on the risk of VRUs crashes and crash frequencies and injuries. Basically, risk factors are classified into two categories: i) site-level risk factors that are linked to a certain facility, and ii) area-level risk factors that are linked within areawide geography.

2.2.1 Site-Level VRUs-Related Risk Factors

Varying risk factors have been used in different site-level VRU crash studies as presented in Table 2-3. In general, site-level risk factors can be categorized into three groups: i) design/infrastructure characteristics (e.g., speed limits, median type, and on/off-road bike lanes) ii) intersection and segment characteristics (e.g., bus stop density, sidewalk presence, paved/unpaved shoulders, traffic control devices, and the number of lanes), iii) individual's characteristics (e.g., visibility (fluorescent clothing), vehicle's size and age, VRU age), and iv) demographic and socioeconomic characteristics (e.g., employment (Qin and Ivan 2001)and population density, and land use). **Table 2-3** shows risk factors influencing VRU safety at the site-level. Many site-level studies used population and employment density, posted speed limits, and bus stop density.

Table 2-3: List of Site-level VRU Risk Factors

Risk Factor Category	Previous Studies	Risk Factors
Design/Infrastructure characteristics	Strauss, Miranda-Moreno, and Morency (Strauss, Miranda-Moreno, and Morency 2013; Wei and Lovegrove 2013); Miranda-Moreno, Strauss, and Morency (Miranda-Moreno, Strauss, and Morency 2011; Y. Wang and Kockelman 2013); Wei and Lovegrove (Miranda-Moreno, Strauss, and Morency 2011; Y. Wang and Kockelman 2013); Karl Kim, Pant, and Yamashita (Karl Kim, Pant, and Yamashita 2010); DaSilva, Smith, and Najm	Bus-stop density; Posted speed limit; Off-road bike lanes; Paved/unpaved shoulders; Number of lanes; Roadway lighting conditions; Intersection density; Motorized traffic volume; Shoulder width; Median type; Work Zones; Paved/unpaved sidewalks; On-street parking; Traffic control type; Marked/unmarked crosswalks; Total lane kilometers, Bicycle lane

	(DaSilva, Smith, and Najm 2003); Karen Dixon et al. (Karl Kim, Pant, and Yamashita 2010); Risley (Risley 1985); Marshall, Garrick, and Hansen (Marshall, Garrick, and Hansen 2008); Hamann and Peek-Asa (Hamann and Peek-Asa 2013; Teschke et al. 2012); McMahon et al. (McMahon et al. 1999; Karen Dixon, et al. 2015); Aziz, Ukkusuri, and Hasan (Aziz, Ukkusuri, and Hasan 2013); Zegeer et al. (Zegeer et al. 2001); (Ulfarsson, Kim, and Booth 2010; Sullivan and Flannagan 2002; Haleem, Alluri, and Gan 2015); Karl Kim, Pant, and Yamashita (Karl Kim, Pant, and Yamashita 2010); Wei and Lovegrove (Karl Kim, Pant, and Yamashita 2010); Fitzpatrick, Avelar, and Turner (Fitzpatrick,	kilometers; Arterial–local intersection percentage

	Avelar, and Turner 2018); Karen Dixon et al. (Karen Dixon, et al. 2015); Obaidat et al. (Obaidat et al. 2007); Shaw et al. (Shaw et al. 2016); McMahon et al. (McMahon et al. 1999); Marshall, Garrick, and Hansen (Marshall, Garrick, and Hansen 2008); C. Lee and Abdel-Aty (C. Lee and Abdel-Aty 2005); Wei and Lovegrove (C. Lee and Abdel-Aty 2005); Zegeer et al. (Zegeer et al. 2001); Wei and Lovegrove (Wei and Lovegrove 2013)	
Demographic and socioeconomic characteristics	Siddiqui, Abdel-Aty, and Choi (Siddiqui, Abdel-Aty, and Choi 2012a); J. Lee, Abdel-Aty, and Jiang (J. Lee, Abdel-Aty, and Jiang 2015; Ukkusuri, Hasan, and Aziz 2011); Karl Kim, Pant, and Yamashita (Karl Kim, Pant, and Yamashita 2010); Siddiqui, Abdel-Aty, and Choi (Siddiqui,	Population; Employment density; Median household income; Land use; Non-motorized traffic volume; Living under the poverty level; Job count

	Abdel-Aty, and Choi 2012a); Narayanamoorthy, Paleti, and Bhat (Siddiqui, Abdel-Aty, and Choi 2012a); Amoh-Gyimah, Saberi, and Sarvi (Amoh-Gyimah, Saberi, and Sarvi 2016); Karl Kim, Pant, and Yamashita (Karl Kim, Pant, and Yamashita 2010); DaSilva, Smith, and Najm (DaSilva, Smith, and Najm 2003); Karl Kim, Pant, and Yamashita (Karl Kim, Pant, and Yamashita 2010)	
Individual's characteristics (VRUs/drivers)	Stoker et al. (Karl Kim, Pant, and Yamashita 2010); Rodgers (Rodgers 1995); Harruff, Avery, and Alter-Pandya (Harruff, Avery, and Alter-Pandya 1998); Luoma, Schumann, and Traube (Luoma, Schumann, and Traube 1996; DiMaggio and Durkin 2002); C. Lee and Abdel-Aty (C. Lee and Abdel-Aty 2005);	Pedestrian distraction (using cellphone); VRU Age; Visibility (wearing visible clothing); Driver age; Driver gender; Driver distraction; VRU gender; Vehicle size; Number of vehicle occupants; Crossing from non-crosswalk locations; VRU blood alcohol concentration (BAC)

	DiMaggio and Durkin (C. Lee and Abdel-Aty 2005); Wiechel and Guenther (Wiechel and Guenther 1989); Atkins et al. (Atkins et al. 1988); Das and Sun (S. Das and Sun 2015); Ernst (Ernst 2004); C. Lee and Abdel-Aty (C. Lee and Abdel-Aty 2005; Huemer 2018b; Öström and Eriksson 2001)	

2.2.2 Area-level VRUs Related Risk Factors

Similar to site-level risk factors, a variety of risk factors have been studied regarding their effects on VRU crashes at the area-level. **Error! Reference source not found.** lists the area-evel risk factors used in previous studies. It shows plenty of studies focused on the land use effect on VRU safety. Many other pedestrian and bicyclist risk factors were highlighted such as population density, VRU age, and lower-income. A considerable amount of the studies were conducted at the area-level to explore VRU crash-related features (Narayanamoorthy, Paleti, and Bhat 2013; Parkin, Wardman, and Page 2007; Wei and Lovegrove 2013; Abdel-Aty et al. 2013; (Gladhill and Monsere 2012) Chen 2015; Hamann and Peek-Asa 2013; Strauss, Miranda-Moreno, and Morency 2013; Miranda-Moreno, Strauss, and Morency 2011; Siddiqui, Abdel-Aty, and Choi 2012b; (Wei and Lovegrove 2013; P. Chen 2015) (Wei and Lovegrove 2013; P. Chen 2015). Many area-level studies (Narayanamoorthy, Paleti, and Bhat 2013; Parkin, Wardman, and Page 2007) use data aggregated at the census tract level, and some other studies

(M. Abdel-Aty et al. 2013; Wei and Lovegrove 2013) used TAZ-level aggregated data. Few studies (Gladhill and Monsere 2012) used the grid layout areas. ***Table 2-4*** shows common risk factors influencing VRU safety at an aggregated level (area-level).

Table 2-4: List of Area-level VRU Risk Factors

Category	**Previous Studies**	**Risk Factors**
Traffic characteristics	Wier et al. (Wier et al. 2009; C. Lee and Abdel-Aty 2005; Loukaitou-Sideris, Liggett, and Sung 2007; M. Abdel-Aty et al. 2013); Hamann and Peek-Asa (Wier et al. 2009; C. Lee and Abdel-Aty 2005; Loukaitou-Sideris, Liggett, and Sung 2007; Abdel-Aty et al. 2013); Kaplan and Giacomo Prato ((Kaplan and Giacomo Prato 2015)), Jacobsen (Jacobsen 2003; Elvik 2009; Prato et al. 2016); Abdel-Aty et al. (M. Abdel-Aty et al. 2013; J. Lee, Abdel-Aty, and Jiang 2015); Siddiqui, Abdel-Aty, and Choi (Siddiqui, Abdel-Aty, and Choi	Vehicle traffic volume; Speed limit; Walking/biking trips; Truck percentage

	2012a); Demetriades (Siddiqui, Abdel-Aty, and Choi 2012a); Abdel-Aty et al. (M. Abdel-Aty et al. 2013); Wier et al. (Wier et al. 2009); Karen Dixon, et al. (Karen Dixon, et al. 2015); Retting (R. A. Retting 1993)	
Landuse characteristics	Amoh-Gyimah, Saberi, and Sarvi (R. A. Retting 1993); Ukkusuri, Hasan, and Aziz (Ukkusuri, Hasan, and Aziz 2011; K Kim, Brunner, and Yamashita 1953); X. Wang et al. (X. Wang et al. 2016); Wier et al. (Wier et al. 2009; Siddiqui, Abdel-Aty, and Choi 2012a); Y. Wang and Kockelman (Y. Wang and Kockelman 2013); R. Noland and Quddus (R. Noland and Quddus 2004; LaScala, Gerber, and Gruenewald 2000; M.	Mixed land use, Density of public schools, Income level

	Abdel-Aty et al. 2013; Roberts, Norton, and Taua 1996)	
Demographic and socioeconomic characteristics	LaScala, Gerber, and Gruenewald (R. Noland and Quddus 2004; LaScala, Gerber, and Gruenewald 2000; Abdel-Aty et al. 2013; Roberts, Norton, and Taua 1996); Abdel-Aty et al. (M. Abdel-Aty et al. 2013; Greene-Roesel, Diogenes, and Ragland 2007; Miranda-Moreno, Strauss, and Morency 2011; Loukaitou-Sideris, Liggett, and Sung 2007); Loukaitou-Sideris, Liggett, and Sung (Loukaitou-Sideris, Liggett, and Sung 2007; Siddiqui, Abdel-Aty, and Choi 2012b; R. B. Noland, Klein, and Tulach 2013; J. Lee, Abdel-Aty, and Jiang 2015); J. Lee, Abdel-Aty, and Choi (Loukaitou-Sideris, Liggett,	Population density; Number of licensed drivers; Density of minority households, Poor neighborhoods; Vehicle ownership; Unemployment rate; Household income; Percentage of the low-income population; Education percentage; Crime Density; Household size; Vehicle ownership

	and Sung 2007; Siddiqui, Abdel-Aty, and Choi 2012b; R. B. Noland, Klein, and Tulach 2013; J. Lee, Abdel-Aty, and Jiang 2015); (J. Lee et al. (J. Lee et al. 2015); Ukkusuri, Hasan, and Aziz (Ukkusuri, Hasan, and Aziz 2011); LaScala, Gerber, and Gruenewald (LaScala, Gerber, and Gruenewald 2000); Cottrill and Thakuriah (Cottrill and Thakuriah 2010); McMahon et al. (McMahon et al. 1999); (McMahon et al. 1999); Qin and Ivan (Qin and Ivan 2001)	
Roadway geometry/Infrastructure	Jacobsen (Jacobsen 2003; Elvik 2009; Prato et al. 2016); Pucher, Komanoff, and Schimek (Pucher, Komanoff, and Schimek 1999); Wei and Lovegrove (Wei and Lovegrove 2013; P. Chen	Presence of bike paths; Traffic signal density; Number of pedestrian crossings; Off-arterial bicycle routes; Signal density; Presence of parking signs; Number of vehicle trips; Minor and major arterial length; Road

	2015), Abdel-Aty et al. (M. Abdel-Aty et al. 2013; J. Lee, Abdel-Aty, and Jiang 2015); Chen ((P. Chen 2015); X. Wang et al. (X. Wang et al. 2016); (Y. Wang and Kockelman 2013); Moeinaddini, Asadi-Shekari, and Zaly Shah (Moeinaddini, Asadi-Shekari, and Zaly Shah 2014); Y. Zhang et al. (Y. Zhang et al. 2015); Dai and Jaworski (Dai and Jaworski 2016); Cai et al. (Cai et al. 2016); Guo et al. ((Guo et al. 2017); Demetriades (Demetriades 2004); Abdel-Aty et al. (M. Abdel-Aty et al. 2013); Wier et al. (Wier et al. 2009).	density, Percentage of 3-leg intersections; Average intersection spacing; Transit stop density; School access; Sidewalk density; Street network size; Roadway length; Density of major roads; Number of intersections; Clustered road networks; Segments with fixed gradients; Distance to transit trips, VRU and driver age; Signalized intersections density; Sidewalk length, Local road density; Network pattern (irregular, grid)

2.2.3 Summary

Studies mentioned in **Table 2-3** and **Table 2-4** involved a wide range of variables connected to site-level and area-level VRU crash risk. Many studies only used a limited number of risk factors due to data unavailability and/or data collection complexity. As a result, concluding reliable inferences from the studied variables from both sites and areas becomes a challenging task. Besides, some site-level variables (i.e., median type, and work zone sites) and area-level variables (i.e., signalized intersections density, and several pedestrian crossings) have been tested in a few numbers of studies, recommending more intensive research to be conducted to reach more comprehensive conclusions, since contradicting results have been detected. Lastly, in view of the fact that pedestrian and bicycle crash-related variables are continuous in nature, for example, the percentage of bike lanes that provide continuous, shared side path density, and sidewalk density, such accurate variables can be conveniently defined and collected in area-level studies.

2.3 Primary Responsible Party in VRU Crashes

Safety policies are put together and translated into traffic safety rules by transportation officials. For instance, drivers are demanded to drive below or at speed limits and should maintain their vehicles at a complete stop at the stoplight. Such rules are set to assure safety for all users on the road. However, drivers violate these rules sometimes and are involved in crashes. For example, 58% of traffic crash fatalities resulted from intoxicated driving and speeding (National Highway Traffic Safety Administration (NHTSA) 2016). To mitigate the frequency and severity of traffic crashes, there is an urgent need to understand the crash contributing factors. In fact, researchers have done a considerable amount of research on crash contributing factors. However, few have considered the human behavior factor in crashes and how it influences crash severity based on the primary responsible party.

In light of research considering the pedestrian responsibility of crashes, pedestrians' rule violations that were commonly found to be associated with crashes are as follows: i) failing to yield the right of way, ii) disobeying traffic signals, and iii) intoxicated pedestrians (Baltes 1998;

Preusser et al. 2002; Oxley et al. 2005; Stutts, Hunter, and Pein 1996). Wootton analyzed pedestrian crashes and stated that pedestrian behavior, alcohol use (by pedestrians and drivers), and poor pedestrian visibility at night coupled with violation of driver expectations were the most significant causes of pedestrian-vehicle crashes (Isaac Adam Wootton 2006). An interesting conclusion was driven by a study in Vancouver, Canada, which stated that the distribution of crash responsibility between the several parties (pedestrian, motorist, or both) varied with the enforcement laws and the local safety culture (Cinnamon, Schuurman, and Hameed 2011). Ulfarsson and colleagues conducted a study to predict the primary responsible party. The authors stated that parties found responsible were pedestrians (59%), drivers (32%), and both road users (9%) (Ulfarsson, Kim, and Booth 2010). Kim et al. concluded that when the pedestrian is responsible, a higher fatality risk occurs in the crash compared to crashes where the motor-vehicle driver is the primary responsible party (J.-K. Kim et al. 2008; 2010). Some studies concluded that commonly, pedestrians are more responsible party compared with drivers such as (Preusser et al. 2002; C. Lee and Abdel-Aty 2005). However, other studies contradict such a statement (J.-K. Kim et al. 2008; Karl Kim, Brunner, and Yamashita 2008).

Some other studies focused on analyzing crashes were only the primary responsible party is the non-motorized road user such as (Islam and Jones 2014). Salon and McIntyre (2018) have recently published a study where the factors related to pedestrian and bicyclist's crash severity were dependent on the primary responsible party. It is hypothesized that the occurrence of one crash is more related to the characteristics of the primarily responsible party involved in that crash (C. Lee and Abdel-Aty 2005). Lee and Abdel-Aty (C. Lee and Abdel-Aty 2005) studied pedestrian-vehicle crashes at signal-controlled intersections in Florida and found that intoxicated pedestrians were more likely to get involved in nighttime crashes than intoxicated motor-vehicle drivers. When in fact, motor-vehicle drivers were more correlated to crashes occurring at controlled intersections. Similarly, the severity outcome of one crash could also be associated with the characteristics of the primarily responsible party. However, in a study of pedestrian injury patterns, the authors concluded that more severe crashes happened when drivers violated speed limits compared to crashes where drivers were inattentive (Damsere-Derry et al. 2010). Factors influencing the crash severity of crashes where pedestrian are the primary responsible party, in urban and rural locations in the State of Alabama indicated that the studied variables had different influencing effects between rural and urban pedestrian responsible accidents (Islam and Jones 2014).

Regarding research in bicycle-vehicle crashes, many errors are highlighted for the different driver or bicyclist error types. 56% of bicycle-car crashes in Germany occurred when turning at an intersection or entering the road from a private property were found to occur because the motor-vehicle driver is the primary responsible party (Schreck 2017). Rasanen & Summala (1998), showed that regarding crash statistics, bicyclists' rule-violations is behind crashes involving bicyclists, and in 45% of those crashes with bicyclists as the primary responsible party, errors are common as follows: i) wrong path cycling (in 17.1% of crashes), ii) bicycling too fast for conditions (in 7.5% of crashes), and iii) errors while entering fluent traffic (in 6.9% of crashes). While in single-bike crashes, police reports have four main common errors; i) cycling too fast for conditions (in 17.0% of crashes), ii) alcohol usage (in 15.9% of crashes), iii) cycling on the wrong path (in 4.6% of crashes), and iv) cycling without using lights and other technical issues (in 3.6% of crashes). In a study by Huemer (2018), results showed that 55% of bicyclist related crashes occurred where bicyclists cycled on the wrong cycling path, and 29% of bicyclists did not use their light while cycling at night (Huemer 2018a). According to the modeling procedure, the primary responsible party was used as the model's dependent variable such in (C. Lee and Abdel-Aty 2005; Karl Kim, Brunner, and Yamashita 2008; Ulfarsson, Kim, and Booth 2010; G. Zhang, Yau, and Zhang 2014), and used as an explanatory variable such in (J.-K. Kim et al. 2008; G. Zhang, Yau, and Zhang 2014; Haleem, Alluri, and Gan 2015).

For research focusing on the primary responsibility among both VRUs, a study (R. Schneider, Stefanich, and Corsi 2015) provided an analysis of VRUs crashes that occurred in Wisconsin.

Previous studies of VRUs being the primary responsible party of the crash occurrence are limited. Furthermore, there is no noteworthy research explaining how an explanatory variable may affect responsible and non-responsible VRUs within the same crash. It is expected that the new knowledge of different crash severity outcomes pertaining to the responsible party may enhance the process of informing highly rewarding safety policies and fruitful pedestrian and bicyclist's awareness.

2.4 Dominant Factors of VRU's Crashes

Many studies that shed light on dominant factors related to pedestrian and bicycle crashes have examined roadway geometric characteristics such as the number of lanes, median type, speed limits, and speed ratio (i.e., the ratio of speed from crash data over the posted speed limit). Morency et al. confirmed the association between wider roads and higher pedestrian-related crash frequency, since a wider road may encourage drivers to speed and jeopardize pedestrians. Designated right-turn lanes and nearby driveway crossings were associated with higher pedestrian crash risk, while median crossing refuges were associated with lower pedestrian crash risk at intersections (10). Quiet streets, gentle slopes, and the absence of streetcar tracks are some design features associated with lower bicyclist crash risk (11). Moreover, cycle tracks (11, 12), traffic diverters, and local streets tend to separate cyclists from the moving traffic, leading to fewer crashes (11). Roads with more traffic signals, street parking signs, and automobile trips are associated with more frequent bicycle crashes. Roadways with a speed limit of 35 mph and intersection density are positively related to the likelihood of pedestrian and bicyclist crashes (11). Additionally, Cai et al. 2016 stated that traffic analysis zones (TAZs) with longer sidewalk lengths, more pedestrians, and more employment are more susceptible to have pedestrian crashes, while TAZs with longer sidewalk lengths, more employment, and higher population density are more likely to have bicyclist crashes (13). Road speed limit also found to be affecting bicycle crash frequency, as Siddiqui and colleagues, 2012 stated that highways with speed limit >35 mph are more likely to have bicyclist crashes. The density of signalized intersections, arterial, and local road proportion, sidewalk length, are positively correlated with pedestrian and/or bicyclist-motor vehicle crashes (14). While controlling for exposure variables, several studies have identified specific pedestrian facilities to be negatively associated with pedestrian crashes, including median refuge islands and rectangular rapid flashing beacons (4).

Some studies have explored how behaviors are related to pedestrian and bicyclist crashes. Helmet usage, travel programs such as routes to school, wearing reflective clothing, and education related to safety among bicyclists have been associated with fewer bicyclist fatalities. Pedestrians and bicyclists crossing a red-light signal or using mobile devices, and motor-vehicle drivers turning right on red without waiting for other road users to cross are some of the most important safety behaviors studied (15). Vehicle speed impact on pedestrian fatality risk reported that car speed positively and strongly affects fatality risk among pedestrians. Above 20 mph, small increases in speed produce relatively large increases in pedestrian injury severity (16, 17).

Other studies have identified exposure as an important variable associated with pedestrian and bicyclist crashes. There are a variety of exposure measures in the literature (i.e., using Census journey to work data as an exposure proxy variable (18)), but this concept is commonly represented using pedestrian, bicyclist, and automobile counts (19, 20). Several studies showed the relationship between the number of pedestrian and bicyclist crashes and pedestrian and bicyclist activity levels. Results confirm that the relationship is not linear (commonly referred to as "safety in numbers" effect): pedestrian or bicycle crash risk (e.g., crashes per crossing or per trip) decreases with the increase in walking or cycling (19 – 22). (21, 23). One challenge for using this important variable in safety analyses is that few jurisdictions have sufficient pedestrian or bicyclist count data, resulting in the use of proxy variables to represent exposure.

Regarding economic, demographic, and social characteristics,

Table 2-5 summarizes factors stated in some previous studies.

Table 2-5: Summary of Economic, Demographic, and Social Characteristics Influencing Pedestrian and Bicycle Crash Frequency

Author name, year	**Economic, demographic, and social characteristics**	**Emphasis**
Siddiqui et al., 2012 (24)	Total population, the proportion of the uneducated population, land use (presence of restaurants and bars), and park coverage	Pedestrians
Nashad, 2016 (25)	Number of dwelling units, population density, total employment and percentage of households with zero or one car ownership Vehicle-Miles Travelled (VMT), middle-aged (25-64) and male drivers, neighborhoods with large retail and residential land use,	Pedestrians and bicyclists

	high vehicular traffic movements, high employment and population density, low-income, and high minority environmental justice areas and races	
Lee and Abdel-Aty, 2005 (26)	Hotel room density, number of walking/biking, population density, and school enrollment density, the proportion of industrial employment, low- income, and high minority environmental justice areas and races	Pedestrians
Loukaitou et al., 2007 (27)	Age < 18 years old, neighborhoods with large retail and residential land uses, high vehicular traffic movements, and high employment and population density	Pedestrians and bicyclists
Nordback et al., 2014 (28)	Percentage of households without access to private vehicles	Bicyclists
Schneider et al., 2017 (22)	Economic, demographic, and social characteristics	Pedestrians and bicyclists
Morency et al., 2012 (10)	Neighborhoods with large retail and residential land use, high vehicular traffic movements, and high employment and population density	Pedestrians

2.5 Road User Crash Fault Assessment

Within the scope of injury severity analyses, huge efforts have discovered the relationships between crash injury severity and roadway characteristics, human factors, and traffic operation environments (Savolainen et al. 2011). In general, human factors are considered to be the most prevalent factors contributing to crashes, followed by roadway environment and vehicle factors (H. Zhang 2010). many factors have been of interest since the analysis results enhanced the driver-focused educational programs (M. A. Abdel-Aty and Abdelwahab 2000). In addition to the commonly used personnel features (e.g., gender, age, driving experience, and blood alcohol concentration (BAC)), crashes were also identified to be related to the driver faults (Karl Kim and Li 1996; Ichikawa, Nakahara, and Taniguchi 2015; Walter and Studdert 2015; Islam, Jones, and Dye 2014; Zhao, Wang, and Jackson 2019; Penmetsa, Pulugurtha, and Duddu 2017; Duddu, Penmetsa, and Pulugurtha 2018; R. Schneider, Stefanich, and Corsi 2015). Limited research considered studying the effect of fault status on the injury severity level, and the factors affecting the fault of a road user.

(Yu et al. 2019) investigated a variety of crash influencing factors of at-fault out-of-state drivers. It was identified that the influencing variables for crashes caused by out-of-state drivers are car ownership, speeding, and driving under the influence (DUI) for the intersection crashes and being an old driver, speeding, and dark roadway environment for roadway segment crashes. (J. Lee, Abdel-Aty, and Choi 2014) analyzed the relationship between the number of at-fault drivers and their residence zonal characteristics. It was concluded that not only roadway/traffic factors affect the crash occurrence, but also several demographic and socioeconomic characteristics of residence zones where the at-fault drivers live. Yet, there was no detailed exploration carried out for the specific influencing factors affecting a specific road user to be responsible for a crash, and no investigation directed to study if the crash injury severity is affected by a specific road user being at-fault.

Concerning the analytical methodologies used for the analysis and taking into account the discrete outcomes besides the ordinal characteristics of the crash severity outcomes, ordered logistic regression models (e.g., ordered logit/probit models) were used most commonly. For example, (Karl Kim, Brunner, and Yamashita 2008) utilized techniques of the logistic regression model to examine factors conjoined with severe and fatal crashes. (G. Zhang, Yau, and Zhang 2014) formulated a stepwise logistic regression model to identify risk factors that impact pedestrian and driver fault status in pedestrian-motor vehicle crash analysis in China.

Yet, the key assumption of the ordinal regression models called “parallel odds assumption” or “parallel regression assumption” is that explanatory variables have a persistent effect throughout different crash severity outcomes. Therefore, to relax this limitation, the partial proportional odds (PPO) models were developed ((Peterson and Harrell 1990) and were implemented to evaluate pedestrian injury severity by (Sasidharan and Menéndez 2014); fault status was studied together with other pedestrian and driver characteristics. The results show that the PPO models provide more detailed knowledge of the contributing factors than the MNL and the ordered logit models. Besides that, the PPO model outperformed the other models based on the information-theoretic approach. The same conclusion has been drawn by other researchers (Qin, Wang, and Cutler 2013); the authors compare the performance of the MNL and the mixed logit (ML) models with the PPO model, to evaluate the effect of multiple determinants on the severity of crashes involving large trucks. (Penmetsa, Pulugurtha, and Duddu 2017) also utilized PPO models to investigate the injury severity of not-at-fault drivers in two-vehicle crashes. The study results showed that the data was ineffectual with the assumption of the ordered probit model; the proportional odds assumption, therefore the PPO model was adopted. The results were consistent with previous research about the influencing factors that affect the at-fault drivers’ injury severity. **Table 2-6** shows different adopted techniques of crash responsibility assignment in previous research.

Table 2-6: A Summary of the Existing Methods Used for Assigning the Responsible Road User In A Crash In Previous Empirical Studies.*

(Study author and year)	Data source	Summary of methods used in determining fault
Kim, Brunner, and Yamashita (2008)	Police-reported crash data.	Based on law enforcement (i.e., pedestrian jaywalking laws), the fault is determined by deciding on the party that received a ticket/citation.
Kim, Brunner, and Yamashita (2008)		The fault is determined by specially trained investigators who are dispatched to the crash scene, and they typically search for the action or behavior of a certain party that caused the accident to occur.
Ulfarsson, Kim, and Booth (2010)		The police officer assigns the fault to the party based on who acted negligently or is in other ways found to have caused the crash.
(Zhao, Wang, and Jackson 2019)		Drivers were considered ''at-fault'' if they were given any verbal/written warning, infraction, or arrest/summons. If no action was taken by the officer, the driver was considered not-at-fault.

Walter and Studdert (2015)		The fault is determined based on citations issued for crash responsible drivers and recorded in the police report.
Kim et al. (1998)		Police officers assign the fault party and record the information in the crash report based on the officer's narrative summary of the event.
Kim and Li (1996)		The fault is determined by investigating officers and then reported on crash report forms. Based on a logistic model that explains fault among motorists as a function of various. The crash report form is to ascertain who is at fault and that although individual fields such as driver's license number, birth date, vehicle identification number, and so forth may be sources of error.
Adanu et al. (2017)		The fault is determined based on the investigation done by the officer who completed the crash report.
Spainhour and Wootton (2007)	Police-reported crash data and case review data stemmed from manual case reviews of multiple crash data sources.	The fault is determined by the investigating officers which are then reported on crash report forms. Florida department of transportation (DOT) currently uses an algorithm to assign fault. FDOT thereby presumes that the

		individual in the first section is at fault unless a citation was given to drivers or pedestrians in subsequent sections of the crash report, in which case fault is reassigned to the person receiving the citation.
Lee, Abdel-Aty, and Choi (2014)	ZIP code information of road users involved in traffic crashes and police recorded data.	The fault is determined based on the crash investigation done by police officers, who issue citations according to the investigation. Citations are then recorded in the police report which confirms the fault party.
Goh et al. (2014)	Pictures and video recordings captured from CCTV cameras.	The fault is typically made with the aid of pictures and video recordings captured from CCTV installed and is done by police officers and adjusters from the insurance company before an at-fault assessment is made for the purpose of insurance claims.
Russo et al. (2014)	Michigan traffic crash facts (MTCF) data query tool.	The fault is assigned to one driver based on the judgment of the investigating officer, who decides that the at-fault driver must have performed one or more hazardous actions (e.g. speeding, failing to yield, disregarding traffic control devices) that contributed to the crash.

Zhang, Yau, and Zhang (2014)	Traffic accident data.	The fault is determined by police officers, under the following two circumstances: (1) determined by police that he/she should bear the whole responsibility of the accident; (2) determined by police that he/she should bear the main responsibility of the accident. For instance, when crashes occur on roadways without pedestrian facilities, motor vehicles are typically held responsible because pedestrians are commonly considered as a vulnerable group. Conversely, pedestrians are mostly determined to be liable.
Islam and Jones (2014)	Police-reported crash data filtered using the critical analysis reporting environment (CARE) software system.	Dataset already has the at-fault party assigned in the original police-reported crash database.
Schneider, Stefanich, and Corsi (2015)	“MV4000” Crash report forms.	Since the police do not assign “fault” for crashes in Wisconsin, the study used the detailed narrative on the MV4000 forms, and the type of citation helped as well to interpret which party or parties the police officer viewed as being primarily responsible for the crash.

Ichikawa, Nakahara, and Taniguchi (2015)	Police-reported crash data and driving exposure data.	The fault is determined by police investigators at the scene of the motor vehicle crash.
Ratrout et al. (2017)	Police-reported crash data, as well as the crash, driver, and vehicle-related data collected from police stations.	Drivers at fault and not at fault were separated and investigated through factor analysis for 19 parameters related to their background and knowledge of traffic signs.
Penmetsa, Pulugurtha, and Duddu (2017)	Crash data obtained from the Highway Safety Information System (HSIS) which included information related to accident, roadway, vehicle, and occupant.	For each vehicle involved in a crash, the crash reports provide three contributing factors (which indirectly explain the traffic rule the driver violated) that led to the crash. If the driver has not committed any traffic violation, a value of zero is provided under the contributing factor variable, implying that the driver is not at fault in the crash.
Das et al. (2018)	Police-reported crash data which contains crash, roadway geometry, and vehicle-related data.	Based on the crash event investigated by the police officer who records the fault party in the report.

Yu et al. (2019)	Police-reported crash data obtained from the FDOT Crash Analysis Reporting (CAR) system, driver information data extracted from the CAR system, and Traffic analysis zone (TAZ) shapefile obtained from the US Census Bureau.	The fault is determined through using the license address state variable "ADRSTATE" for at-fault drivers, which is coded in the crash data.
(Islam and Hossain 2019)	Police reported crash database filtered using the (CARE) software system.	Dataset already has the at-fault party assigned in the original police-reported crash database by the police officer who completed the report at the scene.

*Note that the mentioned studies are ordered by the usage of the common data source.

2.6 References

Abdel-Aty, Mohamed A., and Hassan T. Abdelwahab. 2000. "Exploring the Relationship between Alcohol and the Driver Characteristics in Motor Vehicle Accidents." *Accident Analysis & Prevention* 32 (4): 473–82. https://doi.org/10.1016/S0001-4575(99)00062-7.

Abdel-Aty, Mohamed, Jaeyoung Lee, Chowdhury Siddiqui, and Keechoo Choi. 2013. "Geographical Unit Based Analysis in the Context of Transportation Safety Planning." *Transportation Research Part A: Policy and Practice* 49 (March): 62–75. https://doi.org/10.1016/j.tra.2013.01.030.

Amoh-Gyimah, Richard, Meead Saberi, and Majid Sarvi. 2016. "Macroscopic Modeling of Pedestrian and Bicycle Crashes: A Cross-Comparison of Estimation Methods." *Accident Analysis & Prevention* 93 (August): 147–59. https://doi.org/10.1016/j.aap.2016.05.001.

Atkins, R. M., W. H. Turner, R. B. Duthie, and B. R. Wilde. 1988. "Injuries to Pedestrians in Road Traffic Accidents." *BMJ (Clinical Research Ed.)* 297 (6661): 1431–34.

Aziz, H.M. Abdul, Satish V. Ukkusuri, and Samiul Hasan. 2013. “Exploring the Determinants of Pedestrian–Vehicle Crash Severity in New York City.” *Accident Analysis & Prevention* 50 (January): 1298–1309. https://doi.org/10.1016/j.aap.2012.09.034.

Baltes, Michael. 1998. “Descriptive Analysis of Crashes Involving Pedestrians in Florida, 1990-1994.” *Transportation Research Record: Journal of the Transportation Research Board* 1636 (January): 138–45. https://doi.org/10.3141/1636-22.

Blaizot, Stéphanie, Francis Papon, Mohamed Mouloud Haddak, and Emmanuelle Amoros. 2013. “Injury Incidence Rates of Cyclists Compared to Pedestrians, Car Occupants and Powered Two-Wheeler Riders, Using a Medical Registry and Mobility Data, Rhône County, France.” *Accident Analysis & Prevention* 58 (September): 35–45. https://doi.org/10.1016/j.aap.2013.04.018.

Bu, Fanping, Ryan Greene-Roesel, Mara Chagas Diogenes, and David R. Ragland. 2007. “Estimating Pedestrian Accident Exposure: Automated Pedestrian Counting Devices Report,” March. https://escholarship.org/uc/item/0p27154n.

Cai, Qing, Mohamed Abdel-Aty, Jaeyoung Lee, and Naveen Eluru. 2017. “Comparative Analysis of Zonal Systems for Macro-Level Crash Modeling.” *Journal of Safety Research* 61 (June): 157–66. https://doi.org/10.1016/j.jsr.2017.02.018.

Cai, Qing, Jaeyoung Lee, Naveen Eluru, and Mohamed Abdel-Aty. 2016. “Macro-Level Pedestrian and Bicycle Crash Analysis: Incorporating Spatial Spillover Effects in Dual State Count Models.” *Accident Analysis & Prevention* 93 (August): 14–22. https://doi.org/10.1016/j.aap.2016.04.018.

Cameron, M.H. 1982. “A Method of Measuring Exposure to Pedestrian Accident Risk.” *Accident Analysis & Prevention* 14 (5): 397–405. https://doi.org/10.1016/0001-4575(82)90019-7.

Chen, Peng. 2015. “Built Environment Factors in Explaining the Automobile-Involved Bicycle Crash Frequencies: A Spatial Statistic Approach.” *Safety Science* 79 (November): 336–43. https://doi.org/10.1016/j.ssci.2015.06.016.

Chu, Xuehao. 2009. “Measuring Injury Risks from Motor Vehicle Crashes with an Integrated Approach.” *Journal of Transportation Safety & Security* 1 (3): 190–202. https://doi.org/10.1080/19439960903149645.

Cinnamon, Jonathan, Nadine Schuurman, and S. Morad Hameed. 2011. “Pedestrian Injury and Human Behaviour: Observing Road-Rule Violations at High-Incident Intersections.” Edited by Tom Denson. *PLoS ONE* 6 (6): e21063. https://doi.org/10.1371/journal.pone.0021063.

Cottrill, Caitlin D., and Piyushimita (Vonu) Thakuriah. 2010. “Evaluating Pedestrian Crashes in Areas with High Low-Income or Minority Populations.” *Accident Analysis & Prevention* 42 (6): 1718–28. https://doi.org/10.1016/j.aap.2010.04.012.

Damsere-Derry, James, Beth E. Ebel, Charles N. Mock, Francis Afukaar, and Peter Donkor. 2010. “Pedestrians’ Injury Patterns in Ghana.” *Accident Analysis & Prevention* 42 (4): 1080–88. https://doi.org/10.1016/j.aap.2009.12.016.

Das, Subasish, and Xiaoduan Sun. 2015. “Factor Association with Multiple Correspondence Analysis in Vehicle–Pedestrian Crashes.” *Transportation Research Record: Journal of the Transportation Research Board* 2519 (1): 95–103. https://doi.org/10.3141/2519-11.

Demetriades, D. 2004. “Pedestrians Injured by Automobiles: Relationship of Age to Injury Type and Severity 1.” *Journal of the American College of Surgeons* 199 (3): 382–87. https://doi.org/10.1016/j.jamcollsurg.2004.03.027.

DiMaggio, Charles, and Maureen Durkin. 2002. "Child Pedestrian Injury in an Urban Setting: Descriptive Epidemiology." *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine* 9 (1): 54–62.

Duddu, Venkata R., Praveena Penmetsa, and Srinivas S. Pulugurtha. 2018. "Modeling and Comparing Injury Severity of At-Fault and Not at-Fault Drivers in Crashes." *Accident Analysis & Prevention* 120 (November): 55–63. https://doi.org/10.1016/j.aap.2018.07.036.

Elvik, Rune. 2009. "The Non-Linearity of Risk and the Promotion of Environmentally Sustainable Transport." *Accident Analysis & Prevention* 41 (4): 849–55. https://doi.org/10.1016/j.aap.2009.04.009.

Ernst, M. 2004. "MEAN STREETS 2004 - HOW FAR HAVE WE COME? PEDESTRIAN SAFETY, 1994-2003," November. https://trid.trb.org/view/745471.

"FDOT-BDV24-977-07-Rpt.Pdf." n.d. Accessed December 3, 2018. http://www.fdot.gov/research/Completed_Proj/Summary_TE/FDOT-BDV24-977-07-rpt.pdf.

Fitzpatrick, Kay, Raul Avelar, and Shawn Turner. 2018. "Guidebook on Identification of High Pedestrian Crash Locations," April. https://trid.trb.org/view/1508436.

Gladhill, Kristie, and Christopher M. Monsere. 2012. "Exploring Traffic Safety and Urban Form in Portland, Oregon." *Transportation Research Record: Journal of the Transportation Research Board* 2318 (1): 63–74. https://doi.org/10.3141/2318-08.

Greene-Roesel, Ryan, Mara Chagas Diogenes, and David R. Ragland. 2007. "ESTIMATING PEDESTRIAN ACCIDENT EXPOSURE," 232.

Guler, S. Ilgin, and Offer Grembek. 2016. "Use of Different Exposure Metrics for Understanding Multi-Modal Travel Injury Risk." *International Journal of Transportation Science and Technology* 5 (1): 28–37. https://doi.org/10.1016/j.ijtst.2016.06.004.

Guo, Qiang, Pengpeng Xu, Xin Pei, S.C. Wong, and Danya Yao. 2017. "The Effect of Road Network Patterns on Pedestrian Safety: A Zone-Based Bayesian Spatial Modeling Approach." *Accident Analysis & Prevention* 99 (February): 114–24. https://doi.org/10.1016/j.aap.2016.11.002.

Haleem, Kirolos, Priyanka Alluri, and Albert Gan. 2015. "Analyzing Pedestrian Crash Injury Severity at Signalized and Non-Signalized Locations." *Accident Analysis & Prevention* 81 (August): 14–23. https://doi.org/10.1016/j.aap.2015.04.025.

Hamann, Cara, and Corinne Peek-Asa. 2013. "On-Road Bicycle Facilities and Bicycle Crashes in Iowa, 2007–2010." *Accident Analysis & Prevention* 56 (July): 103–9. https://doi.org/10.1016/j.aap.2012.12.031.

Hankey, Steve, and Greg Lindsey. 2016. "Facility-Demand Models of Peak Period Pedestrian and Bicycle Traffic: Comparison of Fully Specified and Reduced-Form Models." *Transportation Research Record: Journal of the Transportation Research Board* 2586 (1): 48–58. https://doi.org/10.3141/2586-06.

Harruff, R. C., A. Avery, and A. S. Alter-Pandya. 1998. "Analysis of Circumstances and Injuries in 217 Pedestrian Traffic Fatalities." *Accident; Analysis and Prevention* 30 (1): 11–20.

Huemer, Anja Katharina. 2018a. "Motivating and Deterring Factors for Two Common Traffic-Rule Violations of Cyclists in Germany." *Transportation Research Part F: Traffic Psychology and Behaviour* 54 (April): 223–35. https://doi.org/10.1016/j.trf.2018.02.012.

———. 2018b. “Cycling under the Influence of Alcohol in Germany.” *Transportation Research Part F: Traffic Psychology and Behaviour* 56 (July): 408–19. https://doi.org/10.1016/j.trf.2018.05.013.

Ichikawa, Masao, Shinji Nakahara, and Ayako Taniguchi. 2015. “Older Drivers’ Risks of at-Fault Motor Vehicle Collisions.” *Accident Analysis & Prevention* 81 (August): 120–23. https://doi.org/10.1016/j.aap.2015.05.004.

Isaac Adam Wootton. 2006. “Identification of Contributing Factors and Accuracy of Fault Prediction Using Various Sources of Fatal Pedestrian Crash Data in Florida,” 252.

Islam, Samantha, and Steven L. Jones. 2014. “Pedestrian At-Fault Crashes on Rural and Urban Roadways in Alabama.” *Accident Analysis & Prevention* 72 (November): 267–76. https://doi.org/10.1016/j.aap.2014.07.003.

Islam, Samantha, Steven L. Jones, and Daniel Dye. 2014. “Comprehensive Analysis of Single- and Multi-Vehicle Large Truck at-Fault Crashes on Rural and Urban Roadways in Alabama.” *Accident Analysis & Prevention* 67 (June): 148–58. https://doi.org/10.1016/j.aap.2014.02.014.

Jacobsen, P L. 2003. “Safety in Numbers: More Walkers and Bicyclists, Safer Walking and Bicycling.” *Injury Prevention* 9 (3): 205–9. https://doi.org/10.1136/ip.9.3.205.

Kaplan, Sigal, and Carlo Giacomo Prato. 2015. “A Spatial Analysis of Land Use and Network Effects on Frequency and Severity of Cyclist–Motorist Crashes in the Copenhagen Region.” *Traffic Injury Prevention* 16 (7): 724–31. https://doi.org/10.1080/15389588.2014.1003818.

Karen Dixon, Kay Fitzpatrick, Raul Avelar, and Das. 2015. “Analysis of the Shoulder Widening Need on the State Highway System.” 2015. https://www.researchgate.net/publication/323946139_Analysis_of_the_Shoulder_Widening_Need_on_the_State_Highway_System.

Kim, Joon-Ki, Gudmundur F. Ulfarsson, Venkataraman N. Shankar, and Sungyop Kim. 2008. “Age and Pedestrian Injury Severity in Motor-Vehicle Crashes: A Heteroskedastic Logit Analysis.” *Accident Analysis & Prevention* 40 (5): 1695–1702. https://doi.org/10.1016/j.aap.2008.06.005.

Kim, Joon-Ki, Gudmundur F. Ulfarsson, Venkataraman N. Shankar, and Fred L. Mannering. 2010. “A Note on Modeling Pedestrian-Injury Severity in Motor-Vehicle Crashes with the Mixed Logit Model.” *Accident Analysis & Prevention* 42 (6): 1751–58. https://doi.org/10.1016/j.aap.2010.04.016.

Kim, K, I M Brunner, and E Y Yamashita. 1953. “Influence of Land Use, Population, Employment, and Economic Activity on Accidents.” *Transportation Research Record*, 9.

Kim, Karl, I. Made Brunner, and Eric Yamashita. 2008. “Modeling Fault among Accident—Involved Pedestrians and Motorists in Hawaii.” *Accident Analysis & Prevention* 40 (6): 2043–49. https://doi.org/10.1016/j.aap.2008.08.021.

Kim, Karl, and Lei Li. 1996. “Modeling Fault among Bicyclists and Drivers Involved in Collisions in Hawaii, 1986–1991.” *Transportation Research Record: Journal of the Transportation Research Board* 1538 (1): 75–80. https://doi.org/10.1177/0361198196153800110.

Kim, Karl, Pradip Pant, and Eric Yamashita. 2010. “Accidents and Accessibility: Measuring Influences of Demographic and Land Use Variables in Honolulu, Hawaii.” *Transportation Research Record: Journal of the Transportation Research Board* 2147 (1): 9–17. https://doi.org/10.3141/2147-02.

LaScala, Elizabeth A., Daniel Gerber, and Paul J. Gruenewald. 2000. "Demographic and Environmental Correlates of Pedestrian Injury Collisions: A Spatial Analysis." *Accident Analysis & Prevention* 32 (September): 651–58. https://doi.org/10.1016/s0001-4575(99)00100-1.

Lee, Chris, and Mohamed Abdel-Aty. 2005. "Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida." *Accident Analysis & Prevention* 37 (4): 775–86. https://doi.org/10.1016/j.aap.2005.03.019.

Lee, Jaeyoung, Mohamed Abdel-Aty, and Keechoo Choi. 2014. "Analysis of Residence Characteristics of At-Fault Drivers in Traffic Crashes." *Safety Science* 68 (October): 6–13. https://doi.org/10.1016/j.ssci.2014.02.019.

Lee, Jaeyoung, Mohamed Abdel-Aty, Keechoo Choi, and Helai Huang. 2015. "Multi-Level Hot Zone Identification for Pedestrian Safety." *Accident Analysis & Prevention* 76 (March): 64–73. https://doi.org/10.1016/j.aap.2015.01.006.

Lee, Jaeyoung, Mohamed Abdel-Aty, and Ximiao Jiang. 2015. "Multivariate Crash Modeling for Motor Vehicle and Non-Motorized Modes at the Macroscopic Level." *Accident Analysis & Prevention* 78 (May): 146–54. https://doi.org/10.1016/j.aap.2015.03.003.

Lindsey, Greg. n.d. "Modeling Bicyclist Exposure to Risk and Crash Risk: Some Exploratory Studies," 73.

Loukaitou-Sideris, Anastasia, Robin Liggett, and Hyun-Gun Sung. 2007. "Death on the Crosswalk: A Study of Pedestrian-Automobile Collisions in Los Angeles." *Journal of Planning Education and Research* 26 (3): 338–51. https://doi.org/10.1177/0739456X06297008.

Luoma, J., J. Schumann, and E. C. Traube. 1996. "Effects of Retroreflector Positioning on Nighttime Recognition of Pedestrians." *Accident; Analysis and Prevention* 28 (3): 377–83.

Marshall, Wesley E., Norman W. Garrick, and Gilbert Hansen. 2008. "Reassessing On-Street Parking." *Transportation Research Record: Journal of the Transportation Research Board* 2046 (1): 45–52. https://doi.org/10.3141/2046-06.

McMahon, Patrick, Chandler Duncan, J. Stewart, Charles Zegeer, and Asad Khattak. 1999. "Analysis of Factors Contributing to 'Walking Along Roadway' Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 1674 (1): 41–48. https://doi.org/10.3141/1674-06.

Miranda-Moreno, Luis F., Jillian Strauss, and Patrick Morency. 2011. "Disaggregate Exposure Measures and Injury Frequency Models of Cyclist Safety at Signalized Intersections." *Transportation Research Record: Journal of the Transportation Research Board* 2236 (1): 74–82. https://doi.org/10.3141/2236-09.

Moeinaddini, Mehdi, Zohreh Asadi-Shekari, and Muhammad Zaly Shah. 2014. "The Relationship between Urban Street Networks and the Number of Transport Fatalities at the City Level." *Safety Science* 62 (February): 114–20. https://doi.org/10.1016/j.ssci.2013.08.015.

Molino, John A., Jason F. Kennedy, Patches L. Johnson, Pascal A. Beuse, Amanda K. Emo, and Ann Do. 2009. "Pedestrian and Bicyclist Exposure to Risk: Methodology for Estimation in an Urban Environment." *Transportation Research Record: Journal of the Transportation Research Board* 2140 (1): 145–56. https://doi.org/10.3141/2140-16.

Narayanamoorthy, Sriram, Rajesh Paleti, and Chandra R. Bhat. 2013. "On Accommodating Spatial Dependence in Bicycle and Pedestrian Injury Counts by Severity Level."

Transportation Research Part B: Methodological 55 (September): 245–64. https://doi.org/10.1016/j.trb.2013.07.004.
National Highway Traffic Safety Administration (NHTSA). 2016. “NHTSA.” Text. National Highway Traffic Safety Administration (NHTSA). 2016. https://www.nhtsa.gov/.
Noland, Robert B., Nicholas J. Klein, and Nicholas K. Tulach. 2013. “Do Lower Income Areas Have More Pedestrian Casualties?” *Accident Analysis & Prevention* 59 (October): 337–45. https://doi.org/10.1016/j.aap.2013.06.009.
Noland, Robert, and Mohammed Quddus. 2004. “Analysis of Pedestrian and Bicycle Casualties with Regional Panel Data.” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1897: 28–33.
Obaidat, Mohammed Taleb, H. R Al-Masaeid, Oift Al-Haji, and Abdalla Qudah. 2007. “A Knowledge- Based System for Pedestrian’s Crossing Behavior through Video Cameras.” ResearchGate. 2007. https://www.researchgate.net/publication/299136030_A_knowledge-_Based_System_for_Pedestrian's_Crossing_Behavior_through_Video_Cameras.
Öström, Mats, and Anders Eriksson. 2001. “Pedestrian Fatalities and Alcohol.” *Accident Analysis & Prevention* 33 (2): 173–80. https://doi.org/10.1016/s0001-4575(00)00028-2.
Oxley, Jennifer A., Elfriede Ihsen, Brian N. Fildes, Judith L. Charlton, and Ross H. Day. 2005. “Crossing Roads Safely: An Experimental Study of Age Differences in Gap Selection by Pedestrians.” *Accident Analysis & Prevention* 37 (5): 962–71. https://doi.org/10.1016/j.aap.2005.04.017.
Papadimitriou, Eleonora, George Yannis, and John Golias. 2012. “Analysis of Pedestrian Exposure to Risk in Relation to Crossing Behavior.” *Transportation Research Record: Journal of the Transportation Research Board* 2299 (1): 79–90. https://doi.org/10.3141/2299-09.
Parkin, John, Mark Wardman, and Matthew Page. 2007. “Estimation of the Determinants of Bicycle Mode Share for the Journey to Work Using Census Data.” *Transportation* 35 (1): 93–109. https://doi.org/10.1007/s11116-007-9137-5.
Penmetsa, Praveena, Srinivas S. Pulugurtha, and Venkata R. Duddu. 2017. “Examining Injury Severity of Not-At-Fault Drivers in Two-Vehicle Crashes.” *Transportation Research Record: Journal of the Transportation Research Board* 2659 (1): 164–73. https://doi.org/10.3141/2659-18.
Peterson, Bercedis, and Frank E. Harrell. 1990. “Partial Proportional Odds Models for Ordinal Response Variables.” *Applied Statistics* 39 (2): 205. https://doi.org/10.2307/2347760.
Prato, Carlo Giacomo, Sigal Kaplan, Thomas Kjær Rasmussen, and Tove Hels. 2016. “Infrastructure and Spatial Effects on the Frequency of Cyclist-Motorist Collisions in the Copenhagen Region.” *Journal of Transportation Safety & Security* 8 (4): 346–60. https://doi.org/10.1080/19439962.2015.1055414.
Preusser, David F, JoAnn K Wells, Allan F Williams, and Helen B Weinstein. 2002. “Pedestrian Crashes in Washington, DC and Baltimore.” *Accident Analysis & Prevention* 34 (September): 703–10. https://doi.org/10.1016/s0001-4575(01)00070-7.
Pucher, John, Charles Komanoff, and Paul Schimek. 1999. “Bicycling Renaissance in North America?” *Transportation Research Part A: Policy and Practice* 33 (September): 625–54. https://doi.org/10.1016/s0965-8564(99)00010-5.

Qin, Xiao, and John N. Ivan. 2001. "Estimating Pedestrian Exposure Prediction Model in Rural Areas." *Transportation Research Record: Journal of the Transportation Research Board* 1773 (1): 89–96. https://doi.org/10.3141/1773-11.

Qin, Xiao, Kai Wang, and Chase E. Cutler. 2013. "Logistic Regression Models of the Safety of Large Trucks." *Transportation Research Record: Journal of the Transportation Research Board* 2392 (1): 1–10. https://doi.org/10.3141/2392-01.

Raford, Noah, and David Ragland. 2004. "Space Syntax: Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety." *Transportation Research Record: Journal of the Transportation Research Board* 1878 (1): 66–74. https://doi.org/10.3141/1878-09.

Raford, Noah, Park Street, and David R Ragland. 2006. "Pedestrian Volume Modeling for Traffic Safety and Exposure Analysis: The Case of Boston, Massachusetts," 24.

Rasmussen, Benjamin, Gabe Rousseau, and William Lyons. 2013. "Estimating the Impacts of the Nonmotorized Transportation Pilot Program: Developing a New Community-Wide Assessment Method," 14.

Retting, Richard A. 1993. "A Review of Fatal Injuries to Pedestrians Induced by Urban Truck Crashes." *Proceedings: Association for the Advancement of Automotive Medicine Annual Conference* 37: 117–28.

Risley, John S. 1985. Review of *Review of MOTOR VEHICLE ACCIDENT RECONSTRUCTION AND CAUSE ANALYSIS. SECOND EDITION*, by Rudolf Limpert. *Jurimetrics* 25 (3): 335–39.

Roberts, I, R Norton, and B Taua. 1996. "Child Pedestrian Injury Rates: The Importance of 'Exposure to Risk' Relating to Socioeconomic and Ethnic Differences, in Auckland, New Zealand." *Journal of Epidemiology & Community Health* 50 (2): 162–65. https://doi.org/10.1136/jech.50.2.162.

Rodgers, Gregory B. 1995. "Bicyclist Deaths and Fatality Risk Patterns." *Accident Analysis & Prevention* 27 (2): 215–23. https://doi.org/10.1016/0001-4575(94)00063-R.

Saha, Dibakar, Priyanka Alluri, Albert Gan, and Wanyang Wu. 2018. "Spatial Analysis of Macro-Level Bicycle Crashes Using the Class of Conditional Autoregressive Models." *Accident Analysis & Prevention* 118 (September): 166–77. https://doi.org/10.1016/j.aap.2018.02.014.

Salon, Deborah. 2016. "Estimating Pedestrian and Cyclist Activity at the Neighborhood Scale." *Journal of Transport Geography* 55 (July): 11–21. https://doi.org/10.1016/j.jtrangeo.2016.06.023.

Sasidharan, Lekshmi, and Mónica Menéndez. 2014. "Partial Proportional Odds Model—An Alternate Choice for Analyzing Pedestrian Crash Injury Severities." *Accident Analysis & Prevention* 72 (November): 330–40. https://doi.org/10.1016/j.aap.2014.07.025.

Savolainen, Peter T., Fred L. Mannering, Dominique Lord, and Mohammed A. Quddus. 2011. "The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives." *Accident Analysis & Prevention* 43 (5): 1666–76. https://doi.org/10.1016/j.aap.2011.03.025.

Schneider, Robert J., Offer Grembek, and Matthew Braughton. 2013. "Pedestrian Crash Risk on Boundary Roadways: University Campus Case Study." *Transportation Research Record: Journal of the Transportation Research Board* 2393 (1): 164–73. https://doi.org/10.3141/2393-19.

Schneider, Robert J., Todd Henry, Meghan F. Mitman, Laura Stonehill, and Jesse Koehler. 2012. "Development and Application of Volume Model for Pedestrian Intersections in San

Francisco, California." *Transportation Research Record: Journal of the Transportation Research Board* 2299 (1): 65–78. https://doi.org/10.3141/2299-08.
Schneider, Robert, Joseph Stefanich, and Larry Corsi. 2015. "Wisconsin Pedestrian and Bicycle Crash Analysis: 2011-2013," 251.
Schreck, Benjamin. 2017. "Cycling and Designing for Cyclists in Germany: Road Safety, Guidelines and Research." *Transactions on Transport Sciences* 8 (1): 44–57. https://doi.org/10.5507/tots.2017.007.
Shaw, John W., Madhav V. Chitturi, Youngjun Han, William Bremer, and David A. Noyce. 2016. "Bicyclist and Pedestrian Safety in Work Zones: Recent Advances and Future Directions." In . https://trid.trb.org/view/1392530.
Siddiqui, Chowdhury, Mohamed Abdel-Aty, and Keechoo Choi. 2012a. "Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes." *Accident Analysis & Prevention* 45 (March): 382–91. https://doi.org/10.1016/j.aap.2011.08.003.
———. 2012b. "Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes." *Accident Analysis & Prevention* 45 (March): 382–91. https://doi.org/10.1016/j.aap.2011.08.003.
Stoker, Philip, Andrea Garfinkel-Castro, Meleckidzedeck Khayesi, Wilson Odero, Martin N. Mwangi, Margie Peden, and Reid Ewing. 2015. "Pedestrian Safety and the Built Environment: A Review of the Risk Factors." Edited by Jack L. Nasar. *Journal of Planning Literature* 30 (4): 377–92. https://doi.org/10.1177/0885412215595438.
Strauss, Jillian, Luis F. Miranda-Moreno, and Patrick Morency. 2013. "Cyclist Activity and Injury Risk Analysis at Signalized Intersections: A Bayesian Modelling Approach." *Accident Analysis & Prevention* 59 (October): 9–17. https://doi.org/10.1016/j.aap.2013.04.037.
———. 2014. "Multimodal Injury Risk Analysis of Road Users at Signalized and Non-Signalized Intersections." *Accident Analysis & Prevention* 71 (October): 201–9. https://doi.org/10.1016/j.aap.2014.05.015.
Stutts, Jane C, William W Hunter, and Wayne E Pein. 1996. "Pedestrian Crash Types: 1990s Update." *TRANSPORTATION RESEARCH RECORD*, 7.
Sullivan, John M., and Michael J. Flannagan. 2002. "The Role of Ambient Light Level in Fatal Crashes: Inferences from Daylight Saving Time Transitions." *Accident Analysis & Prevention* 34 (4): 487–498.
Teschke, Kay, M. Anne Harris, Conor C. O. Reynolds, Meghan Winters, Shelina Babul, Mary Chipman, Michael D. Cusimano, et al. 2012. "Route Infrastructure and the Risk of Injuries to Bicyclists: A Case-Crossover Study." *American Journal of Public Health* 102 (12): 2336–43. https://doi.org/10.2105/AJPH.2012.300762.
Tobey, H. N. 1983. *Pedestrian Trip Making Characteristics and Exposure Measures.* Washington, D.C.: Federal Highway Administration, Office of Safety and Traffic Operations Research and Development.
Ukkusuri, Satish, Samiul Hasan, and H. M. Abdul Aziz. 2011. "Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency." *Transportation Research Record: Journal of the Transportation Research Board* 2237 (1): 98–106. https://doi.org/10.3141/2237-11.
Ulfarsson, Gudmundur F., Sungyop Kim, and Kathleen M. Booth. 2010. "Analyzing Fault in Pedestrian–Motor Vehicle Crashes in North Carolina." *Accident Analysis & Prevention* 42 (6): 1805–13. https://doi.org/10.1016/j.aap.2010.05.001.

Walter, Simon J, and David M Studdert. 2015. “Relationship between Penalties for Road Traffic Infringements and Crash Risk in Queensland, Australia: A Case-Crossover Study.” *International Journal of Epidemiology* 44 (5): 1722–30. https://doi.org/10.1093/ije/dyv148.

Wang, Xuesong, Junguang Yang, Chris Lee, Zhuoran Ji, and Shikai You. 2016. “Macro-Level Safety Analysis of Pedestrian Crashes in Shanghai, China.” *Accident Analysis & Prevention* 96 (November): 12–21. https://doi.org/10.1016/j.aap.2016.07.028.

Wang, Yiyi, and Kara M. Kockelman. 2013. “A Poisson-Lognormal Conditional-Autoregressive Model for Multivariate Spatial Analysis of Pedestrian Crash Counts across Neighborhoods.” *Accident Analysis & Prevention* 60 (November): 71–84. https://doi.org/10.1016/j.aap.2013.07.030.

Wei, Feng, and Gordon Lovegrove. 2013. “An Empirical Tool to Evaluate the Safety of Cyclists: Community Based, Macro-Level Collision Prediction Models Using Negative Binomial Regression.” *Accident Analysis & Prevention* 61 (December): 129–37. https://doi.org/10.1016/j.aap.2012.05.018.

“Welcome to ROSA P | Bicycle Crash Risk: How Does It Vary and Why - 35055 | US Transportation Collection.” n.d. Accessed December 3, 2018. https://rosap.ntl.bts.gov/view/dot/35055.

Wiechel, John F., and Dennis A. Guenther. 1989. “Review of Pedestrian Safety Research in the United States.” In . https://doi.org/10.4271/890757.

Wier, Megan, June Weintraub, Elizabeth H. Humphreys, Edmund Seto, and Rajiv Bhatia. 2009. “An Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications for Land Use and Transportation Planning.” *Accident Analysis & Prevention* 41 (1): 137–45. https://doi.org/10.1016/j.aap.2008.10.001.

Yu, Rongjie, Yin Zheng, Yong Qin, and Mohamed Abdel-Aty. 2019. “Crash Injury Severity Analyses with Multilevel Thresholds of Change Modelling Approach for At-Fault out-of-State Drivers.” *Journal of Transportation Safety & Security*, March, 1–18. https://doi.org/10.1080/19439962.2019.1579290.

Zegeer, Charles V., J. Richard Stewart, Herman Huang, and Peter Lagerwey. 2001. “Safety Effects of Marked Versus Unmarked Crosswalks at Uncontrolled Locations: Analysis of Pedestrian Crashes in 30 Cities.” *Transportation Research Record: Journal of the Transportation Research Board* 1773 (1): 56–68. https://doi.org/10.3141/1773-07.

Zhang, Guangnan, Kelvin K.W. Yau, and Xun Zhang. 2014. “Analyzing Fault and Severity in Pedestrian–Motor Vehicle Accidents in China.” *Accident Analysis & Prevention* 73 (December): 141–50. https://doi.org/10.1016/j.aap.2014.08.018.

Zhang, Hong. 2010. “Identifying and Quantifying Factors Affecting Traffic Crash Severity in Louisiana,” 113.

Zhang, Yuanyuan, John Bigham, David Ragland, and Xiaohong Chen. 2015. “Investigating the Associations between Road Network Structure and Non-Motorist Accidents.” *Journal of Transport Geography* 42 (January): 34–47. https://doi.org/10.1016/j.jtrangeo.2014.10.010.

Zhao, Shanshan, Kai Wang, and Eric Jackson. 2019. “Evaluation of Not-At-Fault Assumption in Quasi-Induced Exposure Method Using Traffic Crash Data at Varied Geographical Levels.” *Transportation Research Record: Journal of the Transportation Research Board*, April, 036119811984103. https://doi.org/10.1177/0361198119841036.

Chapter 3 : Methodologies

3.1 Traditional Crash Prediction Models (CPMs)

Crash Predicting Models (CPMs) are an effective approach to exploring the relationship between crash frequency or crash severity and a set of predictors from the statistical perspective. Once the relationship is established, the mean crash count or the probability of an injury type can be estimated. It is anticipated that the explanatory variables are not only statistically correlated but are logically related to crash occurrence. Such a regression method assumes the error as random noise, and the mean can be represented as the true value around which observations fluctuate. A review is provided on two categories of CPMs: crash frequency models and crash severity models. Only widely implemented models are covered in the scope of this review. Additionally, a brief discussion regarding a limited number of studies, in which the crash count is predicted for a specific primary responsible party, is provided. Later, the deficiencies of the previous research are brought to light and the contribution to the existing literature is defined.

3.1.1 Crash Frequency Models

Crash frequency modeling focuses on establishing a quantitative relationship between crash count and contributing factors based on the statistical significance unveiled from the data. Due to the non-negative integer nature of crash count data, generalized linear models (GLMs) are used in place of linear regression models. Erlander et al. (Erlander, Gustavsson, and Lárusson 1969) were the first to apply a GLM, Poisson regression model, to predict crashes. After that, the Poisson regression model has been widely applied in crash frequency research.

Although the Poisson model can model the crash frequency, its predicting performance usually compromises due to a variety of issues pertaining to the crash count data. Typical issues include over-dispersion, low sample mean, and injury-severity/crash-type correlation. Failure to account for any of these issues would lead to biased coefficient estimates and inaccurate conclusions regarding the properties of the data population.

Different issues may arise due to different characteristics of the crash count data. Over-dispersion can be caused by the omitting of relevant variables. Over dispersed crash data admit more variance than expected under the assumed probabilistic distribution. The issue arises when the mean and variance components are defined by the same parameter of a model such as Poisson where the mean equals the variance. The low-sample-mean issue occurs when some roadway facilities have few observed crashes, usually with excessive zero responses in the crash frequency data. The injury-severity/crash-type correlation could be present due to the fact that different types of crashes or injuries that happen at the same location are correlated because they are the collective outcomes of observed and unobserved effects. Modeling the correlated variables separately may prevent the revelation of crucial relationships, especially the understanding of how different crash types/severities relate to each other.

Based on the Poisson model, various models have been developed and applied to handle these issues. Overall, popular crash frequencies models include Poisson, Poisson-Gamma regression (commonly known as the negative binomial (NB) model), Poisson-Lognormal, and zero-inflated NB model (Qin, Ivan, and Ravishanker 2004; Cheng et al. 2017; El-Basyouny and Sayed 2009; Y. Wang and Kockelman 2013; Lord and Miranda-Moreno 2008; Lord 2006; Park and Lord 2007; Mothafer, Yamamoto, and Shankar 2016; Wei and Lovegrove 2013; Dong et al. 2014; Raihan et al. 2019; Liu et al. 2018; Poch and Mannering 1996; Venkataraman et al. 2011; Lord, Washington, and Ivan 2005; Liu et al. 2018; V. Shankar and Mannering 1996; Prato et al. 2016; Kaplan and Giacomo Prato 2015; Lord and Mannering 2010; Miranda-Moreno, Strauss, and Morency 2011).

3.1.1.1 Base Model: Poisson Model

The conventional base model for analyzing crash frequency is the Poisson model (Lord and Mannering 2010). The Poisson model assumes that the probability of having a specific integral number of crashes follows the Poisson distribution given the mean crash frequency, while the mean crash frequency is defined as a function of independent variables. The Poisson regression model is limited in that the variance of the data is restricted to be equal to the mean. Since both the variance and mean are defined by the same parameter, the Poisson regression model cannot handle over-dispersion. In 1953, a study (K Kim, Brunner, and Yamashita 1953), conducted to study the relationship between population, employment, land use, echometric outcome, and pedestrian-vehicle crashes. Both the Poisson model and the NB model were constructed, and the results of the study showed the preference of the NB model over the Poisson model when there is a need to control for the over-dispersion. Chen applied Poisson and NB models to study the safety effects of bike lanes. The author showed that the Poisson model was used for modeling bicycle crashes were no overdispersion was detected. The conclusion stated that the installation of bike lanes does not increase crash likelihood (L. Chen et al. 2012). In (J. Wang, Huang, and Zeng 2017), a study that aimed at exploring the effect of zonal factors on VRUs crash risk, authors showed that NB model is superior to the Poisson model since over-dispersion occurred in the data because of omission of important variables or due to measurement errors.

Even though the Poisson model is useful for modeling crash outcomes, its assumption of the variance equal to the mean is usually violated when analyzing crash count data (Lord and Mannering 2010). Besides, Poisson and NB model are two models that are unable to account for the unobserved heterogeneity.

3.1.1.2 Models Addressing Over-Dispersion

The Poisson model has been extended to handle the over-dispersion in the crash count data. Two such extended Poisson models that have been widely used are the negative binomial (NB) regression model and the Poisson-lognormal (PLN) regression model. Both models modify the error term of the Poisson model by making it follow some distributions, gamma distribution in the NB model, and normal distribution in the PLN model (Lord and Mannering 2010). Because of the modified error term, the variance of each model is greater than the corresponding mean. Hence, both models can account for over-dispersed crash count data. Schneider et al. adopted the NB model in (R. J. Schneider, Ryznar, and Khattak 2004) for pedestrian-vehicle crash prediction to account for the overdispersion in the crash data. Whereas in (Wei and Lovegrove 2013) the same model was considered to control for overdispersion in data analyzed for developing bicycle CPMs. Lee et al. proposed the use of the Poisson log-normal model as an alternative of the Poisson model for the over-dispersion in the crash data collected for modeling crashes for different transportation modes including VRUs (J. Lee, Abdel-Aty, and Jiang 2015).

3.1.1.3 Models Addressing Low Sample Mean

The low sample mean of the crash count data is usually characterized by the preponderance of an excess number of zero observed crashes in a dataset (Lord and Mannering 2010). Zero-inflated models have been proposed to deal with this issue to improve the estimation accuracy. Zero-inflated models assume there are two types of zero observations: inherently safe conditions imply zero crashes by nature (structural zeros), while inherently unsafe conditions imply zero crashes by chance (sampling zeros) (4). The zero-inflated model inflates the number of zeros by adding zeros from the crash-free state with a count data process characterized by a Poisson or NB distribution, referred to as zero-inflated Poisson (ZIP) model or zero-inflated NB (ZINB) model.

Lord et al. support the fact that ZIP and ZINP models are suitable to account for the excess zeroes that researchers observe through analyzing crash counts (Lord, Washington, and Ivan 2005). Also, a study (Pour et al. 2012)studied the performance of NB, ZIP, and ZINB models on pedestrian-vehicle crash with excess zeroes. The drawn conclusion is that the ZIP model outperformed the other models and is suitable for analyzing pedestrian crashes with a high number of zeroes resulting from sites without any pedestrian activity. Shankar and colleagues, presented a study (V. N. Shankar et al. 2003) on pedestrian-vehicle crash modeling, using models based on negative binomial and mixing distributions. The study results showed ZIP is preferred when modeling pedestrian-vehicle crashes.

3.1.1.4 Models Addressing Injury-Severity/Crash-Type Correlation

According to (Anastasopoulos et al. 2012), crashes must take into account the correlations among crash severity levels and among crash types (i.e., rear-end, Sideswipe) instead of solely modeling each crash category. El-Basyouny (El-Basyouny 2011) concluded that the failure to account for such correlations may be due to unobserved error terms or in the case the analysis resulted in omitted variables. Also, univariate modeling considering independent crash counts may lead to inaccurate results.

Bivariate or multivariate models are applied when the issue of injury-severity and crash-type correlation arises. The crash counts of different injury-severities or crash-types need to be modeled simultaneously, as they are not independent due to shared unobserved factors. The correlation of different injury-severities or crash-types is explicitly modeled by the correlation matrix in the specification of bivariate/multivariate models. By accounting for the correlation, the multivariate count models can provide more accurate estimation and therefore more accurate predictions. Bivariate models jointly model two injury-severities or crash-types, while multivariate models model more than two at the same time. In crash frequency studies, bivariate/multivariate models include bivariate/multivariate Poisson models (J. Ma and Kockelman 2006) (X. Ye et al. 2009) and bivariate/multivariate Poisson-lognormal models (MVPLN) (Park and Lord 2007; J. Ma and Kockelman 2006; J. Lee, Abdel-Aty, and Jiang 2015; Aguero-Valverde and Jovanis 2009; K. Wang et al. 2017). The MVPLN regression model could address overdispersion (where the variance is larger than the mean) in the data.

Given this, crash data showed that crashes between vehicles and pedestrians and crashes between vehicles and bicyclists are highly correlated and have common significant variables affecting both road users equally. Recently, Cheng, Gill, Vo, et al. (Cheng et al. 2018) presented a comprehensive analysis for the estimation of pedestrian and bicyclist crash counts at the TAZ level. Their bivariate Dirichlet process mixture model accounts for the unobserved heterogeneity by combining the strengths of the bivariate specification to include the correlation among crash modes.

Various issues related to the crash count data have been discussed along with models that can mitigate those issues. In summary, the nature of the data guides the selection of a model. Crash frequency, for example, can be assumed to follow a Poisson distribution. When the variance of the crash count is larger than the mean, crash data are said to be over-dispersed. Over-dispersed count data are usually modeled with a negative binomial or Poisson-lognormal distribution. When a dataset includes an excessive number of sites with zero crashes, alternative models such as the zero-inflated models should be considered. When possible correlations between crash types or crash severities exist, bivariate/multivariate Poisson/Poisson-lognormal should be applied to account for such correlation.

3.1.2 Crash Severity Models

Equally if not more important is the task of identifying the contributing factors and their impacts on crash injury severities. The methodologies and techniques for crash severity modeling, like its crash count modeling counterpart, are diverse. But unlike crash count which is a non-negative integer that can change from zero to a large figure, the injury severity outcome has a finite number of alternatives (e.g., a KABCO scale). In economics, discrete choice models describe, explain, and predict choices between two or more discrete alternatives. Moving from simple to complex, from weak to robust, the methodological evolvement of crash severity modeling benefits tremendously from the development of econometrics and from travel demand models where highway route choice and transportation model choice are typical applications for a discrete choice model.

Frequently used statistical methodologies for analyzing crash injury severity include (but not restricted to) multinomial logit (MNL) model (Tay et al. 2011; Ulfarsson and Mannering 2004), mixed logit (ML) model, and ordered logit (OL) or ordered probit (OP) models (Kockelman and Kweon 2002; C. Lee and Abdel-Aty 2005; Zahabi et al. 2011). Both the MNL and ML models assume that the severities are unordered, while the OL/OP models rely on the ordered-severity assumption.

3.1.3.1 Multinomial Logit (MNL) Model

The MNL model is one of the most applied discrete-outcome modeling approach used to estimate the impact of different variables on crash severities (Ulfarsson and Mannering 2004; J.-K. Kim et al. 2007; V. Shankar and Mannering 1996; Tay et al. 2011; J. Lee et al. 2018). The MNL model was first used in the transportation field by McFadden (D. McFadden 1972). After that, many researchers applied this method to model crash severities due to its flexibility in allowing variables to have a dual effect (concave and convex) which navigate to the upper and lower severity levels (J.-K. Kim et al. 2007; Ulfarsson and Mannering 2004).

The MNL model has been extensively applied to model the severity of VRU crashes. Tay et al. (Tay et al. 2011) applied the MNL model to study how a series of factors influence pedestrian-vehicle crash severities in Korea, such as roadway environment, traffic control devices, weather conditions, and pedestrian/driver and vehicle characteristics. Results showed that fatal and serious crashes were associated with crashes involving heavy vehicles, drivers who were under influence of alcohol, male drivers who are under 65 years, pedestrians who are 65 years or more or female pedestrians, roads with high posted speed limits, and inclement weather conditions are a set of factors influencing the probability of each pedestrian injury severity level. Çelik and Oktay (Çelik and Oktay 2014) analyzed two years of pedestrian crash data in New Mexico using the MNL model and revealed that many factors increase the probability of a pedestrian to be involved in fatal crashes (e.g. presence of pedestrian crosswalks, drivers over the age of 65, primary-educated drivers). The finding is beneficial to the development of countermeasures from pedestrians' side. In another study (Kane and Haile 2015), the authors explored the contributing factors that affect injury severity of pedestrian-rail and vehicle-rail crashes using the MNL model. The study showed that foggy weather, open space development areas, and the presence of trucks are factors that are more likely causatives of severe injury and fatal crashes.

One advantage of the MNL model is that it allows the explanatory variables related to one injury severity, as well as their parameter estimations, to vary. The MNL model should be an appropriate model when possibilities of different injury severities are related to different contributing factors or are affected differently by the same factor. Yamamoto et al. (Yamamoto, Hashiji, and Shankar 2008) argued that non-ordinal models may offer unbiased estimates of the parameters, especially in the situation of crash underreporting. Ye and Lord (F. Ye and Lord 2011) examined the influence of crash underreporting on the estimation of crash severities and found that the MNL model was not immune to this issue. The authors suggested setting fatal crashes as the baseline in the MNL model to minimize the bias (F. Ye and Lord 2011).

3.1.3.2 Mixed Logit (ML) Model

Similar to the MNL model, the ML model also assumes the severities are unordered. But unlike the MNL model, the ML model can handle unobserved data heterogeneity which suggests the parameters may vary across different observations. Disregarding data heterogeneity may lead to bias and inefficient statistical inferences (76). The mixed logistic (ML) model overcomes this limitation by allowing parameters to be random.

The ML model can approximate any random utility model (Daniel McFadden and Train 2000), though known of its high flexibility. Though the ML model was discovered a long time ago, it has been extensively used only over the last decade (Sebastien 2008). The model popularity increased due to its ability to address the limitations of the MNL model. The ML model allows for heterogeneous effects and correlation in unobserved factors (Heiss 2016).

J.-K. Kim et al. (J.-K. Kim et al. 2010) applied the ML model to analyze pedestrian injury severity in pedestrian-vehicle crashes. The authors stated that the use of the ML model enhances the interpretation of the model results compared to the MNL model and heteroscedastic logistic model presented in (J.-K. Kim et al. 2008). Furthermore, the ML model predicts the mean and standard deviation values of the probabilities, which is not offered by the MNL model. In (Moore et al. 2011), the authors presented models of bicycle-vehicle injury severity at intersections. The study conducted a MNL model and ML model intending to compare the influence of each of the studied variables. Zhou et al. adopted the ML model to account for the potential unobserved heterogeneity in the effects of the studied characteristics on the collision nature and the vehicles involved in the hit-and-run crashes.

3.1.3.3 Ordered Logit (OL) Model

In 1998, a study (Bhat and Pulugurta 1998) initiated the use of ordered logit to explore the contrast among vehicle ownership decisions. Afterward, researchers studying crash risk factors started applying ordered logit/probit (OL/OP) models, because like vehicle ownership decisions the pedestrian and bicycle injury severity is often reported as an ordered variable. That is, fatality is the highest order while property damage is the lowest. The OL/OP models distinguish the inherent ordering in severity outcomes, hence, become the workhorse for injury severity analysis through the literature. Note the OL model is different from the OP model only in the distribution of the error term.

Given that injury severity levels are ordinal, the OL/OP models have been widely applied to study the relationship between contributing factors and the crash severity outcome. In (O'Donnell and Connor 1996), the authors applied OL/OP intending to predict the severity of motor-vehicle crashes. The conclusion stated that both models were found alike. Factors related to severe injuries among old drivers were explored in (Khattak Aemal J. et al. 2002) using the OL model. Zahabi et al. (Zahabi et al. 2011) applied the OL model to explore the effect of multiple factors (e.g. vehicle type, road connectivity, and vehicle movement) on pedestrian-vehicle injury severity. Conclusions drawn from the estimated model stated that crashes occurring at a signalized intersection resulted in higher injury severity levels to pedestrians compared to bicyclists. Besides, the through movement was found to be associated with sustaining an injury. An important assumption linked with the OL model is the assumption of an equal relationship between each pair of severity categories, commonly known as the proportional odds assumption. The OL model gained popularity and was widely used since it can accommodate the ordered nature of crash severity. The OL model has also been used to model crash injury severity sustained on low-volume rural roads (Prato, Rasmussen, and Kaplan 2014) and pedestrian-rail crash injury severity (Khattak and Tung 2015).

However, the stumbling block of the traditional ordered response model is that the model enforces intense restrains on the threshold parameters' structure, for example, the traditional OL/OP model constrains the thresholds to have fixed values across crashes (Eluru, Bhat, and Hensher 2008).

Three crash severity models have been discussed above. In summary, the MNL model is most widely recognized as an unordered discrete outcome model because it relaxes the effects of similar contributing factors across all injury severities. The ML model has been gaining ground due to its ability to account for unobserved heterogeneity. The OL/OP model is a widely popular ordered discrete outcome model due to its ability to account for the ordinal nature of crash injury severity and its easy estimation procedure.

3.1.3.4 Critical Issues

Previous research studies highlighted the key factors associated with crash frequency and severity. However, the interrelationships among the variables have not been well discovered and research still has not provided an efficient amount of information regarding the direction and strength of the causal effects. During the last three decades, SEM has been considered a valuable tool for researchers. SEM showed the capability to handle complex relationships between both exogenous and endogenous variables. Furthermore, SEM permits the exploration of the direction and strength of the causal effects between the exogenous and endogenous variables, plus it provides the ability to introduce latent variables with unobserved variables. It was previously introduced in many research fields (e.g. travel behavior, natural science, transportation safety, and others (Kuppam and Pendyala 2001; Ulleberg and Rundmo 2003; Hayduk 1987; J.-Y. Lee, Chung, and Son 2008; Golob 2003).

3.1.3 Structural Equation Model (SEM)

Structural equation modeling (SEM) technique is well known in the academic literature due to its ability to capture complex relationships existed between variables, as well as its flexibility with introducing latent variables to accommodate for unobserved variables that are believed to be important in a study. However, the application of SEM in transportation studies is limited, though SEM has been applied in disciplines such as educational, psychological, and political science research (Golob 2003). Since 1980, SEM has been initially used to model travel behavior and was known as the flexible linear-in-parameter multivariate statistical modeling technique (Golob 2003). This technique has not been widely adopted in road traffic safety-related research, especially in pedestrian and bicyclist crash modeling.

SEM enables researchers to test their hypothesis through a hybrid technique involving factor analysis and path analysis (Weston and Gore 2006). The uniqueness of this modeling is considering not only measured variables, but also latent variables which are not directly measured. Mainly, SEM comprises two key elements: measurement and structural models. The measurement model shows the relationship between the observed variables and the latent constructs, whereas, the structural model shows the relationship between the latent constructs. So, when both elements are joined, the full structural equation model is obtained (Weston and Gore 2006).

Following is a summary of studies conducting different models incorporated in SEM. Exploratory factor analysis (EFA) is a tool to investigate the variables incorporated in the dataset, mainly it is responsible for providing a factor structure and is beneficial to the researcher in terms of reducing the number of exploratory variables. Confirmatory factor analysis (CFA) confirms the structure delivered by the EFA and goes beyond variable reduction, as it determines how the selected latent factors in the structural model are measured by the variables in the X-measurement model. More interesting results can be derived from the path analysis which is considered a special case of SEM. Path analysis contains path diagrams, covariates and correlations, and most importantly the direct, indirect, and total effects. It differs from SEM in that it only deals with observed variables with each variable having a single indicator.

3.1.3.1 Exploratory Factor Analysis (EFA)

Towards contributing in measuring transportation facilities' safety level, a study adopted SEM technique to capture the interrelations between several variables (i.e., roadway geometry, driver, and vehicle type-related variables) and latent factor traffic accident size, which is an important index to measure safety level of transportation facilities and is created from the following observed variables: number of injuries, number of fatalities, number of involved vehicles and number of damaged vehicles. Initially, factor analysis was performed on the observed variables aiming at categorizing the variables under multiple factors depending on factor loadings, afterward, the observed variables' correlation matrix was estimated to finally develop the SEM. This SEM suggested that among the three developed latent variables - road, environment, and driver-related latent factors - accident size was highly affected by road-related factors (J.-Y. Lee, Chung, and Son 2008). Schorr and Hamdar (2014) adopted the SEM technique with a plan to develop an intersection safety propensity index for signalized and unsignalized intersections. In this study, it was argued that developing a singular value index provides a better method for ranking intersection safety as compared with the use of multiple criteria such as several vehicles, total injuries, and total fatalities involved in an accident. Due to the absence of a priori knowledge about the factor structure, the authors used factor analysis as an initial step during the model development process, then tested multiple SEMs having different structures and the best fit model according to the root mean square error of approximation (RMSEA) was carefully chosen among the statistically significant converging models (Schorr and Hamdar 2014).

3.1.3.2 Confirmatory Factor Analysis (CFA)

In a research on the relationship between land use and travel behavior, Van Acker and colleagues (2007) have used a priori model from previous literature as a base for creating multiple measurement models through CFA to verify the ability of the measured variables to correctly measure the constructed latent factors. Afterward, a structural regression model (SRM) is created to show the relationships among the latent variables and reach the final SEM with standardized regression weights of the tested variables in a relationship with travel behavior. The major conclusion drawn by the authors is that socio-economic characteristics are the most influential group of characteristics affecting travel behavior (Van Acker, Witlox, and Van Wee 2007). A study aiming to explore the relationship between road accessibility (measured by a latent factor influenced by bus route length, road length, number of intersections, and number of dead ends (Karl Kim, Pant, and Yamashita 2010)) and motor-vehicle crash severity, utilized CFA with a priori model to establish the structural equation model.

The authors have also conducted their study based on a combination of two approaches, starting with CFA to develop the measurement model and following it by the structural model (or causal model) to expose the causal relationships among the studied latent variables (Hassan and Abdel-Aty 2013). In (2013), Hassan and Abdel-Aty adopted SEM to study driver's behavior under reduced visibility conditions. The authors used a different approach to applying SEM in this study; EFA to determine the number and nature of underlying factors-latent factors are also constructed. Afterward, SEM was used to explain the intricate relationships among both variables: the manifest and latent variables. Additionally, the SEM technique can handle a considerable number of exogenous and endogenous variables at once (Hassan and Abdel-Aty 2011).

3.1.3.3 Hybrid of factor analysis and path analysis models

An SEM may be a combination of three advanced statistical analysis techniques, namely: confirmatory factor analysis CFA, path analysis-using observed variables- or exploratory factor analysis EFA, and hybrid models-using path analysis along with latent variables- which in other words provides integration between factor analysis and path analysis (Asparouhov and Muthén 2009). In a study in the traffic safety area, Wang and Qin (2014) adopted the SEM technique to study the influence of driver characteristics, highway geometry, roadway conditions, and environmental factors on single-vehicle crash severity. The influence of the abovementioned variables was tested through three latent variables: collision force, vehicle operating speed before collision occurrence, and severity index. The path analysis model was first conducted to serve as a point of reference with only observed variables since the authors argued that even with a priori assumption, the studied variables may have an indirect relationship with crash outcomes. Then, three SEMs were presented with one, two, and three latent variables, respectively (K. Wang and Qin 2014). The authors concluded that to predict the severity of single-vehicle crashes, using the two latent factors speed and collision force provides the most meaningful estimates.

Hamdar et al., (2008) intended to develop aggressiveness propensity index (API), a latent quantity representing environmental, situational, and driving behavior variables, for intersections by adopting the SEM technique through three approaches. First, the measurement model was developed using CFA which confirmed five factors, instead of six factors that were built from the hypothesized model. Second, the SEM was built to illustrate each dimensions' worth in increasing/decreasing the aggressive driving pattern and eventually the API. Third, after the equations were developed from the SEM, an EFA was implemented and applied on 10 intersections for validation. The major lesson from this study is that the final SEM reveals that the driver's tendency to aggressive behavior during driving might be impacted by factors like several heavy vehicles, traffic volume, and several pedestrians in an intersection (Hamdar, Mahmassani, and Chen 2008).

Najaf and colleagues (2018) utilized SEM to study the complex relationships between diverse urban form characteristics and roadway safety. Basically, the analysis was initiated by an EFA followed by a CFA to develop a model showing the interrelationships amidst the set of independent variables (exogenous variables) affected by other variables through the constructed latent factors and to reveal these variable's covariance structure. Then, path analysis (PA) models were constructed to illustrate the relationship between the constructed latent factors, mediators (that mediate the indirect effect of independent variables on dependent variables), and dependent variables (endogenous variables). Urban traffic areas experienced a safer environment where there is a more job-housing balance, a more polycentric design, and less low-density sprawl among the area's different tracts (Najaf et al. 2018). Choo and Mokhtarian (2007) studied causal relationships between travel, telecommunications, land use, economic activity, and sociodemographic characteristics, through conducting path analysis, followed by building an SEM.

3.2 Summary

Multiple crash count and severity models are discovered in the literature. It is understood that data and the purpose of the study affect the methodology selection. However, many critical issues may arise through the analyses of crash data (e.g., overdispersion, underreporting) that cannot be effectively handled by one method. Hence, triaging data issues and selecting the method for achieving the ultimate research objectives will occur through the course of this dissertation. Also, research will consider incorporating econometric models that enable studying unobserved variables, due to the lack/difficulty of data collection or absence of important variables. For instance, the SEM model technique will be used for its flexibility and ability to detect the intricate relationships among exogenous and endogenous variables.

3.3 References

Aguero-Valverde, Jonathan, and Paul P. Jovanis. 2009. "Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking." *Transportation Research Record: Journal of the Transportation Research Board* 2136 (1): 82–91. https://doi.org/10.3141/2136-10.

Anastasopoulos, Panagiotis Ch., Venky N. Shankar, John E. Haddock, and Fred L. Mannering. 2012. "A Multivariate Tobit Analysis of Highway Accident-Injury-Severity Rates." *Accident Analysis & Prevention* 45 (March): 110–19. https://doi.org/10.1016/j.aap.2011.11.006.

Asparouhov, Tihomir, and Bengt Muthén. 2009. "Exploratory Structural Equation Modeling." *Structural Equation Modeling: A Multidisciplinary Journal* 16 (3): 397–438. https://doi.org/10.1080/10705510903008204.

Bhat, Chandra R., and Vamsi Pulugurta. 1998. "A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions." *Transportation Research Part B: Methodological* 32 (1): 61–75.

Çelik, Ali Kemal, and Erkan Oktay. 2014. "A Multinomial Logit Analysis of Risk Factors Influencing Road Traffic Injury Severities in the Erzurum and Kars Provinces of Turkey." *Accident Analysis & Prevention* 72 (November): 66–77. https://doi.org/10.1016/j.aap.2014.06.010.

Chen, Li, Cynthia Chen, Raghavan Srinivasan, Claire E. McKnight, Reid Ewing, and Matthew Roe. 2012. "Evaluating the Safety Effects of Bicycle Lanes in New York City." *American Journal of Public Health* 102 (6): 1120–27. https://doi.org/10.2105/ajph.2011.300319.

Cheng, Wen, Gurdiljot Singh Gill, Ravi Dasu, Meiquan Xie, Xudong Jia, and Jiao Zhou. 2017. "Comparison of Multivariate Poisson Lognormal Spatial and Temporal Crash Models to Identify Hot Spots of Intersections Based on Crash Types." *Accident Analysis & Prevention* 99 (February): 330–41. https://doi.org/10.1016/j.aap.2016.11.022.

Cheng, Wen, Gurdiljot Singh Gill, Tom Vo, Jiao Zhou, and Taha Sakrani. 2018. "Use of Bivariate Dirichlet Process Mixture Spatial Model to Estimate Active Transportation-Related Crash Counts." *Transportation Research Record: Journal of the Transportation Research Board*, July, 036119811878279. https://doi.org/10.1177/0361198118782797.

Dong, Chunjiao, David B. Clarke, Xuedong Yan, Asad Khattak, and Baoshan Huang. 2014. "Multivariate Random-Parameters Zero-Inflated Negative Binomial Regression Model: An Application to Estimate Crash Frequencies at Intersections." *Accident Analysis & Prevention* 70 (September): 320–29. https://doi.org/10.1016/j.aap.2014.04.018.

El-Basyouny, Karim. 2011. “New Techniques for Developing Safety Performance Functions.” University of British Columbia. https://doi.org/10.14288/1.0063049.

El-Basyouny, Karim, and Tarek Sayed. 2009. “Collision Prediction Models Using Multivariate Poisson-Lognormal Regression.” *Accident Analysis & Prevention* 41 (4): 820–28. https://doi.org/10.1016/j.aap.2009.04.005.

Eluru, Naveen, Chandra R. Bhat, and David A. Hensher. 2008. “A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes.” *Accident Analysis & Prevention* 40 (3): 1033–54. https://doi.org/10.1016/j.aap.2007.11.010.

Erlander, S., J. Gustavsson, and E. Lárusson. 1969. “Some Investigations on the Relationship between Road Accidents and Estimated Traffic.” *Accident Analysis & Prevention* 1 (1): 17–64. https://doi.org/10.1016/0001-4575(69)90004-9.

Golob, Thomas F. 2003. “Structural Equation Modeling for Travel Behavior Research.” *Transportation Research Part B: Methodological* 37 (1): 1–25. https://doi.org/10.1016/s0191-2615(01)00046-7.

Hamdar, Samer H., Hani S. Mahmassani, and Roger B. Chen. 2008. “Aggressiveness Propensity Index for Driving Behavior at Signalized Intersections.” *Accident Analysis & Prevention* 40 (1): 315–26. https://doi.org/10.1016/j.aap.2007.06.013.

Hassan, Hany M., and Mohamed A. Abdel-Aty. 2011. “Analysis of Drivers’ Behavior under Reduced Visibility Conditions Using a Structural Equation Modeling Approach.” *Transportation Research Part F: Traffic Psychology and Behaviour* 14 (6): 614–25. https://doi.org/10.1016/j.trf.2011.07.002.

———. 2013. “Exploring the Safety Implications of Young Drivers’ Behavior, Attitudes and Perceptions.” *Accident Analysis & Prevention* 50 (January): 361–70. https://doi.org/10.1016/j.aap.2012.05.003.

Hayduk, Leslie A. 1987. *Structural Equation Modeling with LISREL: Essentials and Advances*. JHU Press.

Heiss, Florian. 2016. “Discrete Choice Methods with Simulation.” *Econometric Reviews* 35 (4): 688–92. https://doi.org/10.1080/07474938.2014.975634.

Kane, Martin R, and Elias Haile. 2015. “Analyzing Severity of Vehicle Crashes at Highway-Rail Grade Crossings: Multinomial Logit Modeling” 54 (2): 19.

Kaplan, Sigal, and Carlo Giacomo Prato. 2015. “A Spatial Analysis of Land Use and Network Effects on Frequency and Severity of Cyclist–Motorist Crashes in the Copenhagen Region.” *Traffic Injury Prevention* 16 (7): 724–31. https://doi.org/10.1080/15389588.2014.1003818.

Khattak Aemal J., Pawlovich Michael D., Souleyrette Reginald R., and Hallmark Shauna L. 2002. “Factors Related to More Severe Older Driver Traffic Crash Injuries.” *Journal of Transportation Engineering* 128 (3): 243–49. https://doi.org/10.1061/(ASCE)0733-947X(2002)128:3(243).

Khattak, Aemal, and Li-Wei Tung. 2015. “Severity of Pedestrian Crashes at Highway-Rail Grade Crossings” 54 (2): 10.

Kim, Joon-Ki, Sungyop Kim, Gudmundur F. Ulfarsson, and Luis A. Porrello. 2007. “Bicyclist Injury Severities in Bicycle–Motor Vehicle Accidents.” *Accident Analysis & Prevention* 39 (2): 238–51. https://doi.org/10.1016/j.aap.2006.07.002.

Kim, Joon-Ki, Gudmundur F. Ulfarsson, Venkataraman N. Shankar, and Sungyop Kim. 2008. “Age and Pedestrian Injury Severity in Motor-Vehicle Crashes: A Heteroskedastic Logit

Analysis." *Accident Analysis & Prevention* 40 (5): 1695–1702. https://doi.org/10.1016/j.aap.2008.06.005.
Kim, Joon-Ki, Gudmundur F. Ulfarsson, Venkataraman N. Shankar, and Fred L. Mannering. 2010. "A Note on Modeling Pedestrian-Injury Severity in Motor-Vehicle Crashes with the Mixed Logit Model." *Accident Analysis & Prevention* 42 (6): 1751–58. https://doi.org/10.1016/j.aap.2010.04.016.
Kim, K, I M Brunner, and E Y Yamashita. 1953. "Influence of Land Use, Population, Employment, and Economic Activity on Accidents." *Transportation Research Record*, 9.
Kim, Karl, Pradip Pant, and Eric Yamashita. 2010. "Accidents and Accessibility: Measuring Influences of Demographic and Land Use Variables in Honolulu, Hawaii." *Transportation Research Record: Journal of the Transportation Research Board* 2147 (1): 9–17. https://doi.org/10.3141/2147-02.
Kockelman, Kara Maria, and Young-Jun Kweon. 2002. "Driver Injury Severity: An Application of Ordered Probit Models." *Accident Analysis & Prevention* 34 (3): 313–21. https://doi.org/10.1016/s0001-4575(01)00028-8.
Kuppam, Arun R, and Ram M Pendyala. 2001. "A Structural Equations Analysis of Commuters' Activity and Travel Patterns," 22.
Lee, Chris, and Mohamed Abdel-Aty. 2005. "Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida." *Accident Analysis & Prevention* 37 (4): 775–86. https://doi.org/10.1016/j.aap.2005.03.019.
Lee, Jaeyoung, Mohamed Abdel-Aty, and Ximiao Jiang. 2015. "Multivariate Crash Modeling for Motor Vehicle and Non-Motorized Modes at the Macroscopic Level." *Accident Analysis & Prevention* 78 (May): 146–54. https://doi.org/10.1016/j.aap.2015.03.003.
Lee, Jaeyoung, Shamsunnahar Yasmin, Naveen Eluru, Mohamed Abdel-Aty, and Qing Cai. 2018. "Analysis of Crash Proportion by Vehicle Type at Traffic Analysis Zone Level: A Mixed Fractional Split Multinomial Logit Modeling Approach with Spatial Effects." *Accident Analysis & Prevention* 111 (February): 12–22. https://doi.org/10.1016/j.aap.2017.11.017.
Lee, Ju-Yeon, Jin-Hyuk Chung, and Bongsoo Son. 2008. "Analysis of Traffic Accident Size for Korean Highway Using Structural Equation Models." *Accident Analysis & Prevention* 40 (6): 1955–63. https://doi.org/10.1016/j.aap.2008.08.006.
Liu, Chenhui, Mo Zhao, Wei Li, and Anuj Sharma. 2018. "Multivariate Random Parameters Zero-Inflated Negative Binomial Regression for Analyzing Urban Midblock Crashes." *Analytic Methods in Accident Research* 17 (March): 32–46. https://doi.org/10.1016/j.amar.2018.03.001.
Lord, Dominique. 2006. "Modeling Motor Vehicle Crashes Using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter." *Accident Analysis & Prevention* 38 (4): 751–66. https://doi.org/10.1016/j.aap.2006.02.001.
Lord, Dominique, and Fred Mannering. 2010. "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives." *Transportation Research Part A: Policy and Practice* 44 (5): 291–305. https://doi.org/10.1016/j.tra.2010.02.001.
Lord, Dominique, and Luis F. Miranda-Moreno. 2008. "Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-

Gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective." *Safety Science* 46 (5): 751–70. https://doi.org/10.1016/j.ssci.2007.03.005.

Lord, Dominique, Simon P. Washington, and John N. Ivan. 2005. "Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory." *Accident Analysis & Prevention* 37 (1): 35–46. https://doi.org/10.1016/j.aap.2004.02.004.

Ma, Jianming, and Kara Kockelman. 2006. "Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity." *Transportation Research Record: Journal of the Transportation Research Board* 1950 (January): 24–34. https://doi.org/10.3141/1950-04.

McFadden, D. 1972. "CONDITIONAL LOGIT ANALYSIS OF QUALITATIVE CHOICE BEHAVIOR." *WORKING PAPER INSTITUTE OF URBAN AND REGIONAL*, no. 199/. https://trid.trb.org/view.aspx?id=235187.

McFadden, Daniel, and Kenneth Train. 2000. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics* 15 (5): 447–70. https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1.

Miranda-Moreno, Luis F., Jillian Strauss, and Patrick Morency. 2011. "Disaggregate Exposure Measures and Injury Frequency Models of Cyclist Safety at Signalized Intersections." *Transportation Research Record: Journal of the Transportation Research Board* 2236 (1): 74–82. https://doi.org/10.3141/2236-09.

Moore, Darren N., William H. Schneider, Peter T. Savolainen, and Mohamadreza Farzaneh. 2011. "Mixed Logit Analysis of Bicyclist Injury Severity Resulting from Motor Vehicle Crashes at Intersection and Non-Intersection Locations." *Accident Analysis & Prevention* 43 (3): 621–30. https://doi.org/10.1016/j.aap.2010.09.015.

Mothafer, Ghasak I.M.A., Toshiyuki Yamamoto, and Venkataraman N. Shankar. 2016. "Evaluating Crash Type Covariances and Roadway Geometric Marginal Effects Using the Multivariate Poisson Gamma Mixture Model." *Analytic Methods in Accident Research* 9 (March): 16–26. https://doi.org/10.1016/j.amar.2015.11.001.

Najaf, Pooya, Jean-Claude Thill, Wenjia Zhang, and Milton Greg Fields. 2018. "City-Level Urban Form and Traffic Safety: A Structural Equation Modeling Analysis of Direct and Indirect Effects." *Journal of Transport Geography* 69 (May): 257–70. https://doi.org/10.1016/j.jtrangeo.2018.05.003.

O'Donnell, C.J., and D.H. Connor. 1996. "Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice." *Accident Analysis & Prevention* 28 (6): 739–53. https://doi.org/10.1016/s0001-4575(96)00050-4.

Park, Eun Sug, and Dominique Lord. 2007. "Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity." *Transportation Research Record: Journal of the Transportation Research Board* 2019 (1): 1–6. https://doi.org/10.3141/2019-01.

Poch, Mark, and Fred Mannering. 1996. "Negative Binomial Analysis of Intersection-Accident Frequencies." *Journal of Transportation Engineering* 122 (2): 105–13. https://doi.org/10.1061/(ASCE)0733-947X(1996)122:2(105).

Pour, Mehdi Hossein, Joewono Prasetijo, Ahmad Shukri Yahaya, and Seyed Mohammad Reza Ghadiri. 2012. "Modeling Vehicle-Pedestrian Crashes With Excess Zero Along Malaysia Federal Roads." *Procedia - Social and Behavioral Sciences*, SIIV-5th International Congress - Sustainability of Road Infrastructures 2012, 53 (October): 1216–25. https://doi.org/10.1016/j.sbspro.2012.09.970.

Prato, Carlo Giacomo, Sigal Kaplan, Thomas Kjær Rasmussen, and Tove Hels. 2016. "Infrastructure and Spatial Effects on the Frequency of Cyclist-Motorist Collisions in the Copenhagen Region." *Journal of Transportation Safety & Security* 8 (4): 346–60. https://doi.org/10.1080/19439962.2015.1055414.

Prato, Carlo Giacomo, Thomas Kjær Rasmussen, and Sigal Kaplan. 2014. "Risk Factors Associated with Crash Severity on Low-Volume Rural Roads in Denmark." *Journal of Transportation Safety & Security* 6 (1): 1–20. https://doi.org/10.1080/19439962.2013.796027.

Qin, Xiao, John N. Ivan, and Nalini Ravishanker. 2004. "Selecting Exposure Measures in Crash Rate Prediction for Two-Lane Highway Segments." *Accident Analysis & Prevention* 36 (2): 183–91. https://doi.org/10.1016/S0001-4575(02)00148-3.

Raihan, Md Asif, Priyanka Alluri, Wensong Wu, and Albert Gan. 2019. "Estimation of Bicycle Crash Modification Factors (CMFs) on Urban Facilities Using Zero Inflated Negative Binomial Models." *Accident Analysis & Prevention* 123 (February): 303–13. https://doi.org/10.1016/j.aap.2018.12.009.

Schneider, Robert J, Rhonda M Ryznar, and Asad J Khattak. 2004. "An Accident Waiting to Happen: A Spatial Approach to Proactive Pedestrian Planning." *Accident Analysis & Prevention* 36 (2): 193–211. https://doi.org/10.1016/s0001-4575(02)00149-5.

Schorr, Justin P., and Samer H. Hamdar. 2014. "Safety Propensity Index for Signalized and Unsignalized Intersections: Exploration and Assessment." *Accident Analysis & Prevention* 71 (October): 93–105. https://doi.org/10.1016/j.aap.2014.05.008.

Sebastien, Lecocq. 2008. "Variations in Choice Sets and Identification of Mixed Logit Models: Monte Carlo Evidence." 2008. http://agris.fao.org/agris-search/search.do?recordID=LV2016028261.

Shankar, Venkataraman, and Fred Mannering. 1996. "An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity." *Journal of Safety Research* 27 (3): 183–94. https://doi.org/10.1016/0022-4375(96)00010-2.

Shankar, Venkataraman N., Gudmundur F. Ulfarsson, Ram M. Pendyala, and MaryLou B. Nebergall. 2003. "Modeling Crashes Involving Pedestrians and Motorized Traffic." *Safety Science* 41 (August): 627–40. https://doi.org/10.1016/s0925-7535(02)00017-6.

Tay, Richard, Jaisung Choi, Lina Kattan, and Amjad Khan. 2011. "A Multinomial Logit Model of Pedestrian–Vehicle Crash Severity." *International Journal of Sustainable Transportation* 5 (4): 233–49. https://doi.org/10.1080/15568318.2010.497547.

Ulfarsson, Gudmundur F., and Fred L. Mannering. 2004. "Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents." *Accident Analysis & Prevention* 36 (2): 135–47. https://doi.org/10.1016/s0001-4575(02)00135-5.

Ulleberg, Pål, and Torbjørn Rundmo. 2003. "Personality, Attitudes and Risk Perception as Predictors of Risky Driving Behaviour among Young Drivers." *Safety Science* 41 (5): 427–43. https://doi.org/10.1016/S0925-7535(01)00077-7.

Van Acker, Veronique, Frank Witlox, and Bert Van Wee. 2007. "The Effects of the Land Use System on Travel Behavior: A Structural Equation Modeling Approach." *Transportation Planning and Technology* 30 (4): 331–53. https://doi.org/10.1080/03081060701461675.

Venkataraman, Narayan S., Gudmundur F. Ulfarsson, Venky Shankar, Junseok Oh, and Minho Park. 2011. "Model of Relationship between Interstate Crash Occurrence and Geometrics: Exploratory Insights from Random Parameter Negative Binomial

Approach." *Transportation Research Record: Journal of the Transportation Research Board* 2236 (1): 41–48. https://doi.org/10.3141/2236-05.

Wang, Jie, Helai Huang, and Qiang Zeng. 2017. "The Effect of Zonal Factors in Estimating Crash Risks by Transportation Modes: Motor Vehicle, Bicycle and Pedestrian." *Accident Analysis & Prevention* 98 (January): 223–31. https://doi.org/10.1016/j.aap.2016.10.018.

Wang, Kai, John N. Ivan, Nalini Ravishanker, and Eric Jackson. 2017. "Multivariate Poisson Lognormal Modeling of Crashes by Type and Severity on Rural Two Lane Highways." *Accident Analysis & Prevention* 99 (February): 6–19. https://doi.org/10.1016/j.aap.2016.11.006.

Wang, Kai, and Xiao Qin. 2014. "Use of Structural Equation Modeling to Measure Severity of Single-Vehicle Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2432 (1): 17–25. https://doi.org/10.3141/2432-03.

Wang, Yiyi, and Kara M. Kockelman. 2013. "A Poisson-Lognormal Conditional-Autoregressive Model for Multivariate Spatial Analysis of Pedestrian Crash Counts across Neighborhoods." *Accident Analysis & Prevention* 60 (November): 71–84. https://doi.org/10.1016/j.aap.2013.07.030.

Wei, Feng, and Gordon Lovegrove. 2013. "An Empirical Tool to Evaluate the Safety of Cyclists: Community Based, Macro-Level Collision Prediction Models Using Negative Binomial Regression." *Accident Analysis & Prevention* 61 (December): 129–37. https://doi.org/10.1016/j.aap.2012.05.018.

Weston, Rebecca, and Paul A. Gore. 2006. "A Brief Guide to Structural Equation Modeling." *The Counseling Psychologist* 34 (5): 719–51. https://doi.org/10.1177/0011000006286345.

Yamamoto, Toshiyuki, Junpei Hashiji, and Venkataraman N. Shankar. 2008. "Underreporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models." *Accident Analysis & Prevention* 40 (4): 1320–29. https://doi.org/10.1016/j.aap.2007.10.016.

Ye, Fan, and Dominique Lord. 2011. "Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit." *Transportation Research Record: Journal of the Transportation Research Board* 2241 (1): 51–58. https://doi.org/10.3141/2241-06.

Ye, Xin, Ram M. Pendyala, Simon P. Washington, Karthik Konduri, and Jutaek Oh. 2009. "A Simultaneous Equations Model of Crash Frequency by Collision Type for Rural Intersections." *Safety Science* 47 (3): 443–52. https://doi.org/10.1016/j.ssci.2008.06.007.

Zahabi, Seyed Amir H., Jillian Strauss, Kevin Manaugh, and Luis F. Miranda-Moreno. 2011. "Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2247 (1): 81–90. https://doi.org/10.3141/2247-10.

Chapter 4 : Pedestrian and Bicyclist Corridor Crash Analysis

4.1 Introduction

This chapter focus is analyzing crashes involving pedestrians and bicyclists, -vulnerable roadway users (VRUs)-, that are negatively correlated with roadway factors, and positively correlated with environmental and socioeconomic factors. Specific variables representing these factors are often correlated, making it difficult to accurately characterize relationships between individual variables and pedestrian and bicyclist safety. The statistical methods previously used are aimed to construct models that represent the direct relationships between explanatory and dependent variables. However, the causes of crashes often involve intricate relationships among multiple variables, which may not be adequately captured.

In this study, the SEM approach is adopted by integrating exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) which is a special case of SEM to establish the relationship between pedestrian and bicyclist crashes and explanatory variables. Specifically, the following research questions were explored: what are the most important latent variables associated with the frequency of crashes involving VRUs? Among the key latent variables, what combinations of measurable variables provide the most significant representation of these key latent variables? The interrelationships among explanatory variables may be better understood by applying the structural equation model (SEM) technique, as SEM is generally viewed as a combination of factor analysis and path analysis. This highly flexible model structure is capable of representing the complex interrelationship among exogenous and endogenous variables through the inclusion of "unobserved" or latent variables. Specifically, SEM can handle correlations between explanatory variables that represent similar concepts and have overlapped impacts on the dependent variable(s). The benefits of this study may help community planners, transportation researchers, and policymakers with a better understanding of the intricate interrelationship of the influential factors contributing to VRUs road crashes.

4.2 Data and Corridor Selection Process

This study follows Cai and colleagues' recommendation to further study the common unobserved factors affecting pedestrian and bicyclist crashes (Cai et al. 2016). In contrast to previous studies that primarily focused on predicting pedestrian and bicyclist crashes at specific locations (e.g., intersections), this study focuses on a sample of 200 one-mile-long highway corridors in Wisconsin. **Figure 4-1** illustrates the corridor selection process. The corridors are in the areas with at least 100 residents per square mile, generally including cities, suburbs, and villages but excluding rural areas in Wisconsin. Although spatial diversity is desirable, the focus was on urbanized areas since these areas tend to have higher volumes of pedestrians and bicyclists and more pedestrian and bicyclist crashes. Among the 200 study corridors, most are located in Southeast Wisconsin; 115 had at least one reported pedestrian crash and 67 had at least one reported bicycle crash.

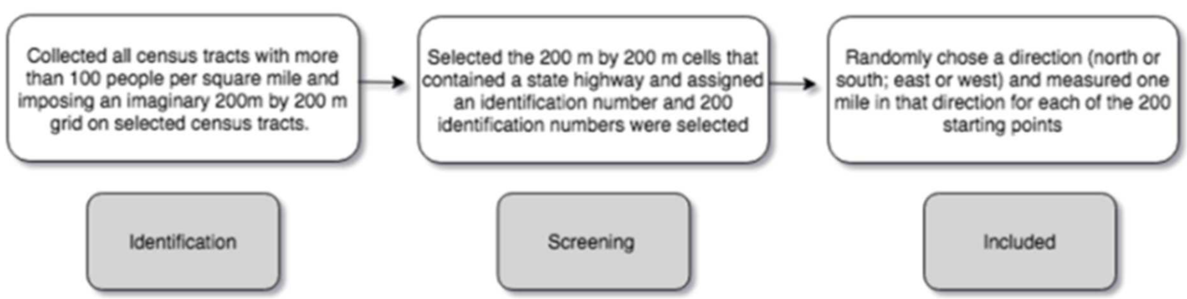


Figure 4-1: Corridor Selection Process

This study examined the frequency of pedestrian and bicyclist crashes reported to police between 2011 and 2015 in each study corridor. These data were gathered from the Wisconsin Department of Transportation (WisDOT) WisTransPortal Database and only included crash records with latitude and longitude coordinates. Explanatory variables were collected from multiple databases including the WisDOT highway inventory, US Environmental Protection Agency (EPA) Smart Location Database, US Census Topologically Integrated Geographic Encoding and Referencing (TIGER/Line) dataset, and Google Maps and Google Street View imagery. Explanatory variables included exposure-related variables (e.g., annualized average daily traffic (AADT)), roadway segment characteristics (e.g., motor vehicle AADT, the average number of through lanes, and posted speed limit), roadway intersection characteristics (e.g., number of residential/non-residential driveways, number of signalized/un-signalized intersections, number of right-turn/left-turn lanes on state highway approaches to all intersections), and socioeconomic data from surrounding census tracts. None of the study corridors had pedestrian or bicyclist counts, so proxy variables were used to represent pedestrian and bicyclist exposure, such as the percentage of workers who regularly walked or bicycled to work, population density, and job density in the surrounding neighborhoods (based on census block groups). **Table 4-1** caries summary statistics and description of the dataset.

Table 4-1: Description and Summary Statistics of The Corridor Variables (N=200)

Notation	**Description**	**Coding**	**Mean (Standard Deviation) or Percentage**
Wisconsin Information System for Local Roads (WISLR)			
Ped_1115	Number of pedestrian crashes (2011-2015)	Continuous	1.91 (3.4)
Bike_1115	Number of Bicyclist crashes (2011-2015)	Continuous	0.85 (1.69)
Google Maps and Google Street View Imagery			
High_Spd_Lmt	Posted speed limit higher than 35 mph	1 = Yes 0 = No	1 = 46% 0 = 54%

Pav_Shoulder	Percentage of corridor covered by paved shoulders on both sides (shoulder on only one side for full length = 0.5)	1 = Yes 0 = No	1 = 72 % 0 = 28 %
Bike lanes	Percentage of corridor covered by designated bike lanes on both sides (bike lane on only one side for full length = 0.5)	Continuous	0.09 (0.26)
Sidewalk	Percentage of corridor covered by sidewalks on both sides (sidewalk on only one side for full length = 0.5)	Continuous	0.39 (0.44)
Sidepath	Percentage of corridor covered by side paths on both sides (side path on only one side for full length = 0.5)	Continuous	0.05 (0.19)
Unsignalized	Unsignalized intersections along the corridor	Continuous	5 (4)
Mid_Block	Marked midblock crosswalks across the state highway along the corridor	Continuous	0.04 (0.24)
TWLTL	Percentage of corridor length with a two-way left-turn lane	Continuous	0.06 (0.16)
US Census TIGER/Line dataset			
Log_AADT	Natural log of the average of all Annualized Average Daily Traffic (AADT) volume counts along the corridor	Continuous	9.359 (0.66)
Walk	Transportation mode used to travel to work (walking)	Continuous	0.027 (0.030)
Bike	Transportation mode used to travel to work (biking)	Continuous	0.006 (0.012)
Employ_Density	Gross employment density (jobs/acre) on unprotected land	Continuous	2.18 (3.81)
Edu_Less_H	Percentage of educational attainment for the population 25 years and over: less than high school	Continuous	0.09 (0.07)
EPA/SLD			
Total_Veh0	Percentage of population with zero car ownership	Continuous	0.07 (0.07)

Low_Wage	Percentage of workers earning $1250/month or less (home location), 2010 decennial Census	Continuous	0.28 (0.04)
Poverty	poverty status in the past 12 months by disability status by employment status for population 20 to 64 years for whom poverty status is determined (percentage)	Continuous	0.12 (0.10)

4.3 Methodology

The primary interest of using SEM lies in the test of its theoretical construct which variables and their relationships. As shown in **Figure 4-2** (31), an SEM model can be depicted in a path diagram consisting of boxes and circles, which are connected by arrows. Observed variables are usually represented by square or rectangular boxes (i.e., TWLTL), while unobserved or latent variables are usually represented by circles or eclipses (i.e., Low Social Status). A directional arrow (or path) in the model usually indicates a statistical dependence, in which the variable at the tail of the arrow causes the variable at the point. A double-headed arrow does not represent such a statistical dependence, but an indication of correlation between variables. Through the x-measurement model for exogenous variables, a y-measurement model for endogenous variables, and the structural model between latent variables, SEM can differentiate between direct, indirect, and total effects between variables. By combining the structural model with measurement models, SEM expresses the regression effects of exogenous "independent" variables on the endogenous "dependent" ones, as well as, expressing autocorrelation "effects between endogenous variables." For more details see (Schumacker and Lomax 2004).

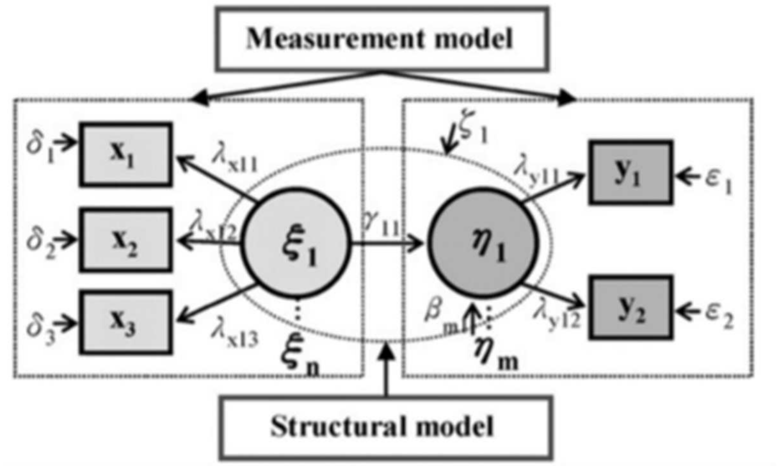


Figure 4-2: Example of A Structural Equation Model (Variable Definitions are Shown In Table 4-2)

The formulation of SEM in Equation 1 suggests a structure between the covariances between observed variables (36):

$$\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} A_y & 0 \\ 0 & A_x \end{bmatrix} \begin{bmatrix} \eta \\ \xi \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \delta \end{bmatrix} \qquad \text{Eq. 4-1}$$

Table 4-2: SEM elements

Model	Variable	Variable Description
	x	q x 1 column vector of observed exogenous variables
	y	p x 1 column vector of observed endogenous variables
	ξ	n x 1 column vector of latent exogenous variables
	η	m x 1 column vector of latent endogenous variables
Measurement	δ	q x 1 column vector of measurement error terms for observed variables x
	ε	p x 1 column vector of measurement error terms for observed variables x
	A_x	The matrix (q x n) of structural coefficients for latent exogenous variables to their observed indicator variables
	A_y	The matrix (q x n) of structural coefficients for latent endogenous variables to their observed indicator variables

Structural	Γ	The matrix (m x n) of regression effects for exogenous latent variables to endogenous latent variables
	β	The coefficient matrix (m x m) of direct effects between endogenous latent variables
	ς	m x 1 column vector of error terms

The model goodness of fit can be measured by the comparative fit index (CFI) (Bentler 1990) in Eq. 4-1 and the root mean square error of approximation (RMSEA) Browne and Cudeck 1992) in Eq. 4-2.

$$CFI = 1 - \frac{\tau\ est.model}{\tau\ indep.model} \qquad \text{Eq. 4-2}$$

Where,

τ indep. model = $X^2_{indep.model}$ - $df_{indep.model}$

τ est. model = $X^2_{est.model}$ - $df_{est.model}$

$$\text{Estimated RMSEA} = \sqrt{\frac{\tau\ est.model}{Ndf_{model}}} \qquad \text{Eq. 4-3}$$

Where,

τ_i: the degree of misspecification of the model

X^2: chi-square statistic

df: the degree of freedom

N: sample size

CFI index is calculated using $\chi 2$ statistics for two models: the target and the baseline models; and measures how better the model fits with a comparison to the baseline model. The baseline model includes means and variances of the observed variables in addition to the covariances of the observed exogenous variables. Both indices (root mean square error of approximation (RMSEA) and comparative fit index (CFI)) assume that the target model is approximately correct, but CFI carries another assumption that the baseline model is also correct. CFI is based on the assumption that all latent variables are uncorrelated and performs well even when the sample size is small

(Tabachnick and Fidell 2006). Values for CFI range between 0 and 1, with a value closer to 1 indicates a better fit. Root mean square error of approximation (RMSEA)-which is a function of chi-square and degree of freedom- measures the difference between the observed and predicted values (Tiadatara 2009). A value of less than 0.08 indicates a good fit model.

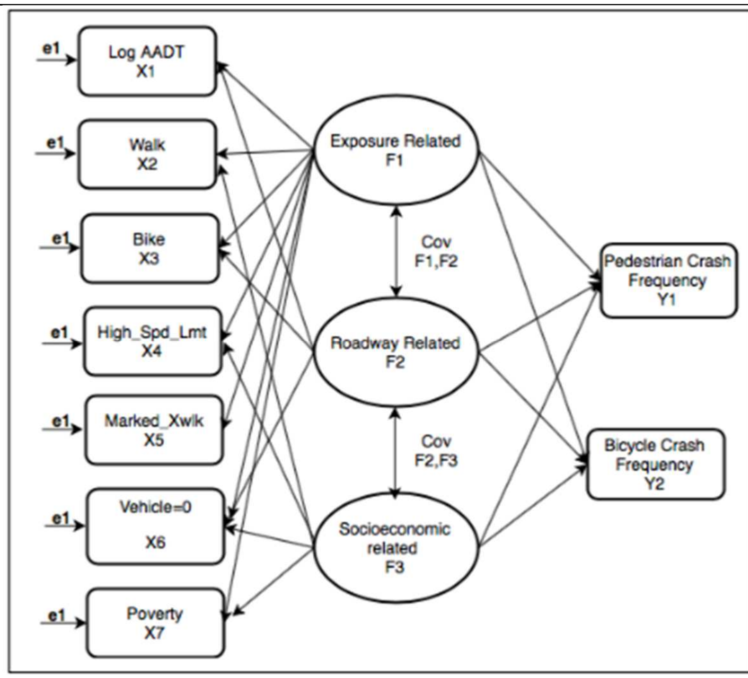
4.3.1 SEM Model Specification, Estimation and Evaluation

4.3.2 Exploratory Factor Analysis (EFA)

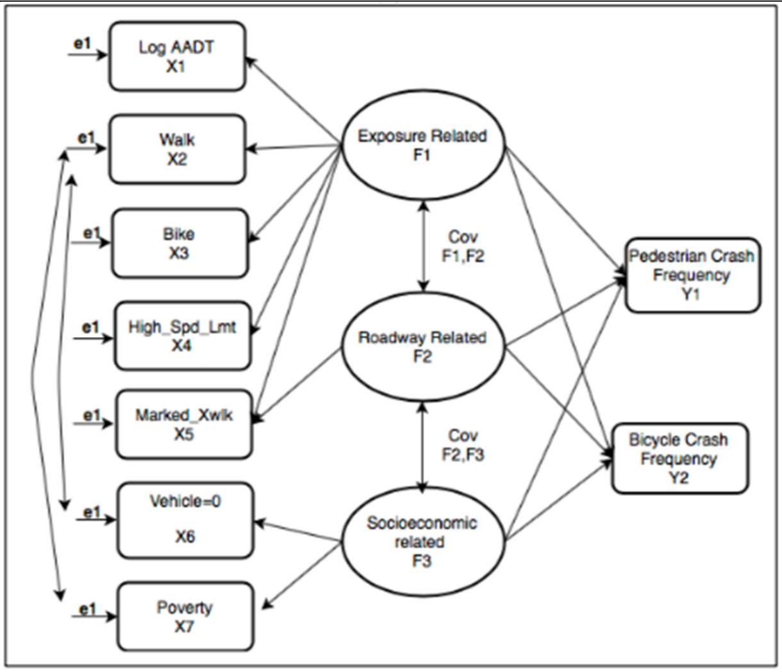
A hypothesized model assessed the relative factors affecting pedestrian and bicyclist crash risk **(Figure 4-3)** and indicated that exposure, roadway, and socioeconomic should be used as latent variables that connect exogenous and endogenous variables. The behavioral-related latent variable is not included since the dataset lacks behavioral input variables. The three latent variables (oval shape) are predictors of the number of crashes that involve either pedestrians or bicyclists (square shape) taking place on the study segments. The latent variables are allowed to correlate. A substantive theoretical model does not exist, so the exploratory factor analysis (EFA) is used to obtain the empirical factor model and explore the structural portion of SEM. EFA assumes that every observed variable is an indication or a measurement of a latent variable **(Figure 4-3a).** EFA is usually performed as a precursor to confirmatory factor analysis (CFA) (Pett, Lackey, and Sullivan 2003) which confirms theoretically valid relationships **(Figure 4-3b).** Researchers vary in terms of sample size recommendations for factor analysis. A sample size of 200 and the sample-to-variable ratio of 3:1 kept this study within the acceptable ranges for applying factor analysis (39).

To test the ability to apply factor analysis using this study's data, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was applied. It is hypothesized that the correlation matrix is an identity matrix so that Bartlett's Test of Sphericity to test this hypothesis (Hamdar, Mahmassani, and Chen 2008) was used. Bartlett's test of sphericity resulted in a P-value of 0.02

(<0.5 is recommended), and a KMO index value of 0.8091 (>0.6 is recommended); both are considered meritorious (Costello and Osborne 2005). The strength of the linear relationship between two variables is crucial. Many variables in the study dataset were highly correlated (i.e., the correlation between the percentage of high wage and total vehicles of two or more is 0.79), and these variables were not used together in the same model. A threshold of 0.5 was accepted as a correlation coefficient between the set of variables chosen for the analysis (39, 40).



(a)



(b)

Figure 4-3: Illustration of the Conceptual Distinction Between EFA (a) and CFA (b)

The number of latent variables can be evaluated using the visual tool called the Scree test. The Scree test showed a clear drop between the third and fourth components, meaning the most suitable number of factors lies between three and four factors. Additionally, goodness of fit indices in the three-factor EFA model show acceptable values (RMSEA = 0.000; CFI = 1.064).

The estimation method for factor loading coefficients, which measure the strength between observed and latent variables, relies on data quality. The Maximum likelihood (ML) or principal axis factoring (PAF) method is recommended, depending on whether the data is normally or significantly non- normally distributed (Wier et al. 2009). The ML estimation method was chosen after variables were standardized through the "scale function" and the rotation method is variance maximizing (varimax) rotation. EFA was conducted using the first part of the sample size of 100 corridors. **Table 4-3** shows the results of the exploratory factor analysis. Despite the cross-loading that appeared in the (High_Spd_Lmt) variable, the observed variables that are highly correlated with factors show distinctive characteristics.

(High_Spd_Lmt), (Bikelane), (Pav_Shoulder), (Sidewalk), and (Unsignalized) are highly correlated with F1 which can be called pedestrian and bicycle-oriented roadway. (Walk), (Bike), (Employ_Density), and (log_AADT) are highly correlated with F2 which can be called exposure. (Edu_Less_H), (Total_Veh0), (Low_Wage), and (Poverty) are highly correlated with F3 which can be called low social status. Hence, three factors – exposure, social status, and pedestrian and bicycle-oriented roadway – were constructed from the observed variables in the data collection. Factor loadings >0.4 are in bold.

Table 4-3: EFA Loadings for Measurement Models (N=100)

Variable	Factors		
	Pedestrian and Bicycle-Oriented Roadway (F1)	Exposure (F2)	Low Social Status (F3)
High_Spd_Lmt	**-0.55**	**-0.40**	-0.28
Bikelane	**0.41**	0.11	0.11
Pav_Shoulder	0.38	-0.11	-0.06
Sidewalk	**0.63**	0.04	0.17
Sidepath	-0.06	0.01	0.01
Unsignalized	**0.74**	0.14	0.09
Mid_Block_	0.03	0.02	0.00
TWLTL	-0.25	0.20	0.22
Walk	0.10	**0.69**	-0.02
Bike	0.12	**0.65**	0.17
Employ_Density	0.14	**0.54**	0.26
log_AADT	0.16	**0.48**	0.11
Edu_Less_H	-0.09	0.02	**0.78**
Total_Veh0	0.1	0.16	**0.86**
Low_Wage	0.15	0.27	**0.71**
Poverty	0.14	0.32	**0.90**

4.3.2.1 Confirmatory factor analysis

The CFA model was analyzed using the remaining observations. CFA is often used to evaluate a prior theory or hypothesis such as the number of factors, types of factors, whether or not the factors are correlated, and which observed variables are indicators of which factor. Now, given the EFA results, CFA helps cross-validate the structure as well as the factor loading since EFA is purely data-driven. Prior knowledge informs that the presence of ped/bike-friendly facilities, percent of the working population, AADT, walking/biking, and gross employment density are considered to be related to pedestrian and bicyclist exposure. However, the high score of factors loading in the EFA suggested that high-speed limit is also a strong indicator of exposure and thus, the high-speed limit was used as an indicator for the latent factor exposure in CFA. By contrast, EFA indicated paved shoulder has a low correlation with pedestrian and bicycle-oriented roadway or any of the three factors but it was kept in CFA because of prior knowledge.

The exploratory factor structure suggests that a CFA is fitted based on three latent variables in the X measurement model (Figure 4-3b). The fourth latent variable was added following the similar concepts in (32, 33). The fourth latent variable in the Y measurement model (Figure 4-3b) is the endogenous latent variable, so-called “Crash Index”, which is measured by pedestrian crashes and bicyclist crashes. **Figure 4-4** illustrates the resulting SEM with all latent variables. All variables were significant at a 5% level, and non- significant variables were removed (e.g., Mid_Block and Sidepath).

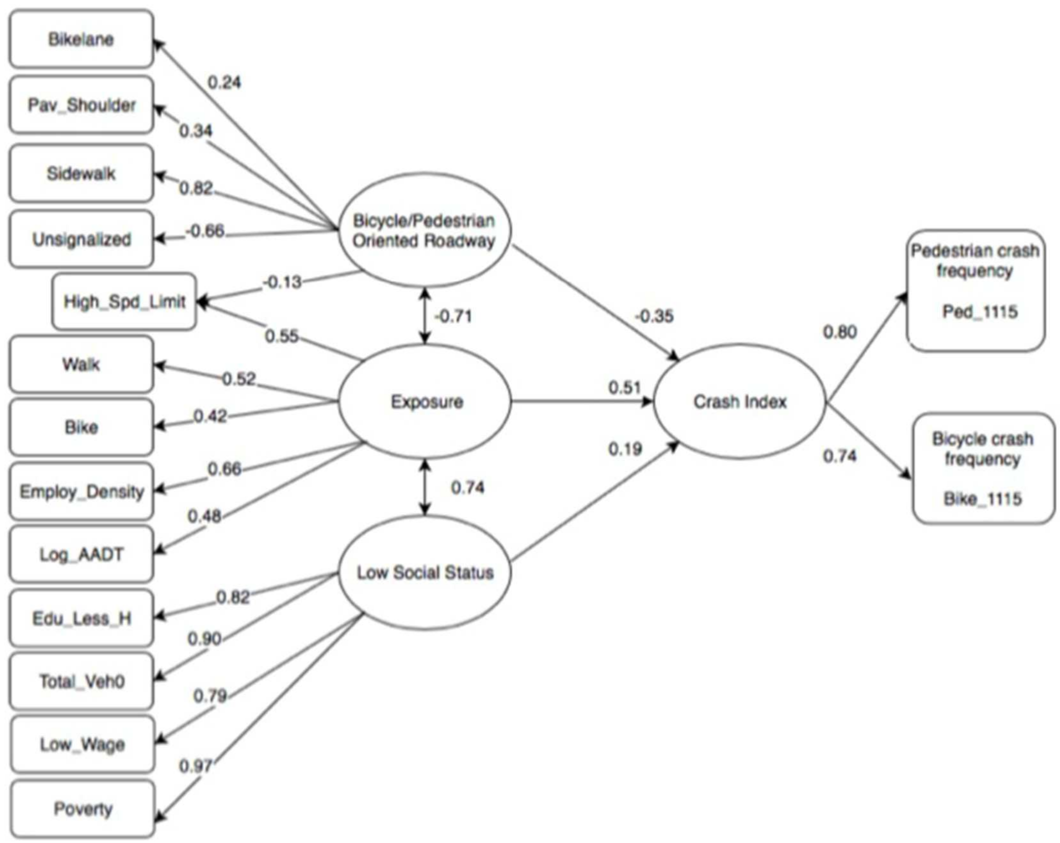


Figure 4-4: Final Structural Equation Model

The overall fit of the model and the significance of some model parameters were evaluated. Both the RMSEA and CFI indices are within the cut-off values of (0.064) and (0.930), respectively. Hence, the model does fit despite the result of the chi-square test.

4.4 Results

The SEM technique enhances safety studies with its ability to build a structure among variables (e.g., pedestrian and bicyclist safety studies at intersections). The ability to include multiple endogenous measures (e.g., pedestrian crashes, bicyclist crashes) is a benefit because it results in a more informative framework. SEM also guides with safety-related data collection, and it highlights pertinent variables that can be gathered to represent important latent variables. Several models were tested to identify a statistically significant model.

The final SEM displays standardized parameters for all coefficients. The structural model which can be viewed as a standard regression equation using standardized parameters include latent exogenous variables bicycle/pedestrian-oriented roadway, exposure, and low social status and latent endogenous variable crash index. The regression coefficients show that the crash index is strongly and positively influenced by exposure latent variable (coefficient = 0.51), moderately and negatively affected by bicycle/pedestrian-oriented roadway (coefficient = -0.35), and weakly and positively affected by the low social status (coefficient = 0.19). Regarding the measurement models, the x-measurement model implies that sidewalk coverage along the corridor, bike lane, and paved shoulder coverage are good features of a bicycle/pedestrian-oriented roadway. Bike lanes, paved shoulders, and sidewalks may lead to higher exposure for pedestrians and bicyclists, but they also provide designated space for these VRUs and may decrease the likelihood of crashes. The y-measurement model implies pedestrian and bicycle crash count is strong and positive measures of the crash index.

The low social status latent variable was positively and highly influenced by many variables (e.g., low educational level, and low wage). The results show a positive effect between lower educational level and crash frequency: well-educated residents may have had more driver education training and may be more aware of road safety and the consequences of crashes. People who live in lower-income neighborhoods may travel more by walking and bicycling due to limited resources for automobile travel. Lower rates of car ownership may be positively related to crash frequency through increased pedestrian and bicyclist exposure. It is also possible that areas with higher-income residents may have environments that are more conducive to biking and walking (e.g., more high-quality pedestrian and bicycle infrastructure).

Looking at the exposure latent variable, walking or biking as a transportation mode, in addition to employment density and log_AADT positively affect the pedestrian and bicyclist's exposure to traffic, hence, lead to more crashes. The high-speed variable was having dual citizenship, meaning that it was correlated to both the bicycle/pedestrian-oriented roadway and the latent exposure variable. However, the results show that it is positively related to the exposure variable (0.55), but negatively related to a bicycle/pedestrian-oriented roadway latent variable. As speed increases, pedestrian and bicycle crashes may be more likely because drivers may not detect pedestrians and bicyclists on the sides of the road as well and longer stopping distances are needed to avoid collisions. In contrast, high traffic volumes may increase traffic congestion and reduce motor vehicle speeds, leading to an overall reduction in crash frequency.

The results contain similar conclusions from previous studies. Exceeding the speed limit showed an increase in the probability of being involved in a crash (32, 33). It was significant in bicycle/pedestrian-oriented roadway (-0.13) and exposure latent variables (0.55) but had a higher impact on exposure latent variables. This underscores the value of SEM since it can clarify complex relationships between variables. A unique conclusion is derived from the correlation between two exogenous latent variables, showing the high positive correlation between low social status and exposure. Owning zero vehicles (shows a low social status) will increase the individual's exposure and therefore increase his/her crash index leading to more crash involvement. Also, the presence of paved shoulders tends to decrease the crash index (indirectly) by improving the road design for pedestrians and bicyclists. Paved shoulders provide additional space for pedestrians and bicyclists outside of travel lanes, even the elevation between the shoulder and the roadway, and reduce the presence of gravel or sand that may contribute to bicyclist crashes.

4.5 References

Bentler, P.M. 1990. “Fit Indexes, Lagrange Multipliers, Constraint Changes and Incomplete Data in Structural Models.” *Multivariate Behavioral Research* 25 (2): 163–72. https://doi.org/10.1207/s15327906mbr2502_3.

Browne, Michael W., and Robert Cudeck. 1992. “Alternative Ways of Assessing Model Fit.” *Sociological Methods & Research* 21 (2): 230–58. https://doi.org/10.1177/0049124192021002005.

Cai, Qing, Jaeyoung Lee, Naveen Eluru, and Mohamed Abdel-Aty. 2016. “Macro-Level Pedestrian and Bicycle Crash Analysis: Incorporating Spatial Spillover Effects in Dual State Count Models.” *Accident Analysis & Prevention* 93 (August): 14–22. https://doi.org/10.1016/j.aap.2016.04.018.

Costello, Anna B, and Jason Osborne. 2005. “Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis.” *Exploratory Factor Analysis* 10 (7): 10.

Hamdar, Samer H., Hani S. Mahmassani, and Roger B. Chen. 2008. “Aggressiveness Propensity Index for Driving Behavior at Signalized Intersections.” *Accident Analysis & Prevention* 40 (1): 315–26. https://doi.org/10.1016/j.aap.2007.06.013.

Pett, Marjorie, Nancy Lackey, and John Sullivan. 2003. *Making Sense of Factor Analysis*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc. https://doi.org/10.4135/9781412984898.

Schumacker, Randall E., and Richard G. Lomax. 2004. *A Beginner’s Guide to Structural Equation Modeling*. 2nd ed. Lawrence Erlbaum Associates. http://gen.lib.rus.ec/book/index.php?md5=77afe29f322c8b7a656a3ee94ecd6f2d.

Tabachnick, Barbara G., and Linda S. Fidell. 2006. *Using Multivariate Statistics (5th Edition)*. Needham Heights, MA, USA: Allyn & Bacon, Inc.

Tiadatara, Den Iwan. 2009. “Joseph F. Hair, William C. Black, Barry J. Babin, Rolph E. Anderson Multivariate Data Analysis 7th Edition 2009.” https://www.academia.edu/6955544/Joseph_F._Hair_William_C._Black_Barry_J._Babin_Rolph_E._Anderson_Multivariate_Data_Analysis_7th_Edition_2009.

Wier, Megan, June Weintraub, Elizabeth H. Humphreys, Edmund Seto, and Rajiv Bhatia. 2009. “An Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications for Land Use and Transportation Planning.” *Accident Analysis & Prevention* 41 (1): 137–45. https://doi.org/10.1016/j.aap.2008.10.001.

Chapter 5 : Party At-Fault Assignment And Analysis

5.1 Introduction

Pedestrian and bicyclist safety emerge as the top safety concerns with the increase in mass motorization of our society, as vulnerable road users (VRU) are more likely to suffer from serious injuries and fatalities when involved in a crash. For each trip, pedestrians are 1.5 times more likely to be killed in a crash than vehicle occupants (CDC, 2019). Identifying important risk factors of pedestrian-vehicle crashes will help address the existing safety deficiencies and facilitate the development of proactive countermeasures. However, crashes are complex events in which each party involved could share different responsibilities: a driver, a VRU, or both parties can be at fault. The term "fault", or alternative terms such as "primarily responsible" (R. Schneider, Stefanich, and Corsi 2015; Young and Salmon 2015; Romano, Voas, and Camp 2017), "culpability" (Dorn and af Wåhlberg 2019), is commonly used in this line of research to describe which party is liable for a traffic crash. Determining who is at-fault and identifying the underlying factors are tremendously beneficial for understanding the possible crash circumstances and causes and thus, instrumental to the design of appropriate safety solutions.

One of the primary sources regarding the possible fault party is the police crash report, in addition to roadways, environmental, personal, and crash-related factors. In this chapter, the first objective is to explore appropriate crash characteristics and contents in the crash report that assist in determining the fault party and to propose procedural guidance on investigating and assigning a certain party to be at-fault. The second objective is to examine the association between the at-fault party and the crash injury risk and by quantifying the effect of the roadway, environmental, personal, and crash-related factors on the judgment and decision-making of the at-fault party.

Specifically, the investigation of crash reports contains two parts: first, data fields and narrative section in a crash report are carefully screened for the indication of the party at fault; second, machine learning and statistical models (e.g. Z-test, XGboost, and multinomial logistic (MNL)) are employed to quantify the effect of certain factors on a party at-fault and corresponding crash injury severities. The findings are anticipated to provide crucial input for developing and deploying effective enforcement, education, and engineering solutions that target to improve safety behavior and actions of drivers, VRUs, or both.

5.2 Literature Review

The literature review is divided into three major sections: 1) methods for assigning fault party, 2) relevant methods for classifying crash injury severity, and 3) relevant and key findings from previous studies for the at-fault parties.

5.2.1 Methods for Assigning Party At-Fault

Several studies used crash reports where the at-fault party is assigned by the police officer at crash time, the process involved referring back to the crash report as the target data source (i.e., (Islam and Jones 2014; Islam and Hossain 2019).

A few research obtained the needed information for fault assignment from the police narrative (R. Schneider, Stefanich, and Corsi 2015). The authors stated that since the police do not assign "fault" for crashes in Wisconsin, the study utilized the detailed narrative on the MV4000 crash forms, along with the type of citation issued at the time of the crash. This method helped to interpret which party or parties the police officer viewed as being primarily responsible for the crash.

While other researchers used the available data fields in the crash report for gathering the key factors (C. Lee and Abdel-Aty 2005). Lee and Abdel-Aty (2005), investigated the effects of average traffic volume at intersections (that is not readily available in crash reports) on pedestrian crashes. In their study, the authors assigned the fault party based on personal characteristics. Pedestrian crashes were classified into two types: 1) crashes at driver's fault, and 2) crashes at pedestrian's fault. The premise of this classification is that crashes at driver's fault are more associated with driver characteristics whereas crashes at pedestrian's fault are more associated with pedestrian characteristics. In crashes where the driver received a traffic violation citation, the crash was classified as crashes at the driver's fault. Else, crashes were classified as crashes at pedestrian's fault (C. Lee and Abdel-Aty 2005).

Regarding the research based on obtaining evidence on the fault party from the picture and video recording. Currie and Logan (2014), took into account key driver, vehicle, roadway, and environmental characteristics that are hypothesized to influence the probability of bus drivers being deemed at-fault in bus-involved accidents, and excluded certain driver-specific details (e.g., risk perception, and educational level). The methodology used is considered to be less biased compared to police records and self-reported data (Goh et al. 2014).

Researchers also utilized crash tools to gather information about the crash circumstances such as the PBCAT tool. Das and colleagues (2020), associated the PBCAT tool with a framework developed for applying ML models to classify pedestrian crash types and assign the fault to different parties (i.e., drivers, pedestrians, both, none/undefined) (Das, le, and Dai 2020).

In summary, different State department of transportation (DOT's) assign the fault party within the crash report at the time of the crash, based on the issued citation in case the citation is relevant to the crash occurrence and any action that is associated with the crash.

Police narrative included in the crash form, together with the type of citation to decide which party is primarily responsible for the crash, the pedestrian and bicycle crash analysis (PBCAT) tool, pictures and video recordings captured by closed-circuit television (CCTV) cameras have been used by researchers to enhance the process of determining the fault party in a crash. **Table 5-1** summarizes the existing methods for assigning an at-fault party in a crash.

Table 5-1: A summary of the existing methods used for assigning the responsible road user in a crash by previous empirical studies.

Study Author and (Year)	Data Source	Summary of Methods Used in Determining Fault
Kim, Brunner, and Yamashita (2008)	Police-reported crash data.	Based on law enforcement (i.e., pedestrian jaywalking laws), the fault is determined by deciding on the party that received a ticket/citation.
Kim, Brunner, and Yamashita (2008)		The fault is determined by specially trained investigators who are dispatched to the crash scene, and they typically search for the action or behavior of a certain party that caused the accident to occur.
Ulfarsson, Kim, and Booth (2010)		The police officer assigns the fault to the party based on who acted negligently or is in other ways found to have caused the crash.
(Zhao, Wang, and Jackson 2019)		Drivers were considered ‘‘at-fault’’ if they were given any verbal/written warning, infraction, or arrest/summons. If no action was taken by the officer, the driver was considered not-at-fault.
Walter and Studdert (2015)		The fault is determined based on citations issued for crash responsible drivers and recorded in the police report.

Kim et al. (1998)		Police officers assign the fault party and record the information in the crash report based on the officer’s narrative summary of the event.
Kim and Li (1996)		The fault is determined by investigating officers and then reported on crash report forms. Based on a logistic model that explains fault among motorists as a function of various variables. The crash report form is to ascertain who is at fault and that although individual fields such as driver’s license number, birth date, vehicle identification number, and so forth may be sources of error.
Adanu et al. (2017)		The fault is determined based on the investigation done by the officer who completed the crash report.
Spainhour and Wootton (2007)	Police-reported crash data and case review data stemmed from manual case reviews of multiple crash data sources.	The fault is determined by the investigating officers which are then reported on crash report forms. Florida department of transportation (DOT) currently uses an algorithm to assign fault. FDOT thereby presumes that the individual in the first section is at fault unless a citation was given to drivers or pedestrians in subsequent sections of the crash report, in which case fault is reassigned to the person receiving the citation.
Lee, Abdel-Aty, and Choi (2014)	ZIP code information of road users involved in traffic crashes and police recorded data.	The fault is determined based on the crash investigation done by police officers, who issue citations according to the

		investigation. Citations are then recorded in the police report which confirms the fault party.
Goh et al. (2014)	Pictures and video recordings captured from CCTV cameras.	The fault is typically made with the aid of pictures and video recordings captured from CCTV installed and is done by police officers and adjusters from the insurance company before an at-fault assignment is made for insurance claims.
Russo et al. (2014)	Michigan traffic crash facts (MTCF) data query tool.	The fault is assigned to one driver based on the judgment of the investigating officer, who decides that the at-fault driver must have performed one or more hazardous actions (e.g. speeding, failing to yield, disregarding traffic control devices) that contributed to the crash.
Zhang, Yau, and Zhang (2014)	Traffic accident data.	The fault is determined by police officers, under the following two circumstances: (1) determined by police that he/she should bear the whole responsibility of the accident; (2) determined by police that he/she should bear the main responsibility of the accident. For instance, when crashes occur on roadways without pedestrian facilities, motor vehicles are typically held responsible because pedestrians are commonly considered as a vulnerable group. Conversely, pedestrians are mostly determined to be liable.

Islam and Jones (2014)	Police-reported crash data filtered using the critical analysis reporting environment (CARE) software system.	Dataset already has the at-fault party assigned in the original police-reported crash database.
Schneider, Stefanich, and Corsi (2015)	MV4000 Crash report forms.	Since the police do not assign "fault" for crashes in Wisconsin, the study used the detailed narrative on the MV4000 forms, and the type of citation helped as well to interpret which party or parties the police officer viewed as being primarily responsible for the crash.
Ichikawa, Nakahara, and Taniguchi (2015)	Police-reported crash data and driving exposure data.	The fault is determined by police investigators at the scene of the motor vehicle crash.
Ratrout et al. (2017)	Police-reported crash data, as well as the crash, driver, and vehicle-related data collected from police stations.	Drivers at fault and not at fault were separated and investigated through factor and principal component analysis (PCA) for 19 parameters related to their background and knowledge of traffic signs.
Penmetsa, Pulugurtha, and Duddu (2017)	Crash data obtained from the Highway Safety Information System (HSIS) which included information related to accident, roadway, vehicle, and occupant.	For each vehicle involved in a crash, the crash reports provide three contributing factors (which indirectly explain the traffic rule the driver violated) that led to the crash. If the driver has not committed any traffic violation, a value of zero is provided under the contributing factor variable, implying that the driver is not at fault in the crash.

Das et al. (2018)	Police-reported crash data which contains crash, roadway geometry, and vehicle-related data.	Based on the crash event investigated by the police officer who records the fault party in the report.
Yu et al. (2019)	Police-reported crash data obtained from the FDOT Crash Analysis Reporting (CAR) system, driver information data extracted from the CAR system, and traffic analysis zone (TAZ) shapefile obtained from the US Census Bureau.	The fault is determined through using the license address state variable "ADRSTATE" for at-fault drivers, which is coded in the crash data.
Islam and Hossain (2019)	Police reported crash database filtered using the (CARE) software system.	Dataset already has the at-fault party assigned in the original police-reported crash database by the police officer who completed the report at the scene.
Das, le, and Dai (2020)	Pedestrian and bicycle crash analysis tool(PBCAT).	PBCAT tool was associated with a framework developed for applying ML models to classify pedestrian crash types and assign the fault to different parties (i.e., drivers, pedestrians, both, none/undefined).

Note: Studies are ordered by the common data source.

5.2.2 Methods for Classification

Diverse methods were applied for predicting the probabilities of being one of the multiple categories of fault parties. Zhang and colleagues (2014) analyzed pedestrian-vehicle crash reports to identify significant risk factors connected with pedestrian and driver liability/fault in a crash using logistic regression models. The authors identified risk factors that impact pedestrian and driver fault status in pedestrian-motor vehicle crashes. The results showed that traffic crashes in which pedestrians were deemed at fault, resulted in serious and fatal injuries. Regarding pedestrians' liability, individual characteristics such as being manufacturing workers and farmers, vehicle features such as being in a crash that involved general use vehicles, road conditions such as pedestrian separation structures, and roadways with traffic lights or signs, and during rush hour, are all risk factor associated with the pedestrian being more likely to be at-fault in pedestrian-vehicle crashes (G. Zhang, Yau, and Zhang 2014).

Penmetsa, Pulugurtha, and Duddu (2017) utilized the partial proportional odds (PPO) model for the analysis of the severity outcome of not-at-fault drivers in two-vehicle crashes. Exceeding the speed limit, reckless driving, and going the wrong way are the three traffic rule violations of at-fault drives that are more likely to result in severe injuries to not-at-fault drivers compared to disregarding traffic signals/signs/markings. Additionally, crashes involving two and three traffic violations are 68% and 186% more likely to result in severe injuries compared to crashes involving at-fault drivers with only one traffic rule violation (Penmetsa, Pulugurtha, and Duddu 2017).

Few studies have adopted artificial intelligence (AI) approaches. For an instant, Das et. al. evaluated three machine learning algorithms, including support vector machine (SVM), random forest (RF), and the extreme gradient boosting (XGBoost) algorithm which is an implementation of the gradient boosted decision tree (GBDT) technique.

The authors aimed at investigating at-fault motorcycle riders, to classify crash types from pedestrian crash typing data and identify factors associated with the driver, pedestrian, both, or no-fault in a crash (Das, le, and Dai 2020). The analysis results showed that 45% and 42% of the crashes involved drivers and pedestrians at-fault, respectively. Furthermore, the results highlighted that if a pedestrian was crossing the roadway and was struck by a turning vehicle, the driver was most likely (79%) to be at-fault; if the vehicle was not turning, the pedestrian was at-fault in (81%)

of the crashes. Pedestrians darting out into the roadway were deemed at fault in 100% of the crashes.

Classification and regression trees (CART) have also been used in the research investigating the categorization of the at-fault party. Mohaymany et. al. (2010) adopted the CART method combined with the quasi-induced exposure concept, to study the target variable; driver status. Driver status involved two classes; at-fault drivers, and not-at-fault drivers. The results showed that drivers who are younger than 28 years old, whose driving license is type 2 -a driving license that is for driving with passenger car and light vehicles -and whose driving experience is less than two years are most probably responsible for overtaking crashes (Mohaymany, Kashani, and Ranjbari 2010). Yan and Radwan (2006), also adopted the same methodology. The analysis results showed that the youngest driver group (< 21 years) shows the largest crash propensity, and the 32–75 years drivers driving large size vehicles have a larger crash propensity compared to those driving passenger vehicles (Yan and Radwan 2006).

Other researchers implemented Factor analysis (FA) and principal component analysis (PCA) to analyze crash variables linked with at-fault and not-at-fault drivers. The authors implemented logit models to quantify the effects of the influencing variables.

The results revealed that driver's experience and knowledge of traffic signs for chauffeurs have a positive impact on reducing fault behavior of drivers (Ratrout et al. 2017).

Multiple correspondence analysis (MCA) technique, which is an extension of the correspondence analysis (CA) was adopted by Jalayer et. al. (2016) to identify roadway/environmental, motorcycle and motorcycle related variables influencing motorcycle-involved crashes with motorcyclists being at fault. According to the obtained results, the main contributing factors to at-fault motorcycle-involved crashes are light conditions, time of day, driver condition, roadway curvature and grade, and weather conditions (Jalayer and Zhou 2016).

5.2.3 Findings from Previous Studies for the Party At-Fault

Researchers used crash reports that assigned the fault party, to conduct research concerning factors associated with the frequency and severity of crashes involving at-fault and not-at-fault parties (i.e., (Kim et al. 1998; Spainhour and Wootton 2007; Kim, Brunner, and Yamashita 2008; Ulfarsson, Kim, and Booth 2010; G. Zhang, Yau, and Zhang 2014; Islam, Jones, and Dye 2014; Ichikawa, Nakahara, and Taniguchi 2015; Das et al. 2018; Yu et al. 2019; Islam and Hossain 2019). In general, human factors are considered to be the most prevalent factors contributing to crashes, followed by roadway environment and vehicle factors (H. Zhang 2010). Many factors have been of interest since the analysis results enhanced the driver-focused educational programs (Abdel-Aty and Abdelwahab 2000).

In addition to the commonly used personnel features (e.g., gender, age, driving experience, and blood alcohol concentration (BAC)), crashes were also identified to be related to the driver faults (Kim and Li 1996; Ichikawa, Nakahara, and Taniguchi 2015; Walter and Studdert 2015; Islam, Jones, and Dye 2014; Zhao, Wang, and Jackson 2019; Penmetsa, Pulugurtha, and Duddu 2017; Duddu, Penmetsa, and Pulugurtha 2018; R. Schneider, Stefanich, and Corsi 2015).

Limited research considered studying the effect of fault status on the injury severity level, and the factors affecting the fault of a road user. (Yu et al. 2019) investigated a variety of crash influencing factors of at-fault out-of-state drivers. It was identified that the influencing variables for crashes caused by out-of-state drivers are car ownership, speeding, and driving under the influence (DUI) for the intersection crashes and being an old driver, speeding, and dark roadway environment for roadway segment crashes.

(J. Lee, Abdel-Aty, and Choi 2014) analyzed the relationship between the number of at-fault drivers and their residence zonal characteristics. It was concluded that not only roadway/traffic factors affect the crash occurrence, but also several demographic and socioeconomic characteristics of residence zones where the at-fault drivers live. Yet, there was no detailed exploration carried out for the specific influencing factors affecting a specific road user to be responsible for a crash, and no investigation directed to study if the crash injury severity is affected by a specific road user being at-fault. Hence, this study aims at bridging the gap by analyzing the crash injury severity caused by drivers and VRUs separately. The special contributing factors leading a specific road user to be at-fault of the crash would be identified while the importance of each contributing factor on both, the pedestrian and the driver's fault would be examined and discussed as well.

The study of Lee and colleagues (2005) analyzed vehicle-pedestrian crashes at intersections in Florida over 4 years, 1999–2002, and considered using an ordered probit model to investigate the likelihood of pedestrian injury severity. The results highlighted findings categorized regarding the fault party.

For instance, the authors concluded that middle-age (25–64) and male drivers are more involved in crashes as causers than other driver groups and that crashes occur more frequently at the intersections with other traffic control (e.g. stop signs, yield signs, etc.) in urban areas when non-intoxicated drivers are driving passenger cars at night. Additionally, the proportion of fatal vehicle-pedestrian crashes was higher when the driver was normal in condition.

Whereas, In the case of crashes at pedestrian’s fault, the crashes occur more frequently under similar conditions to crashes at driver’s fault, but also the undivided and wide (i.e. a greater number of lanes) intersections (C. Lee and Abdel-Aty 2005). Also, an interesting finding showed that the condition of the pedestrian involved in fatal crashes was mostly unknown which is mainly due to the lack of input on that category in the police report.

5.3 Methodologies

In this study, Z-test, XGboost, and multinomial logistic regression are used to serve the primary goal which is to understand if the crash severity is affected by the fault status, as well as investigating the risk factors affecting the crash severity outcome for at-fault and no-fault parties. Selecting the suitable method relies on the underlying objectives of the study, which as far as the study goals were descriptive and predictive mining to distinguish the severe crashes-related factors and obtain the highest potential accuracy. Below is a summary of the methodologies.

5.3.1 Z-Test Concerning Injury Severity Proportion

In this analysis, the Z-test for proportions was selected as the statistical test to indicate if a particular variable of the newly created roadway, driver, pedestrian, and bicyclist-related pedestrian and bicyclist-related variables has higher (fatal (K) and severe (A) injury) proportion is significantly different than the proportion of (fatal (K) and severe (A) injury) injuries for the population. The test was conducted in RStudio using the “prop. test” function at a 95% confidence level. Note that the formula of the Z-test statistics **Z** = $\hat{p} - p$ pq/n

Eq. 5-1 is valid when sample size (n) is large enough; np, nq should be ≥ 5. In case of a small sample size (such in “NMTSFQ.REFL.LTNG.HLMT” variable in Error! Reference ource not found. to Error! Reference source not found., the Fisher Exact probability test is used for comparing the two proportions.

$$Z = \frac{\hat{p}-p}{\sqrt{pq/n}} \qquad \textbf{Eq. 5-1}$$

Where:

- $\hat{p}$: sample proportion;
- p: population proportion;
- q: 1-p;
- n: sample size.

5.3.2 Gradient Boosting Decision Tree (GBDT)

GBDT is an iterative decision tree algorithm, proposed by Friedman (Breiman 1984) at Stanford University. It includes multiple additive trees (MATs), where each tree is considered a weak learner, and the following tree is trained based on the error of the former tree (Friedman 2001). The algorithm can conquer the over-fitting problem of an individual tree by merging hundreds of weak decision trees. In the GBT model, a basis function $f(x)$ describes a response variable *y* in a function of summation of weighted basis functions for individual trees as follows:

$$f(x) = \sum_{n=1}^{m} \beta_n b(x; \gamma_n) \qquad \textbf{Eq. 5-2}$$

Where:

- $b(x; \gamma_n)$ is the basis function for individual tree n; m is the total number of trees;
- γ_n is the split variable; and β_n is the estimated parameter that minimizes the loss function, *L(y, f(x))*.

Hence, the output is constructed by cumulating the conclusions from all trees from this algorithm. GBDT owns a few advantages, including the ability to find non-linear transformations, the ability to handle skewed variables without requiring transformations, computational robustness, and high scalability (Industrial Networks and Intelligent Systems 2018). Also, the strength of the GBDT algorithm arises from its ability to recognize feature combinations. It is one of the most powerful techniques for classification and producing prediction models -generalized by allowing an arbitrary distinguishable loss function- a form of an ensemble of weak prediction models (Haroon 2017). For further details on the thorough description of gradient boosting trees from a statistical learning point of view, refer to the paper by Guelman (Guelman 2012).

5.3.3 Extreme Gradient Boosting Decision Tree (XGboost)

The analysis in this study attempted the extreme gradient boosting decision tree technique (XGboost). XGboost is one of the four common tree-based ensemble learning techniques; random forest (RF), GBDT, and Adaboost with a decision tree (C. Wang et al. 2019). XGboost is an extreme Gradient Boosting technique in applied machine learning which encapsulates the high performance and better speed over the other classification and regression models (Kandan et al. 2019).

Gradient boosting methods are found to help in classification and regression models, via predicting errors of previous trees using ensemble learning to boost both, model performance and accuracy. The data was split into training and test datasets, where our model was trained in the assigned trained dataset and then tested in the test dataset to obtain the prediction accuracy. R software was used to build the model in the "xgboost" package. Regarding accuracy validation, the k-fold cross-validation technique was used. The final model had (1000) trees.

5.3.4 Multinomial Logit (MNL) Model

Even though each model used in the crash severity-related literature has its advantages, it is noted that MNL models are the most widely used approach used to identify the relationship between the dependent and independent variables (Chen and Fan 2018). Shankar and Mannering (1996) provided a detailed discussion of MNL models. The MNL model is a discrete choice model which deals with three or more levels of the response variable, without taking into consideration the order of the levels. It is common to categorize crash severity level into five discrete categories: a) property damage only, b) possible injury, c) evident injury, d) disabling injury, and e) fatality.

Using these five discrete crash severity categories, one can develop a statistical model that may be used to predict the crash likelihood of a specific severity level. **Eq. 5-3** displays the probability of a crash n with an i severity level:

$$P_n\,(\mathrm{i}) = \mathrm{P}\,(U_{in} \geq U_{jn})\ \forall\ \mathrm{j} \neq \mathrm{i} \qquad \textbf{Eq. 5-3}$$

Where:

- P_n (i): the probability of crash n to occur with a severity level of i;
- P: probability;
- U_{in}: function to determine the utility of a crash n to occur resulting severity level of i.

The linear function of U_{in} maybe demonstrated as in **Eq. 5-4**.

$$U_{in} = \beta_i X_n + \varepsilon_{in} \qquad \textbf{Eq. 5-4}$$

Where:

- X_n: explanatory variable's vector which determines the crash severity;
- β_i: estimable coefficient vector for the injury outcome i, using the standard maximum likelihood methods (Shankar and Mannering 1996; Tay et al. 2011);
- $\beta_i X_n$: is an observable component;
- ε_{in}: error term, unobserved component, and is assumed to be independently distributed -accounts for the unobserved factors affecting crash severity-.

Combining **Eq. 5-3** and **Eq. 5-4, Eq. 5-5** may be formed as:

$$P_n\,(i) = P\big(\beta_i X_n - \beta_j X_n \geq \varepsilon_{jn} - \varepsilon_{in}\big)\ \forall\ \mathrm{j} \neq \mathrm{i} \qquad \textbf{Eq. 5-5}$$

Replacing the error term by a generalized extreme value (GEV) form. Assuming the GEV, the MNL severity model can be obtained as in **Eq. 5-6.**

$$P_n\,(i) = exp(\beta_i X_n) / \textstyle\sum_j^0 exp(\beta_j X_n) \qquad \textbf{Eq. 5-6}$$

5.4 VRU Crash Data Collection and Processing

The analysis involves crash reports in the DT4000 form, generated in the State of Wisconsin for the years 2017 and 2018. Furthermore, especially regarding the crash risk, the effects of the party at fault on the crash severity outcome were investigated.

5.4.1 Data Collection

Data for this study were provided by the Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin–Madison through the WisTransportal system ("The WisTransPortal System" n.d.). The data was generated form crash reports in the DT4000 form and represents pedestrian and bicyclist-vehicle crashes which occurred in the State of Wisconsin between 2017 and 2018.

Pedestrian/bicyclist-vehicle traffic crashes were identified via a pedestrian flag and a bicycle flag assigned by the Wisconsin Department of Motor vehicle (DMV). Overall, the dataset contains (4,025) crashes, of these, (60) crashes did not include any VRU (pedestrians and bicyclists only), and (324) crashes recorded with the motorist sustaining more severe injury than the non-motorist. The remaining sample size is (N=3,641) crashes involving at least one VRU. In this study, the analysis considers pedestrian and bicyclist's Injury severity. Among the total number of (3,641) crashes, 948 (26.03%) were injury type C; 1,992 (54.71%) were injury Type B; 581 (15.95%) were injury Type A, and 120 (3.29%) were fatal injury crashes. To collect sufficient observations for all levels of crash severity, crash severity levels were recategorized into three classes as follows: fatal injury which is type (K), type (A); severe injury, and type (B) and type (C) are combined in a new category; evident and possible injury (B+C).

In a crash involving non-motorists, the non-motorists can be either unit 1 or 2. According to the crash data user guide, "[1,2] Denotes unit-level information, where a unit is any vehicle, bicycle, pedestrian, or equipment involved in a crash. Unit level element names in the data file are appended with “1” or “2”, representing the first or second unit involved in the crash”.

When more than two units are involved including non-motorists, a non-motorist may be coded as neither unit 1 or 2. However, the study is limited to analyzing the actions and behavior of the first two units. There are often more than one crash contributing factor in data fields such as circumstances, driver actions, behavior, so-called multi-valued elements, and denoted as [A, B, C]. [1,2][A,B] denotes combined unit level and multi-valued elements. For example, DRVRPC1A and DRVRPC1B describe the first two contributing factors listed for the driver of the first unit on the DT4000 crash report.

Therefore, when necessary, the CONCATENATE function is used in the analysis to join data from Unit 1 and 2 -a unit is a driver, a pedestrian, or a bicyclist-, as well as from A, B, C, etc. For the same data types (i.e., DRVRDOIN 1 and DRVRDOIN 2), concatenation is done to join the two text strings into one text string (DRVRDOIN 1,2) and to reduce the resulted values of a specific data field. A filter may be used to filter values of the DRVRDOIN 1 data field when Role 1 is a driver, and the same way when Role 2 is a driver, values of DRVRDOIN 2 data field are filtered, creating (DRVRDOIN 1,2) data field. A string can be a text, number, or a Boolean value.

Whereas, for data fields that take multiple values (i.e., ROADCOND A, ROADCOND B, ROADCOND C), direct concatenation separated by a comma is approached to create (ROADCOND A, B, C) data field. Note that the number of attributes provided in each element varies and is based on the minimum set of data elements recommended by the Model Minimum Uniform Crash Criteria (MMUCC) standard (National Highway Traffic Safety Administration 2017). After concatenation, attributes with small percentage values (e.g., <1%) are not analyzed or presented and displayed separately as one category through the total percentage value (i.e., Total including other combinations). Note that multi-value attributes (i.e., BIKE, UT TRK) are a result of applying the concatenation function to the data field attributes.

A few of the most important data sources in crash reports are the citation field (STATNM [1, 2] [A, B, C, D], action and circumstances-related fields (i.e., NMTACT, DRVRPC, DNMFTR, DRVRDOIN), and the crash description box where the police officer provide a narrative that indicates the units and the circumstances of the crash, with the eyewitness testimony if available continued to be used in the DT4000 form. The focus of this chapter is directed towards the fault party assignment; whereas, data accuracy is handled later on in **Chapter 6**.

5.4.2 Data Processing

Based on the empirical results of the existing literature the variables are chosen, while any record with missing information is dropped. Data processing, cleaning, and missing values detection were completed by using the "tidyr" and "dplyr" packages from the "Tidyverse" collection of R packages designed for data science, after assembling the data. Components of the crash dataset were categorized under four factors that were later used in the analysis: a) Driver-related factors, such as age, gender, safety equipment used, impairment; b) Pedestrian and bicyclist-related factors, such as age, gender, safety equipment used, impairment; c) Roadway-related factors, such as intersection type, available traffic control device (TCD), horizontal and vertical road terrain, roadway surface type and condition, trafficway division, light condition; d) Crash/vehicle-related factors, such as posted speed limit, vehicle type; and e) environmental-related factors, such as weather, light condition, etc. **Table 5-2** to **Table 5-6** provide summary statistics of the possible contributing factors based on 3,614 VRU-involved motor vehicle crashes.

Table 5-2: Summary statistics of possible driver-related factors influencing the injury severity

Variable	Indication	Injury severity									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
DNMFTR [1, 2] [A, B]	Any relevant condition of the individual (motorist or non-motorists) that is directly related to the crash.										
	NORM	9	0.25%	249	6.84%	1269	34.85%	544	14.94%	2071	56.88%
	NO OBS	42	1.15%	128	3.52%	428	11.76%	237	6.51%	835	22.93%
	UI MDA	7	0.19%	33	0.91%	41	1.13%	25	0.69%	106	2.91%
	Other values	62	1.70%	171	4.70%	254	6.98%	142	3.90%	629	17.28%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
DRVRPC [1, 2] [A, B]	The actions by the driver that may have contributed to the crash, based on the judgment of the law enforcement officer investigating the crash.										
	FTY	8	0.22%	109	2.99%	562	15.44%	243	6.67%	922	25.32%
	NO	61	1.68%	249	6.84%	790	21.70%	362	9.94%	1462	40.15%
	ID	3	0.08%	16	0.44%	55	1.51%	33	0.91%	107	2.94%
	Other values	48	1.32%	207	5.69%	585	16.06%	310	8.52%	1150	31.59%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
DRVRDOIN [1, 2]	The controlled maneuver for this motor vehicle prior to the beginning of the sequence of events.										
	GO STR	70	1.92%	302	8.29%	663	18.21%	249	6.84%	1284	35.27%
	LT TRN	5	0.14%	65	1.79%	339	9.31%	171	4.70%	580	15.93%

	RT TRN	3	0.08%	25	0.69%	243	6.67%	151	4.15%	422	11.59%
	BACKING	1	0.03%	9	0.25%	44	1.21%	21	0.58%	75	2.06%
	Other values	41	1.13%	180	4.94%	703	19.31%	356	9.77%	1280	35.15%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SFTYEQP[1,2]	The restraint equipment in use at the time of the crash.										
	NONE	6	0.16%	29	0.80%	137	3.76%	53	1.46%	225	6.18%
	SH/LP	81	2.22%	372	10.22%	1294	35.54%	622	17.08%	2369	65.06%
	UNKN	16	0.44%	102	2.80%	352	9.67%	153	4%	623	17.11%
	UNTYPE	0	0.00%	4	0.11%	22	0.60%	9	5%	35	5.71%
	Other values	1	0.03%	7	0.19%	20	0.55%	13	0.36%	41	1.13%
	348 (9.56%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ALCFLAG	Indicates whether law enforcement suspected that at least one driver or non-motorists involved in the crash had used alcohol. This includes both alcohol use under the legal limit and at or over the legal limit.										
	UNKN	13	0.36%	81	2.22%	309	8.49%	180	4.94%	583	16.01%
	Y	35	0.96%	106	2.91%	113	3.10%	45	1.24%	299	8.21%
	N	72	1.98%	394	10.82%	1570	43.12%	723	19.86%	2759	75.78%
	2759 (75.78%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
DRUGLFAG*	Indicates whether law enforcement suspected that at least one driver or non-motorists involved in the crash had used drugs (Y/N/UNKN).										
	UNKN	19	0.52%	95	2.61%	336	9.23%	194	5.33%	644	17.69%

	Y	11	0.30%	14	0.38%	11	0.30%	8	0.22%	44	1.21%
	2953 (81.10%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
Sex [1, 2]	The sex of the non-motorists involved in a crash.										
	F	25	0.69%	190	5.22%	749	20.57%	359	9.86%	1323	36.34%
	M	80	2.20%	307	8.43%	947	26.01%	409	11.23%	1743	47.87%
	UNKN	15	0.41%	84	2.31%	296	8.13%	180	4.94%	575	15.79%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
Age [1, 2]	The age of the non-motorists involved in a crash in years.										
	<30	48	1.32%	254	6.98%	807	22.16%	424	11.65%	1533	42.10%
	30-64	57	1.57%	253	6.95%	913	25.08%	412	11.32%	1635	44.91%
	≥65	15	0.41%	74	2.03%	272	7.47%	112	3.08%	473	12.99%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Table 5-3: Summary statistics of possible pedestrian and bicyclist/VRU-related factors influencing the injury severity

Variable	**Indication**	**Injury severity**									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
DNMFTR [1, 2] [A, B]	Any relevant condition of the individual (motorist or non-motorists) that is directly related to the crash.										
	PHY IMP	10	0.27%	21	0.56%	38	1.04%	24	0.65%	92	2.53%

	NORM	5	0.12%	125	3.42%	636	17.45%	272	7.47%	1037	28.47%
	NO OBS	20	0.54%	41	1.13%	128	3.52%	65	1.79%	254	6.96%
	OTHR	86	2.36%	395	10.85%	1191	32.70%	588	16.14%	2259	62.04%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTLOC [1, 2]	The location of the non-motorists with respect to the roadway at the time of the crash.										
	ATI MX	15	0.41%	170	4.67%	779	21.40%	456	12.52%	1420	39.00%
	ATI NX	10	0.27%	78	2.14%	244	6.70%	103	2.83%	435	11.95%
	ATI UM	4	0.11%	45	1.24%	171	4.70%	86	2.36%	306	8.40%
	BIKE LN	0	0.00%	4	0.11%	44	1.21%	14	0.38%	62	1.70%
	NAI NX	65	1.79%	202	5.55%	454	12.47%	162	4.45%	883	24.25%
	SHLDR	10	0.27%	21	0.58%	67	1.84%	31	0.85%	129	3.54%
	Other values	16	0.44%	61	1.68%	233	6.40%	96	2.64%	406	11.15%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTACT [1, 2]	The actions/circumstances of the non-motorists that may have contributed to the crash, based on the judgment of the law enforcement officer investigating the crash.										
	NF TRFC, DK CLTH, DISREG, FC TRFC, IM XING, SUDDEN	31	0.85%	163	4.48%	487	13.38%	265	7.27%	946	25.98%
	NO IMPR	21	0.58%	155	4.26%	751	20.63%	348	9.56%	1275	35.02%
	Other values	68	1.87%	263	7.22%	754	20.71%	335	9.20%	1420	39.00%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
	The action of a non-motorist immediately prior to a crash.										

NMTPRIOR [1, 2]	A TRFC	0	0.00%	8	0.22%	46	1.26%	16	0.44%	70	1.92%
	RDWY OT	17	0.47%	5	0.14%	23	0.63%	8	0.22%	37	1.46%
	SIDE WK	0	0.00%	0	0.00%	11	0.30%	1	0.03%	12	0.33%
	W TRFC	1	0.03%	4	0.11%	19	0.52%	7	0.19%	30	0.85%
	XING	6	0.16%	62	1.70%	180	4.94%	72	1.98%	331	8.79%
	WAITING	3	0.08%	37	1.02%	70	1.92%	35	0.96%	147	3.98%
	Other values	93	2.55%	465	12.77%	1643	45.12%	809	22.22%	3014	82.67%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTSFQ [1, 2] [A, B]*	The safety equipment in use by the operator non-motorist at the time of the crash.										
	NONE	95	2.61%	453	12.44%	1483	40.73%	714	19.61%	2745	75.39%
	LTNG/REFL, HLMT	3	0.08%	52	1.43%	244	6.70%	75	2.06%	374	10.27%
	Other values	22	0.60%	76	2.09%	265	7.28%	159	4.37%	522	14.34%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
Sex [1, 2]	The sex of the non-motorists involved in a crash.										
	F	33	0.91%	217	5.96%	681	18.70%	343	9.42%	1274	34.99%
	M	87	2.39%	360	9.89%	1305	35.84%	595	16.34%	2347	64.46%
	UNKN	0	0.00%	4	0.11%	6	0.16%	10	0.27%	20	0.55%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
Age [1, 2]	The age of the non-motorists involved in a crash in years.										
	<30	24	0.66%	240	6.59%	1095	30.07%	500	13.73%	1859	51.06%
	30-64	61	1.68%	266	7.31%	735	20.19%	364	10.00%	1426	39.17%
	≥65	35	0.96%	75	2.06%	162	4.45%	84	2.31%	356	9.78%

	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Table 5-4: Summary statistics of possible roadway-related factors influencing the injury severity

Variable	**Indication**	**Injury severity**									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
TRFCWAY [1, 2]	Indication of whether or not the trafficway for this vehicle is divided and whether it serves one-way or two-way traffic.										
	UNDIV	65	1.79%	402	11.04%	1393	38.26%	621	17.06%	2481	68.14%
	DIV NO	31	0.85%	55	1.51%	184	5.05%	70	1.92%	340	9.34%
	OW	1	0.03%	13	0.36%	50	1.37%	22	0.60%	86	2.36%
	Other values	23	0.63%	111	3.05%	365	10.02%	235	6.45%	734	20.16%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
TOTLANES [1, 2]	1 Lane	1	0.03%	17	0.47%	139	3.82%	63	1.73%	220	6.04%
	2 Lanes	69	1.90%	394	10.82%	1311	36.01%	634	17.41%	2408	66.14%
	3 Lanes	8	0.22%	27	0.74%	101	2.77%	43	1.18%	179	4.92%
	>3 Lanes	42	1.15%	143	3.93%	441	12.11%	208	5.71%	834	22.91%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RDWYPC [A, B, C]	Factors of the road which may have contributed to the crash.										
	NONE	114	3.13%	549	15.08%	1884	51.74%	904	24.83%	3451	94.78%

	RSC	3	0.08%	6	0.16%	29	0.80%	17	0.47%	55	1.51%
	OTHR	3	0.08%	26	0.71%	79	2.17%	27	0.74%	135	3.71%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RDCOND [A, B, C]	The roadway surface condition at the time and place of a crash.										
	DRY	97	2.66%	476	13.07%	1699	46.66%	784	21.53%	3056	83.93%
	WET	16	0.44%	81	2.22%	252	6.92%	133	3.65%	482	13.24%
	OTHR	7	0.19%	24	0.66%	41	1.13%	31	0.85%	103	2.83%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ROADHOR [1, 2]	The horizontal road terrain at the point of impact.										
	LT	2	0.05%	9	0.25%	29	0.80%	8	0.22%	48	1.32%
	RT	6	0.16%	6	0.16%	22	0.60%	11	0.30%	45	1.24%
	ST	109	2.99%	544	14.94%	1830	50.26%	864	23.73%	3347	91.93%
	Other values	3	0.08%	22	0.60%	111	3.05%	65	1.79%	201	5.52%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ROADVERT [1, 2]	The vertical road terrain at the point of impact.										
	DN	14	0.38%	38	1.04%	43	1.18%	22	0.60%	117	3.21%
	CST	0	0.00%	4	0.11%	21	0.58%	3	0.08%	28	0.77%
	LVL	14	0.38%	42	1.15%	64	1.76%	25	0.69%	145	3.98%
	SAG	28	0.77%	84	2.31%	128	3.52%	50	1.37%	290	7.96%
	UP	42	1.15%	126	3.46%	192	5.27%	75	2.06%	435	11.95%
	Other values	22	0.62%	287	7.89%	1544	42.40%	773	21.24%	2626	72.13%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

INTTYPE	The type of intersection in which a crash occurred. An intersection consists of two or more roadways that intersect at the same level.										
	5	1	0.03%	2	0.05%	2	0.30%	13	0.36%	27	0.74%
	4 WAY	25	0.69%	222	6.10%	222	23.29%	426	11.70%	1521	41.77%
	L	0	0.00%	1	0.03%	1	0.11%	2	0.05%	7	0.19%
	RAB	0	0.00%	1	0.03%	1	0.36%	5	0.14%	19	0.52%
	T	7	0.19%	48	1.32%	48	6.59%	132	3.63%	427	11.73%
	Y	0	0.00%	4	0.11%	4	0.14%	3	0.08%	12	0.33%
	Other values	87	2.39%	303	8.32%	1714	23.92%	367	10.08%	1628	44.72%
	NA	1590 (43.67%)									
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
TRFCCNTL [1, 2]	The type of TCD applicable to this motor vehicle at the crash location.										
	NONE	89	2.44%	335	9.20%	878	24.11%	327	8.98%	1629	44.74%
	STOP	7	0.19%	65	1.79%	387	10.63%	210	5.77%	669	18.37%
	TS OP	3	0.08%	39	1.07%	265	7.28%	127	3.49%	434	11.92%
	Other values	21	0.59%	142	3.90%	462	12.69%	284	7.80%	909	24.97%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
LOCTYPE	The type of location at which a crash occurred.										
	I	33	0.91%	286	7.85%	1143	31.39%	589	16.18%	2051	56.33%
	N	87	2.39%	295	8.10%	849	23.32%	359	9.86%	1590	43.67%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Table 5-5: Summary statistics of possible significant crash/vehicle-related factors influencing the injury severity

Variable	Indication	Injury severity									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
VEHTYPE [1, 2]	Specific category for the type of vehicle which was involved in a crash.										
	CAR	58	1.59%	291	7.99%	746	20.49%	396	10.88%	1491	40.95%
	SUV	20	0.55%	71	1.95%	138	3.79%	87	2.39%	316	8.68%
	UT TRK	17	0.47%	48	1.32%	75	2.06%	35	0.96%	175	4.81%
	P VAN	4	0.11%	13	0.36%	42	1.15%	21	0.58%	80	2.20%
	Other values	21	0.58%	158	4.34%	991	27.22%	409	11.23%	1579	43.36%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SPEEDFLAG	Flag indicating whether speed was a factor in a crash.										
	Y	8	0.22%	36	0.99%	29	0.80%	11	0.30%	84	2.31%
	3557 (97.69%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
POSTSPD[1,2]	The posted/statutory speed limit for a motor vehicle at the time of the crash. A value of 77 indicates not Applicable.										
	<35 mph	51	1.40%	422	11.59%	1541	42.32%	769	21.12%	2783	76.44%
	35 to 50 mph	32	0.88%	107	2.94%	319	8.76%	115	3.16%	573	15.74%
	>50 mph	37	1.02%	52	1.43%	132	3.63%	64	1.76%	285	7.83%

	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SCHZONE	Flag indicating whether a crash occurred in an active school zone.										
	Y	2	0.05%	10	0.27%	46	1.26%	30	0.82%	88	2.42%
	3553 (97.58%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
CONSZONE	Y	3	0.08%	6	0.16%	17	0.47%	6	0.16%	32	0.88%
	3553 (97.58%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Table 5-6: Summary Statistics of Possible Significant environmental-related Factors Influencing the injury severity

Variable	**Indication**	**Injury severity**									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
Environmental-Related Factors											
WTCOND	The prevailing atmospheric conditions that existed at the time of the crash.										
	CLDY	36	0.99%	139	3.82%	502	13.79%	243	6.67%	920	25.27%
	CLEAR	68	1.87%	367	10.08%	1272	34.94%	590	16.20%	2297	63.09%
	RAIN	9	0.25%	42	1.15%	141	3.87%	72	1.98%	264	7.25%
	Other values	7	0.19%	33	0.91%	77	2.11%	43	1.18%	160	4.39%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
LGTCOND	The type/level of light that existed at the time of the motor vehicle crash.										

	DAWN	37	1.02%	55	1.51%	84	2.31%	20	0.55%	196	5.38%
	DUSK	6	0.16%	12	0.33%	55	1.51%	19	0.52%	92	2.53%
	DAY	35	0.96%	298	8.18%	1381	37.93%	666	18.29%	2380	65.37%
	DARK	7	0.19%	18	0.49%	60	1.65%	27	0.74%	112	3.08%
	LITE	35	0.96%	197	5.41%	403	11.07%	210	5.77%	845	23.21%
	DK/UN	0	0.00%	1	0.03%	9	0.25%	6	0.16%	16	0.44%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ENVPC[A,B,C]	Environmental conditions which may have contributed to the crash.										
	GLARE	3	0.08%	18	0.49%	54	1.48%	18	0.49%	93	2.55%
	NONE	106	2.91%	503	13.81%	1769	48.59%	845	23.21%	3223	88.52%
	OBSTR	0	0.00%	13	0.36%	39	1.07%	15	0.41%	67	1.84%
	WTHR	10	0.27%	40	1.10%	110	3.02%	61	1.68%	221	6.07%
	Other values	1	0.03%	7	0.19%	18	0.49%	9	0.25%	35	0.96%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Table 5-2 to

Variable	Indication	Injury severity									
		K		A		B		C		Subtotal	
		N	%	N	%	N	%	N	%	N	%
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
VEHTYPE [1, 2]	Specific category for the type of vehicle which was involved in a crash.										
	CAR	58	1.59%	291	7.99%	746	20.49%	396	10.88%	1491	40.95%
	SUV	20	0.55%	71	1.95%	138	3.79%	87	2.39%	316	8.68%
	UT TRK	17	0.47%	48	1.32%	75	2.06%	35	0.96%	175	4.81%
	P VAN	4	0.11%	13	0.36%	42	1.15%	21	0.58%	80	2.20%
	Other values	21	0.58%	158	4.34%	991	27.22%	409	11.23%	1579	43.36%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SPEEDFLAG	Flag indicating whether speed was a factor in a crash.										
	Y	8	0.22%	36	0.99%	29	0.80%	11	0.30%	84	2.31%
	3557 (97.69%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
POSTSPD[1,2]	The posted/statutory speed limit for a motor vehicle at the time of the crash. A value of 77 indicates not Applicable.										
	<35 mph	51	1.40%	422	11.59%	1541	42.32%	769	21.12%	2783	76.44%

	35 to 50 mph	32	0.88%	107	2.94%	319	8.76%	115	3.16%	573	15.74%
	>50 mph	37	1.02%	52	1.43%	132	3.63%	64	1.76%	285	7.83%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SCHZONE	Flag indicating whether a crash occurred in an active school zone.										
	Y	2	0.05%	10	0.27%	46	1.26%	30	0.82%	88	2.42%
	3553 (97.58%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
CONSZONE	Y	3	0.08%	6	0.16%	17	0.47%	6	0.16%	32	0.88%
	3553 (97.58%) blanks										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Table 5-6 described the selected crash variables. Mostly, (44.91%) of drivers are between 30-64 years old, a fair part (47.87%) of them are males, 65.06% of them used shoulder and lap belt, 8.21% of them are drunk drivers, and 25.32% of them failed to yield the ROW. Regarding the vehicles involved in the pedestrian-vehicle crashes, the table indicates that (40.95%) of them are passenger vehicles.

Whereas, for the pedestrians and bicyclists engaged in vehicle crashes, more than half (51.06 %) of them are < 30 years. Concerning the locations were the pedestrian and bicycle-vehicle crashes occurred, the surface of most of them (83.93%) are dry, most of them (55.26%) are controlled intersections and a vast majority (91.93%) of them are straight road. Concerning the environmental conditions accompanied by the pedestrian and bicycle-vehicle crashes, (63.09%) of the crashes happened under clear weather conditions. The largest part (99.12%) and (97.58%) of crashes did not occur in working areas and around school zones, respectively.

5.4.3 Data Sources for Assigning Party At-fault

The assignment of the party at-fault is based on three main sources: i) pedestrian and bicycle law violation and driver citation, ii) conditions/circumstance contributing to a crash (for both drivers, and pedestrians and bicyclists) that strongly suggest responsibilities for an undesirable event, mistake, or defective, and iii) police narrative including eyewitness statement if available.

5.4.3.1 Citations

Wisconsin State legislations ("Wisconsin State Legislations"), which includes comprehensive rules of the right of way (ROW) were followed. For instance, rules regarding pedestrians' crossing location, pedestrian signal status and walk/no walk actions, and vehicle drivers stopping or leave standing any vehicle in places such as within an intersection or on a crosswalk.

R. J. Schneider and Sanders (2015) stated that studies examining public understanding of crosswalk laws suggested that crosswalk laws may be confusing, counterintuitive, or possibly inappropriate for the local driving culture. The authors stated that there is an essential need for enhanced education and a need to build enforcement strategies that will complement and support the conducted engineering treatments. It is vital to learn how non-motorists' law differs from the social norm. For instance, any pedestrian, bicyclist, or electric mobility device rider crossing not at the crosswalk shall yield the ROW to vehicles upon the roadway.

Crash reports record the issued citations which helped in suggesting but not yet determining the responsible party. Types of citations were reviewed (related or unrelated to the crash) issued for the drivers. Unrelated citations are excluded from the considered citations. Crash-related citations were issued in 25 crashes (30%). R. Schneider, Stefanich, and Corsi (2015) expressed that the low number of crash relevant stations maybe because citations are not given to fatally wounded pedestrians.

Citations issued for the crash varied in types, such as Failure to signal when changing lanes or turning, Failure to drive within a specified lane/driving in between two lanes, speeding, and running a red light/a stop sign. Some observations record other citation types issued even without being related to the crash scene such as citations for an expired license. Additionally, in some crash reports police narrative indicate that the driver violated a traffic law, but the driver was not issued a citation for the violation. Error! Reference source not found. describes the reviewed citations from the data field "STATNM [1, 2] [A, B, C, D] ordered by their relevancy.

5.4.3.2 Conditions, Circumstances, and Actions Contributing to A Crash

Driver behavioral data is a rich source of information that gained the interest of many researchers. Wang and Qin (2015) stated that driver characteristics and behavior appear to have a great influence on the error severity outcome. Moreover, the authors concluded that male drivers (≤ 25) as against other age groups except for women (≥ 55) have a significantly higher probability of making severe mistakes.

The key driver and non-motorists' behavior factors are reported on the crash report by police officers at the crash scene. For drivers, those behaviors may be viewed as risky, careless/inattentive, and reckless driver behavior, red light violation, and uncertain driving. Whereas, non-motorists' behavior examples are failing to obey the traffic control device, darting into the roadway, wearing dark clothes, disobeying a traffic control person. driver behavioral data are used to make decisions to allocate crash prevention resources at the local, state, and national levels (National Safety Council (NSC), 2017).

5.4.3.3 Police Narrative

The language used by the police officer to provide valuable data in the report is referred to as the police narrative. It is a formal story-telling process necessary to focus investigative strategy. Mostly, it is joined with a diagram that sketches the crash scene. Multiple information may be extracted from police narrative reports, such as the location of the unit at the time of the crash, and a not to scale sketch of the roadway/intersection/curve showing the entire crash scene. Such details assist the police officer to elaborate on how the crash may occur. Regarding information extraction from police narrative reports, the process must be done manually since narrative reports are very noisy compared to other types of text data, due to spelling errors and typos (Chau, Xu, and Chen 2002). Note that police narrative includes eyewitness testimony if available.

Recent studies of eyewitnesses and human memory have suggested that eyewitness evidence is much like trace evidence left at a crime scene ("Eyewitness-Public-20091105.Pdf" n.d.). Further, according to the crash investigation regulations, the police officer investigating the crash should preserve statements from drivers and eyewitnesses for each crash. Eyewitnesses review is supplied within the narrative dedicated space in the police crash report, provides some extra details about the crash occurrence. In some cases, the individual (driver or pedestrian) confesses a statement about his/her fault leading to the crash occurrence.

The following is an example of a police narrative with eyewitness testimony: "*On January 11th, 2017 I was dispatched to a car vs pedestrian accident just south of the xxx-traffic circle. Upon contact with the pedestrian, he stated that he was walking westbound across xxx St in the crosswalk when he was struck by a vehicle. When we made contact with unit #1 driver, she stated that she was traveling southbound on xxx St in the outside lane when the vehicle also traveling southbound in the inside lane stopped. She did not know why it stopped and only slowed down to* *see the pedestrian right before she struck him with her veh. damage is to the right front hood. Several individuals witnessed this incident and stated that same story as both the pedestrian and the unit #1 driver. All individuals were identified via WI photo dl. The pedestrian was looked at by OFD and received treatment on the scene but refused medical transport and was given a ride home by another officer. No citations were issued*".

Also, an example for a police narrative without an eyewitness testimony: "*Unit 1 was n/b on xxx. The pedestrian was returning from getting the newspaper from the box across the street from his house. It appears the pedestrian was right at the edge of the road when he was struck by unit 1. The pedestrian was wearing blue clothing. It was dark outside and there were no street or yard lights. Driver unit 1 said he never saw the pedestrian even after striking him. He pulled over a few houses up and turned around to see what he struck. he did not see anything. He turned around and went back to the area and still did not see anything. He continued down the road to a business that had a lit parking lot.*"

5.5 Analysis and Assignment of Party At-fault

5.5.1 Crash data fields-based query and results

The analysis was initiated by querying data fields in the dataset of 2017-2018 pedestrian and bicycle crashes and considering specific attributes that showed a relation to a driver or a VRU being at-fault in a crash. This query resulted in creating four data fields; driver citation, VRU citation, driver actions/circumstances, and VRU actions/circumstances, that were extracted from two data sources; "citations" and "actions, conditions, and circumstances" shown in **Table 5-7.**

Table 5-7: Crash data fields involved in the party at fault assignment.

Data Source	Extracted information	The specific information

Citations		The field name and citation types considered in suggesting the fault party.	STATNM [1, 2] [A, B, C, D]: the statute number of the violation for which a driver was cited. Driver citations (e.g., 341.04 (3), 343.05, 344.62 (1)), and VRU citations (e.g., 346.804, 347.489(1), 347.489 (2)). Check Error! Reference source not found. for a full list of citations considered for driver and VRU violations that may be related to crashes.
Actions, Conditions, and Circumstances	**Actions**	NMTACT [1, 2]: the actions/circumstances of the non-motorists that may have contributed to the crash, based on the judgment of the law enforcement officer investigating the crash.	FTY (failing to yield the ROW); SUDDEN (sudden movement into traffic); IM XING (improper crossing of the roadway); INATTV (inattentive driving); DISREG (disregarding traffic control device); DK CLTH (wearing dark clothes); IM XING (improper crossing of roadway/jaywalking).
		DRVRDOIN [1, 2]: the controlled maneuver for this motor vehicle prior to the beginning of the sequence of events, such as turn on red.	RTOR (turn on red); NO PASS (violating no passing zone).
	Conditions and Circumstances	DNMFTR [1, 2] [A, B]: any relevant condition of the individual (motorist or	PHY IMP (physical impairment); UI MDA (being under the influence of medications/alcohol/drugs).

		non-motorists) that is directly related to the crash.	

5.5.2 Narrative Text Classification of Party at-Fault

Machine learning techniques such as random forests (RF) can extract features and rank their importance from unstructured text data such as crash narrative. In **Figure 5-1**, the most 20 important features extracted along with the (text length) feature are illustrated. The RF model found that the following keywords best describe the driver at-fault crashes and VRU at-fault crashes: fail, back, left, shoulder, jump, legal_cross, run, crosswalk, infront, shoulder, dark, red, sudden, no_see, fled, dart, path, walk, and run_across. The “textlength” keyword is also considered important because it is an engineered feature that was considered a good feature in differentiating between driver at-fault and VRU at-fault crashes.

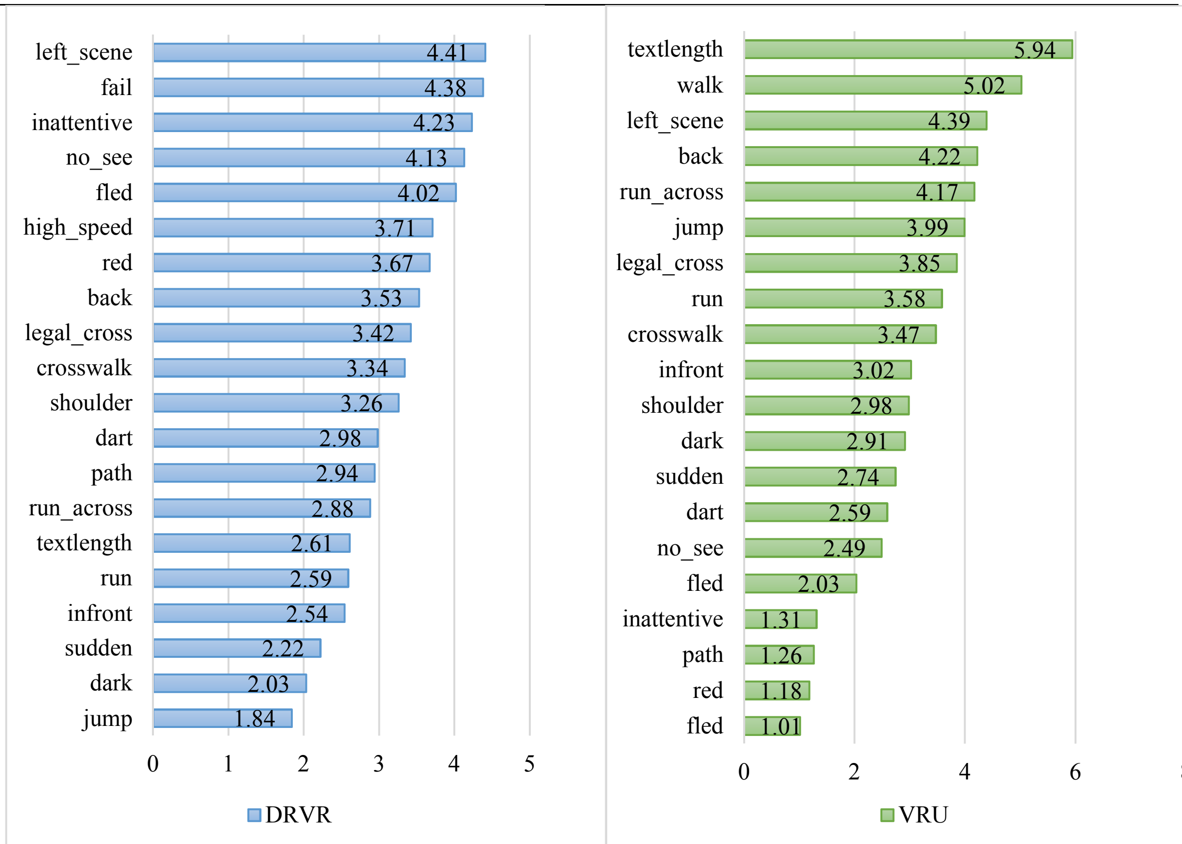


Figure 5-1: Variable importance plot of the RF model's frequently used words

The words fled_scene and "not_see" were considered important for categorizing driver at-fault crashes because they reflect the actions related to the driver being responsible of the crash, as they often appear in phrases such as "***fled the scene*** *without rendering aid*", and "*stated she did* ***not see****/was blinded by the sun*". Likewise, "backing", "turn_red", "rate" indicated a possible cause of driver at-fault crashes due to the driver's responsibility of paying attention to VRU in a crosswalk or on the sidewalk in front of a driveway, such in "***backing*** *out of his driveway*", "*make a right* ***turn on red*** *onto xxx ave and struck the bicycle*", and "*at a high rate of speed when it struck unit 2 who was walking*". The keyword "backing" was excluded from word normalization so that it is not consolidated with "back" into back since the later gives a different meaning compared to “backing”. “back” appeared to be associated with the VRU as in “***back*** *pain and did seek medical attention*”, and “it struck the **back** of the bicyclist”. “driveway” indicated a possible fault of the driver in a crash due to crashes occurring with VRUs while exiting/entering the **driveway**. This keyword occurred through the narratives as: “*backing E/B out of the* ***driveway****”, and “left turn into the* ***driveway***”.

The variable importance plot in **Figure 5-1** also presented the most important keywords implying the VRU’s fault in a crash. The keyword “run_into” appeared to be the most important in terms of predicting a VRU fault crash, due to the improper action by the VRU as reported by the police officer. It has appeared through the narrative as “*pedestrian* ***ran into*** *traffic chasing after*”, and as ” *pedestrian crossed in front of the parked vehicle and* ***ran into*** *the side of*”. Having a have clear view on how terms/keywords appear throughout the corpus of narratives aided in the process of narrative text processing.

Such as in this case, where the term “out” is removed so that the phrase “*run out into*” is converted to “run_into” after stemming and using the bigrams and calculated together with the “run_into” keyword when it appears in the narrative such “*unit 2 (pedestrian) that crossed in between parked traffic and running into the roadway”.*

Figure 5-1 also presented that the engineered feature “textlength” indicated crashes that occurred due to VRU fault. With looking into this feature through the narrative and counting the characters of the narratives, this feature showed a relationship with VRU at-fault crash narratives. “sudden” as appeared in ”***suddenly*** *went into traffic*”, “run” as appeared in “*unit 2 (pedestrian) was running sb and* ***ran*** *into the street mid block*”, “infront” as appeared in “*ran* ***infront*** *of unit one agaist the traffic light”*, and “dart” as appeared in “*unit two which is a pedestrian. unit two* ***darted*** *into traffic crossing*”, indicated VRU at-fault crashes. Such crashes occurred due to unexpected and quick movement of the VRU, which is considered in most cases to trigger an immediate and hurried reaction by the driver leading to a crash. “stand” and “not_in” reflected the VRU location of first responders in secondary crashes. At the first glance, both keywords do not show an association with the VRU location; however, the following are examples of how these two location-related keywords appeared in the narratives: *“was **standing** in the roadway in the middle of the street*”, and “*pedestrian was **not in** a crosswalk”*. “dark_cloth” immediately seems to characterize a VRU at-fault crash, as it was in almost all cases associated with “clothes/clothing” in crash narratives where the VRU is decided at-fault. For instance*, “in **dark clothing** and he didn't see him”,* and *“didn't see unit #2 because she was wearing **dark clothing**”*.

Through the manual narrative revision, some scenarios were found to be confusing due to the lack of a specific definition of a driver and VRU at-fault crash, especially when a narrative involves an eyewitness testimony that leans towards showing that the driver is at-fault even if not. Hence, a specific standard was followed in determining the fault party manually to determine a crash was a result of the fault of a specific party. For instance, a driver violating a specific traffic rule that was not included in the field attributes mentioned in **Error! Reference source not found.** (i.e., failure to yield the ROW), was considered a driver at-fault crash. Otherwise, if the narrative did not involve such violation even if a citation related to the crash was issued, it will not be considered a driver at-fault crash since the citation is considered in the fault assignment as a sperate data source.

The following narratives do not indicate any of the two fault parties were responsible for the crash, and therefore may be debated not to be deemed as a driver at-fault crash and VRU at-fault crash, respectively.

- *"unit 2 was riding her bike in the crosswalk on xxx st at w. xxx ave. unit 1 was traveling eastbound on x and struck unit 2. driver of unit 1 said he was unable to stop in enough time".*
- *"operator of unit two is a bicycle. operator does not remember which direction he was riding his bike prior to the accident. operator does not remember which direction the striking vehicle was coming from. based on the location of the bicycle and operator after the crash the vehicle could possibly have been traveling northbound on n xxx St. it is still unknown which direction the bicyclist was riding.".*

Such scenarios express how challenging is the task of correctly classifying the crash's responsible fault party crashes for a human being, let alone machines. Even so, the fact that the model classified these crashes reveals its sensitivity.

5.5.3 Party at-Fault Analysis and Assignment

The analysis was initiated by querying data fields in the dataset of 2017-2018 pedestrian and bicycle crashes and considering specific attributes that showed a relation to a driver or a VRU being at-fault in a crash. This query resulted in creating four data fields; driver citation, VRU citation, driver actions/circumstances, and VRU actions/circumstances, that were extracted from two data sources; "citations" and "actions, conditions, and circumstances" shown in **Table 5-7.**

Regarding the police narrative part in a crash report, a manual review was performed to classify the crash party at fault. It is noted that the text data in the narrative can be very noisy due to spelling errors and typos in police scenarios, which lowers the data quality and make the text less applicable to be used in automated processing techniques such as natural language processing (NLP). The hand-labeled data have two purposes: i) establishing gold labels that represent reliable ground-truth values, and ii) creating manually annotating training examples to train a classifier.

After collecting the information from the explanatory variables from the abovementioned data sources. Regarding the dummy coding, each of the explanatory variables from the three data sources was coded as follows:

- Citations: coded (1) if a citation related to the crash is issued for any party and coded (0) if no citation was issued. Check Error! Reference source not found. for the types of citations issued for VRUs and drivers that are believed to be related to the studied crashes;
- Conditions (i.e., physical impairment, being under influence of medications/alcohol/drugs, wearing dark clothes); circumstances and actions (i.e., failing to yield the ROW, inattentive driving, improper crossing by a pedestrian/bicyclist) contributing to a Crash: coded as 1 if any party has a condition that might have contributed to the crash issued for any party and coded as 0 if no citation was issued.
- Police Narrative: coded (1) if the narrative deems the party to be at fault, and coded (0) if not. Here are some example keywords which influence Police narrative fault party decision, such as "crosswalk, fail, path, walk, fled, back, no_see, left, legal_cross, dart, run, run_across, infront, shoulder, scene, inattentive, jump, dark, sudden".

The query based on the appropriate attributes of two crash data fields citations" and "actions, conditions, and circumstances" return four possible categories: DRVR at-fault, VRU at-fault, BOTH at-fault and NONE at-fault for all 3,641 crash cases. Similarly, the manual review of crash narratives assigned these cases to one of the four categories. **Table 5-8** shows the confusion matrix of the results between the crash data fields-based query and manual review of narratives.

Table 5-8: Classification confusion matrix of the crash data fields and narratives.

	Crash Narrative					**Metrics**		
Data Field	DRVR at-fault	VRU at-fault	BOTH at-fault	NONE at-fault	**Grand Total**	Consistency	NONE cases	Complementary
DRVR at-fault	587	198	5	555	**1345**	0.743		0.936
VRU at-fault	85	335		1145	**1565**	0.797		
BOTH at-fault	55	4		30	**89**			
NONE at-fault	601			41	**642**		0.176	
Grand Total	**1328**	**537**	**5**	**1771**	**3641**			
Metrics								
Consistency	0.807	0.623						
NONE cases				0.486				
Complementary	0.988							

It is clear from the confusion matrix in **Table 5-8** that the crash party at-fault assignment may not be the same between data fields-based query and manual annotation. The discrepancies are plausibly caused by 1) the inconsistencies in documenting the crash scenarios. For example, relevant circumstances or actions were checked in the crash data fields but not documented or not properly documented in the narratives, or vice versa; and 2) the inconsistencies between the judgment of the reviewers who assigned crash party at-fault and the relevant data fields that indicate the crash party at-fault.

Table 5-8 also shows a few metrics that enhance the knowledge about how crash data fields are considered complementary to the manual narrative and shows the consistency of each technique. Regarding consistency, more than 80% of driver at-fault cases based on the manual review were identified using data fields. Also, for the VRU at-fault 62.3% of manually reviewed VRU at-fault cases were classified correctly by using data fields. Looking at data fields, 74.3% of the cases confirmed by data field query as driver at-fault cases were also classified correctly by the manual review, also 79.7% of VRU at-fault casers were classified by data field were correctly classified by manual review.

For NONE at-fault cases, if narratives were manually reviewed, only 48.6% can be classified correctly. In other words, 1771 NONE observations cannot be decided by the manual review. Whereas, using data fields, only 17.6% of NONE cases cannot be decided as NONE through data fields. To understand how data fields, complement to manual narrative review and vice versa, the complementary metric was calculated. **Table 5-8** shows that 98.8% of manually reviewed cases can be confirmed by data fields. Whereas, cases that cannot be confirmed by data fields but can be defined by manual narrative are 93.6%. Hence, among 642 cases that cannot be confirmed by data fields, the manual review almost classified 100% as a driver at fault. This analysis supports the use of "at least" for the three sources for assigning party at-fault; it also confirms that data field and narrative can complement each other, enhancing the proposal of using them together not separately as inconsistency still exists.

To consolidate the inconsistent information from the citation, actions/circumstances, and crash narrative, it is determined that if at least "Yes" is checked for one party and no "Yes" from the other, the party is considered the Party at-fault. If "Yes" is checked for both parties, then both parties are at fault. If none of the parties are checked with "Yes", then, no party is at fault or the party at-fault is unknown. Afterward, the fault party assignment process illustrated in **Figure 5-2**. For text mining results and discussion, refer to **Appendix B**: Police Narrative Text Mining Process. Results of the final fault assignment which involved other data field attributes and following the proposed guideline in **Figure B1** are shown in **Table 5-9**.

Table 5-9: Fault assignment results

	Driver at-Fault	**VRU at-Fault**	**Unknown Fault**	**Both Parties at-Fault**	**Total**
N	1743	1481	40	378	3641
%	47.87%	40.65%	1.10%	10.38%	100%

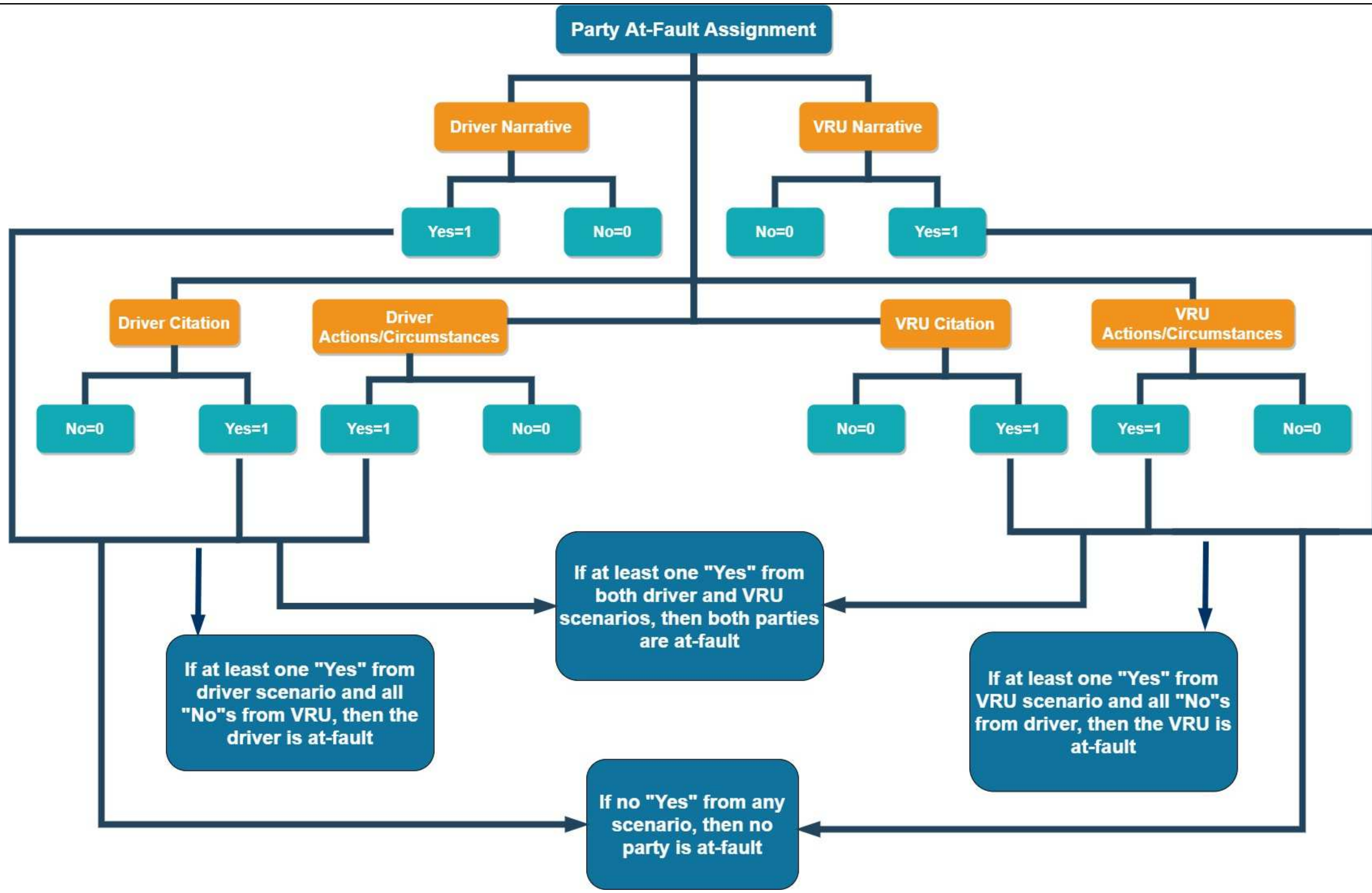


Figure 5-2: An illustration of the designed "party at-fault assignment" guideline

5.6 Crash Injury Severity Analysis by Party at-Fault

In this section, the results of the Z-test and XGboost analysis of variable importance for predicting driver at-fault crashes and VRU at-fault crashes are discussed. Moreover, the MNL model analysis results of variables influencing each specific injury severity will be discussed, for both types of crashes; driver at-fault crashes and VRU at-fault crashes.

5.6.1 Z-test analysis

The results of the z analysis can be found in **Table 5-10** to **Table 5-11**. These results show the proportion of fatal and severe injury crashes, and evident and possible injury crashes (non-severe) by each of the driver-related, VRU-related, roadway-related, crash-related, and environmental-related variables. The table identifies the crash variable that has a significantly different proportion of fatal and severe injury versus non-severe injury crashes using the z-test for proportions. The classification of the VRU injury severity level is determined based on the hypothesis that each variable influence different severity levels in different crash types (driver at-fault crashes, VRU at-fault crashes, both at-fault crashes, none/unknown at-fault crashes) which helps in determining the models created later for the different crash types. A $(-)$ symbol implies crash variables with a significantly lower proportion of fatal and severe injuries at a 95% confidence level. Whereas, a $(+\,+)$ symbol implies crash variables with a significantly higher proportion of fatal and severe injuries at a 95% confidence level.

Table 5-10: Summary of the Z-test results for driver at-fault crashes

Variable indication	Variable symbol	Fatal and severe injury (K+A) crash%	Evident and possible injury crash (B+C) %	Sig. result of the z-test *	Sample size %
Driver-related factors					
Driver action (failed to yield ROW)	DRVRPC.FTY	1.54%	5.41%		49.13%
Driver action (other)	DRVRPC.OTHER	5.58%	24.69%	–	46.26%
Driver age (<30)	DRVRAGE.LESS.30	3.95%	16.51%	–	51.10%
Driver age (30-64)	DRVRAGE.BETWEEN.30.64	4.04%	17.39%		44.98%
Driver age (≥65)	DRVRAGE.MORE.THAN.65	2.77%	0.85%	++	60.55%
Driver sex (male)	DRVR.SEX.M	2.94%	9.70%	–	50.55%
Driver sex (female)	DRVR.SEX.F	3.08%	13.90%		50.82%
Driver safety equipment (shoulder/lap belt)	DRVR.SFTYEQP.SH.LP	5.22%	23.51%		47.37%
Non-motorist/VRU-related factors					
Non-motorist Safety equipment (NONE)	NMTSFQ.NONE	6.98%	29.47%	–	78.52%
Non-motorist Safety equipment (lights, reflectors, helmet)	NMTSFQ.REFL.LTNG.HLMT	0.14%	0.38%		39.58%
Non-motorist action prior crash (no improper action)	NMTACT.NO.IMPR	0.60%	1.59%		4.97%
Non-motorist age (<30)	NMT.AGE.LESS.30	2.64%	10.13%		49.26%
Non-motorist age (30 to 64)	NMT.AGE.30.64	2.55%	10.82%		52.42%
Non-motorist age (≥65)	NMT.AGE.65.MORE	2.77%	0.85%	++	49.81%
Roadway-related factors					
Vertical road terrain (level)	ROADVERT.LVL	5.36%	20.82%	–	50.18%
Trafficway division (divided with barrier)	TRFCWAY.DIV.BAR	4.23%	18.98%	–	71.49%
Vehicle type (Utility truck)	VEHTYPE.UT.TRK	0.44%	1.73%		45.40%

Vehicle type (car)	VEHTYPE.CAR	4.15%	16.81%		51.35%
Vehicle type (sport utility vehicle-SUV)	VEHTYPE.SUV	0.88%	3.49%		50.48%
Total number of lanes (>3)	TOTLANES.MORE.3	2.36%	10.02%		49.56%
Total number of lanes (two)	TOTLANES.2	5.66%	24.11%	–	48.07%
Horizontal road terrain (straight)	ROADHOR.ST	5.60%	21.72%	–	50.23%
Traffic control (traffic signal)	TRFCCNTL.TS.OP	1.79%	7.25%		50.85%
Traffic control (stop sign)	TRFCCNTL.STOP	0.69%	2.44%		51.12%
Road condition (dry)	RDCOND.DRY	4.89%	18.54%	–	50.96%
Intersection type (4-way)	INTTYPE.4WAY	2.42%	9.45%		50.88%
Intersection type (T)	INTTYPE.T.PED.DT	0.63%	2.33%		55.96%
Road surface (black/bituminous)	SURFTYPE.BLACK.PED.DT	3.90%	14.64%	–	51.61%
Crash/vehicle-related factors					
Controlled maneuver (going straight)	DRVRDOIN.GO.STR	2.00%	8.21%		49.60%
Speed limit (<35) mph	POSTSPD.LESS.35	6.07%	30.60%	–	48.56%
Speed limit (30-50)	POSTSPD.35.TO.50	1.90%	5.16%		45.25%
Speed limit (≥50) mph	POSTSPD.MORE.50	2.88%	1.24%	++	52.63%
Environmental-related factors					
Light condition (light)	LGTCOND.LITE	1.59%	7.80%		51.43%
Light condition (day)	LGTCOND.DAY	3.38%	12.58%	–	49.87%
Light condition (dawn)	LGTCOND.DAWN	0.14%	0.63%		27.18%
Weather condition (rain)	WTCOND.RAIN	0.55%	2.47%		52.88%
Weather condition (clear)	WTCOND.CLEAR	3.57%	14.12%		51.27%
Weather condition (cloudy)	WTCOND.CLDY	1.51%	5.99%		49.91%
Total Crashes including other variable attributes		335	1408		1743

Table 5-11: Summary of the Z-test results for VRU at-fault crashes

Variable indication	Variable symbol	Fatal and severe injury (K+A) crash%	Evident and possible injury crash (B+C) %	Sig. result of the z-test *	Sample size %
Driver-related factors					
Driver action (failed to yield ROW)	DRVRPC.FTY	1.04%	3.95%		35.34%
Driver action (other)	DRVRPC.OTHER	5.69%	23.37%	–	44.42%
Driver age (<30)	DRVRAGE.LESS.30	1.54%	8.10%		24.07%
Driver age (30-64)	DRVRAGE.BETWEEN.30.64	4.23%	17.80%	–	46.25%
Driver age (≥65)	DRVRAGE.MORE.THAN.65	0.44%	1.92%		39.45%
Driver sex (male)	DRVR.SEX.M	1.48%	6.34%		31.32%
Driver sex (female)	DRVR.SEX.F	2.00%	9.06%		33.14%
Driver safety equipment (shoulder/lap belt)	DRVR.SFTYEQP.SH.LP	4.89%	10.30%		41.53%
Non-motorist/VRU-related factors					
Non-motorist Safety equipment (NONE)	NMTSFQ.NONE	0.58%	1.40%		4.26%
Non-motorist Safety equipment (lights, reflectors, helmet)	NMTSFQ.REFL.LTNG.HLMT	0.14%	0.33%		35.42%
Non-motorist action prior crash (no improper action)	NMTACT.NO.IMPR	6.26%	26.20%	–	73.42%
Non-motorist age (<30)	NMT.AGE.LESS.30	1.57%	7.09%		33.37%
Non-motorist age (30 to 64)	NMT.AGE.30.64	1.51%	6.45%		31.22%
Non-motorist age (≥65)	NMT.AGE.65.MORE	0.44%	1.92%		32.45%
Roadway-related factors					
Vertical road terrain (level)	ROADVERT.LVL	3.16%	14.03%		32.96%
Trafficway division (divided with barrier)	TRFCWAY.DIV.BAR	0.47%	0.00%		1.44%

Vehicle type (Utility truck)	VEHTYPE.UT.TRK	0.33%	1.37%		35.63%
Vehicle type (car)	VEHTYPE.CAR	2.20%	10.60%		31.36%
Vehicle type (sport utility vehicle-SUV)	VEHTYPE.SUV	0.66%	2.22%		33.33%
Total number of lanes (>3)	TOTLANES.MORE.3	2.09%	8.16%		40.99%
Total number of lanes (two)	TOTLANES.2	5.03%	20.57%	–	41.33%
Horizontal road terrain (straight)	ROADHOR.ST	3.32%	14.50%		32.76%
Traffic control (traffic signal)	TRFCCNTL.TS.OP	1.18%	4.67%		32.92%
Traffic control (stop sign)	TRFCCNTL.STOP	0.33%	1.73%		33.63%
Road condition (dry)	RDCOND.DRY	2.80%	11.65%		31.42%
Intersection type (4-way)	INTTYPE.4WAY	1.57%	6.23%		33.45%
Intersection type (T)	INTTYPE.T	0.27%	1.29%		29.53%
Road surface (black/bituminous)	SURFTYPE.BLACK	2.31%	9.09%		31.73%
Crash/vehicle-related factors					
Controlled maneuver (going straight)	DRVRDOIN.GO.STR	1.26%	5.38%		32.27%
Speed limit (<35) mph	POSTSPD.LESS.35	5.36%	25.60%	–	41.00%
Speed limit (30-50)	POSTSPD.35.TO.50	1.70%	5.30%		44.89%
Speed limit (≥50) mph	POSTSPD.MORE.50	0.82%	1.90%		34.74%
Environmental-related factors					
Light condition (light)	LGTCOND.LITE	0.88%	5.27%		33.68%
Light condition (day)	LGTCOND.DAY	2.25%	7.94%		31.85%
Light condition (dawn)	LGTCOND.DAWN	0.99%	0.22%		42.72%
Weather condition (rain)	WTCOND.RAIN	0.30%	1.37%		29.33%
Weather condition (clear)	WTCOND.CLEAR	2.09%	8.98%		32.09%
Weather condition (cloudy)	WTCOND.CLDY	0.85%	4.04%		32.54%
Total Crashes including other variable attributes		288	1193		1481

In summary, **Table 5-10** to **Table 5-11** presents crash circumstances by prevalence and shows sample sizes and percentage of injury severity levels for each variable's attribute. For driver at-fault crash results showed in **Table 5-10**, crashes involving drivers acted in actions other than failure to yield the ROW, drivers younger than 30 years, male drivers, non-motorists negligence of using safety equipment, straight and level road terrains, traffic ways divided with a barrier, two-lane roadways, dry-surface and roadways, bituminous roadways, and roadways with speed limits (<35) mph showed a significantly lower percentage of severe versus non-severe injuries. Whereas, crashes involving drivers and pedestrians/bicyclists (≥65) years, and those crashes occurred on roadways with speed limits of (≥50) mph showed a significantly higher percentage of severe versus non-severe injuries. Such variables are of a high influence on driver at-fault crashes.

While for VRU at-fault crashes, the results in **Error! Reference source not found.** showed hat the following variables showed a significantly lower percentage of severe versus non-severe injuries, highlighting the importance of capturing the effects of these variables on VRU at-fault crashes: crashes involving drivers acted in actions other than failure to yield the ROW, drivers aged between 30 and 64 years, VRUs involved in crashes with no improper actions, crashes occurred on two-lane roadways, and those occurred on roadways with a speed limit of (<35) mph.

5.6.2 XGboost Decision Tree Results

The raw Wisconsin pedestrian-vehicle crash data is randomly split into training and test sets, and the crash injury severity of driver fault crashes and VRU fault crashes are fit into the XGboost model. Seventy-three explanatory variables are checked for influencing crash severity. Twenty variables were found to significantly influencing variables of the crash type severity were found, and the most important twenty contributing variables to the severity level of a crash type are summarized. **Table 5-12** and **Table 5-13** present the variable ranking in ascending order.

Table 5-12: Variable Level of Importance for Each Severity Level for Driver Fault Crash Types

P (Fatal (K) Crash)			P (Severe Injury (A) Crash)			P (Evident and Possible Injury (B+C) Crash)		
Variable	**Score**	**Cumulative %**	**Variable**	**Score**	**Cumulative %**	**Variable**	**Score**	**Cumulative%**
Speed limit (35 to50 mph)	100.00	37	**Speed limit (35 to50 mph)**	100.0	17	**Speed limit (35 to50 mph)**	100.00	38
Trafficway division (divided with no traffic barrier)	31.56	49	Driver age (<30)	68.7	29	**Trafficway division (divided with no traffic barrier)**	29.36	49
Driver action (other)	**14.12**	**54**	**Driver action (other)**	**42.1**	**37**	**Driver action (other)**	**15.62**	**54**
Driver age (<30)	12.82	59	Trafficway division (divided with no traffic barrier)	37.7	43	Total number of lanes (3)	14.00	60
Total number of lanes (3)	12.07	63	**Total number of lanes (3)**	32.6	49	Driver age (<30)	13.78	65
Light condition (day)	11.62	68	Driver age (≥65)	32.4	54	Horizontal road terrain (straight)	12.64	70
Traffic control (traffic signal)	11.59	72	Light condition (day)	26.1	59	Driver age (≥65)	12.02	74
Vertical road terrain (level)	9.72	76	Total number of lanes (one)	25.6	63	Road condition (dry)	9.62	78
Traffic control (stop sign)	8.84	79	Traffic control (traffic signal)	24.6	68	Light condition (day)	9.24	81
Driver age (≥65)	7.98	82	Weather condition (cloudy)	22.2	72	Vertical road terrain (level)	8.06	84

Road condition (dry)	7.25	85	**Road condition (dry)**	21.3	75	Driver action (failed to yield ROW)	7.43	87
Controlled maneuver (going straight)	7.13	87	Horizontal road terrain (straight)	18.4	78	Traffic control (traffic signal)	6.28	89
Driver action (failed to yield ROW)	5.51	89	Vertical road terrain (level)	16.4	81	Total number of lanes (one)	5.25	91
Intersection type (4 way)	5.29	91	**Light condition (dark)**	16.3	84	**Light condition (dark)**	4.42	93
Driver sex (male)	4.46	93	**Controlled maneuver (going straight)**	16.1	87	**Controlled maneuver (going straight)**	4.40	95
Total number of lanes (one)	4.40	95	Driver action (failed to yield ROW)	15.8	90	Weather condition (cloudy)	3.78	96
Light condition (dark)	4.34	96	Driver sex (male)	15.5	92	Intersection type (4 way)	3.49	97
Weather condition (cloudy)	4.17	98	Traffic control (stop sign)	15.0	95	Traffic control (stop sign)	2.35	98
Horizontal road terrain (straight)	3.85	99	Intersection type (4 way)	14.5	98	Driver sex (male)	2.32	99
Road surface type (black/bituminous)	2.19	100	**Road surface type (black/bituminous)**	14.2	100	**Road surface type (black/bituminous)**	2.19	100

Table 5-13: Variable Level of Importance for Each Severity Level for VRU Fault Crash Types

P (Fatal (K) Crash)			**P (Severe Injury (A) Crash)**			**P (Evident and Possible Injury (B+C) Crash)**		
Variable	**Score**	**Cumulative %**	**Variable**	**Score**	**Cumulative %**	**Variable**	**Score**	**Cumulative%**
Safety equipment (lights, reflectors, helmet)	100.00	45	**Safety equipment (lights, reflectors, helmet)**	100.00	44	**Safety equipment (lights, reflectors, helmet)**	100.00	36
Total number of lanes (one)	13.23	51	**Trafficway (divided with barrier)**	16.40	51	**Trafficway (divided with barrier)**	26.04	45
Trafficway (divided with barrier)	12.95	56	**Non-motorist action prior crash (no improper action)**	12.21	56	**Non-motorist action prior crash (no improper action)**	18.27	52
Non-motorist age (30 to 64)	10.26	61	**Non-motorist age (30 to 64)**	10.01	60	**Non-motorist age (30 to 64)**	15.31	58
Speed limit (35 to 50) mph	10.11	66	Total number of lanes (one)	9.52	65	Light condition (light)	14.65	63
Vehicle type (Utility truck)	9.93	70	**Speed limit (35 to 50) mph**	8.89	68	**Speed limit (35 to 50) mph**	14.13	68
Light condition (light)	8.81	74	**Vehicle type (sport utility vehicle-SUV)**	8.77	72	**Vehicle type (sport utility vehicle-SUV)**	11.39	72
Non-motorist action prior crash (no improper action)	8.71	78	Non-motorist age (≥65)	8.74	76	Traffic control (stop sign)	10.54	76
Non-motorist age (≥65)	7.14	81	Non-motorist age (<30)	7.08	79	Intersection type (4 way)	10.41	80

Weather condition (rain)	6.69	84	**Weather condition (rain)**	6.53	82	Horizontal road terrain (straight)	9.95	83
Traffic control (stop sign)	6.37	87	Horizontal road terrain (straight)	6.01	85	Total number of lanes (one)	7.75	86
Horizontal road terrain (straight)	4.53	89	Light condition (light)	5.88	87	Weather condition (rain)	6.98	89
Non-motorist age (<30)	4.13	91	Vertical road terrain (level)	5.00	89	Weather condition (clear)	6.89	91
Vertical road terrain (level)	4.07	93	Weather condition (clear)	4.62	91	Non-motorist age (≥65)	5.63	93
Vehicle type (car)	3.33	94	**Safety equipment (none)**	4.13	93	**Safety equipment (none)**	4.34	95
Weather condition (clear)	3.22	96	**Speed limit (>50) mph**	3.94	95	**Speed limit (>50) mph**	3.26	96
Speed limit (>50) mph	2.70	97	Intersection type (4 way)	3.53	96	Non-motorist action prior crash (in roadway-other)	3.13	97
Safety equipment (none)	2.55	98	Traffic control (stop sign)	3.30	98	Vehicle type (car)	3.13	98
Intersection type (4 way)	2.48	99	Vehicle type (car)	2.51	99	Non-motorist age (<30)	2.96	99
Non-motorist action prior crash (in roadway-other)	2.18	100	**Non-motorist action prior crash (in roadway-other)**	2.45	100	Vertical road terrain (level)	2.41	100

Table 5-12 and Table 5-13 represented variable importance scores for both driver at-fault crashes and VRU at-fault crashes. Variable importance in a tree is estimated by a score that reflects the contribution this variable makes in predicting the crash severity in the predictor variable and measured based on the several instances the variable is used as a splitter, and the enhancement this tree provides for the mean squared error by the splits of the variable. Then, the importance score is accumulated and calculated for a single tree over the ensemble of trees, and the average summation value is scaled to obtain a score of (100) for the most important variable. Accordingly, the scaled average value for each variable reflects the variable's importance. A variable with a high importance value denotes that the variable contributes significantly to the prediction (Friedman 2001). It is noted from the driver fault section of the table, that the first six variables' contribution to the injury severity prediction in driver fault crashes reached as high as 70%. For fatal injury and (evident and possible injury) levels of crash severity, the variable with the great emphasis is "Trafficway division (divided with no traffic barrier)", meaning that trafficway division with no barrier at the crash scene compared with the remaining variables provides the greatest participation in the explanation of fatal and (evident and possible injury) crashes.

Also, "Driver action (other)"is the third variable with the greatest emphasis on all levels of crash severity, accounting for 54%, 42.1%, and 15.62% of the emphasis that driver action participates too, respectively. The absolute cumulative participation of the variables is labeled "Cumulative %" in **Table 5-12 and Table 5-13**, denoting the collective contribution of each of the variables. Whereas, in crashes resulting in severe injuries, the second most contributing factor is" driver age <30", being responsible for 29% in explaining injury crashes. **Table 5-12 and Table 5-13** also show that variables are different in describing and affecting the crash outcome.

Additionally, a variable demonstrating high importance for a level of crash severity may be less essential for another. For example, “Horizontal road terrain (straight)” is the six most essential variable in predicting crashes that were evident and possible injuries occurred yet is less essential in predicting fatal and severe injury crashes. Even though, “speed limit (30 to 50 mph)” is the most effectual variable in all crash severity levels. “Trafficway division (divided with no traffic barrier) and Driver action (other)” contribute significantly in predicting crash severity of all levels.

Other interesting findings include a) “Driver age (<30)” provide higher contribution when explaining crashes resulting in severe injuries but less contribution when clarifying evident and possible injury (B+C) crashes; b) “Light condition (day)” provides a great contribution in explaining fatal and severe crashes but slight contribution for explaining evident and possible injury (B+C) crashes; c) “Trafficway division (divided with no traffic barrier)” provides a great contribution in explaining fatal and evident and possible injury (B+C) crashes compared to explaining severe injury crashes; and d) “Speed limit (35 to 50 mph)”, “Trafficway division (divided with no traffic barrier)”, “Driver action (other)”, “Driver age (<30)”, and “Total number of lanes (>3)”, significantly influence all crash severity levels. It is clear that the XGboost technique efficiently pointed out the variables that significantly influence crash severity levels and ranked them per their degree of significance.

5.6.3 Crash Variables in the Injury Severity Models

The following analysis carries out injury severity models for VRUs involved in driver fault crashes and VRU crashes, provide a comprehensive marginal effect analysis for the most significant and highly important variables. The marginal effect analysis offers a supplementary detailed explanation of how every variable participates in all crash severity levels. **Table 5-14 and Table 5-16** show the estimated coefficients of each variable involved in the MNL model using for driver at-fault crash variables, and VRU at-fault crash variables, respectively. The marginal effects of each significant factor on the likelihood of each injury-severity class are reported in **Table 5-15 and Table 5-17**.

Table 5-14: Injury Severity Model of Driver At-Fault Crash Variables

Variable	**Code**	**Severe Injury (A) Crash**		**Evident and Possible Injury (B+C) Crash**	
		Coef.	P-value	Coef.	P-value
Intercept		3.05	0.02	4.89	0.00
Speed limit (35 to50 mph)	POSTSPD.30.TO.50	-3.76	0.01	-4.38	0.00
Total number of lanes (>3)	TOTLANES.MORE.3	-2.31	0.09	-2.31	0.07
Driver age (<30 mph)	DRVR.AGE.UNDER.30	-2.93	0.08	-2.24	0.09
Driver action (other)	DRVRPC.ID	2.88	0.06	2.86	0.05
Trafficway division (divided with no traffic barrier)	TRFCWAY.DIV.NO	-2.40	0.19	-3.54	0.04

Table 5-15: Marginal Effects of Injury Severity Variables for Driver At-Fault Crashes

Variable	**P (Fatal (K) Crash)**	**P (Severe Injury (A) Crash)**	**P (Evident and Possible Injury (B+C) Crash)**
Speed limit (35 to50 mph)	0.9169	-0.1346	-0.7822
Total number of lanes (>3)	0.5001	-0.1136	-0.3864
Driver age (<30)	0.5198	-0.2015	-0.3183
Driver action (other)	-0.6207	0.1432	0.4775
Trafficway division (divided with no traffic barrier)	0.7095	-0.0245	-0.684

The analysis in both **Table 5-14** and **Table 5-15** Error! Reference source not found.showed the variables selected for VRU severity rate prediction in driver fault crash type.

Crashes occurred on roadways with a speed limit of (35 to 50 mph) among the other two groups of speed limit considered for the study (i.e., <35 and >50 mph) showed that pedestrians and bicyclists are less likely to sustain severe injuries.

The same finding applies to crashes that occurred on roadways with more than three lanes. Driver age and actions showed an influence as well on VRU severity rate prediction in driver fault crash type. Both **Table 5-14** and **Table 5-15** showed that pedestrians and bicyclists are less likely to sustain severe injuries when the driver is younger than 30 years old (<30).

Whereas, if the driver is inattentive, careless, or acted erratically, pedestrians and bicyclists are more likely to sustain severe injuries. Pedestrian and bicyclist actions immediately before the crash showed an influence on their injury severity rate in **Table 5-14** and **Table 5-15** showed that pedestrians and bicyclists are less likely to sustain severe injuries if they were crossing a roadway divided with no traffic barrier, compared to other traffic division levels considered in this study such as two-way undivided trafficways, and one-way traffic roadways.

Table 5-16: Injury Severity Model for VRU At-Fault Crash Variables

Variable	**Code**	**Severe Injury (A) Crash**		**Evident and Possible Injury (B+C) Crash**	
		Coef.	P-value	Coef.	P-value
Intercept		2.59	0.00	4.26	0.00
Light condition (light)	LGTCOND.LITE	---	---	-1.22	0.02
Trafficway (divided with barrier)	TRFCWAY.DIV.BAR	-1.61	0.01	-1.66	0.00
Speed limit (35 to 50) mph	POSTSPD.35.TO.50	-0.76	0.08	-1.05	0.01
Safety equipment (lights and reflectors)	NMTSFQ.REFL.LGT	-1.67	0.00	-2.09	0.00

Non-motorist age (30 to 64)	NMT.AGE.30.TO.64	-0.93	0.22	-1.27	0.06
Vehicle type (sport utility vehicle-SUV)	VEHTYPE.SUV	0.76	0.05	0.63	0.09
Non-motorist action prior crash (no improper action)	NMTPRIOR.NO.IMPR	-1.38	0.01	-1.38	0.00
Total number of lanes (two)	TOTLANES.2	-0.90	0.03	-0.64	0.11

Table 5-17: Marginal Effects of Injury Severity Variables for VRU At-Fault Crashes

Variable	**P (Fatal (K) Crash)**	**P (Severe Injury (A) Crash)**	**P (Evident and Possible Injury (B+C) Crash)**
Light condition (light)	0.0108	0.0419	-0.0527
Trafficway (divided with barrier)	0.0153	-0.0191	0.0037
Speed limit (35 to 50) mph	0.0093	-0.0466	0.0372
Safety equipment (lights and reflectors)	0.0187	-0.0713	0.0525
Non-motorist age (30 to 64)	0.0113	-0.0542	0.0429
Vehicle type (sport utility vehicle-SUV)	0.0063	-0.0361	0.0298
Non-motorist action prior crash (no improper action)	-0.0015	0.0128	-0.0113
Total number of lanes (two)	0.0175	-0.0060	0.0114

The analysis in both **Table 5-16** and **Table 5-17** showed the variables selected for VRU severity rate prediction in driver fault crash type, where:

i) Four variables are roadway-related; light condition (light), trafficway division (divided with a traffic barrier), posted speed limit 35 to 50), and a total number of lanes (two). **Table 5-16** presents that crashes occurred during the night on street lit roadways did not show an influence on severely injured pedestrians and bicyclists; however, **Table 5-17** shows that pedestrians and bicyclists involved in crashes in such roadway circumstances are more likely to sustain severe injuries. This can be because those groups of VRU may feel more confident while crossing a street lit roadway, even in dark conditions.

Regarding roadways divided with a traffic barrier and roadways with speed limits of 35 to 50 mph, both **Table 5-16** and **Table 5-17** showed that pedestrians and bicyclists involved in crashes on such roadways are more likely to sustain severe injuries, compared to other traffic way division categories (i.e., divided with a painted median, undivided, divided with no barrier) and speed limit categories (<35 mph and >50 mph) considered in this study. Considering the total number of lanes dedicated to traffic in the roadway, VRU involved in crashes on roadways with two through lane are less likely to be severely injured, as shown in **Table 5-16** and **Table 5-17**.

ii) Three variables are VRU-related; safety equipment usage (lights, reflectors, helmet), age (30 to 64 years), and action before the crash (no improper action). Safety equipment usage was studied in two categories (whether equipment such as lights, reflectors, helmet are used, and whether not), also another main category of non-motorist action was considered; (no improper action). **Table 5-16** and **Table 5-17** showed that all three VRU-related variables lower the crash injury severity. In other words, pedestrians and bicyclists of age between 30 to 64, those who did not act improperly, and those who use a type of safety equipment are less likely to sustain severe injuries in VRU fault crash types.

iii) Vehicle type (sport utility vehicle-SUV) also showed an influence. Both, **Table 5-16** and **Table 5-17** presented that pedestrians and bicyclists involved in crashes with drivers driving an SUV are more likely to be severely injured.

5.7 References

Abdel-Aty, Mohamed A., and Hassan T. Abdelwahab. 2000. "Exploring the Relationship between Alcohol and the Driver Characteristics in Motor Vehicle Accidents." *Accident Analysis & Prevention* 32 (4): 473–82. https://doi.org/10.1016/S0001-4575(99)00062-7.

Breiman, Leo. 1984. *Classification and Regression Trees / Leo Breiman ... [and Others]*. Wadsworth Statistics/Probability Series. Belmont, Calif.: Wadsworth International Group.

CDC. 2019. "CDC Works 24/7." Centers for Disease Control and Prevention. July 2, 2019. https://www.cdc.gov/index.htm.

Chau, Michael, Jennifer J Xu, and Hsinchun Chen. 2002. "Extracting Meaningful Entities from Police Narrative Reports," 5.

Chen, Zhen, and Wei (David) Fan. 2018. "A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity in North Carolina." *International Journal of Transportation Science and Technology*, October. https://doi.org/10.1016/j.ijtst.2018.10.001.

Das, Subasish, Anandi Dutta, Karen Dixon, Lisa Minjares-Kyle, and George Gillette. 2018. "Using Deep Learning in Severity Analysis of At-Fault Motorcycle Rider Crashes." *Transportation Research Record: Journal of the Transportation Research Board*, September, 036119811879721. https://doi.org/10.1177/0361198118797212.

Das, Subasish, Minh le, and Boya Dai. 2020. "Application of Machine Learning Tools in Classifying Pedestrian Crash Types: A Case Study," May.

Dorn, L., and A.E. af Wåhlberg. 2019. "Behavioural Culpability for Traffic Accidents." *Transportation Research Part F: Traffic Psychology and Behaviour* 60 (January): 505–14. https://doi.org/10.1016/j.trf.2018.11.004.

Duddu, Venkata R., Praveena Penmetsa, and Srinivas S. Pulugurtha. 2018. "Modeling and Comparing Injury Severity of At-Fault and Not at-Fault Drivers in Crashes." *Accident Analysis & Prevention* 120 (November): 55–63. https://doi.org/10.1016/j.aap.2018.07.036.

"Eyewitness-Public-20091105.Pdf." n.d. Accessed July 23, 2019. https://www.doj.state.wi.us/sites/default/files/2009-news/eyewitness-public-20091105.pdf.

Franklin, James. 2005. "The Elements of Statistical Learning: Data Mining, Inference and Prediction." *The Mathematical Intelligencer* 27 (2): 83–85. https://doi.org/10.1007/BF02985802.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." In . https://doi.org/10.1214/aos/1013203451.

Goh, Kelvin, Graham Currie, Majid Sarvi, and David Logan. 2014. "Factors Affecting the Probability of Bus Drivers Being At-Fault in Bus-Involved Accidents." *Accident Analysis & Prevention* 66 (May): 20–26. https://doi.org/10.1016/j.aap.2013.12.022.

Guelman, Leo. 2012. "Gradient Boosting Trees for Auto Insurance Loss Cost Modeling and Prediction." *Expert Systems with Applications* 39 (3): 3659–67. https://doi.org/10.1016/j.eswa.2011.09.058.

Haroon, Danish. 2017. *Python Machine Learning Case Studies: Five Case Studies for the Data Scientist*. Apress.

Ichikawa, Masao, Shinji Nakahara, and Ayako Taniguchi. 2015. "Older Drivers' Risks of at-Fault Motor Vehicle Collisions." *Accident Analysis & Prevention* 81 (August): 120–23. https://doi.org/10.1016/j.aap.2015.05.004.

Industrial Networks and Intelligent Systems. 2018. New York, NY: Springer Berlin Heidelberg.

Islam, Samantha, and Akhter B Hossain. 2019. "INJURY SEVERITY ANALYSIS OF OLDER DRIVER AT-FAULT CRASHES AT UNSIGNALIZED INTERSECTIONS: ALABAMA CASE STUDY," 6.

Islam, Samantha, and Steven L. Jones. 2014. "Pedestrian At-Fault Crashes on Rural and Urban Roadways in Alabama." *Accident Analysis & Prevention* 72 (November): 267–76. https://doi.org/10.1016/j.aap.2014.07.003.

Islam, Samantha, Steven L. Jones, and Daniel Dye. 2014. "Comprehensive Analysis of Single- and Multi-Vehicle Large Truck at-Fault Crashes on Rural and Urban Roadways in Alabama." *Accident Analysis & Prevention* 67 (June): 148–58. https://doi.org/10.1016/j.aap.2014.02.014.

Jalayer, Mohammad, and Huaguo Zhou. 2016. "A Multiple Correspondence Analysis of At-Fault Motorcycle-Involved Crashes in Alabama: At-Fault Motorcycle-Involved Crashes in Alabama." *Journal of Advanced Transportation* 50 (8): 2089–99. https://doi.org/10.1002/atr.1447.

Jelil, Radhia Abd. 2019. "Optimization of Atmospheric Plasma Surface Modification Process Using Decision Trees" 14 (4): 7.

Kandan, G Mani, Sarilya Jaiswal, Rahul Mishra, and Mrs Steffina Muthukumar. 2019. "Analysis and Prediction of Road Accidents By Graphical Visualization Using Machine Learning" 6 (1): 4.

Kim, Karl, I. Made Brunner, and Eric Yamashita. 2008. "Modeling Fault among Accident—Involved Pedestrians and Motorists in Hawaii." *Accident Analysis & Prevention* 40 (6): 2043–49. https://doi.org/10.1016/j.aap.2008.08.021.

Kim, Karl, and Lei Li. 1996. "Modeling Fault among Bicyclists and Drivers Involved in Collisions in Hawaii, 1986–1991." *Transportation Research Record: Journal of the Transportation Research Board* 1538 (1): 75–80. https://doi.org/10.1177/0361198196153800110.

Kim, Karl, Lei Li, James Richardson, and Lawrence Nitz. 1998. "Drivers at Fault." *Journal of Safety Research* 29 (3): 171–79. https://doi.org/10.1016/S0022-4375(98)00013-9.

Landauer, Thomas K, Peter W. Foltz, and Darrell Laham. 1998. "An Introduction to Latent Semantic Analysis." *Discourse Processes* 25 (2–3): 259–84. https://doi.org/10.1080/01638539809545028.

Lee, Chris, and Mohamed Abdel-Aty. 2005. "Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida." *Accident Analysis & Prevention* 37 (4): 775–86. https://doi.org/10.1016/j.aap.2005.03.019.

Lee, Gain. 2017. "Identifying Characteristics of Online Civil Complaints by Text Mining and Their Influence on the Crash Risk.Pdf." 2017. https://scholarworks.unist.ac.kr/bitstream/201301/22568/1/Identifying%20characteristics%20of%20online%20civil%20complaints%20by%20text%20mining%20and%20their%20influence%20on%20the%20crash%20risk.pdf.

Lee, Jaeyoung, Mohamed Abdel-Aty, and Keechoo Choi. 2014. "Analysis of Residence Characteristics of At-Fault Drivers in Traffic Crashes." *Safety Science* 68 (October): 6–13. https://doi.org/10.1016/j.ssci.2014.02.019.

Mohaymany, Afshin Shariat, Ali Tavakoli Kashani, and Andishe Ranjbari. 2010. "Identifying Driver Characteristics Influencing Overtaking Crashes." *Traffic Injury Prevention* 11 (4): 411–16. https://doi.org/10.1080/15389588.2010.485283.

National Highway Traffic Safety Administration. 2017. "MMUCC." Text. NHTSA. August 24, 2017. https://www.nhtsa.gov/mmucc-1.

National Safety Council (NSC). 2017. "Undercounted Is Underinvested: How Incomplete Crash Reports Impact Efforts to Save Lives." https://trid.trb.org/view/1465448.

Nayak, Richi, Noppadol Piyatrapoomi, and Justin Weligamage. 2010. "Application of Text Mining in Analysing Road Crashes for Road Asset Management." In *Engineering Asset Lifecycle Management*, edited by Dimitris Kiritsis, Christos Emmanouilidis, Andy Koronios, and Joseph Mathew, 49–58. London: Springer London. https://doi.org/10.1007/978-0-85729-320-6_7.

Patra, Braja Gopal, Vahed Maroufy, Babak Soltanalizadeh, Nan Deng, W. Jim Zheng, Kirk Roberts, and Hulin Wu. 2020. "A Content-Based Literature Recommendation System for Datasets to Improve Data Reusability – A Case Study on Gene Expression Omnibus (GEO) Datasets." *Journal of Biomedical Informatics* 104 (April): 103399. https://doi.org/10.1016/j.jbi.2020.103399.

Penmetsa, Praveena, Srinivas S. Pulugurtha, and Venkata R. Duddu. 2017. "Examining Injury Severity of Not-At-Fault Drivers in Two-Vehicle Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2659 (1): 164–73. https://doi.org/10.3141/2659-18.

Rao, Shivani, and Avinash Kak. 2011. "Retrieval from Software Libraries for Bug Localization: A Comparative Study of Generic and Composite Text Models," 10.

Ratrout, Nedal T., Shakhawat Chowdhury, Uneb Gazder, and Syed Masiur Rahman. 2017. "Characterization of Crash-Prone Drivers in Saudi Arabia – A Multivariate Analysis." *Case Studies on Transport Policy* 5 (1): 134–42. https://doi.org/10.1016/j.cstp.2016.11.001.

Romano, Eduardo, Robert B. Voas, and Bayliss Camp. 2017. "Cannabis and Crash Responsibility While Driving below the Alcohol per Se Legal Limit." *Accident Analysis & Prevention* 108 (November): 37–43. https://doi.org/10.1016/j.aap.2017.08.003.

Ruthven, Ian, and Mounia Lalmas. 2003. "A Survey on the Use of Relevance Feedback for Information Access Systems." *The Knowledge Engineering Review* 18 (2): 95–145. https://doi.org/10.1017/S0269888903000638.

Salton, G., A. Wong, and C. S. Yang. 1975. "A Vector Space Model for Automatic Indexing." *Communications of the ACM* 18 (11): 613–20. https://doi.org/10.1145/361219.361220.

Salton, Gerard, and Christopher Buckley. 1988. "Term-Weighting Approaches in Automatic Text Retrieval." *Information Processing & Management* 24 (5): 513–23. https://doi.org/10.1016/0306-4573(88)90021-0.

Šarić, Frane, Goran Glavaš, Mladen Karan, Jan Šnajder, and Bojana Dalbelo Bašić. 2012. "TakeLab: Systems for Measuring Semantic Text Similarity." In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of* *the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, 441–448. Montréal, Canada: Association for Computational Linguistics. https://www.aclweb.org/anthology/S12-1060.

Schneider, Robert J., and Rebecca L. Sanders. 2015. "Pedestrian Safety Practitioners' Perspectives of Driver Yielding Behavior across North America." *Transportation Research Record: Journal of the Transportation Research Board* 2519 (1): 39–50. https://doi.org/10.3141/2519-05.

Schneider, Robert, Joseph Stefanich, and Larry Corsi. 2015. "Wisconsin Pedestrian and Bicycle Crash Analysis: 2011-2013," 251.

Shankar, Venkataraman, and Fred Mannering. 1996. "An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity." *Journal of Safety Research* 27 (3): 183–94. https://doi.org/10.1016/0022-4375(96)00010-2.

Spainhour, Lisa K., and Isaac A. Wootton. 2007. "Modeling Fault in Fatal Pedestrian Crashes by Using Various Data Sources." *Transportation Research Record: Journal of the Transportation Research Board* 2002 (1): 64–71. https://doi.org/10.3141/2002-08.

Tay, Richard, Jaisung Choi, Lina Kattan, and Amjad Khan. 2011. "A Multinomial Logit Model of Pedestrian–Vehicle Crash Severity." *International Journal of Sustainable Transportation* 5 (4): 233–49. https://doi.org/10.1080/15568318.2010.497547.

"The WisTransPortal System." n.d. Accessed July 22, 2019. https://transportal.cee.wisc.edu/.

Turney, P. D., and P. Pantel. 2010. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37 (February): 141–88. https://doi.org/10.1613/jair.2934.

Ulfarsson, Gudmundur F., Sungyop Kim, and Kathleen M. Booth. 2010. "Analyzing Fault in Pedestrian–Motor Vehicle Crashes in North Carolina." *Accident Analysis & Prevention* 42 (6): 1805–13. https://doi.org/10.1016/j.aap.2010.05.001.

Vrbanec, Tedo, and Ana Meštrović. 2020. "Corpus-Based Paraphrase Detection Experiments and Review." *Information* 11 (5): 241. https://doi.org/10.3390/info11050241.

Walter, Simon J, and David M Studdert. 2015. "Relationship between Penalties for Road Traffic Infringements and Crash Risk in Queensland, Australia: A Case-Crossover Study." *International Journal of Epidemiology* 44 (5): 1722–30. https://doi.org/10.1093/ije/dyv148.

Wang, Chen, Lin Liu, Chengcheng Xu, and Weitao Lv. 2019. "Predicting Future Driving Risk of Crash-Involved Drivers Based on a Systematic Machine Learning Framework." *International Journal of Environmental Research and Public Health* 16 (3). https://doi.org/10.3390/ijerph16030334.

Wang, Kai, and Xiao Qin. 2015. "Exploring Driver Error at Intersections: Key Contributors and Solutions." *Transportation Research Record: Journal of the Transportation Research Board* 2514 (1): 1–9. https://doi.org/10.3141/2514-01.

"Wisconsin State Legislature Home Page." n.d. Accessed May 30, 2020. https://legis.wisconsin.gov/.

Yan, Xuedong, and Essam Radwan. 2006. "Analyses of Rear-End Crashes Based on Classification Tree Models." *Traffic Injury Prevention* 7 (3): 276–82. https://doi.org/10.1080/15389580600660062.

Young, Kristie L., and Paul M. Salmon. 2015. "Sharing the Responsibility for Driver Distraction across Road Transport Systems: A Systems Approach to the Management of Distracted Driving." *Accident Analysis & Prevention* 74 (January): 350–59. https://doi.org/10.1016/j.aap.2014.03.017.

Yu, Rongjie, Yin Zheng, Yong Qin, and Mohamed Abdel-Aty. 2019. "Crash Injury Severity Analyses with Multilevel Thresholds of Change Modelling Approach for At-Fault out-of-State Drivers." *Journal of Transportation Safety & Security*, March, 1–18. https://doi.org/10.1080/19439962.2019.1579290.

Zhang, Guangnan, Kelvin K.W. Yau, and Xun Zhang. 2014. "Analyzing Fault and Severity in Pedestrian–Motor Vehicle Accidents in China." *Accident Analysis & Prevention* 73 (December): 141–50. https://doi.org/10.1016/j.aap.2014.08.018.

Zhang, Hong. 2010. "Identifying and Quantifying Factors Affecting Traffic Crash Severity in Louisiana," 113.

Zhang, Xu, Eric Green, Mei Chen, and Reginald R. (Reg) Souleyrette. 2019. "Identifying Secondary Crashes Using Text Mining Techniques." *Journal of Transportation Safety & Security*, April, 1–21. https://doi.org/10.1080/19439962.2019.1597795.

Zhao, Shanshan, Kai Wang, and Eric Jackson. 2019. "Evaluation of Not-At-Fault Assumption in Quasi-Induced Exposure Method Using Traffic Crash Data at Varied Geographical Levels." *Transportation Research Record: Journal of the Transportation Research Board*, April, 036119811984103. https://doi.org/10.1177/0361198119841036.

Chapter 6 : Evaluating Data Quality for Pedestrian & Bike Crashes

6.1 Background

Traffic safety continues to be considered a major public issue as the number of motor vehicle crash injuries and deaths increase. Based on statistics shared by the World Health Organization (WHO), traffic accidents result in 1.35 million deaths yearly worldwide. Data also showed that 37,133 deaths occurred from motor vehicle crashes in the year 2017 (National Highway Traffic Safety Administration, 2017). With yearly increases in travel and no improvement over the current safety performance, fatalities and injuries could increase by 50 percent by 2020 (NHTSA, 2020). Unfortunately, among these crashes, vehicle crashes involving pedestrians/bicyclists are a source of serious concern as such crashes have a high severity level and a unique inevitability. In 2017 there were 5,977 pedestrians and 783 bicyclists killed in crashes with motor vehicles in the United States. This group of vulnerable road users (VRUs) account for a growing share of total US traffic fatalities. In 2017, pedestrians and bicyclists represented 18.2% of total traffic fatalities. Pedestrian and bicyclist fatalities increased by 32 percent in the ten years between 2009 and 2018.

Data from Wisconsin DMV shows that between the years 2017 and 2018, an average of 879 bicycle crashes occurred, 5 bicyclists were killed, and 792 bicyclists were injured. Moreover, 1519 pedestrian crashes occurred, 56 pedestrians were killed, and 1425 pedestrians were injured on average during the same period. (“Wisconsin DMV Official Government Site - Final Year-End Crash Statistics” 2018).

One of the initiatives to improve traffic safety in Wisconsin was to revise the previous MV4000 crash reporting form into a new DT4000 form.

The forms were switched statewide on January 1, 2017. Some issues with the previous MV4000 crash form included poor reporting of roadway curvature, no data field indicating driver distraction, no specification of the exact traffic barrier, safety equipment used by the individual (motorist and non-motorists), limited information about non-motorist characteristics, and imprecise location of the non-motorists at the time of the crash. The new DT4000 crash form incorporated new crash elements and more detailed attributes. It is clear that the DT4000 crash form captures more details about the crash circumstances. This study examines the value of the newly added data fields in the DT4000 driver crash report form. The question is, in what ways do these new attributes add significant value to the previously used data fields?

There were three major goals for this study: (i) investigate how completely new DT4000 crash form data fields were filled out by law enforcement, with a focus on pedestrian/bicycle-vehicle crash-related data fields to enhance the knowledge about the conditions/circumstances that may have contributed to these crashes; (ii) assess the changes in attributes of the common data fields that are recategorized in the DT4000 crash form to acknowledge the enhancements accompanied in the DT4000 crash form of the common data fields, and (iii) examine if new and recategorized data fields in the DT4000 crash form enhances the pedestrian/bicycle injury severity model accuracy. This information can provide examples of how the new DT4000 crash data can ultimately be applied to make the Wisconsin transportation system safer, especially for pedestrians and bicyclists.

6.2 Data Collection and Processing

In Wisconsin, 284,082 crashes occurred during the two-year (2017-2018) period. The distribution of pedestrians, bicyclists, and drivers involved, are listed based on the injury severity resulted from the crash in **Table 6-1** below.

Table 6-1: Injury severity distribution of road users involved in crashes between 2017-2018

Road User / Injury Severity	Drivers only		At least a Pedestrian		At least a bicyclist		Subtotal	
	N	(%)	N	(%)	N	(%)	N	(%)
K (fatal injury)	898	0.36%	110	0.04%	12	0.00%	1020	0.41%
A (suspected serious injury)	4692	1.88%	458	0.18%	149	0.06%	5299	2.12%
B (suspected minor injury)	22574	9.05%	1080	0.43%	958	0.38%	24612	9.86%
C (possible injury)	26505	10.62%	579	0.23%	392	0.16%	27476	11.01%
O (no apparent injury)	190798	76.47%	141	0.06%	148	0.06%	191087	76.59%
Subtotal	245467	98.39%	2368	0.95%	1659	0.66%	249494	100.00%

Crashes involving pedestrians and bicyclists in Wisconsin have been used for this study. The crash data were downloaded from WisTransportal Crash Retrieval Facility in the old MV4000 crash report format and new DT4000 crash report format. Comparing the same crash data coded in different forms allows us to determine if the new or expanded data elements in DV4000 provide value-added information for gaining a better understanding of pedestrian/bike crashes in Wisconsin. A total of 4,025 crashes were retrieved after applying the following query:

“SELECT * FROM DTCRPRD.SUMMARY_COMBINED C WHERE C.CRSHDATE BETWEEN TO_DATE('2017-JAN','YYYY-MM') AND LAST_DAY(TO_DATE('2018-DEC','YYYY-MM')) AND (C.DEERFLAG IS NULL OR UPPER(C.DEERFLAG) != 'Y') AND (C.BIKEFLAG = 'Y' OR C.PEDFLAG = 'Y') AND C.LOCTYPE IN ('I','N') ORDER BY C.DOCTNMBR”

In a crash involving a non-motorist, the non-motorists can be either unit 1 or 2. According to the crash data user guide, “[1,2] Denotes unit-level information, where a unit is any vehicle, bicycle, pedestrian, or equipment involved in a crash. Unit level element names in the data file are appended with “1” or “2”, representing the first or second unit involved in the crash”. When more than two units are involved including non-motorists, a non-motorist may be coded as neither unit 1 or 2. **Table 6-2** shows how the data fields are processed and analyzed, by referring to the ROLE [1, 2] data field indicating driver, pedestrian, and bicyclist roles in a crash.

Table 6-2: Types of Person Involved in A Crash

ROLE2-DT	ROLE1-DT							
	Bicyclist		Driver		Pedestrian		Total	
	N	%	N	%	N	%	N	%
Bicyclist			1152	31.64%			1152	31.64%
Driver	344	9.45%			121	3.32%	465	12.77%
Pedestrian			2024	55.59%			2024	55.59%
Total	344	9.45%	3176	87.23%	121	3.32%	3641	100.00%

The data shows that:

- The bicyclist involved in a crash with a driver is entered as unit 1 in 344 (9.45%) cases, and as unit 2 in 1152 (31.64%) cases in a total of 1496 (41.08%) bicyclist crashes.

- The driver involved in a crash with non-motorists is entered as unit 1 in 3176 (87.23%) cases and as unit 2 in 465 (12.77%) cases.
- The pedestrian involved in a crash with a driver is entered as unit 1 in 121 (3.32%) cases, and as unit 2 in 2024 (55.59%) cases in a total of 2145 (58.91%) pedestrian crashes.

The results presented in **Table 6-2** showed the importance of analyzing the pedestrian crash dataset separately from the bicyclist crash dataset since there is no obvious rule followed to report a pedestrian, a bicyclist, a driver as unit 1, or unit 2. It might be the fact that a road user (driver, pedestrian, or bicyclist) is entered as unit 1 when he/she is at fault. This study is limited to analyzing the actions and behavior of the first two units one of which is a non-motorist. Also, some crashes involve more than two units and pedestrians/bicyclists were also reported in units other than unit 1 and unit 2. However, in this study the analyzed crashes are pedestrian and bicyclist crashes that include two units only.

Distinguishing the role of unit 1 and unit 2 as a driver or a non-motorist is also essential for us to determine the level of injury severity sustained by a non-motorist during a motor vehicle crash. For any analysis regarding injury severity, the primary interest is in the ones that a non-motorist sustained a more severe injury, which excludes 324 crashes where the motorist sustained more severe injury than the non-motorist and 60 crashes where both units are drivers-vehicle to vehicle crashes.

The selected data fields, their indication, and whether they include more detailed attributes or are a new addition to the DT4000 crash form, are shown in

Table 6-3. Note that (blank) value denotes that the field was left blank/missed and was not filled with any value, whereas; BLANK value denotes that the field involves an option to report a blank value if the field is not related to the situation (i.e., VEHDMG [1, 2] is a field in the

MV4000 crash form dataset which involves several attributes identifying the extent to which the damage affects the vehicles' operability, and a BLNK attribute that can be filled when the vehicle damage was not investigated).

Table 6-3: A List of the Selected Data Fields for the Analysis

MV4000 Crash Form	DT4000 Crash Form	Indication	Description of the Change
Roadway Level			
ROADHOR	ROADHOR [1,2]	Horizontal Road Terrain	More detailed attributes
ROADVERT	ROADVERT [1,2]	Vertical Road Terrain	
ROADCOND	RDCOND [A,B,C]	Road Surface Condition	
TRFCWAY	TRFCWAY [1,2]	Trafficway Description	
RLTNRDWY	RLTNRDWY	Location Of First Harmful Event	
ACCDLOC	LOCTYPE	Crash Location Type	
TRFCNTL [1,2]	TRFCCNTL [1,2]	Traffic Control Device (TCD) In Effect	
---	SURFTYPE [1,2]	Road Surface Type	A New data field
---	TOTLANES [1,2]	Total Number Of Lanes	
---	RLTNTRWY	Crash Location With Respect To Trafficway	
---	INTTYPE	Intersection Type Where The Crash Occurred	
---	TRFCINOP [1,2]	Status Of The TCD	
---	RLTNJNIC	Crash Occurrence Within An Interchange Area	
---	RLTNJNLC	Crash Occurrence In A Junction/Interchange Area	
Environmental Level			
WTHRCOND	WTCOND [A, B]	Prevailing Atmospheric Conditions	More detailed attributes
LGTCOND	LGTCOND	Light Conditions	

HWYPC [1,2]	RDWYPC [A, B, C]	Apparent Factors Of The Road/ Highway	
---	ENVPC[A,B,C]	Contributing Environmental Conditions	A New data field
Driver Level			
DRVRPC [1,2]	DRVRPC [1,2] [A,B,C,D]	Driver Contributing Actions/Circumstances	More detailed attributes
DRVRDO [1,2]	DRVRDOIN [1,2]	Controlled Maneuver By The Driver	
SAFETY [1,2]	SFTYEQP [1, 2]	Safety Equipment Used By The Driver	
---	RACE [1,2]	Driver Race	A New data field
---	TEENDRVR	Teen Driver	
---	DISTFLAG	Distraction/Inattentive Driving Flag	
---	DNMFTR [1,2] [A,B]	Individual Condition Relevant To The Crash	
Pedestrian Level			
NMTACT [1,2] [A,B]	NMTACT [1,2] [A,B]	Pedestrian Actions/Circumstances Contributing To The Crash	More detailed attributes
NMTLOC [1,2]	NMTLOC [1,2]	Pedestrian Location With Respect To The Roadway	
---	NMTSFQ [1,2] [A,B]	Safety Equipment Used By The Pedestrian	A New data field
	DNMFTR [1,2] [A,B]	Individual Condition Relevant To The Crash	
	NMTPRIOR [1,2]	Pedestrian Actions Immediately Prior To The Crash	
Bicyclist Level			
NMTACT [1,2] [A,B]	NMTACT [1,2] [A,B]	Bicyclist Actions/Circumstances Contributing To The Crash	More detailed attributes
NMTLOC [1,2]	NMTLOC [1,2]	Bicyclist Location With Respect To The Roadway	
---	NMTSFQ [1,2] [A,B]	Safety Equipment Used By The Bicyclist	A New data field

	DNMFTR [1,2] [A,B]	Individual Condition Relevant To The Crash	
	NMTPRIOR [1,2]	Bicyclist Actions Immediately Prior To The Crash	
Crash Level			
ACCDTYPE	MOSTHARM [1,2]	Events Resulting In The Most Severe Injury	More detailed attributes
SPEEDFLAG	SPEEDFLAG	Vehicle Speeding Status	
HITRUN	HITRUN	Hit And Run	
---	SCHZONE	School Zone	A New data field
Vehicle Level			
VEHTYPE [1,2]	VEHTYPE [1,2]	Vehicle Type Involved In The Crash	More detailed attributes
VEHDMG [1,2]	VEHDMG [1,2]	Extent Of Vehicle Damage	

Often times, there is more than one crash contributing factor in data fields such as circumstances, driver actions, behavior, so-called multi-valued elements, and denoted as [A, B, C]. [1,2][A,B] denotes combined unit level and multi-valued elements. For example, DRVRPC1A and DRVRPC1B describe the first two contributing factors listed for the driver of the first unit on the DT4000 crash report. Therefore, when necessary, the "*concatenate*" function in Excel is used in the analysis to join data from Unit 1 and 2 -a unit is a driver, a pedestrian, or a bicyclist, as well as from A, B, C, etc. Concatenation is the operation of joining character strings end-to-end and a string can be a text, number, or a Boolean value.

For the same data types (i.e., DRVRDOIN 1 and DRVRDOIN 2), concatenation is done to join the two text strings into one text string (DRVRDOIN 1,2). A filter may be used to filter values of the DRVRDOIN 1 data field when Role 1 is a driver, and the same way when Role 2 is a driver, values of DRVRDOIN 2 data field are filtered, creating (DRVRDOIN 1,2) data field. Whereas, for data fields that take multiple values (i.e., ROADCOND A, ROADCOND B, ROADCOND C), direct concatenation separated by a comma is approached to create (ROADCOND A, B, C) data field. Note that the number of attributes provided in each element varies and is based on the minimum set of data elements recommended by the Model Minimum Uniform Crash Criteria (MMUCC) standard (National Highway Traffic Safety Administration 2017).

After concatenation, many new attribute values may be created due to the combination of strings; thus, attributes with small percentage values (e.g., <1%) are not analyzed separately but as one category (i.e., Total including other combinations). Note that multi-value attributes (i.e., BIKE, UT TRK) are a result of applying the concatenation function to the data field attributes.

6.3 Exploratory Data Analysis

This section presents exploratory data analysis, including the Univariate and Multi-variate analysis of a selected list of MV4000 and DT4000 crash form data fields. The Uni-variate analysis involves all the data fields that are hypothesized to carry useful information and compare informative data fields in both crash forms. The descriptive statistics based on the Uni-variate analysis highlight the attribute values that are overrepresented in the crash data. Subsequently, Multi-variate analysis is carried out following the cross-classification method.

6.3.1 Univariate and Multi-variate Analysis

This section involves preliminary data analysis, including a comprehensiveness comparison between a selected list of MV4000 and DT4000 crash form data fields. This analysis involves all the data fields that are hypothesized to carry useful information and compare informative data fields in both crash forms. Followed by descriptive statistics that highlight the attribute values that are overrepresented in the crash data. Subsequently, supplementary analysis is carried out following the cross-classification method. The percentage of combined values from two or more attributes in the crash data is presented and the dependence and association between different attributes are explored. For the definition of data fields and their corresponding attributes in MV4000 and DT4000 refer to **Appendix C**: **.**

6.3.1.1A Comparison Between MV4000 and DT4000 Crash Form Data Fields

This section analyses and discusses single variables/data fields, or univariate analysis, that are identified as the most relevant fields to pedestrian and bicycle-related crashes and are believed to help with understanding the circumstances associated with such crashes.

Descriptive statistics were generated for the new data fields in the DT4000 form, to check the completion and the distribution of field attributes. In case data fields exist in both MV4000 and DT4000 but with different attributes, attributes were mapped between the two forms, and then separate descriptive statistics for each was generated. In some cases, a contingency table is generated with rows being MV4000 attributes and columns being DT4000 or vice versa. Descriptive statistics are expected to display different distribution and patterns of a data field if the attributes have been changed. The changes will lead to the identification of new and extra specific patterns, conditions, and circumstances contributing to a crash. Also, it includes studying data fields that are roadway, environment, driver, pedestrian, bicyclist, crash, and vehicle-related. **Error! Reference source not found.** illustrates the analysis process of the MV4000 and DT4000 rash forms data fields.

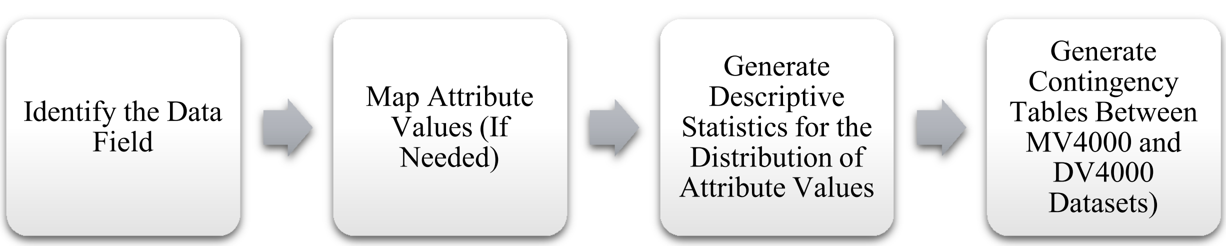


Figure 6-1: Analysis Process of MV4000 and DT4000 Crash Form Data Fields

Observing the comprehensive comparative analysis of the selected crash variables, the potential significant data fields are discussed below.

For road surface conditions, 13.24% of the crashes occurred in wet roadway surface conditions. Viewing the type of trafficway division, 68.14%, and 9.59% of the crashes occurred in two-way undivided highways (UNDIV), and divided highways without a traffic barrier (DIV NO), respectively. Concerning the total number of lanes in a roadway where a crash took place, roadways with two and four total number of lanes contribute together to 80.31% of the total number of crashes.

Reporting the location type in MV4000 and DT4000 crash forms is consistent. 56.33% of crashes occurred at intersections, whereas, 43.67% of them occurred at non-intersection/midblock locations. Taking into account the intersection type in which the crash occurred, 41.77% and 11.73% of the crashes occurred at 4-way stop intersections and T intersections, respectively. For the traffic control device (TCD) in effect at the time of the crash, 16.11% ((59.78% locations lacking a TCD (NONE) excluding 43.67% non-intersection locations (N)) of the crashes are reported as lacking a TCD (NONE). 23.85% and 12.55% of the crashes occurred at traffic signal controlled (TS OP) and at stop sign-controlled (STOP/SS) locations, respectively.

6.3.1.2 Enhanced analysis of selected MV & DT variables

This section focuses on the multi-variate analysis (MVA) or so-called correspondence modeling, which is used to account for confounding effects and detect the presence of significant association through allowing for the association between two or more categorical data fields. Selected attributes of each data field are listed for each data field as rows and columns, and the joint cell of each row and column value is tested. The lower right-hand corner value contains the sum of either the row or column marginal frequencies, which both must be equal to N.

This type of analysis can be used to test hypothetical frameworks to decide whether or not effects are present and can be further used to create new data fields based on useful relationships defined between row and column data fields. This section analyses and discusses a selection of two or more data fields presented in a 2X3 or more format adopted from the MV4000 and DT4000 crash forms, that were found useful - during the comprehensiveness comparison analysis process - for understanding the circumstances associated with vehicle crashes involving pedestrians and bicyclists. The following relationships are investigated using data fields that were found to provide constructive evidence of an effect on crashes through the Uni-variate analysis.

6.3.1.2.1 Action-Location Relationships

This relationship may reveal the common actions of drivers and how they are affected by the pedestrians' location. The non-motorist location affects the decision-making process of the driver, which influences the drivers' actions. Many researchers studied pedestrian location, and driver actions separately, however, the aim is testing if a certain non-motorist location accompanied with a certain driver action increases the odds of a more severely injured non-motorist involved in a vehicle crash (Mitman, Ragland, and Zegeer 2008; Schneider and Stefanich 2016; Kemnitzer et al. 2019). (C. Lee and Abdel-Aty 2005) studied drivers' behaviors and proposed more intensive driver education and restrictive traffic regulations targeted at middle-aged male drivers in an intention to reduce pedestrian crashes. Adding to that, as the authors examined pedestrian behaviors and location, they highlighted the value of improving pedestrian designated areas for travel, as a result of recognizing that a high percentage of crashes occur at marked and unmarked crosswalks.

Referring to **Table 6-4** to **Table 6-9**, both behavior and non-behavior related data fields were studied. Considering pedestrians' actions, darting into roads seen to be pedestrians' most common action before the crash. Hence, education messages specifically designed for pedestrians may be proposed to minimize such behavioral actions. The studied variable combinations shown in this section may be categorized as i) non-behavior (e.g., driver movement and pedestrian/bicyclist location); and ii) behavior and non-behavior (e.g., pedestrian/bicyclist action with pedestrian/bicyclist location). Finding driver movements associated with pedestrian/bicyclist location may be addressed by restricting a specific turn, adding dedicated left-turn signal phases, enhancing intersection lighting, and reconfiguring lanes.

6.3.1.2.2 Roadway Characteristics Relationships

Studying the separate effect of roadway characteristics on VRUs' injury severity is common among researchers (i.e., Dong et al. 2019; Chen and Fan 2018). Roadway characteristics (such as interstate, junction, and roadway profile), and environmental characteristics (such as light condition and weather condition) have significant effects on the injury severities of VRU involved crashes (Dong et al. 2019). However, studying the effect of a combination of roadway characteristics may enhance the knowledge about the crash circumstances.

Table 6-10 to

Table 6-14 shows multiple studied combinations of roadway characteristics, which help in recommending different actions to help minimize their effect on crashes. For instance, findings linked to poor roadway lighting provide an opportunity to suggest different roadway visibility enhancement actions or installing tools that provide drivers with a better vision that help in recognizing pedestrians and bicyclists at different roadway locations.

6.3.1.2.3 Driver Actions-Roadway Characteristics Relationships

Researchers have attempted to explain drivers' actions/behavior at signalized crosswalks as roadway-related characteristics (i.e., Hunter, Srinivasan, and Martell 2012; Kutela and Teng 2019). Hence, it pleads for another question to be asked about how driver action may be affected by different roadway characteristics. **Table 6-15** to

Table 6-17, show drivers' contributing actions associated with crashes occurred within and at a specific location of an interchange area. Roadway characteristics can be a source of information for studying drivers' actions at the time of the crash.

Crashes occurring at small exit angles indicate driver fatigue issues (Wootton and Spainhour 2004). Downgrades and two-way divided roads can be associated with drivers tend to speed, and drivers are likely to take less effective avoidance maneuvers when driving in dark roadways with and without streetlights. (J.-K. Kim et al. 2008). Significant increase in the frequency of looking for pedestrians when they encountered advance yield markings (Fisher and Garay-Vega 2012). A very strong inverse correlation, with low yield rates on high-speed roadways, suggests that drivers, tend not to give pedestrians' right-of-way on marked crosswalks (Bertulis and Dulaski 2014).

6.3.1.2.4 Driver Actions-Pedestrian/Bicyclist Actions Relationships

An analysis of driver and pedestrian actions in pedestrian crashes also found that the most pr evalent combination of driver and pedestrian maneuvers in both fatal and injury crashes is Drivin g Straight Ahead (driver action) and Crossing Not at Intersection (pedestrian action).

Making a left

turn (driver action) and Crossing at the Intersection (pedestrian action) was the second most com mon combination of driver and pedestrian action("VDOT_Pedestrians_Crash_Assessment_2014-2018.Pdf" n.d.; Sheykhfard and Haghighi 2019). Hence, these findings support the need to study driver-non-motorist actions related to the crash.

Table 6-18 and

Table 6-21, show the relationship between driver condition and pedestrian/bicyclist action, as well as, driver movement relationship with pedestrian/bicyclist actions. Drivers tend to slow down when pedestrians are not looking at the approaching drivers, also drivers were also found to stop more often when approach velocity was low (Katz, Zaidel, and Elgrishi 1975).

Drivers' behavior in proximity of pedestrians is likely to be statistically significantly less aggressive when the approach velocity is lower, curbside parking is not allowed, when a crosswalk exists, and when the street involves a higher number of pedestrians crossing (Obeid et al. 2017). Hence, the aim is to study the driver-pedestrian interaction to propose and evaluate safety measures and traffic calming techniques.

6.3.1.2.5 Pedestrian/Bicyclist Location-Roadway Characteristics Relationships

Roadway characteristics and their association with pedestrian/bicycle-vehicle crashes have long been studied by researchers (i.e., (Schneider et al. 2010; Schneider, Grembek, and Braughton

2013; Kim 2019; Morrison et al. 2019). The use of knowledge on the non-motorists location at the time of the crash in the company of different roadway characteristics may enrich the investigation of pedestrian/bicycle-vehicle crashes.

Table ***6-22*** and

Table ***6-23*** examine pedestrian and bicyclist locations at different types of intersections. Engineering decisions can be informed by the non-motorists location. For instance, if crashes occurred not at crosswalk locations at a 4-way intersection, road marking and signs could be added at that location. Also, education messages to motorists, pedestrians, and bicyclists can emphasize looking for pedestrians located in the minor road which is one of the intersection arms before making a left turn. Other education messages to motorists, can be focused on yielding the ROW when located in the minor road of a 4-way intersection.

6.3.1.2.6 Pedestrian/Bicyclist Action-Roadway Characteristics Relationships

To effectively reduce pedestrian/bicycle-vehicle crashes and improve non-motorist safety, it is important to identify why, where, and how these crashes occurred. Hence, the work of past researchers in identifying non-motorist actions at the time of the crash (i.e., (Brewer et al. 2006; Shi et al. 2007; Sun, Sun, and Shan 2019; Pelé, Deneubourg, and Sueur 2019; Yue et al. 2020) was followed, which is influenced by roadway characteristics. The pedestrian speed at an unsignalized crosswalk in a roadway is higher than that of a signalized crossing because the former conflicts are more intense (Shi et al. 2007).

and **Error! Reference source not found.**, show pedestrian and bicyclist actions associated with the TCD type.

Engineering decisions can be applied after examining pedestrian and bicyclist actions associated with specific roadway characteristics. For instance, if crash locations without any TCD, showed that pedestrians or bicyclists suddenly moved in the roadway, median island, markings, or a TCD could be added at such locations.

6.3.1.2.7 Environmental Conditions - Roadway Characteristics Relationships

Fountas and colleagues concluded that the effect of lighting characteristics on driving behavior depends on other environmental factors, in particular weather conditions. Also, the authors stated that it should be noted that the most pronounced effect of the pedestrian involvement indicator on serious and fatal injuries is identified in the model reflecting darkness and poor weather on unlighted roadways, whereas the least pronounced effect is observed in the model reflecting daylight and poor weather (Fountas et al. 2020).

Per the collected data, a non-negligible percentage of pedestrian/bicycle crashes occurred under certain weather and light conditions (roadway-related condition). Also, previous research studied the relationship between crashes and environmental conditions associated with specific roadway characteristics.

For instance, the effect of road shoulder and weather conditions on crashes was studied by (Kordani, Shirini, and Yazdani 2019). **Table 6-24** to

Table ***6-26*** shows multiple relationships, from prevailing atmospheric conditions-type/level of light to prevailing atmospheric conditions-trafficway division, and lastly road surface type with the condition. Discovering that bicyclist crashes are more likely to occur in rainy weather conditions on undivided roadways can lead to roadway division modifications.

Table 6-4: Driver Movement and Pedestrian Location

DRVRDO [1, 2]-MV	**NMTLOC [1, 2]-PED-MV**							
	(1) in crosswalk		(2) in roadway		(blank)		Total, including other location combination	
	N	%	N	%	N	%	N	%
GO STR	**176**	**4.83%**	**483**	**13.27 %**	560	15.38%	1291	35.46%
LT TRN	**300**	**8.24%**	71	1.95%	195	5.36%	**579**	**15.90%**
RT TRN	**144**	**3.95%**	35	0.96%	295	8.10%	481	13.21%
Total, including other driver action combinations	895	24.58%	1005	27.60%	1529	41.99%	3641	100.00%

DRVRDOIN [1, 2]-DT	**NMTLOC [1, 2]-PED-DT**							
	ATI MX		NAI NX		(blank)		Total, including other location combination	
	N	%	N	%	N	%	N	%
GO STR	**176**	**4.83%**	**369**	**10.14 %**	521	14.31%	1271	34.91%
LT TRN	**300**	**8.24%**			193	5.30%	**579**	**15.90%**
RT TRN	**144**	**3.95%**			293	8.05 %	497	13.16%
Total, including other driver action combinations	895	24.58%	663	18.21%	**1496**	**41.09 %**	3641	100.00%

DT4000 form dataset is more detailed in terms of addressing the non-motorists location (i.e., crosswalk marking associated with the intersection). **Table 6-4** showed consistent information between both crash forms. For instance, a total of 15.90% of the crashes occurred while the driver was making a left turn and the pedestrian is in a crosswalk, and (4.83%, 8.24%, and 3.95%) of crashes occurred in a crosswalk, were associated with the driver going straight, taking a left turn, and taking right turn, respectively.

MV4000 database showed that (4.83%) of crashes occurred while the driver was going straight and the non-motorists located in a crosswalk, regardless of the crosswalk marking status. Whereas the DT4000 dataset showed that this percentage of these crashes occurred while the driver was going straight, and the pedestrian located at an intersection with marked crosswalk. According to the State of Wisconsin's law (Wisconsin State, 2020), the driver(s) must yield the right-of-way (ROW) to the pedestrian(s) who already started crossing an intersection on a "walk" signal or a green light if there is no walk signal.

Also, the driver(s) must yield the ROW to the pedestrian(s) who started crossing within a marked/unmarked crosswalk at an intersection where there are no traffic lights or traffic control signals. The MV4000 dataset showed that 13.27% of the crashes occurred while the driver was going straight, and the pedestrian was located in the roadway/at a midblock. Though, the DT4000 dataset showed that 10.14% of these crashes involved pedestrians located on the roadway, not in a marked crosswalk. The percentage of crashes occurred in the roadway (based on MV4000) was recategorized under different "in roadway" categories in (DT4000 dataset); (Not At Intersection-On Roadway, Not In Marked Crosswalk, NAI NX), (Not At Intersection-On Roadway, Crosswalk Availability Unknown, NAI UN), and (Not At Intersection-In Marked Crosswalk, NAI MX).

Table 6-5: Driver Movement and Bicyclist Location

DRVRDO [1, 2]-MV	NMTLOC [1, 2]-BIKE-MV							
	(1) in crosswalk		(2) in roadway		(blank)		Total, including other bicyclist locations	
	N	%	N	%			N	%
GO STR	251	6.90%	**471**	**12.97%**	454	12.57%	1291	35.61%
LT TRN	126	3.47%	151	4.15%	233	6.49%	579	16.01%
RT TRN	**208**	**5.73%**	113	3.12%	113	3.14%	481	13.28%
Total, including other driver action combinations	870	23.82%	1096	30.05%	1260	34.75%	3641	100.00%

DRVRDOIN [1, 2]-DT	NMTLOC [1, 2]-BIKE-DT									
	ATI MX		ATI NX		NAI NX		(blank)		Total, including other bicyclist locations	
	N	%	N	%	N	%	N	%	N	%
GO STR	125	3.76%	137	3.76%	**120**	**3.30%**	750	20.60%	1271	34.90%
LT TRN	67	1.84%					386	10.60%	579	15.90%
RT TRN	**179**	**4.91%**	28	0.77%	9	0.24%	186	5.11%	479	13.16%
Total, including other driver action combinations	525	14.42%	279	7.66%	220	6.04%	**2145**	**58.91%**	3641	100.00%

5.73% of the crashes occurred while drivers were taking a right turn (RT TRN), and while the bicyclist is in a crosswalk (1) regardless if the crosswalk is marked or not. The DT4000 form showed that 4.91% of this percentage of crashes occurred while the bicyclist was located at an intersection and in a marked crosswalk (ATI MX). Referring to Wisconsin Statute 346.23, at a controlled intersection with a "stop" signal, the driver must yield the ROW to the bicyclist(s) crossing at a crosswalk when the bicyclist(s) has started crossing the crosswalk with a green light or a "walk" signal.

12.97% of the crashes occurred while the driver was going straight (GO STR), and the bicyclist was crossing the roadway from the right where there is no intersection (2). The DT4000 form specified that 3.30% of this percentage of crashes occurred while the bicyclist was not located in an intersection and was not in a marked crosswalk (NAI NX). Regarding the ROW, according to Wisconsin Statute (Wisconsin State, 2020) bicycles operate under the same laws as other legal vehicles on the road. The percentage of crashes occurred in the roadway (based on MV4000) was recategorized under different "in roadway" categories in (DT4000 dataset); (Not At Intersection-On Roadway, Not In Marked Crosswalk, NAI NX), (Not At Intersection-On Roadway, Crosswalk Availability Unknown, NAI UN), and (Not At Intersection-In Marked Crosswalk, NAI MX).

Table 6-6: Pedestrian Action and Pedestrian Location

NMTACT [1, 2]-PED-MV	**NMTLOC [1, 2]-PED-MV**					
	(1) in crosswalk		(2) in roadway		TOTAL Including other pedestrian locations	
	N	%	N	%	N	%
Actions other than walking facing/not facing traffic, disregarding signal, darting in the road, and wearing dark clothes (6)	58	1.61%	79	2.17%	262	7.24%
Darting into roadway (3)	277	7.63%	**289**	**7.96%**	1108	30.50%
TOTAL Including other pedestrian action combinations	**895**	**24.58%**	1005	27.62%	3641	100.00%

NMTACT [1, 2]-PED-DT	**NMTLOC [1, 2]-PED-DT**					
	ATI MX		NAI NX		TOTAL Including other pedestrian locations	
	N	%	N	%	N	%
DISREG	50	1.37%			58	1.59%
IM XING			72	1.98%	99	2.72%
NF TRFC	75	2.06%			155	4.26%
NO IMPR	453	12.44%			969	19.12%
SUDDEN			**174**	**4.78%**	270	7.42%
TOTAL Including other pedestrian action combinations	**895**	**24.58%**	663	18.21%	3641	100.00%

The DT4000 form showed more specific information regarding the pedestrian location; (24.58%) of crashes occurred at intersections specifies that the pedestrian was located in marked crosswalks (ATI MX). Whereas, the same percentage was presented in the MV4000 form as crashes occurring at crosswalks (1), regardless of the marking status. The DT4000 form lays down an option for no improper action by the pedestrian (NO IMPR) which is more useful than a blank field in the MV4000 field which may mean that the crash does not involve pedestrians or indicate that there is no specific action to be reported.

Table 6-7: Bicyclist Action and Bicyclist Location

NMTACT [1, 2]-BIKE-MV	NMTLOC [1, 2]-BIKE-MV							
	(1) in crosswalk		(2) in roadway		(3) not in roadway		Total, including other bicyclist locations	
	N	%	N	%	N	%	N	%
Actions other than walking facing/not facing traffic, disregarding signal, darting in the road, and wearing dark clothes (6)	**364**	**10.00%**	**469**	**12.88%**	167	4.59%	**1056**	**29.00%**
Darting into roadway (3)	60	1.65%	87	2.39%	17	0.47%	**172**	**4.72%**
Total, including other bicyclist action combinations	870	23.89%	1096	30.10%	295	8.10%	3641	100.00%

NMTACT [1, 2]-BIKE-DT	NMTLOC [1, 2]-BIKE-DT									
	ATI MX		ATI NX		ATI UM		NAI NX		Total, including other bicyclist locations	
	N	%	N	%	N	%	N	%	N	%
NO IMPR	220	6.04%	68	1.87%	62	1.70%	72	1.98%	579	15.90%
SUDDEN	**60**	**1.65%**							**164**	**4.50%**
TOTAL Including other bicyclist action combinations	525	14.42%	279	7.66%	160	4.39%	220	6.04%	3641	100.00%

The MV4000 form presents that within 29.00% of the crashes, bicyclists involved in these crashes acted with improper actions other than darting into road, disregarding signal, or walking facing/not facing traffic. Such actions involve but are not limited to wearing dark clothes (DK CLTH), crossing improperly/jaywalking (IM XING), failed to yield (F YIELD), and passing improperly (IM PASS). Also, it shows that 4.72% of the crashes involved bicyclists who darted in the roadway (3). In 12.88% and 10.00% of these crashes, bicyclists were located in the roadway and the crosswalk, respectively. As discussed previously, the DT4000 form enhanced the location and action data fields.

Table 6-8: Pedestrian Prior Action and Pedestrian Location

NMTLOC [1, 2]-PED-DT	NMTPRIOR [1, 2]-PED-DT							
	RDWY OT		WAITING		XING		Total, including other pedestrian action combinations	
	N	%	N	%	N	%	N	%
ATI MX			81	2.22%	**760**	**20.87%**	895	24.58%
ATI NX					93	2.55%	156	4.28%
ATI UM					115	3.16%	146	4.01%
NAI NX	129	3.54%			**315**	**8.65%**	663	18.21%
Total, including other pedestrian location combinations	195	5.36%			1352	37.13%	3641	100%

In 20.87% of the crashes, the pedestrian was crossing the roadway while located at an intersection with a marked crosswalk (ATI MX). Whereas in 8.65% of all crashes, the pedestrian was not located at an intersection, in a roadway but not in a marked crosswalk (NAI NX). Totally, 37.13% of the crashes involved pedestrians crossing the roadway at different locations.

Table 6-9: Bicyclist Prior Action and Bicyclist Location

NMTLOC [1, 2]-BIKE-DT	NMTPRIOR [1, 2]-BIKE-DT							
	SIDE WK		W TRFC		XING		Total, including other bicyclist action combinations	
	N	%	N	%	N	%	N	%
ATI MX	110	3.02%			**337**	**9.26%**	**525**	**14.42%**
ATI NX			87	2.39%	98	2.69%	**279**	**7.66%**
ATI UM					83	2.28%	160	4.39%
NAI NX			82	2.25%			220	6.04%
Total, including other bicyclist location combinations	248	6.81%	306	8.40%	612	16.81%	3641	100%

Regarding bicyclist's locations and actions prior to the crash;

- 14.42% of the crashes involved bicyclists located at intersections in marked crosswalks (ATI MX), were 9.26% of this percentage was crossing the roadway (XING),
- 7.66% of crashes occurred while bicyclists were crossing at intersections but no in crosswalks (ATI NX),

Table 6-10: Whether a Crash Occurred Within an Interchange/Junction Area and the Specific Location

RLTNJNLC-DT	**RLTNJNIC-DT**					
	N		Y		TOTAL including unknown and blank values for if the crash occurred in an interchange area	
	N	%	N	%	N	%
INR	343	9.42%	61	1.68%	405	11.12%
INT	**1341**	**36.83%**	297	8.16%	1657	45.51%
NJ	**1379**	**37.87%**	28	0.77%	1417	38.92%
TOTAL including other location areas related to a junction/interchange	**3183**	**87.42%**	411	11.29%	3641	100.00%

The first harmful event leading to crashes occurring at non-interchange areas-or, not interchange related- add up to 87.42%, where 36.83% and 37.87% of these crashes located at intersections and non-junction locations. A non-interchange-related crash means that the location of the crash was not next to an interchange and did not result from an action related to the movement of traffic units through an interchange.

Table 6-11: Types of TCD and Intersection in the Roadway

TRFCCNTL [1, 2]-MV	**INTTYPE-DT**					
	4 WAY		T		Total, including other intersection type and total lane combinations	
	N	%	N	%	N	%
NONE	239	6.56%	149	4.09%	1629	44.74%
SS	198	5.44%	90	2.47%	342	9.39%
TS OP	792	21.75%			1016	27.90%
Total, including other TCD type combinations	**1521**	**41.77%**	427	11.73%	3641	100.00%

TRFCCNTL [1, 2]-DT	**INTTYPE-DT**					
	4 WAY		T		Total, including other intersection type and total lane combinations	
	N	%	N	%	N	%
NONE	239	6.56%	149	4.09%	1629	44.75%
STOP	198	5.44%			342	9.39%
TS OP	**792**	**21.75%**			1016	27.90%
Total, including other TCD type combinations	**1521**	**41.77%**	427	11.73%	3641	100.00%

Knowing that 56.33% of crashes occurred at intersection locations (refer to **Error! Reference source not found.**) and looking at the type of TCD at the intersection along with the intersection type, completes that information gathered about these intersection-related crashes. 41.77% of these crashes occurred at 4-way intersections. Among these 41.77% crashes, 21.75% occurred at 4-way traffic signal-controlled intersections. Consistent information may be gathered from the MV4000 data fields. the data shows that studying crashes at 4-way traffic signal-controlled intersections may provide useful information regarding intersection-related crashes.

Table 6-12: Type of TCD, Intersection Type, and Total Number of Lanes

TRFCCNTL [1, 2]-DT	INTTYPE-DT, TOTLANES [1, 2]-DT							
	4 WAY, 2		4 WAY, 4		T, 2		Total, including other intersection type and total lane combinations	
	N	%	N	%	N	%	N	%
NONE	153	4.34%			107	2.94%	1629	44.75%
STOP	156	4.28%					333	9.14%
TS OP	**322**	**8.84%**	245	6.73%			982	26.97%
Total, including other TCD type combinations	846	23.24%	349	9.59%	278	7.64%	3641	100.00%

Investigating the number of lanes with the intersection type, in conjunction with the type of TCD available at the intersection offers extra information about the intersection environment. The relationship indicates that 8.84% of crashes occurred at 4-way, two-lane, traffic signal-controlled intersections.

Table 6-13: Types of Intersection and TCD, and Total Number of Lanes in the Roadway

INTTYPE-DT & TRFCCNTL [1, 2]-DT	TOTLANES [1, 2]-DT					
	2		4		Total, including other total lane combinations	
	N	%	N	%	N	%
4 WAY, STOP	158	4.34%			191	5.24%
4 WAY, TS OP	**350**	**9.61%**	**270**	**7.42%**	763	20.96%
Total, including other intersection type and TCD type combinations	2289	62.87%	686	18.84%	3641	100.00%

Investigating the number of lanes with the intersection type, in conjunction with the type of TCD available at the intersection offers extra information about the intersection environment. The above relationship indicates that 7.42% and 9.61% of crashes occurred at 4-way (4-WAY), traffic signal-controlled intersections (TS OP) with four-lane (4) and two-lane (2) roadways, respectively.

Table 6-14: Roadway Curvature and Grade in the Direction of Vehicle Travel

ROADVERT [1, 2]-MV	**ROADHOR [1, 2]-MV**					
	C		(blank)		Total	
	N	%	N	%	N	%
H	32	0.88%	269	7.39%	301	8.27%
(blank)	109	2.99%	3231	88.74%	3340	91.73%
Total	141	3.87%	3500	96.13%	3641	100.00%

ROADVERT [1, 2]-DT	**ROADHOR [1, 2]-DT**			
	ST		Total, including other horizontal road terrain combinations	
	N	%	N	%
LVL	**3035**	**83.36%**	3151	86.55%
DN	**88**	**2.42%**	100	2.75%
UP	**99**	**2.06%**	82	2.25%
TOTAL including other vertical road terrain combinations	**3222**	**87.84%**	3641	100.00%

DT4000 form data fields show that the vast majority (83.36%) of crashes occurred on straight (ST) and level (LVL) roads in the travel direction of the vehicle involved in the crash. Also, the rest of the crashes occurring on straight roadways (ST) were almost equally distributed between uphill/upgrade (UP) and downhill/downgrade (DN); 2.06% and 2.42%, respectively. Crashes on curves comprised slightly more than 10% (all crashes excluding crashes on straight roadway curvature) of all crashes. Clearly, the new attributes enhance the knowledge about the type of curve, instead of the roadway curvature data field in MV4000 form which generally states that a curve is identified at the crash location (c).

Table 6-15: Driver Contributing Actions and Intersection Type

DRVRPC [1, 2] [A, B, C, D]-MV	**INTTYPE-DT**					
	4 WAY		T		Total, including other intersection type combinations	
	N	%	N	%	N	%
BLANK	775	**21.29%**	196	5.38%	2035	55.89%
FTY	**547**	**15.02%**	**159**	**4.37%**	963	26.45%
Total, including other driver action combinations	1521	41.77%	427	11.73%	3641	100.00%

DRVRPC [1, 2] [A, B, C, D]-DT	**INTTYPE-DT**					
	4 WAY		T		Total, including other intersection type combinations	
	N	%	N	%	N	%
NO	558	15.32%	147	4.03%	1462	40.15%
FTY	**527**	**14.45%**	**151**	**4.14%**	922	25.29%
Total, including other driver action combinations	1521	41.76%	427	11.71%	3641	100.00%

Although attributes of the driver contributing circumstances have been enhanced in the DT4000 form such as the direction of the improper overtaking, the data showed for the most common two driver circumstances were consistent between MV4000 and DT4000 forms. Around 25% of crashes involved a driver who failed to yield the ROW (FTY), and around 4.00% and 14.00% of these crashes occurred at T intersections (T) and 4-way intersections (4 WAY), respectively. Generally, a driver coming to a 4-way stop without traffic signal control must yield the ROW to the person on the right.

Table 6-16: Driver Contributing Actions and Whether A Crash Occurred within an Interchange/Junction Area

DRVRPC [1, 2] [A, B, C, D]-MV	**RLTNJNIC-DT**					
	N		Y		Total	
	N	%	N	%	N	%
BLANK	1814	49.82%	195	5.36%	2035	55.89%
FTY	**798**	**21.92%**	151	4.15%	963	26.45%
ID	104	2.86%			114	3.13%
OTHR	119	3.27%			132	3.63%
Total, including other driver action combinations	3183	87.42%	411	11.29%	3641	100%

DRVRPC [1, 2] [A, B, C, D]-DT	**RLTNJNIC-DT**					
	N		Y		Total	
	N	%	N	%	N	%
FTY	**759**	**20.83%**	150	4.11%	922	25.29%
NO	1305	35.84%	140	3.84%	1462	40.15%
OTR	87	2.04%	6	0.17%	97	2.67%
ID	97	2.67%	10	0.27%	107	2.94%
Total, including other driver action combinations	3183	87.45%	411	11.27%	3641	100%

An interesting variable to study is presented consistently between MV4000 and DT4000 form. 20.83% and 21.92% of crashes occurred away from an interchange-related area (N) where the driver failed to yield the ROW (FTY), in DT4000 and MV4000 forms respectively.

Table 6-17: Driver Contributing Actions and the Specific Location within an Interchange/Junction Area

DRVRPC [1, 2] [A, B, C, D]-MV	**RLTNJNLC-DT**							
	INR		INT		NJ		Total, including other junction /interchange location combinations	
	N	%	N	%	N	%	N	%
BLANK	**211**	**5.80%**	**822**	**22.58%**	**930**	**25.54%**	2035	55.89%
FTY	**149**	**4.09%**	**675**	**18.54%**	**182**	**5.00%**	1070	29.38%
ID					63	1.73%	126	3.46%
OTHER					88	2.41%	148	4.07%
Total, including other driver action combinations	405	11.12%	1657	45.51%	1417	38.92%	3641	100%

DRVRPC [1, 2] [A, B, C, D]-DT	**RLTNJNLC-DT**							
	INR		INT		NJ		Total, including other junction /interchange location combinations	
	N	%	N	%	N	%	N	%
NO	**156**	**4.28%**	**596**	**7.41%**	**660**	**18.13%**	1462	40.15%
FTY	**132**	**3.62%**	**581**	**15.96%**	**152**	**4.17%**	922	25.33%
Total, including other driver action combinations	405	11.12%	1657	45.51%	1417	38.92%	3641	100%

Interesting variables to study are presented approximately consistently between MV4000 and DT4000 form. Around 5.00%, 16.00%, and 4.00% of crashes occurred at non-junction (NJ), an intersection (INT), and in an intersection-related location (INR), while the driver failed to yield the ROW (FTY), respectively. The values shown in the table showed consistent pattern regarding the location of the first harmful event leading to the crash while the driver did not have any contributing action that may have contributed to the crash (BLANK/NO).

Table 6-18: Driver Condition and Pedestrian Action

DRVRPC [1, 2] [A, B, C, D]-MV	**NMTACT [1, 2] [A, B]-PED-MV**					
	6		BLANK		Total, including other pedestrian action combinations	
	N	%	N	%	N	%
BLANK	**636**	**17.47%**	**894**	**24.55%**	2035	55.89%
FTY	301	8.27%	408	11.21%	963	26.45%
Total, including other driver action combinations	1155	31.72%	1574	43.23%	3641	100.00%

DRVRPC [1, 2] [A, B, C, D]-DT	**NMTACT [1, 2] [A, B]-PED-DT**											
	DISREG		IM XING		NO IMPR		SUDDEN		NF TRFC		Total, including other pedestrian action combinations	
	N	%	N	%	N	%	N	%	N	%	N	%
NO	48	1.32%	44	1.20%	61	1.67%	227	6.23%	42	1.16%	1462	40.15%
FTY					**330**	**9.07%**			47	1.29%	922	25.33%
Total, including other driver action combinations	58	1.59%	99	2.72%	696	19.12%	270	7.42%	155	4.26%	3641	100.00%

The MV4000 form displays that 24.55% and 17.47% of crashes involved a pedestrian that acted in different actions other than darting into road/disregarding traffic signal/walking facing/not facing traffic (6), and a pedestrian with no exact contributing action (BLANK), respectively. however, this information is not useful since it doesn't provide a specific action by the driver and the pedestrian that might be affecting the crash occurrence. Whereas, the DT4000 form data field shows that is more thorough in terms of pedestrian and driver actions. 9.07% of crashes didn't include any pedestrian action while the driver didn't yield the ROW (FTY).

Table 6-19: Driver condition and bicyclist action

DRVRPC [1, 2] [A, B, C, D]-MV	NMTACT [1, 2] [A, B]-BIKE-MV					
	6		BLANK		Total, including other bicyclist action combinations	
	N	%	N	%	N	%
BLANK	**599**	**16.45%**	**1184**	**32.52%**	2035	55.89%
FTY	272	7.47%	578	15.87%	963	26.45%
Total, including other driver action combinations	1056	29.00%	2150	59.05%	3641	100.00%

DRVRPC [1, 2] [A, B, C, D]-DT	NMTACT [1, 2] [A, B]-BIKE-DT					
	NO IMPR		SUDDEN		Total, including other bicyclist action combinations	
	N	%	N	%	N	%
NO	76	2.09%	128	3.51%	1462	40.15%
FTY	**302**	**8.30%**			922	25.33%
Total, including other driver action combinations	579	15.90%	164	4.50%	3641	100.00%

With the bicycle-vehicle crashes, the same pattern is spotted. The MV4000 form displays that 16.45% and 32.52% of crashes involved a bicyclist that acted in different actions other than darting into road/disregarding traffic signal/walking facing/not facing traffic (6), and a bicyclist with no exact contributing action (BLANK), respectively. However, this information is not useful since it doesn't provide a specific action by the driver and the pedestrian that might be affecting the crash occurrence. Whereas, the DT4000 form data field shows that is more thorough in terms of pedestrian and driver actions. 8.30% of crashes didn't include any bicyclist action while the driver didn't yield the ROW (FTY).

Table 6-20: Driver Movement and Pedestrian Action

DRVRDO [1, 2]-MV	NMTACT [1, 2] [A, B]-PED-MV					
	3		6		Total, including other pedestrian action combinations	
	N	%	N	%	N	%
GO STR	89	2.44%	402	11.04%	1291	35.46%
LT TRN	42	1.15%	177	4.86%	579	15.90%
RT TRN	32	0.88%	154	4.23%	481	13.21%
Total, including other driver action combinations	262	7.20%	**1108**	**30.43%**	3641	100.00%

DRVRDOIN [1, 2]-DT	NMTACT [1, 2] [A, B]-PED-DT					
	NO IMPR		SUDDEN		Total, including other pedestrian action combinations	
	N	%	N	%	N	%
GO STR	118	3.24%	**175**	**4.81%**	1284	35.26%
LT TRN	**207**	**5.69%**			579	15.90%
RT TRN	86	2.36%			422	11.59%
Total, including other driver action combinations	696	19.12%	**270**	**7.42%**	3641	100.00%

Actions of non-motorists together with the driver's maneuver prior to the beginning of the sequence of crash events are studied. The DT4000 form shows that more than a quarter of the crashes involving no improper action by the pedestrian, occurred while the driver was making a left turn (NO IMPR-LT TRN, 5.69%). Whereas more than half of the crashes that involved a pedestrian who darted/suddenly moved into the roadway (7.42%), occurred while the driver was going straight and not turning a left/right turn (SUDDEN-GO STR, 4.81%). The MV4000 form presents that 30.43% of crashes reported that the pedestrian acted in different actions (6), other than darting into road, walking facing/not facing traffic, wearing dark clothes, and disregarding a traffic signal. Such actions are improperly standing/working in the roadway, failing to have lights on while walking, wrong way walking, etc.

Table 6-21: Driver Movement and Bicyclist Action

DRVRDO [1, 2]-MV	NMTACT [1, 2] [A, B]-BIKE-MV					
	6		3		Total, including other bicyclist action combinations	
	N	%	N	%	N	%
GO STR	347	9.53%	57	1.57%	1291	35.46%
LT TRN	163	4.48%			579	15.90%
RT TRN	134	3.68%			481	13.21%
Total, including other driver action combinations	**1056**	**29.00%**	172	4.72%	3641	100.00%

DRVRDOIN [1, 2]-DT	NMTACT [1, 2] [A, B]-BIKE-DT					
	NO IMPR		SUDDEN		Total, including other bicyclist action combinations	
	N	%	N	%	N	%
GO STR	**119**	**3.26%**	86	2.36%	1284	35.26%
LT TRN	**112**	**3.07%**			579	15.90%
RT TRN	**138**	**3.79%**			479	13.16%
Total, including other driver action combinations	579	15.90%	164	4.50%	3641	100.00%

The DT4000 form presents that bicyclists who did not act in any improper action at the time of the crash, were encountered almost equally with drivers going straight (NO IMPR-GO STR, 3.26%), taking a left turn (NO IMPR-LT TRN, 3.07%), and taking a right turn (NO IMPR-RT TRN, 3.79%). Whereas, the MV4000 form shows that approximately 30% of the crashes involved reported actions (6) other than darting into road, biking facing/not facing traffic, wearing dark clothes, and disregarding traffic signals. Such actions are improperly standing/working in the roadway, failing to have lights on while biking, wrong-way biking, etc.

Table 6-22: Pedestrian Location and Intersection Type

NMTLOC [1, 2]-PED-MV	INTTYPE-DT					
	4 WAY		T		Total, including other intersection type combinations	
	N	%	N	%	N	%
(1) in crosswalk	**630**	**17.30%**	113	3.10%	**895**	**24.58%**
(2) in roadway	208	5.71%	78	2.14%	1005	27.60%
Total, including other pedestrian location combinations	1521	41.77%	427	11.73%	3641	100.00%

NMTLOC [1, 2]-PED-DT	INTTYPE-DT					
	4 WAY		T		Total, including other intersection type combinations	
	N	%	N	%	N	%
ATI MX	**630**	**17.30%**	113	3.10%	**895**	**24.58%**
ATI NX	82	2.25%			156	4.28%
ATI UM	90	2.47%			146	4.01%
Total, including other pedestrian location combinations	1521	41.77%	427	11.73%	3641	100.00%

- The DT4000 form shows that the greatest percentage of crashes occurring at 4-way intersections (4 WAY), reported that the pedestrian was at an intersection, in a marked crosswalk (4-WAY-ATI MX, 17.30%),
- Overall, 24.58% of reported crashes showed that the pedestrian was located at an intersection, in a marked crosswalk (ATI MX), and pedestrians were equally located at intersections but not in the crosswalk (ATI NX), and at intersections with unmarked/unknown if marked crosswalks (ATI UM). It is clear that in MV4000 form, the same percentage (24.58%) of crashes reported that the pedestrian was located in a crosswalk. However, the information provided doesn't give the same level of detail for the pedestrian location.

Table 6-23: Bicyclist Location and Intersection Type

NMTLOC [1, 2]-BIKE-MV	INTTYPE-DT					
	4 WAY		T		Total, including other intersection type combinations	
	N	%	N	%	N	%
(1) in crosswalk	**475**	**13.05%**	114	3.13%	870	23.89%
(2) in roadway	455	12.50%	161	4.42%	1096	30.10%
(3) not in roadway	73	2.00%			295	8.10%
Total, including other bicyclist location combinations	1521	41.77%	427	11.73%	3641	100.00%

NMTLOC [1, 2]-BIKE-DT	INTTYPE-DT					
	4 WAY		T		Total, including other intersection type combinations	
	N	%	N	%	N	%
ATI MX	**339**	**9.31%**	88	2.42%	525	14.42%
ATI NX	**181**	**4.97%**			279	7.66%
ATI UM	84	2.31%	54	1.48%	160	4.39%
Total, including other bicyclist location combinations	1521	41.77%	427	11.73%	3641	100.00%

Bicyclists showed the same trend as pedestrians in terms of their location in a 4-way intersection, at the time of the crash. In 9.31% of crashes, the bicyclist was located at an intersection in a marked crosswalk (4-WAY-ATI MX). However, 4.97% of 4-way intersection crashes appeared to be associated with bicyclists located at intersections but not in the crosswalk (4-WAY-ATI NX). It is clear that in MV4000 form, the percentage (13.05%%) of crashes reported that the pedestrian was located in a crosswalk (1) doesn't provide the same level of detail for the bicyclist location as in DT4000 form.

Table 6-24: Prevailing Atmospheric Conditions and Type/Level of Light

WTHRCOND -MV	LGTCOND-MV							
	DARK		LIGT		(blank)		Total, including other light condition combinations	
	N	%	N	%	N	%	N	%
CLDY	71	1.95%	205	5.63%	655	17.99%	985	27.05%
CLR	99	2.72%	**462**	**12.69%**	**1601**	**43.97%**	2297	63.09%
RAIN			147	4.04%	96	2.64%	270	7.42%
Total, including other weather condition combinations	196	5.38%	845	23.21%	2389	65.61%	3641	100.00%

WTCOND [A, B]-DT	LGTCOND-DT							
	DARK		DAY		LITE		Total, including other light condition combinations	
	N	%	N	%	N	%	N	%
CLDY			625	17.71%	184	5.05%	920	25.27%
CLEAR	99	2.72%	**1594**	**43.78%**	**462**	**12.69%**	2297	63.09%
RAIN			95	2.61%	141	3.87%	264	7.25%
Total, including other weather condition combinations	196	5.38%	**2380**	**65.37%**	845	23.21%	3641	100.00%

Studying the effect of adverse weather conditions accompanied by poor light conditions, other factors such as driver's cautiousness and non-motorists obeying traffic signs, signals, and police officers appear to be the reason behind noticing fewer crashes with such circumstances. It is clear from examining the relationship between adverse weather conditions accompanied with poor light conditions that in the DT4000 form, 43.78% among 65.37% of crashes occurred during the daylight, were associated with clear weather conditions existing at the time of the crash (DAY-CLEAR). The information is considered consistent in the MV4000 form, as 43.97% of crashes occurred during the daylight, and was associated with clear weather conditions existing at the time of the crash (blank-CLR). Additionally, in both crash forms, the attributes consistently describe the weather and light conditions in the roadway where the crash occurred, i.e., 12.69% of crashes occurred in dark/lighted roadways (LIGT/LITE) and during clear (CLR/CLEAR) weather conditions.

Table 6-25: Trafficway Divided/Undivided with Type of Division and Prevailing Atmospheric Conditions

TRFCWAY - MV	**WTHRCOND -MV**							
	CLOUDY		CLEAR		RAIN		Total, including other weather condition combinations	
	N	%	N	%	N	%	N	%
D/WO	134	3.68%	218	5.99%			387	10.63%
ND	688	18.90%	**1669**	**45.84%**	181	4.97%	**2600**	**71.41%**
OW			110	3.02%			167	4.59%
Total, including other values describing trafficway division	985		2297		270		3641	100.00%

TRFCWAY [1, 2]-DT	**WTCOND [A, B]-DT**							
	CLOUDY		CLEAR		RAIN		Total, including other weather condition combinations	
	N	%	N	%	N	%	N	%
DIV NO	121	3.32%	192	5.28%			349	9.59%
OW			84	2.30%			138	3.78%
UNDIV	**602**	**16.52%**	**1596**	**43.84%**	168	4.62%	**2481**	**68.14%**
Total, including other values describing trafficway division	920	25.22%	2297	63.13%	264	7.23%	3641	100.00%

Both crash forms showed consistent information. For instance, undivided roadway sections showed to be more crash-prone (almost 70%) than other divided roadway sections (UNDIV/ND). Whereas, more than half of them (around 44%) occurred at clear atmospheric conditions (CLEAR). The DT4000 form provides information, showing that 16.52% of crashes occurred under cloudy atmospheric conditions (CLOUDY-UNDIV).

Table 6-26: Road Surface Type and Condition

SURFTYPE [1, 2]	RDCOND [A, B, C]					
	DRY		WET		Total, including other roadway surface condition combinations	
	N	%	N	%	N	%
BLACK	**1843**	**50.62%**	294	8.07%	2194	60.26%
CONC	1018	27.96%	164	4.50%	1217	33.42%
Total, including other roadway surface type combinations	**3056**	**83.93%**	482	13.24%	3641	100.00%%

A major part -50.62%- of crashes that took place on dry (DRY) roadway surface conditions (83.93%), occurred on bituminous road surfaces (BLACK) more than concrete surfaces (CONC). Darker asphalt pavement warms up and helps melt away any snow left on the road surface; hence it is interesting to investigate the reason behind the high percentage of crash rates on dry bituminous roadways.

6.3.1.3 Summary of Multi-variate Analysis of Selected Crash Variables

Regarding the preliminary analysis, for horizontal road terrain, 1.32% and 1.24% of the crashes occurred on roadways curved to the left (LT) and the right (RT), respectively. For vertical road terrain, 2.83% of the crashes occurred on downgrade (DN), 2.25% of the crashes occurred on upgrade roadways (UP), 0.77% of them occurred on hillcrest sections, and 0.19% of them occurred on sag/bottom sections. For road surface conditions, 13.24% of the crashes occurred in wet roadway surface conditions.

Viewing the type of trafficway division, two-way divided and unprotected (painted > 4 feet) median (DIV PNT) 2.36%, divided highway with traffic barrier (DIV BAR) 0.80%, and divided highway median with a barrier (DIV MBR) 1.84%. 68.14%, 9.59%, and 3.79% of the crashes occurred in two-way undivided highways (UNDIV), Divided highways without a traffic barrier (DIV NO), and on highways serving one-way traffic only (OW), respectively. Concerning the total number of lanes in a roadway where a crash took place, roadways with two and four total number of lanes contribute together to 80.31% of the total number of crashes.

Concerning the location of the first harmful event concerning the roadway, 94.04% of the crashes occurred on the roadway (ON), and 2.11% of them occurred on the roadside (R SIDE) occurred on the shoulder as described by the MV4000 crash form. For the specific location of the crash concerning the trafficway, most of the crashes (96.46%) occurred on the trafficway. Reporting the location type in MV4000 and DT4000 crash forms is consistent. 56.33% of crashes occurred at intersections, whereas, 43.67% of them occurred at non-intersection/midblock locations.

Taking into account the intersection type in which the crash occurred, 41.77% and 11.73% of the crashes occurred at 4-way stop intersections and T intersections, respectively. For the traffic control device (TCD) in effect at the time of the crash, 16.11% ((59.78% locations lacking a TCD (NONE) excluding 43.67% non-intersection locations (N)) of the crashes are reported as lacking a TCD (NONE). 23.85% and 12.55% of the crashes occurred at traffic signal controlled (TS OP) and at stop sign-controlled (STOP/SS) locations, respectively.

6.3.2 Injury Severity Distribution by Crash Characteristics

Following is a summary of potential crash variables distributed by injury severity level in **Table 6-24** to **Error! Reference source not found.**. Moreover, after interpreting the Multi-variate analysis results, a group of new variables that showed a significant relationship are chosen for inclusion in the statistical analysis and are analyzed per injury severity level in **Table 6-38** are as follows:

Table 6-27: Descriptive Statistics of the Potential Contributing Driver-Related Crash Variables

Variable	Indication	Injury severity									
		K		A		B		C		Subtotal	
		N	%	N	%	N	%	N	%	N	%
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
Driver-Related Factors											
DNMFTR [1, 2] [A, B]	Any relevant condition of the individual (motorist or non-motorists) that is directly related to the crash.										
	NORM	9	0.25%	249	6.84%	1269	34.85%	544	14.94%	2071	56.88%
	NO OBS	42	1.15%	128	3.52%	428	11.76%	237	6.51%	835	22.93%
	UI MDA	7	0.19%	33	0.91%	41	1.13%	25	0.69%	106	2.91%
	Other values	62	1.70%	171	4.70%	254	6.98%	142	3.90%	629	17.28%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
DRVRPC [1, 2] [A, B]	**The actions by the driver that may have contributed to the crash, based on the judgment of the law enforcement officer investigating the crash**										
	FTY	8	0.22%	109	2.99%	562	15.44%	243	6.67%	922	25.32%
	NO	61	1.68%	249	6.84%	790	21.70%	362	9.94%	1462	40.15%
	ID	3	0.08%	16	0.44%	55	1.51%	33	0.91%	107	2.94%
	Other values	48	1.32%	207	5.69%	585	16.06%	310	8.52%	1150	31.59%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
DRVRDOIN [1, 2]	**The controlled maneuver for this motor vehicle prior to the beginning of the sequence of events**										
	GO STR	70	1.92%	302	8.29%	663	18.21%	249	6.84%	1284	35.27%
	LT TRN	5	0.14%	65	1.79%	339	9.31%	171	4.70%	580	15.93%
	RT TRN	3	0.08%	25	0.69%	243	6.67%	151	4.15%	422	11.59%
	BACKING	1	0.03%	9	0.25%	44	1.21%	21	0.58%	75	2.06%
	Other values	41	1.13%	180	4.94%	703	19.31%	356	9.77%	1280	35.15%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SFTYEQP [1,2]	**The restraint equipment in use at the time of the crash (excluding motorcyclists)**										
	NONE	6	0.16%	29	0.80%	137	3.76%	53	1.46%	225	6.18%
	SH/LP	81	2.22%	372	10.22%	1294	35.54%	622	17.08%	2369	65.06%
	UNKN	16	0.44%	102	2.80%	352	9.67%	153	4%	623	17.11%

	UNTYPE	0	0.00%	4	0.11%	22	0.60%	9	5%	35	5.71%
	Other values	1	0.03%	7	0.19%	20	0.55%	13	0.36%	41	1.13%
	348 (9.56%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

It can be observed from **Table 6-27** that 56.88% and 25.32% of crashes involved drivers who appeared normal (NORM) and failed to yield the ROW (FTY) at the time of the crash, respectively. Regarding the driver's actions that may have contributed to the crash, 35.27%, 15.93%, and 11.59% of drivers involved in crashes were going straight (GO STR), taking a left turn (LT TRN), and taking right turn (RT TRN), respectively. Additionally, 65.06% of drivers used the shoulder and lap belt as a safety constraint at the time of the crash.

Table 6-28: Descriptive Statistics of the Potential Contributing Pedestrian-Related Crash Variables

Variable	**Indication**	**Injury severity**									
		K		A		B		C		Subtotal	
		N	%	N	%	N	%	N	%	N	%
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
Pedestrian-Related Factors											
DNMFTR [1, 2] [A, B]	Any relevant condition of the individual (motorist or non-motorists) that is directly related to the crash.										
	NO OBS	20	0.55%	41	1.13%	76	2.09%	47	1.29%	184	5.05%
	NORM	7	0.19%	163	4.48%	526	14.45%	263	7.22%	959	26.34%
	NORM NO OBS	34	0.93%	58	1.59%	116	3.19%	68	1.87%	276	7.58%
	OTHR	59	1.62%	319	8.76%	1274	34.99%	570	15.66%	2222	61.03%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTLOC [1, 2]	**The location of the non-motorists concerning the roadway at the time of the crash.**										
	ATI MX	14	0.38%	138	3.79%	457	12.55%	286	7.85%	895	24.58%
	ATI NX	8	0.22%	39	1.07%	66	1.81%	43	1.18%	156	4.28%

	ATI UM	4	0.11%	30	0.82%	71	1.95%	41	1.13%	146	4.01%
	NAI NX	61	1.68%	170	4.67%	311	8.54%	121	3.32%	663	18.21%
	SHLDR	9	0.25%	14	0.38%	37	1.02%	18	0.49%	78	2.14%
	Other values	13	0.36%	45	1.24%	97	2.67%	52	1.43%	207	5.68%
	1496 (41.09%) blank values/not pedestrian crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTACT [1, 2]	**The actions/circumstances of the non-motorists that may have contributed to the crash, based on the judgment of the law enforcement officer investigating the crash.**										
	NF TRFC	8	0.22%	21	0.58%	73	2.00%	54	1.48%	156	0.03%
	DISREG	1	0.03%	15	0.41%	25	0.69%	17	0.47%	58	1.59%
	DK CLTH	4	0.11%	8	0.22%	26	0.71%	14	0.38%	52	1.43%
	FC TRFC	3	0.08%	12	0.33%	33	0.91%	23	0.63%	71	1.95%
	IM XING	4	0.11%	23	0.63%	49	1.35%	23	0.63%	99	2.72%
	NO IMPR	17	0.47%	110	3.02%	366	10.05%	203	5.58%	696	19.12%
	SUDDEN	10	0.27%	60	1.65%	128	3.52%	72	1.98%	270	7.42%
	Other values	62	1.71%	187	5.14%	339	9.31%	155	4.26%	743	24.65%
	1496 (41.09%) blank values/not pedestrian crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTPRIOR [1, 2]	The action of a non-motorist immediately prior to a crash.										
	JOGGING	0	0.00%	18	0.49%	38	1.04%	18	0.49%	74	2.03%
	RDWY OT	16	0.00%	3	0.08%	11	0.30%	5	0.14%	19	0.52%
	WAITING	3	0.14%	19	0.52%	32	0.88%	17	0.47%	73	2.00%
	XING	1	0.44%	49	1.35%	91	2.50%	39	1.07%	195	5.36%
	Other values	89	2.42%	347	9.54%	867	23.82%	482	13.24%	1784	49.00%
	1496 (41.09%) blank values/not pedestrian crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTSFQ [1, 2] [A, B]*	The safety equipment in use by the operator non-motorist at the time of the crash.										
	NONE	88	2.42%	369	10.13%	871	23.92%	446	12.25%	1774	48.72%
	HLMT	0	0.00%	3	0.08%	5	0.14%	3	0.08%	11	0.30%*
	LTNG/REFL	0	0.00%	3	0.08%	5	0.14%	3	0.08%	11	0.30%
	Other values	17	0.47%	57	1.58%	149	4.10%	100	2.75%	323	8.87%

	1496 (41.09%) blank values/not pedestrian crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
PEDSEX	The sex of the non-motorists involved in a crash.										
	F	30	0.82%	183	5.03%	451	12.39%	249	6.84%	913	25.08%
	M	79	2.17%	251	6.89%	585	16.07%	306	8.40%	1221	33.53%
	UNKN	0	0.00%	2	0.05%	3	0.08%	6	0.16%	11	0.30%
	1496 (41.09%) blank values/not pedestrian crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
PEDAGE	The age of the non-motorists involved in a crash in years.										
	<30	20	0.55%	176	4.83%	496	11.84%	256	7.03%	948	26.04%
	30-64	56	1.54%	203	5.58%	431	14.17%	242	6.65%	932	7.28%
	≥65	33	0.91%	57	1.57%	112	3.08%	63	1.73%	265	25.30%
	1496 (41.09%) blank values/not pedestrian crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

As shown in **Table 6-28**, 26.34% of pedestrians involved in vehicle crashes appeared normal (NORM), 24.58% of them were located at intersections with marked crosswalks (ATI MX), 7.42% suddenly darted into the roadway at the time of the crash (SUDDEN), and 5.36% crossing the roadway immediately before the crash (XING). Regarding the used safety equipment, 48.72% did not use any (NONE). *DT4000 crash form database mistakenly reported 11 observations, showing that the pedestrian used a helmet as a piece of safety equipment (HLMT). Personal characteristics of involved pedestrians showed that 33.35% are male (M) pedestrians and 26.04% age <30.

Table 6-29: Descriptive Statistics of the Potential Contributing Bicyclist-Related Crash Variables

Variable	Indication	Injury severity									
		K		A		B		C		Subtotal	
		N	%	N	%	N	%	N	%	N	%
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
Bicycle-Related Factors											
DNMFTR [1, 2] [A, B]	NORM	2	0.05%	86	2.36%	745	20.46%	281	7.72%	1114	30.60%
	NORM NO OBS	5	0.14%	24	0.66%	140	3.85%	62	1.70%	231	6.34%
	OTHR	113	3.10%	471	12.94%	1107	30.40%	605	16.62%	2296	63.06%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTLOC [1, 2]	**The location of the non-motorists with respect to the roadway at the time of the crash.**										
	ATI MX	1	0.03%	32	0.88%	322	8.84%	170	4.67%	525	14.42%
	ATI NX	2	0.05%	39	1.07%	178	4.89%	60	1.65%	279	7.66%
	ATI UM	0	0.00%	15	0.41%	100	2.75%	45	1.24%	160	4.39%
	BIKE LN	0	0.00%	4	0.11%	44	1.21%	14	0.38%	62	1.70%
	NAI NX	4	0.11%	32	0.88%	143	3.93%	41	1.13%	220	6.04%
	SHLDR	1	0.03%	7	0.19%	30	0.82%	13	0.36%	51	1.40%
	Other values	3	0.09%	16	0.44%	136	3.73%	44	1.21%	199	5.47%
	2145 (58.91%) blank values/not bicyclist crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTPRIOR [1, 2]	**The action of a non-motorist immediately prior to a crash.**										
	A TRFC	0	0.00%	8	0.22%	46	1.26%	16	0.44%	70	1.92%
	RDWY OT	1	0.03%	2	0.05%	12	0.33%	3	0.08%	18	0.49%
	SIDE WK	0	0.00%	0	0.00%	11	0.30%	1	0.03%	12	0.33%
	W TRFC	1	0.00%	4	0.11%	19	0.52%	7	0.19%	30	0.82%
	XING	5	0.03%	13	0.36%	89	2.44%	33	0.91%	136	3.74%

	Other values	4	0.25%	118	3.25%	776	21.32%	327	8.98%	1230	33.79%
	2145 (58.91%) blank values/not bicyclist crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTACT [1, 2]	**The actions/circumstances of the non-motorists that may have contributed to the crash, based on the judgment of the law enforcement officer investigating the crash.**										
	DISREG	0	0.00%	9	0.25%	51	1.40%	16	0.44%	76	2.09%
	NO IMPR	4	0.11%	45	1.24%	385	10.57%	145	3.98%	579	15.90%
	SUDDEN	1	0.03%	15	0.41%	102	2.80%	46	1.26%	164	4.50%
	Other values	11	0.31%	0	3.99%	953	26.17%	387	10.63%	1496	41.09%
	2145 (58.91%) blank values/not bicyclist crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
NMTSFQ [1, 2] [A, B]	The safety equipment in use by the operator non-motorist at the time of the crash.										
	REFL/LTNG	0	0.00%	5	0.14%	6	0.16%	5	0.14%	16	0.44%
	NONE	7	0.19%	84	2.31%	612	16.81%	268	7.36%	971	26.67%
	HLMT	3	0.08%	41	1.13%	228	6.26%	64	1.76%	336	9.23%
	Other values	110	3.03%	451	12.38%	1146	31.48%	611	16.78%	2318	63.66%
	2145 (58.91%) blank values/not bicyclist crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
BIKESEX	The sex of the non-motorists involved in a crash.										
	F	3	0.08%	34	0.93%	230	6.32%	94	2.58%	361	9.91%
	M	8	0.22%	109	2.99%	720	19.77%	289	7.94%	1126	30.93%
	UNKN	0	0.00%	2	0.05%	3	0.08%	4	0.11%	9	0.25%
	2145 (58.91%) blank values/not bicyclist crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
BIKEAGE	The age of the non-motorists involved in a crash in years.										
	<30	4	0.11%	64	1.76%	599	16.45%	244	6.70%	911	25.02%
	30-64	5	0.14%	63	1.73%	304	92.97%	122	3.35%	494	13.57%
	≥65	2	0.05%	18	0.49%	50	1.37%	21	0.58%	91	2.50%
	2145 (58.91%) blank values/not bicyclist crashes										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Table 6-29 shows that 30.60% of bicyclists involved in vehicle crashes appeared normal (NORM), 14.42% of them were located at intersections with marked crosswalks (ATI MX), 4.50% suddenly darted into the roadway at the time of the crash (SUDDEN), and 3.74% crossing the roadway immediately before the crash (XING). Regarding the used safety equipment, 26.67%% did not use any (NONE), and 9.23% used helmet (HLMT). Personal characteristics of involved bicyclists showed that 30.93% are male (M) pedestrians and 25.02% age <30.

Table 6-30: Descriptive Statistics of the Potential Contributing Roadway-Related Crash Variables

Variable	**Indication**	**Injury severity**									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
Roadway Level Factors											
TRFCWAY [1, 2]	**Indication of whether or not the trafficway for this vehicle is divided and whether it serves one-way or two-way traffic.**										
	UNDIV	65	1.79%	402	11.04%	1393	38.26%	621	17.06%	2481	68.14%
	DIV NO	31	0.85%	55	1.51%	184	5.05%	70	1.92%	340	9.34%
	OW	1	0.03%	13	0.36%	50	1.37%	22	0.60%	86	2.36%
	DIV BAR	23	0.63%	111	3.05%	365	10.02%	235	6.45%	734	20.16%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RDWYPC [A, B, C]	**Factors of the road which may have contributed to the crash.**										
	NONE	114	3.13%	549	15.08%	1884	51.74%	904	24.83%	3451	94.78%
	RSC	3	0.08%	6	0.16%	29	0.80%	17	0.47%	55	1.51%
	OTHR	3	0.08%	26	0.71%	79	2.17%	27	0.74%	135	3.71%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
TOTLANES [1, 2]	1 Lane	1	0.03%	17	0.47%	139	3.82%	63	1.73%	220	6.04%
	2 Lanes	69	1.90%	394	10.82%	1311	36.01%	634	17.41%	2408	66.14%
	3 Lanes	8	0.22%	27	0.74%	101	2.77%	43	1.18%	179	4.92%

	>3 Lanes	42	1.15%	143	3.93%	441	12.11%	208	5.71%	834	22.91%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RDCOND [A, B, C]	**The roadway surface condition at the time and place of a crash.**										
	DRY	97	2.66%	476	13.07%	1699	46.66%	784	21.53%	3056	83.93%
	WET	16	0.44%	81	2.22%	252	6.92%	133	3.65%	482	13.24%
	OTHR	7	0.19%	24	0.66%	41	1.13%	31	0.85%	103	2.83%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RLTNRDWY	**The location of the first harmful event as it relates to its position within or outside the trafficway.**										
	ON	109	2.99%	551	15.13%	1877	51.55%	899	24.69%	3436	94.37%
	R SIDE	4	0.11%	11	0.30%	41	1.13%	21	0.58%	77	2.11%
	OTHR	7	0.19%	19	0.52%	74	2.03%	28	0.77%	128	3.52%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RLTNTRWY	**Identifies the location of a crash with respect to it's relation to a trafficway.**										
	OFF	6	0.16%	12	0.33%	54	1.48%	27	0.74%	99	2.72%
	ON	114	3.13%	567	15.57%	1917	52.65%	914	25.10%	3512	96.46%
	OTHR	0	0.00%	2	0.05%	21	0.58%	7	0.19%	30	0.82%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RLTNJNLC	**The location of the first harmful event of the crash. It identifies the crash's location with respect to the presence in a junction or proximity to components typically in a junction or an interchange area. This field identifies the specific location in a junction or interchange.**										
	INR	9	0.25%	61	1.68%	230	6.30%	105	2.88%	405	11.10%
	INT	25	0.68%	228	6.28%	920	25.27%	484	13.29%	1657	45.52%
	NJ	84	2.31%	274	7.51%	736	20.23%	323	8.88%	1417	38.94%
	OTHR	2	0.05%	16	0.44%	97	2.65%	41	0.91%	148	4.05%
	14 (0.39%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
RLTNJNIC	**The location of the first harmful event of the crash. It identifies the crash's location concerning the presence in a junction or proximity to components typically in a junction or an interchange area. This field identifies if a crash occurred within the Interchange area. (Y/N/UNKN).**										
	N	112	3.08%	514	14.12%	1740	47.79%	817	22.44%	3183	87.42%
	UNKN	0	0.00%	5	0.14%	18	0.49%	10	0.27%	33	0.91%

	Y	8	0.22%	60	1.65%	225	6.18%	118	3.24%	411	11.29%
	14 (0.39%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ROADHOR [1, 2]	**The horizontal road terrain at the point of impact.**										
	LT	2	0.05%	9	0.25%	29	0.80%	8	0.22%	48	1.32%
	RT	6	0.16%	6	0.16%	22	0.60%	11	0.30%	45	1.24%
	ST	109	2.99%	544	14.94%	1830	50.26%	864	23.73%	3347	91.93%
	Other values	3	0.08%	22	0.60%	111	3.05%	65	1.79%	201	5.52%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ROADVERT [1, 2]	**The vertical road terrain at the point of impact.**										
	DN	14	0.38%	38	1.04%	43	1.18%	22	0.60%	117	3.21%
	CST	0	0.00%	4	0.11%	21	0.58%	3	0.08%	28	0.77%
	LVL	14	0.38%	42	1.15%	64	1.76%	25	0.69%	145	3.98%
	SAG	28	0.77%	84	2.31%	128	3.52%	50	1.37%	290	7.96%
	UP	42	1.15%	126	3.46%	192	5.27%	75	2.06%	435	11.95%
	Other values	22	0.62%	287	7.89%	1544	42.40%	773	21.24%	2626	72.13%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
INTTYPE	**The type of intersection in which a crash occurred. An intersection consists of two or more roadways that intersect at the same level.**										
	5	1	0.03%	2	0.05%	2	0.30%	13	0.36%	27	0.74%
	4 WAY	25	0.69%	222	6.10%	222	23.29%	426	11.70%	1521	41.77%
	L	0	0.00%	1	0.03%	1	0.11%	2	0.05%	7	0.19%
	RAB	0	0.00%	1	0.03%	1	0.36%	5	0.14%	19	0.52%
	T	7	0.19%	48	1.32%	48	6.59%	132	3.63%	427	11.73%
	Y	0	0.00%	4	0.11%	4	0.14%	3	0.08%	12	0.33%
	Other values	87	2.39%	303	8.32%	1714	23.92%	367	10.08%	1628	44.72%
	NA	1590 (43.67%)									
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
TRFCCNTL [1, 2]	**The type of traffic control device (TCD) applicable to this motor vehicle at the crash location.**										
	NONE	89	2.44%	335	9.20%	878	24.11%	327	8.98%	1629	44.74%
	STOP	7	0.19%	65	1.79%	387	10.63%	210	5.77%	669	18.37%

	TS OP	3	0.08%	39	1.07%	265	7.28%	127	3.49%	434	11.92%
	Other values	21	0.59%	142	3.90%	462	12.69%	284	7.80%	909	24.97%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
TRFCINOP	**Indicates whether a traffic control device was inoperable or missing at the time of the crash (Y/N/UNKN).**										
	N	117	3.21%	558	15.33%	1900	52.18%	888	24.39%	3463	95.11%
	UNKN	2	0.05%	16	0.44%	47	1.29%	37	1.02%	102	2.80%
	Y	0	0.00%	0	0.00%	6	0.16%	4	0.11%	10	0.27%
	Other values	1	0.04%	7	0.19%	39	1.08%	19	0.52%	66	1.82%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
LOCTYPE	**The type of location at which a crash occurred.**										
	I	33	0.91%	286	7.85%	1143	31.39%	589	16.18%	2051	56.33%
	N	87	2.39%	295	8.10%	849	23.32%	359	9.86%	1590	43.67%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SURFTYPE [1, 2]	CONC	23	0.63%	105	2.88%	357	9.80%	170	4.67%	220	6.04%
	BLACK	34	0.93%	180	4.94%	666	18.29%	318	8.73%	2408	66.14%
	OTHR	63	1.73%	296	8.13%	969	26.61%	460	12.63%	1013	27.82%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Concerning **Table 6-30** implies that 68.14% of trafficways involving vehicle crashes with pedestrians and bicyclists are two-way undivided, 13.24% of the roadways' surface is wet, and 91.93% occurred on straight roadways. 56.33% of crashes occurred at intersection locations, 41.77% of them occurred at 4-way intersections, 44.74% of the crash locations did not involve TCD, with 95.11% of these locations did not indicate inoperable TCD.

Table 6-31: Descriptive Statistics of the Potential Contributing Crash/Vehicle-Related Crash Variables

Variable	Indication	Injury severity									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
Crash/Vehicle Level Factors											
VEHTYPE [1, 2]	**Specific category for the type of vehicle which was involved in a crash.**										
	CAR	58	1.59%	291	7.99%	746	20.49%	396	10.88%	1491	40.95%
	SUV	20	0.55%	71	1.95%	138	3.79%	87	2.39%	316	8.68%
	UT TRK	17	0.47%	48	1.32%	75	2.06%	35	0.96%	175	4.81%
	P VAN	4	0.11%	13	0.36%	42	1.15%	21	0.58%	80	2.20%
	Other values	21	0.58%	158	4.34%	991	27.22%	409	11.23%	1579	43.36%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
VEHDMG [1, 2]	**Identifies the extent to which the damage affects the vehicles operability rather than the cost to repair.**										
	DISABL	4	0.11%	5	0.14%	2	0.05%		0.00%	11	0.30%
	FUNC	10	0.27%	16	0.44%	18	0.49%	1	0.03%	45	1.24%
	MINOR	7	0.19%	35	0.96%	98	2.69%	33	0.91%	173	4.75%
	NO	4	0.11%	31	0.85%	135	3.71%	102	2.80%	272	7.47%
	UNKN	2	0.05%	19	0.52%	50	1.37%	38	1.04%	109	2.99%
	Other Values	1	0.03%	23	0.63%	130	3.57%	58	1.59%	212	5.82%
	2819 (77.42%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ALCFLAG	**Indicates whether law enforcement suspected that at least one driver or non-motorist involved in the crash had used alcohol. This includes both alcohol use under the legal limit and at or over the legal limit.**										
	UNKN	13	0.36%	81	2.22%	309	8.49%	180	4.94%	583	16.01%
	Y	35	0.96%	106	2.91%	113	3.10%	45	1.24%	299	8.21%
	N	72	1.98%	394	10.82%	1570	43.12%	723	19.86%	2759	75.78%
	2759 (75.78%) blank values										

	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
DRUGLFAG*	**Indicates whether law enforcement suspected that at least one driver or non-motorist involved in the crash had used drugs (Y/N/UNKN).**										
	UNKN	19	0.52%	95	2.61%	336	9.23%	194	5.33%	644	17.69%
	Y	11	0.30%	14	0.38%	11	0.30%	8	0.22%	44	1.21%
	2953 (81.10%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SPEEDFLAG	**Flag indicating whether speed was a factor in a crash.**										
	Y	8	0.22%	36	0.99%	29	0.80%	11	0.30%	84	2.31%
	3557 (97.69%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
SCHZONE	**Flag indicating whether a crash occurred in an active school zone.**										
	Y	2	0.05%	10	0.27%	46	1.26%	30	0.82%	88	2.42%
	3553 (97.58%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
CONSZONE	**Flag indicating whether a crash occurred in construction, maintenance, or utility work zone or was related to activity within a work zone.**										
	Y	3	0.08%	6	0.16%	17	0.47%	6	0.16%	32	0.88%
	3609 (99.12%) blank values										
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Regarding vehicle and crash-related variables, **Table 6-31** shows that 40.95% of the vehicles involved in crashes are passenger cars (CAR), 4.75% of involved vehicles' damage extent is minor damage (MINOR). For toxicity, 8.21% of crashes involved alcohol-impaired person (Y).

Table 6-32: Descriptive Statistics of the Potential Contributing environmental-related Crash Variables

Variable	**Indication**	**Injury severity**									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	3.3%	581	16.0%	1992	54.7%	948	26%	3641	100.00%
Environmental Factors											
WTCOND	**The prevailing atmospheric conditions that existed at the time of the crash.**										
	CLDY	36	0.99%	139	3.82%	502	13.79%	243	6.67%	920	25.27%
	CLEAR	68	1.87%	367	10.08%	1272	34.94%	590	16.20%	2297	63.09%
	RAIN	9	0.25%	42	1.15%	141	3.87%	72	1.98%	264	7.25%
	Other values	7	0.19%	33	0.91%	77	2.11%	43	1.18%	160	4.39%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
LGTCOND	**The type/level of light that existed at the time of the motor vehicle crash.**										
	DUSK	37	1.02%	55	1.51%	84	2.31%	20	0.55%	196	5.38%
	LITE	6	0.16%	12	0.33%	55	1.51%	19	0.52%	92	2.53%
	DUSK	35	0.96%	298	8.18%	1381	37.93%	666	18.29%	2380	65.37%
	LITE	7	0.19%	18	0.49%	60	1.65%	27	0.74%	112	3.08%
	DUSK	35	0.96%	197	5.41%	403	11.07%	210	5.77%	845	23.21%
	LITE	0	0.00%	1	0.03%	9	0.25%	6	0.16%	16	0.44%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%
ENVPC[A,B,C]	**Environmental conditions which may have contributed to the crash.**										
	GLARE	3	0.08%	18	0.49%	54	1.48%	18	0.49%	93	2.55%
	NONE	106	2.91%	503	13.81%	1769	48.59%	845	23.21%	3223	88.52%
	OBSTR	0	0.00%	13	0.36%	39	1.07%	15	0.41%	67	1.84%
	WTHR	10	0.27%	40	1.10%	110	3.02%	61	1.68%	221	6.07%

	Other values	1	0.03%	7	0.19%	18	0.49%	9	0.25%	35	0.96%
	Subtotal	120	3.30%	581	15.96%	1992	54.71%	948	26.04%	3641	100.00%

Concerning environmental factors shown in **Table 6-32**, 63.09% and 25.27% of crashes occurred in clear and cloudy weather conditions, respectively. Additionally, 65.37% occurred during the dusk time of the day.

Table 6-33: Descriptive Statistics of the Potential Newly Created Crash Variables (N=3641)**

Variable	**Indication**	**Injury severity**									
		K		**A**		**B**		**C**		**Subtotal**	
		N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
		120	**3.3%**	**581**	**16.0%**	**1992**	**54.7%**	**948**	**26%**	**3641**	**100.00%**
Roadway Characteristics Relationships											
TOTLANES_2 – INTTYPE_4 WAY – TRFCCNTL_TS OP		Two-lane 4-way traffic signal-controlled intersection									
		4	0.11%	43	1.18%	180	4.94%	109	2.99%	336	9.23%
TOTLANES_4 – TRFCWAY_UNDIV		Four-lane undivided trafficway									
		14	0.38%	62	1.70%	193	5.30%	105	2.88%	374	10.27%
TOTLANES_2 – TRFCWAY-UNDIV		Two-traffic lanes without a physical division									
		46	1.26%	229	6.29%	563	15.46%	283	7.77%	1121	30.79%
INTTYPE_4 WAY – TRFCCNTL_TS OP		4-way traffic signal-controlled intersection									
		10	0.27%	109	2.99%	430	11.81%	243	6.67%	792	21.75%
ROADHOR_ST – ROADVERT_LVL		Straight and level graded roadway section									
		97	2.66%	485	13.32%	1644	45.15%	809	22.22%	3035	83.36%
WTCOND_CLEAR – TRFCWAY_UNDIV		Clear weather condition at two-way undivided trafficway									
		40	1.10%	263	7.22%	895	24.58%	398	10.93%	1596	43.83%
RDCOND_DRY – SURFTYPE_BLACK		Dry roadway surface condition with blacktop (bituminous) surface type									
		69	1.90%	313	8.60%	1017	27.93%	444	12.19%	1843	50.62%
LGTCOND_LITE – WTCOND_CLEAR		The dark and lighted roadway at the time of the crash with the clear weather condition									
		19	0.52%	114	3.13%	220	6.04%	109	2.99%	462	12.69%
WTCOND_CLDY – TRFCWAY_UNDIV		Occurred on a two-way undivided trafficway under cloudy weather conditions									
		19	0.52%	86	2.36%	343	9.42%	154	4.23%	602	16.53%

	Driver Actions-Roadway Characteristics Relationships									
INTTYPE_4 WAY – DRVRPC_FTY	The crash occurred at a 4-way intersection while the driver failed to yield the ROW									
	3	0.08%	62	1.70%	320	8.79%	142	3.90%	527	14.47%
RLTNJNIC_N – DRVRPC_FTY	Crash first harmful event is not within an interchange area and the driver failed to yield									
	7	0.19%	93	2.55%	460	12.63%	199	5.47%	759	20.85%
Pedestrian Action and Location Relationships										
NMTACT_PED_NO IMPR – DRVRPC_FTY	Pedestrian did not act improperly but the driver failed to yield the ROW									
	3	0.08%	47	1.29%	193	5.30%	87	2.39%	330	9.06%
NMTPRIOR_PED_XING– NMTLOC_PED_ATI MX	Pedestrian crossing roadway immediately before the crash at an intersection in a marked crosswalk									
	13	0.36%	110	3.02%	396	10.88%	241	6.62%	760	20.87%
INTTYPE_4 WAY – NMTLOC_PED_ATI MX	4-way intersection with a pedestrian located at the intersection in a marked crosswalk									
	10	0.27%	109	2.99%	430	11.81%	243	6.67%	792	21.75%
NMTLOC_ PED_NAI NX – DRVRDOIN_GO STR	Pedestrian not located in an intersection and not in a marked crosswalk with the driver going straight									
	33	0.91%	117	3.21%	159	4.37%	60	1.65%	369	10.13%
Bicyclist Action and Location Relationships										
NMTPRIOR_BIKE_XING – NMTLOC_BIKE_ATI MX	Bicyclist crossing the roadway at an intersection in a marked crosswalk									
	1	0.03%	18	0.49%	206	5.66%	112	3.08%	337	9.26%
INTTYPE_4 WAY – NMTLOC_BIKE_ATI MX	Bicyclist located at the intersection in a marked crosswalk in a 4-way intersection									
	1	0.03%	26	0.71%	214	5.88%	98	2.69%	339	9.31%

Table 6-33 shows that the 16.53% of crashes occurred on a two-way undivided trafficway under cloudy weather conditions (WTCOND_CLDY – TRFCWAY_UNDIV), whereas 43.83% of them occurred on two-way undivided trafficway under clear weather conditions (WTCOND_CLEAR – TRFCWAY_UNDIV). 30.79% of crashes occurred on two-traffic lanes without a physical division, and 83.36% occurred on straight and level graded roadway section (ROADHOR_ST – ROADVERT_LVL). Crashes occurred on a 4-

way traffic signal-controlled intersection, with the pedestrian and bicyclist located at the intersection in a marked crosswalk (INTTYPE_4 WAY – NMTLOC_PED_ATI MX; 21.75%), (INTTYPE_4 WAY – NMTLOC_BIKE_ATI MX, 9.31%), respectively.

** A total of 4025 crashes, excluding 324 crashes where the motorist sustained more severe injury than the non-motorists as the interest is in the severity of injuries sustained by non-motorists involved in a motor vehicle crash. Also, because the analysis is limited to the first two units involved in a crash. There are 60 crashes where the first two units are both drivers; and therefore, their injury severity levels are not counted towards the descriptive statistics of pedestrian/bicycle-vehicle crashes.

6.4 Statistical Tests Concerning Injury Severity Proportion

In this analysis, the Z-test for proportions was selected as the statistical test to indicate if a particular variable of the newly created roadway, driver, pedestrian, and bicyclist-related pedestrian and bicyclist-related variables has higher (fatal (K) and severe (A) injury) proportion is significantly different than the proportion of (fatal (K) and severe (A) injury) injuries for the population. The test was conducted in RStudio using the prop.test () function at a 95% confidence level. Note that the formula of the Z-test statistics **Eq. 6-2** is valid when sample size (*n*) is large enough; *np, nq* should be ≥ 5. In the case of small sample size (such in the "ROADHOR_C-ROADVERT_H" variable in **Table 6-34**, the Fisher Exact probability test is used for comparing the two proportions.

$$z=\frac{\hat{p}-p}{\sqrt{pq/n}} \qquad \textbf{Eq. 6-1}$$

Where:

$\hat{p}$: sample proportion;

p: population proportion;

q: 1-p;

n: sample size;

The results of this analysis can be found in **Table 6-34** and **Table 6-37**. These results show the proportion of fatal injury, severe injury, and non-severe injury including evident and possible injury crashes by each of the newly created variables. The table identifies the crash variable that has a significantly different proportion of fatal and severe injury versus non-severe injury crashes using the z-test for proportions. A $(+\,+)$ symbol implies crash variables with a significantly lower proportion of fatalities and severe injuries at a 95% confidence level, and a $(-)$ symbol implies crash variables with a significantly lower proportion of fatalities and severe injuries at a 95% confidence level.

6.4.1 Pedestrian Crash Variables Using MV4000 Dataset

Table 6-34: Summary of the Z-test for Proportion Results for the Newly Created Pedestrian Crash Variables

Variable Symbol	Variable Indication	Fatal Injury Crash (K)	Severe Injury Crash (A)	Fatal and Severe Injury (K+A) Crash	Evident and Possible Injury Crash (B+C)	Sig. Result of the Z-Test *	Sample Size
Roadway-Environmental-Related							
LGTCOND_LIGT-TRFCWAY_D_WO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier	0.51%	1.12%	1.63%	2.56%		7.11%
LGTCOND_LIGT-TRFCWAY_ND	Streetlight is available at time of the crash in undivided trafficway	0.75%	5.17%	5.92%	14.78%	–	35.14%
LGTCOND_DARK-WTHRCOND_CLDY	No light (dark) is available at time of crash under cloudy weather	0.70%	0.98%	1.68%	1.54%		5.47%
LGTCOND_DARK-WTHRCOND_CLR	No light (dark) is available at time of crash under clear weather	0.79%	1.21%	2.00%	1.59%		6.09%
ROADHOR_C-ROADVERT_H	Curve (not straight) and hill (not level) road terrain	1.72%	8.21%	9.93%	31.70%		70.66%
Driver-Weather Related							
LGTCOND_LIGT-DRVRDO_GO_STR	Streetlight is available at time of crash and driver going straight	0.84%	4.29%	5.13%	6.25%	–	19.32%
DRVRDOIN_GO_STR-NMTLOC_2	A driver going straight and pedestrian located in the roadway	2.05%	7.09%	9.14%	13.38%	–	38.23%
LGTCOND_DARK-DRVRDO_GO_STR	No light (dark) is available at time	1.07%	1.63%	2.70%	1.35%		6.87%

	of crash and driver going straight						
Pedestrian-Weather Related							
LGTCOND_DARK-NMTLOC_2	No light (dark) is available at time of crash and pedestrian located in the roadway	1.35%	1.82%	3.17%	2.52%		9.66%
WTHRCOND_CLR-NMTLOC_2	The crash occurred under clear weather and pedestrian located in the roadway	1.82%	6.57%	8.39%	20.47%	−	48.99%
Pedestrian-Driver Related							
DRVRDO_GO_STR-NMTACT_1	A driver going straight and pedestrian walking not facing traffic	0.28%	1.21%	1.49%	2.52%		6.81%
DRVRDO_GO_STR-NMTACT_6	A driver going straight and pedestrian acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.	1.59%	5.64%	7.23%	10.96%	−	30.88%
Pedestrian-Specific							
NMTLOC_1-NMTACT_6	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.	0.09%	0.51%	0.61%	2.89%		5.94%
Total Cashes (N=2145)		**110**	**434**	**544**	**1601**		

* The significant value (Sig. Result of the Z-Test) is a result of the Z-test of the Difference Between Two Proportions; the proportion of crashes resulting in a fatal (K) and sever (A) injury versus the proportion of crashes resulting in an evident (B) and possible (C) injury for each new variable resulting from a multi-variable analysis, where; − = proportion of crashes resulting in a fatal (K) and sever (A) injury is significantly lower at 95% confidence level, + + = proportion of crashes resulting in a fatal (K) and sever (A) injury is significantly higher at 95% confidence level.

6.4.2 Bicyclist Crash Variables Using MV4000 Dataset

Table 6-35: Summary of the Z-test for Proportion Results for the Newly Created Bicyclist Crash Variables

Variable Symbol	Variable Indication	Fatal Injury Crash (K)	Severe Injury Crash (A)	Fatal and Severe Injury (K+A) Crash	Evident and Possible Injury Crash (B+C)	Sig. Result of the Z-Test *	Sample Size
Roadway-Environmental-Related							
LGTCOND_LIGT-TRFCWAY_D_WO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier	0.00%	0.13%	0.13%	1.20%		3.24%
LGTCOND_LIGT-TRFCWAY_ND	Streetlight is available at time of the crash in undivided trafficway	0.07%	0.94%	1.00%	6.42%	–	18.06%
LGTCOND_DARK-WTHRCOND_CLDY	No light (dark) is available at time of crash under cloudy weather	0.33%	0.00%	0.33%	0.00%		0.80%
LGTCOND_DARK-WTHRCOND_CLR	No light (dark) is available at time of crash under clear weather	1.47%	0.00%	1.47%	0.00%		3.58%
ROADHOR_C-ROADVERT _H	Curve (not straight) and hill (not level) road terrain	0.00%	0.00%	0.00%	1.27%		3.09%
Driver-Weather Related							
LGTCOND_LIGT-DRVRDO_GO_STR	Streetlight is available at time of crash and driver going straight	0.07%	0.53%	0.60%	3.88%		10.90%
DRVRDOIN_GO_STR-NMTLOC_2	A driver going straight and pedestrian located in the roadway	0.27%	2.14%	2.41%	19.79%		54.03%
LGTCOND_DARK-DRVRDO_GO_STR	No light (dark) is available at time of crash and driver going straight	0.00%	0.13%	0.13%	0.74%		2.12%
Bicyclist -Weather Related							
LGTCOND_DARK-NMTLOC_2	No light (dark) is available at time	0.00%	0.27%	0.27%	0.94%		2.94%

	of crash and pedestrian located in the roadway						
WTHRCOND_CLR-NMTLOC_2	The crash occurred under clear weather and pedestrian located in the roadway	0.53%	3.28%	3.81%	29.08%		80.05%
Bicyclist -Driver Related							
DRVRDO_GO_STR-NMTACT_1	A driver going straight and pedestrian walking not facing traffic	0.00%	0.00%	0.00%	0.07%		0.17%
DRVRDO_GO_STR-NMTACT_6	A driver going straight and pedestrian acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.	0.07%	0.80%	0.87%	8.36%		22.46%
Bicyclist -Specific							
NMTLOC_1-NMTACT_6	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.	0.00%	0.60%	0.60%	9.56%		24.73%
Total Cashes (N=1496)		**11**	**145**	**156**	**1340**		

As demonstrated in **Table 6-34**, the most common roadway environmental-related variable that yielded severe fatal and severe pedestrian crashes was when crashes occur on the curve (not straight) and hill (not level) road terrain (ROADHOR_C- ROADVERT_H; 9.93%). The next most crash variable that yielded fatal and severe pedestrian crashes was when a streetlight is available at the time of the crash in an undivided trafficway (LGTCOND_LIGT-TRFCWAY_ND; 5.92%). Drivers going straight while pedestrians are located in the roadway (DRVRDOIN_GO_STR-NMTLOC_2), and when a streetlight is available at the time of crash while drivers going straight

(LGTCOND_LIGT-DRVRDO_GO_STR) were the two most driver weather-related variables that are associated with 9.14%, and 5.13% fatal and severe pedestrian crashes, respectively. Additionally, crashes occurred under clear weather while pedestrians located in the roadway (WTHRCOND_CLR-NMTLOC_2), and crashes occurred on roadways with no available light (dark) and while pedestrians located in the roadway (LGTCOND_DARK-NMTLOC_2), were two most common pedestrian weather-related variables that were responsible for 8.39%, and 3.17% fatal and severe injury pedestrian crashes, respectively. Drivers going straight while pedestrians acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway at the time of the crash (DRVRDO_GO_STR-NMTACT_6), were responsible for 7.23% of the fatal and severe pedestrian crashes. For instance, being In Roadway Improperly (Standing, Lying, Working, Playing), wrong-way walking (W WAY), failure to obey traffic signs, signals, or officer, and being inattentive (talking, eating, Etc.).

However, note that among the aforementioned variables, five crash variables have significantly lower percentages of fatal and severe versus non-severe injury pedestrian crashes; (LGTCOND_LIGT-TRFCWAY_ND), (LGTCOND_LIGT-DRVRDO_GO_STR), (DRVRDOIN_GO_STR-NMTLOC_2), (WTHRCOND_CLR-NMTLOC_2) and (DRVRDO_GO_STR-NMTACT_6) at the 95% confidence level.

While for bicyclist crashes, percentages shown in **Table 6-35** showed a different pattern. Bicyclist crashes contribute with a lower percentage of the total crashes (41.08%), compared to pedestrian crashes. The most common roadway environmental-related variable that yielded severe fatal and severe bicyclist crashes are associated with no available light (dark) in roadways at the time of crash under clear weather conditions (LGTCOND_DARK-WTHRCOND_CLR; 1.47%). Also, crashes occurred when drivers going straight while bicyclists were located in the roadway

(DRVRDOIN_GO_STR-NMTLOC_2), were responsible for 2.41% of fatal and severe bicyclist crashes. Furthermore, concerning bicyclist weather-related variables, 3.81% of crashes that produced fatal and severe injuries occurred under clear weather while bicyclists located in the roadway (WTHRCOND_CLR-NMTLOC_2). Note that among the aforementioned variables, one crash variable has significantly lower percentages of fatal and severe versus non-severe injury bicyclist crashes; available streetlight at time of the crash in an undivided trafficway (LGTCOND_LIGT-TRFCWAY_ND) at the 95% confidence level.

6.4.3 Pedestrian Crash Variables Using DT4000 Dataset

Table 6-36: Summary of the Z-test for Proportion Results for the Newly Created Pedestrian Crash Variables

Variable Symbol	Variable Indication	Fatal Injury Crash (K)	Severe Injury Crash (A)	Fatal and Severe Injury (K+A) Crash	Evident and Possible Injury Crash (B+C)	Sig. Result of the Z-Test *	Sample Size
Roadway-Environmental-Related							
LGTCOND_LITE-TRFCWAY_DIV_NO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier	0.47%	1.12%	1.59%	2.56%		7.04%
LGTCOND_LITE-TRFCWAY_UNDIV	Streetlight is available at time of the crash in undivided trafficway	0.70%	4.99%	5.69%	14.41%	–	34.12%
LGTCOND_DARK-WTCOND_CLDY	No light (dark) is available at time of crash under cloudy weather	0.61%	0.84%	1.45%	1.40%		4.84%
LGTCOND_DARK-WTCOND_CLEAR	No light (dark) is available at time of crash	0.79%	1.21%	2.00%	1.59%		6.09%

	under clear weather						
ROADHOR_LT_RT_CU-ROADVERT_CST_UP_DN_SAG	Curve (not straight) and hill (not level) road terrain	1.72%	8.21%	9.93%	31.70%		70.66%
Driver-Weather Related							
LGTCOND_LITE-DRVRDOIN_GO_STR	Streetlight is available at time of crash and driver going straight	0.84%	4.29%	5.13%	6.25%	–	19.32%
DRVRDOIN_GO_STR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	A driver going straight and pedestrian located in the roadway	1.54%	5.45%	6.99%	10.21%	–	38.23%
LGTCOND_DARK-DRVRDOIN_GO_STR	No light (dark) is available at time of crash and driver going straight	1.07%	1.63%	2.70%	1.35%		6.87%
Pedestrian-Weather Related							
LGTCOND_DARK-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	No light (dark) is available at time of crash and pedestrian located in the roadway	1.26%	1.59%	2.84%	2.10%	+ +	9.66%
WTCOND_CLEAR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	The crash occurred under clear weather and pedestrian located in the roadway	1.31%	4.62%	5.92%	13.01%	–	48.99%
Pedestrian-Driver Related							
DRVRDOIN_GO_STR-NMTACT_NF_TRFC	A driver going straight and pedestrian walking not facing traffic	0.19%	0.37%	0.56%	1.59%		6.81%
DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-	The driver going	2.98%	9.98%	12.96%	20.89%	–	30.88%

DISREG-SUDDEN-DK_CLTH-FC_TRFC	straight and pedestrian acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.						
Pedestrian-Specific							
NMTLOC_ATI_MX-NAI_MX-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.	0.00%	0.47%	0.47%	2.75%		5.94%
Total Cashes (N=2145)		**110**	**434**	**544**	**1601**		

6.4.4 Bicyclist Crash Variables Using DT4000 Dataset

Table 6-37: Summary of the Z-test for Proportion Results for the Newly Created Bicyclist Crash Variables

Variable Symbol	Variable Indication	Fatal Injury Crash (K)	Severe Injury Crash (A)	Fatal and Severe Injury (K+A) Crash	Evident and Possible Injury Crash (B+C)	Sig. Result of the Z-Test *	Sample Size
Roadway-Environmental-Related							
LGTCOND_LITE-TRFCWAY_DIV_NO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier	0.07%	0.07%	0.13%	1.34%		3.58%

LGTCOND_LITE-TRFCWAY_UNDIV	Streetlight is available at time of the crash in undivided trafficway	0.07%	1.47%	1.54%	6.68%		20.01%
LGTCOND_DARK-WTCOND_CLDY	No light (dark) is available at time of crash under cloudy weather	0.00%	0.00%	0.00%	0.27%		0.66%
LGTCOND_DARK-WTCOND_CLEAR	No light (dark) is available at time of crash under clear weather	0.13%	0.07%	0.20%	0.33%		1.29%
ROADHOR_LT_RT_CU-ROADVERT_CST_UP_DN_SAG	Curve (not straight) and hill (not level) road terrain	0.13%	0.33%	0.47%	4.61%		12.36%
Driver-Weather Related							
LGTCOND_LITE-DRVRDOIN_GO_STR	Streetlight is available at time of crash and driver going straight	0.07%	1.07%	1.14%	3.14%		10.42%
DRVRDOIN_GO_STR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	A driver going straight and pedestrian located in the roadway	0.27%	1.54%	1.80%	6.22%		19.52%
LGTCOND_DARK-DRVRDOIN_GO_STR	No light (dark) is available at time of crash and driver going straight	0.00%	0.13%	0.13%	0.74%		2.12%
Bicyclist -Weather Related							
LGTCOND_DARK-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	No light (dark) is available at time of crash and pedestrian located in the roadway	0.00%	0.13%	0.13%	0.60%		1.78%
WTCOND_CLEAR-NMTLOC_ATI_NX-ATI_UL-	The crash occurred under clear	0.13%	0.53%	0.67%	4.48%		12.53%

NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	weather and pedestrian located in the roadway						
Bicyclist -Driver Related							
DRVRDOIN_GO_STR-NMTACT_NF_TRFC	A driver going straight and pedestrian walking not facing traffic	0.00%	0.13%	0.13%	0.40%		1.29%
DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC	A driver going straight and pedestrian acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.	0.40%	4.75%	5.15%	21.86%	–	65.74%
Bicyclist -Specific							
NMTLOC_ATI_MX-NAI_MX-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.	0.07%	2.14%	2.21%	32.55%		84.60%
Total Cashes (N=1496)		**11**	**145**	**156**	**1340**		

The same variables were also examined using the DT4000 dataset, basically, since many variables related to roadway, weather, environment, vehicle, crash, and person involved in the crash have been recategorized and some have different meanings. Percentages presented in **Table 6-36** revealed that the most common roadway environmental-related variable that yielded fatal and severe pedestrian crashes is associated with crashes occurred on the curve (not straight) and hill (not level) road terrain (ROADHOR_LT_RT_CU-ROADVERT_CST_UP_DN_SAG; 9.93%). The next most crash variable that yielded fatal and severe pedestrian crashes was when a streetlight is available at the time of the crash in an undivided trafficway (LGTCOND_LITE-TRFCWAY_UNDIV; 5.69%). This result is consistent with the results gained from analyzing the MV4000 dataset.

Crashes occurred when drivers are going straight while pedestrians are located in the roadway (DRVRDOIN_GO_STR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR), and when a streetlight is available at time of crash while drivers going straight (LGTCOND_LITE-DRVRDOIN_GO_STR) were the two most driver weather-related variables that are associated with 6.99%, and 5.13% fatal and severe pedestrian crashes, respectively. Additionally, crashes occurred under clear weather while pedestrians located in the roadway (WTCOND_CLEAR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR), and crashes occurred on roadways with no available light (dark) and while pedestrians located in the roadway (LGTCOND_DARK- NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR), were two most common pedestrian weather-related variables that were responsible for 5.92%, and 2.84% fatal and severe injury pedestrian crashes, respectively. Drivers going straight while pedestrians acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway at the time of the crash (DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC), were responsible for 12.96% of the fatal and severe pedestrian crashes.

It is also observed that five crash variables have significantly lower percentages of fatal and severe versus non-severe injury pedestrian crashes; (LGTCOND_LITE-TRFCWAY_UNDIV),

(LGTCOND_LITE-DRVRDOIN_GO_STR), (DRVRDOIN_GO_STR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR), (WTCOND_CLEAR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR) and (DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC) at the 95% confidence level. Whereas, one variable showed a significantly higher percentage of fatal and severe versus non-severe injury pedestrian crashes; (LGTCOND_DARK-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR).

While for bicyclist crashes, percentages shown in **Table 6-37** displayed that the most common roadway environmental-related variable that yielded severe fatal and severe bicyclist crashes is associated with streetlight available light at the time of the crash and in an undivided trafficway (LGTCOND_LITE-TRFCWAY_UNDIV; 1.54%). Also, crashes occurred when drivers going straight while bicyclists were located in the roadway (DRVRDOIN_GO_STR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR), were responsible for 1.80% of fatal and severe bicyclist crashes.

Furthermore, concerning bicyclist driver-related variables, 5.15% of crashes that produced fatal and severe injuries occurred while drivers were going straight and bicyclists acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway (DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC). For instance, failure to have lights on when required, Operating without required equipment (bicycle reflectors), failure to obey traffic signs, signals, or officer, and being inattentive (talking, eating, Etc.). The percentages suggest that the (DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC) crash variable has significantly lower percentages of fatal and severe versus non-severe injury bicyclist crashes; at the 95% confidence level.

6.5 Crash Variable Selection Using CHAID

This section shows the results of the CHAID analysis applied to vehicle crashes involving pedestrians and bicyclists. In **Table 6-38**, variables used by the CHAID decision tree technique for analyzing pedestrian crashes are shown. Crash severity is the dependent variable used in the analysis. Crash severity was calculated through excluding any observation where the driver sustained more severe injury than the non-motorists, then the ROLE1, ROLE2 data fields were used to determine injury severity of pedestrians and bicyclists separately. Finally, the injury severity for the crash is calculated using the most severe injury sustained among the crash participants or within 30 days to the involved pedestrian or bicyclist. Any crash where the driver sustained a higher injury severity, is excluded. Also, three injury severity levels have been used to study the effect of variables on each level; K (fatal crash), A (severe injury crash), and B+C (evident/possible injury crash), following the path of many researchers who studied crash injury severity of pedestrians and bicyclists involved in vehicle crashes (i.e., (Kemnitzer et al. 2019)). The distribution of crash severity levels is (K 3.30%, A 16.00%, and B+C 80.7%), hence, the crash severity was reported as fatal and severe injury crash as opposed to evident/possible injury crash is 19.30%.

6.5.1 CHAID Decision Tree of MV4000 Dataset

Table 6-38: Dependent and Independent Variables Including Pedestrian-Related Variable Used to Create the Pedestrian Model Using MV4000 Dataset

Dependent Variable (Abbreviation)	Attributes
Crash injury severity (INJSVR)	Fatality (K)
	Severe Injury (A)

	Evident and Possible (B+C)
Independent Variables	
ROADHOR [1,2]	Horizontal road terrain
ROADVERT [1,2]	Vertical road terrain
RDCOND [A,B,C]	Road surface condition
TRFCWAY [1,2]	Trafficway description
RLTNRDWY	Location of the first harmful event
LOCTYPE	Crash location type
TRFCCNTL [1,2]	The traffic control device (TCD) in effect
SURFTYPE [1,2]	Road surface type
TOTLANES [1,2]	Total number of lanes
RLTNTRWY	Crash location with respect to trafficway
INTTYPE	Intersection type where the crash occurred
TRFCINOP [1,2]	Status of the TCD
WTCOND [A, B]	Prevailing atmospheric conditions
LGTCOND	Light conditions
RDWYPC [A, B, C]	Apparent factors of the road/ highway
ENVPC [A,B,C]	Contributing environmental conditions
DRVRPC [1,2] [A,B,C,D]	Driver contributing actions/circumstances
DRVRDOIN [1,2]	Controlled Maneuver by The Driver
NMTSFQ [1,2] [A,B]	Safety Equipment Used by the Driver
RACE [1,2]	Driver race
TEENDRVR	Teen driver
NMTACT [1,2] [A,B]	Pedestrian actions/circumstances contributing to the crash
NMTLOC [1,2]	Pedestrian location with respect to the roadway
NMTSFQ [1,2] [A,B]	Safety equipment used by the pedestrian
DNMFTR [1,2] [A,B]	Individual condition relevant to the crash
NMTPRIOR [1,2]	Pedestrian actions immediately prior to the crash
PEDAGE	The age of the non-motorists involved in a crash in years.
PEDSEX	The sex of the non-motorists involved in a crash.
MOSTHARM [1,2]	Events resulting in the most severe injury
SPEEDFLAG	Vehicle speeding status
HITRUN	Hit and run
VEHTYPE [1,2]	Vehicle type involved in the crash
LGTCOND_LIGT-TRFCWAY_D_WO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier
LGTCOND_LIGT-TRFCWAY_ND	Streetlight is available at time of the crash in undivided trafficway
LGTCOND_DARK-WTHRCOND_CLDY	No light (dark) is available at time of crash under cloudy weather
LGTCOND_DARK-WTHRCOND_CLR	No light (dark) is available at time of crash under clear weather
ROADHOR_C-ROADVERT_H	Curve (not straight) and hill (not level) road terrain

LGTCOND_LIGT-DRVRDO_GO_STR	Streetlight is available at time of crash and driver going straight
DRVRDOIN_GO_STR-NMTLOC_2	A driver going straight and pedestrian located in the roadway
LGTCOND_DARK-DRVRDO_GO_STR	No light (dark) is available at time of crash and driver going straight
LGTCOND_DARK-NMTLOC_2	No light (dark) is available at time of crash and pedestrian located in the roadway
WTHRCOND_CLR-NMTLOC_2	The crash occurred under clear weather and pedestrian located in the roadway
DRVRDO_GO_STR-NMTACT_1	A driver going straight and pedestrian walking not facing traffic
DRVRDO_GO_STR-NMTACT_6	A driver going straight and pedestrian acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.
NMTLOC_1-NMTACT_6	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.

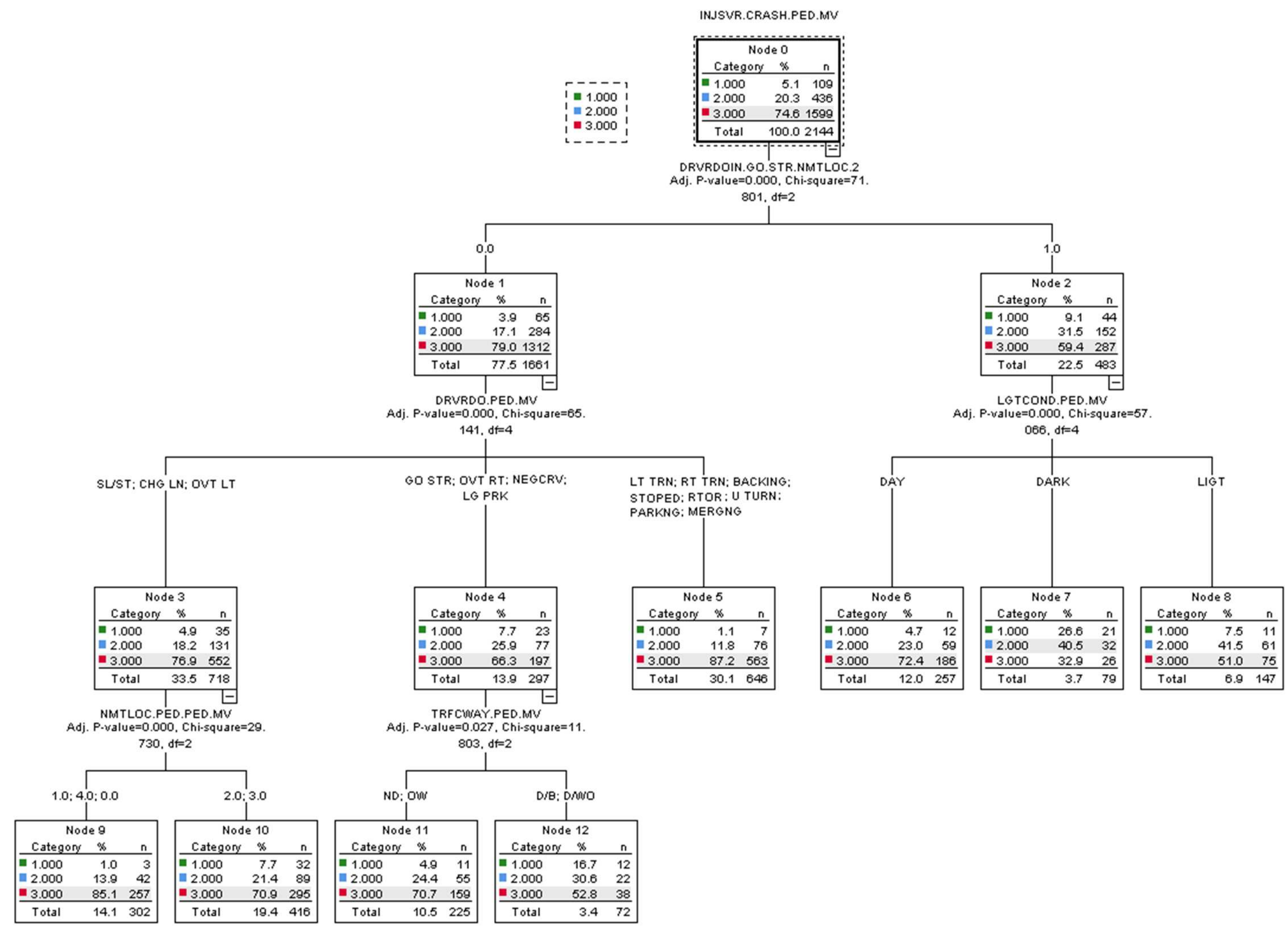


Figure 6-2: CHAID Analysis to Determine Variables that Affect Pedestrians Crash Severity Level Using MV4000 Dataset

The CHAID decision tree for pedestrian crashes is displayed in **Figure 6-2**. Pedestrian crashes resulting in fatal (K), severe injury (A), and evident and possible injury (B+C) were divided into 13 nodes, and 8 terminal nodes. The decision tree structure involves the following five splitting variables:

- A driver going straight with the pedestrian in the roadway (DRVRDOIN.GO.DTR.NMTLOC2)
- Light condition (LGTCOND.PED.MV)
- Driver's action (DRVRDO.PED.MV)
- Trafficway division type (TRFCWAY.PED.MV)
- Pedestrian location (NMTLOC.PED.PED.MV)

The first and top node in the CHAID decision tree output in **Figure 6-2** is "DRVRDOIN.GO.DTR.NMTLOC2", and based on the model, if the driver was going straight and the pedestrian location in the roadway (1.0), the tree predicts 9.1% of fatality crashes and 31.5% of severe injury crashes; if the driver was not going straight and the pedestrian located not in the roadway (0.0), the tree predicts 3.9% of fatality crashes and 17.1% of severe injury crashes.

In the second level of the decision tree, the group including driver was going straight and the pedestrian location in the roadway (1.0) directed to another split in the tree based on the light conditions at the time of the crash. If the crash occurred within the nighttime while streetlights were available (LIGT), the percentage of fatality crash and severe injury crash was 7.5% and 41.5%, respectively. If light conditions were dark and unlit (DARK), the percentage of fatality crash and severe injury crash was 26.6% and 40.5%, respectively. Whereas, if the crash occurred during the daytime (DAY), the percentage of fatality crash and severe injury crash was 4.7% and 23%, respectively.

At the same level, driver action branched from the "DRVRDOIN.GO.DTR.NMTLOC2" node, suggesting the following categorization; first category: slowing/stopped (SL/ST), changing lanes (CHG LN), and overtaking on the left (OVT LT); second category: overtaking on the right (OVT RT), going straight (GO STR), negotiating curve (NEGCRV), and legally parked (LG PRK); third category: making a left turn (LT TRN), making a right turn (RT TRN), backing up (BACKING), stopped in traffic (STOPED), right turn on red (RTOR); U-turn (UTURN), parking maneuver (PARKING), and merging into traffic (MERGING). If the driver's action involved slowing/stopped (SL/ST), changing lanes (CHG LN), or overtaking on the left (OVT LT); the percentage of fatality crash and severe injury crash was 4.9% and 18.2%, respectively. If the driver's action involved overtaking on the right (OVT RT), going straight (GO STR), negotiating curve (NEGCRV), or legally parked (LG PRK), the percentage of fatality crash and severe injury crash was 7.7% and 25.9%, respectively. Whereas, if the driver was making a left turn (LT TRN), making a right turn (RT TRN), backing up (BACKING), stopped in traffic (STOPED), right turn on red (RTOR); U-turn (UTURN), parking maneuver (PARKING), or merging into traffic (MERGING), the percentage of fatality crash and severe injury crash was 1.1% and 11.8%, respectively.

In the third level of the tree, for the group of driver's actions involving overtaking on the right (OVT RT), going straight (GO STR), negotiating curve (NEGCRV), or legally parked (LG PRK), type of trafficway division describing areas designed for motor vehicle operation divided the data into two subgroups: in case of a divided highway with traffic barrier (D/B) and divided highway without traffic barrier (D/WO), the percentages of fatality crash and severe injury crash were 16.7% and 30.6%, respectively. Whereas, in the case of no physical division (ND), or one-way traffic (OW), the percentage of fatality crash and severe injury crash was 4.9% and 24.4%, respectively.

In the same level of the tree, the group of driver's actions involving slowing/stopped (SL/ST), changing lanes (CHG LN), or overtaking on the left (OVT LT) directed to another split in the tree based on pedestrian location. If the pedestrian was located in the roadway (2.0) or not in the roadway (3.0), the percentage of fatality crash and severe injury crash was 7.7% and 21.4%, respectively. Whereas this percentage among pedestrians who were located in a crosswalk (1.0) or on the sidewalk (4.0), dropped down to 1.0% and 13.9% for fatal and severe injury crashes, respectively.

Table 6-39: Dependent and Independent Variables Including Bicyclist-Related Variable Used to Create the Model Using MV4000 Dataset

Dependent Variable (Abbreviation)	**Attributes**
Crash injury severity (INJSVR)	Fatality (K)
	Severe Injury (A)
	Evident and Possible (B+C)
Independent Variables	
ROADHOR [1,2]	Horizontal road terrain
ROADVERT [1,2]	Vertical road terrain
RDCOND A, B, C]	Road surface condition
TRFCWAY [1,2]	Trafficway description
RLTNRDWY	Location of a first harmful event
LOCTYPE	Crash location type
TRFCCNTL [1,2]	The traffic control device (TCD) in effect
SURFTYPE [1,2]	Road surface type
TOTLANES [1,2]	Total number of lanes
RLTNTRWY	Crash location with respect to trafficway
INTTYPE	Intersection type where the crash occurred
TRFCINOP [1,2]	Status of the TCD
WTCOND [A, B]	Prevailing atmospheric conditions
LGTCOND	Light conditions
RDWYPC [A, B, C]	Apparent factors of the road/ highway
ENVPC [A,B,C]	Contributing environmental conditions
DRVRPC [1,2] [A,B,C,D]	Driver contributing actions/circumstances

DRVRDOIN [1,2]	Controlled Maneuver by The Driver
NMTSFQ [1,2] [A,B]	Safety Equipment Used by the Driver
RACE [1,2]	Driver race
TEENDRVR	Teen driver
NMTACT[1,2] [A,B]	Bicyclist actions/circumstances contributing to the crash
NMTLOC [1,2]	Bicyclist location with respect to the roadway
NMTSFQ [1,2] [A,B]	Safety equipment used by the bicyclist
DNMFTR [1,2] [A,B]	Individual condition relevant to the crash
NMTPRIOR [1,2]	Bicyclist actions immediately prior to the crash
BIKESEX	The sex of the non-motorists involved in a crash.
BIKEAGE	The age of the non-motorists involved in a crash in years.
MOSTHARM [1,2]	Events resulting in the most severe injury
SPEEDFLAG	Vehicle speeding status
HITRUN	Hit and run
VEHTYPE [1,2]	Vehicle type involved in the crash
LGTCOND_LIGT-TRFCWAY_D_WO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier
LGTCOND_LIGT-TRFCWAY_ND	Streetlight is available at time of the crash in undivided trafficway
LGTCOND_DARK-WTHRCOND_CLDY	No light (dark) is available at time of crash under cloudy weather
LGTCOND_DARK-WTHRCOND_CLR	No light (dark) is available at time of crash under clear weather
ROADHOR_C- ROADVERT_H	Curve (not straight) and hill (not level) road terrain
LGTCOND_LIGT-DRVRDO_GO_STR	Streetlight is available at time of crash and driver going straight
DRVRDOIN_GO_STR-NMTLOC_2	A driver going straight and pedestrian located in the roadway
LGTCOND_DARK-DRVRDO_GO_STR	No light (dark) is available at time of crash and driver going straight
LGTCOND_DARK-NMTLOC_2	No light (dark) is available at time of crash and pedestrian located in the roadway
WTHRCOND_CLR-NMTLOC_2	The crash occurred under clear weather and pedestrian located in the roadway
DRVRDO_GO_STR-NMTACT_1	A driver going straight and pedestrian walking not facing traffic
DRVRDO_GO_STR-NMTACT_6	A driver going straight and pedestrian acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.
NMTLOC_1-NMTACT_6	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.

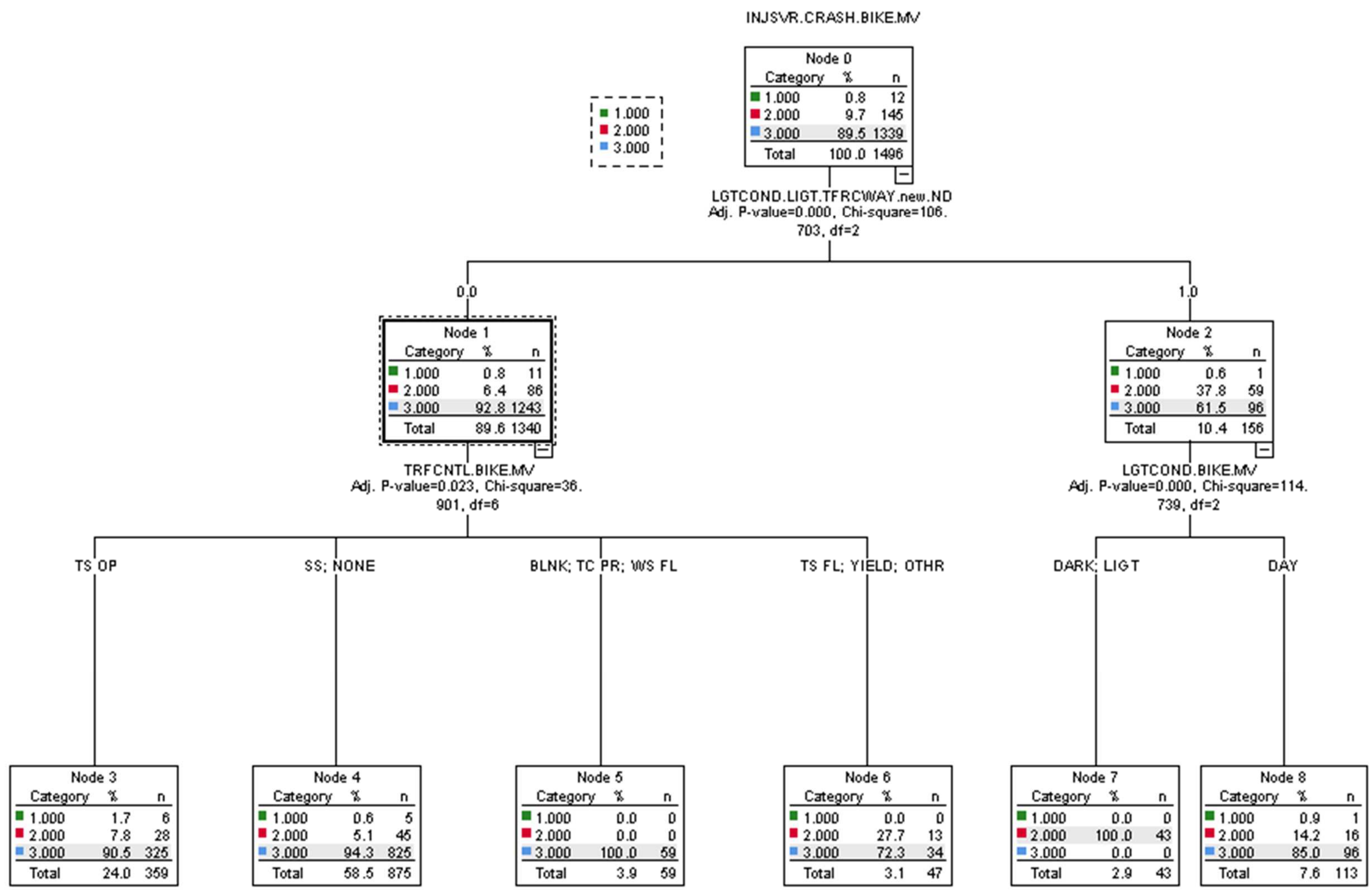


Figure 6-3: CHAID Analysis to Determine Variables that Affect Bicyclists Crash Severity Level Using MV4000 Dataset

The CHAID decision tree for bicyclist crashes is displayed in **Figure 6-3**. Bicyclist crashes resulting in fatal (K), severe injury (A), and evident and possible injury (B+C) were divided into 9 nodes, and 6 terminal nodes. The decision tree structure involves the following three splitting variables:

- Light condition is nighttime with streetlights and trafficway is not physically divided (LGTCOND.LIGT.TRFCWAY.ND)
- Traffic control in effect (TRFCNTL.BIKE.MV)
- Light condition (LGTCOND.BIKE.MV)

The first and top node in the CHAID decision tree output in **Figure 6-3** is "LGTCOND.LIGT.TRFCWAY.ND", and based on the model, if the crash occurred during the nighttime in a location where the light condition involved streetlights and the trafficway is not physically divided (1.0), the tree predicts a percentage of 0.6% and 37.8% of fatal and severe injury crashes, respectively. in the second level, at node 2 "LGTCOND.BIKE.MV", the tree branches to two categories; daytime (DAY), and nighttime/unlit (DARK) and nighttime/with streetlights (LIGT). If the crash occurred during the daytime where the daylight is effective (DAY), the tree predicts 0.9% of fatality crashes and 14.2% of severe injury crashes; if the crash occurred during the nighttime/unlit (DARK) or nighttime/with streetlights (LIGT), the tree predicts 100% of severe injury crashes.

At the same level of the decision tree, the group that did not include crashes occurred during the nighttime in a location where the light condition involved streetlights and the trafficway is not physically divided (0.0), split in the tree based on the traffic control in effect (TRFCNTL.BIKE.MV). If the traffic control device (TCD) in effect includes traffic signal flashing (TS/FL), yield sign (YIELD), or other devices (i.e., RR-xing signal (RRSIG), stop sign with flasher

(SS FL), warning sign (WS)), the percentage of predicted severe injury crashes was 27.7%; if the TCD in effect includes warning sign with flasher (WS FL), a traffic control person was available at the crash location (TC PR), or the filed was left blank, the tree predicts 0.0% of fatality and severe injury crashes; if the TCD involves stop sign (SS), or no TCD available (NONE), the tree predicts 5.1% of severe injury crashes. Whereas, in the case of a traffic signal operating in the crash location (TS OP), the percentage of fatality crash and severe injury crash was 1.7% and 7.8%, respectively.

6.5.2 CHAID Decision Tree of DT4000 Dataset

Table 6-40: Dependent and Independent Variables Including Pedestrian-Related Variable Used to Create the Pedestrian Model Using DT4000 Dataset

Dependent Variable (Abbreviation)	**Attributes**
Crash injury severity (INJSVR)	Fatality (K)
	Severe Injury (A)
	Evident and Possible (B+C)
Independent Variables	
ROADHOR [1,2]	Horizontal road terrain
ROADVERT [1,2]	Vertical road terrain
RDCOND [A,B,C]	Road surface condition
TRFCWAY [1,2]	Trafficway description
RLTNRDWY	Location of the first harmful event
LOCTYPE	Crash location type
TRFCCNTL [1,2]	The traffic control device (TCD) in effect
SURFTYPE [1,2]	Road surface type
TOTLANES [1,2]	Total number of lanes
RLTNTRWY	Crash location with respect to trafficway
INTTYPE	Intersection type where the crash occurred
TRFCINOP [1,2]	Status of the TCD
WTCOND [A, B]	Prevailing atmospheric conditions
LGTCOND	Light conditions
RDWYPC [A, B, C]	Apparent factors of the road/ highway
ENVPC [A,B,C]	Contributing environmental conditions
DRVRPC [1,2] [A,B,C,D]	Driver contributing actions/circumstances
DRVRDOIN [1,2]	Controlled Maneuver by The Driver
NMTSFQ [1,2] [A,B]	Safety Equipment Used by the Driver
RACE [1,2]	Driver race
TEENDRVR	Teen driver

NMTACT [1,2] [A,B]	Pedestrian actions/circumstances contributing to the crash
NMTLOC [1,2]	Pedestrian location with respect to the roadway
NMTSFQ [1,2] [A,B]	Safety equipment used by the pedestrian
DNMFTR [1,2] [A,B]	Individual condition relevant to the crash
NMTPRIOR [1,2]	Pedestrian actions immediately prior to the crash
PEDAGE	The age of the non-motorists involved in a crash in years.
PEDSEX	The sex of the non-motorists involved in a crash
MOSTHARM [1,2]	Events resulting in the most severe injury
SPEEDFLAG	Vehicle speeding status
HITRUN	Hit and run
VEHTYPE [1,2]	Vehicle type involved in the crash
LGTCOND_LITE-TRFCWAY_DIV_NO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier
LGTCOND_LITE-TRFCWAY_UNDIV	Streetlight is available at time of the crash in undivided trafficway
LGTCOND_DARK-WTCOND_CLDY	No light (dark) is available at time of crash under cloudy weather
LGTCOND_DARK-WTCOND_CLEAR	No light (dark) is available at time of crash under clear weather
ROADHOR_LT_RT_CU-ROADVERT_CST_UP_DN_SAG	Curve (not straight) and hill (not level) road terrain
LGTCOND_LITE-DRVRDOIN_GO_STR	Streetlight is available at time of crash and driver going straight
DRVRDOIN_GO_STR-NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	A driver going straight and pedestrian located in the roadway
LGTCOND_DARK-DRVRDOIN_GO_STR	No light (dark) is available at time of crash and driver going straight
LGTCOND_DARK- NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	No light (dark) is available at time of crash and pedestrian located in the roadway
WTCOND_CLEAR- NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	The crash occurred under clear weather and pedestrian located in the roadway
DRVRDOIN_GO_STR-NMTACT_NF_TRFC	A driver going straight and pedestrian walking not facing traffic
DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC	A driver going straight and pedestrian acting other than disregarding signal,

	walking not facing traffic, wearing dark clothes, and darting into roadway
NMTLOC_ATI_MX-NAI_MX-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into roadway

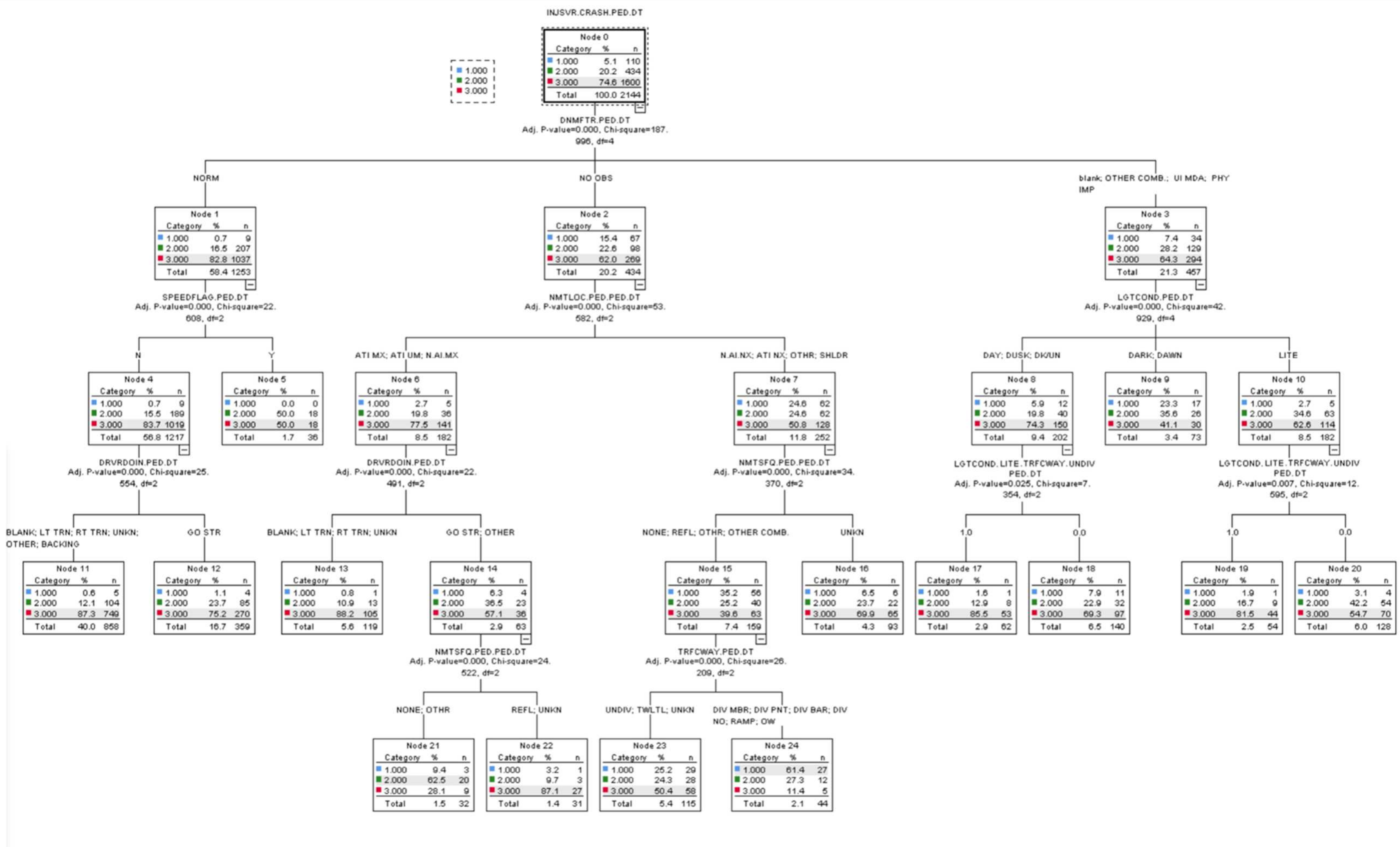


Figure 6-4: CHAID Analysis to Determine Variables that Affect Pedestrians Crash Severity Level Using DT4000 Dataset

The CHAID decision tree for pedestrian crashes is displayed in **Figure 6-4**. Pedestrian crashes resulting in fatal (K), severe injury (A), and evident and possible injury (B+C) were divided into 25 nodes, and 14 terminal nodes. The decision tree structure involves the following seven splitting variables:

- Pedestrian condition relevant to the crash (DNMFTR.PED.DT)
- Light condition (LGTCOND.PED.DT)
- Pedestrian location (NMTLOC.PED.PED.DT)
- Speed flag (SPEEDFLAG.PED.DT)
- Light condition is streetlight at nighttime and trafficway is undivided (LGTCOND.LITE.TRFCWAY.UNDIV)
- Safety equipment used by the pedestrian (NMTSFQ.PED.DT)
- Vehicle controlled maneuver prior to the beginning of the sequence of events (DRVRDOIN.PED.DT)
- Trafficway division (TRFCWAY.PED.DT)

The first and top node in the CHAID decision tree output in **Figure 6-4** is "DNMFTR.PED.DT", and based on the model, it is best to categorize the "DNMFTR.PED.DT" data field in the following categories: pedestrian appeared in normal condition (NORM) in a category, not observed condition (NO OBS) in a category, and involved other combinations (OTHER COMB) such as under influence of medication/drugs/alcohol (UI MDA), physically impaired (PHY IMP), emotional, i.e., depressed, angry, disturbed (EMO), ill/sick/fainted (SICK), confused/disoriented (CONF), paraplegic/restricted to a wheelchair (WCHAIR), blind (BLIND), uses cane/crutches (CANE), or the field is left blank (blank) in another category. If the pedestrian appeared in normal condition (NORM), the tree predicts 0.7% of fatality crashes and 16.5% of severe injury crashes; if the pedestrian condition was not observed (NO OBS), the tree predicts higher percentages of fatality and severe injury crashes; 15.4% and 22.6%, respectively. Whereas, if the pedestrian condition involved other combinations (OTHER COMB.) such as being under influence of medication/drugs/alcohol (UI MDA), physically impaired (PHY IMP), emotional, i.e., depressed, angry, disturbed (EMO), ill/sick/fainted (SICK), confused/disoriented (CONF), paraplegic/restricted to a wheelchair (WCHAIR), blind (BLIND), uses cane/crutches (CANE), or the field is left blank (blank), the tree predicts 7.4% of fatality crashes and the highest percentage of severe injury crashes; 28.2%.

In the first level of the decision tree, the group including other combinations (OTHER COMB.) such as (UI MDA, PHY IMP, EMO, SICK, CONF, WCHAIR, BLIND, CANE), or the field is left blank (blank), directed to another split in the tree based on the light condition existed at the time of the crash. If the crash occurred during the dark time with the lighted roadway (LITE), the percentage of predicted fatality crashes is the least (2.7%) but the predicted severe injury crashes were 34.6%; if the light condition was dark/unlit (DARK) or dawn (DAWN),), the percentage of predicted fatality crashes is the highest (23.3%), and the percentage of predicted severe injury crashes was 35.6%; if the light condition involved daylight (DAY), dusk (DUSK), or dark/unknown lighting (DK/UN), the percentage of predicted fatality crash and severe injury crash was 23.3% and 35.6%, respectively. In the same level of the tree, under the unobserved pedestrian condition (NO OBS) group, the tree split based on the pedestrian location into two categories; pedestrian location involving being at intersection-in marked crosswalk (ATI MX), at intersection-unmarked / unknown if marked crosswalk (ATI UM), or not at intersection-in marked crosswalk (NAI MX) category, and Not At Intersection-On Roadway, Not In Marked Crosswalk (NAI NX), At Intersection-Not In Crosswalk (ATI NX), shoulder/roadside (SHLDR), and other locations (OTHR) such as parking lane/zone (PK LN), bicycle lane (BIKE LN), sidewalk (SDWLK), median/crossing island (MEDIAN), Driveway Access (DRWAY), etc. in another category.

The third node in the first level (NORM) splits based on whether speed was a factor in the crash (SPEEDFLAG). The group including a speed flag (Y) predicts half of the severe injury crashes, whereas, the group not including a speed flag (N) splits based on the vehicle-controlled maneuver, into two categories; drivers going straight (GO STR), and another category involving other maneuvers such as turning left (LT TRN), right turn (RT TRN), and backing (BACKING). The (GO STR) group predicts that the percentage of the fatality and severe injury crashes is 1.1% and 23.7%, respectively. Whereas, the group involving other maneuvers (LT TRN, RT TRN, BACKING) predicts that the percentage of the fatality and severe injury crashes is 0.6% and 12.1%, respectively. At the same level, the light condition group of (LITE) splits based on the crashes occurred when the light condition is streetlight at nighttime and the trafficway is undivided. If the crash occurred in the dark where streetlights are available and trafficway is undivided (Y), the tree predicts 1.9% and 16.7% of the fatality and severe injury crashes, respectively. otherwise, it predicts a slightly higher percentage of fatality and severe injury crashes, 1.9%, and 16.7%, respectively. The light condition group involving (DAY, DUSK, DK/UN) splits also based on the (LGTCOND.LITE.TRFCWAY.UNDIV) and predicts 1.6% and 12.9% of fatal and severe injury crashes in case of crashes involving occurred while streetlight at nighttime is available and trafficway is undivided (1.0).

In the third level, two categories of safety equipment usage split from the group of the (N.AI.NX, ATI.NX, SHLDR, OTHR) group into two categories; the first includes crashes occurred while the pedestrian was not using any safety equipment (NONE), using reflective clothing (REFL), or other equipment (OTHR), and the second involves unknown safety equipment (UNKN). Both groups predict a close percentage of severe injury crashes, 23.7%, and 25.5%, for the first and second groups, respectively. However; the first group (REFL, NONE, OTHR) predicts a high percentage of fatal crashes compared to the second group (UNKN), 35.2%, and 6.5%, respectively.

In the fourth level of the tree, the controlled maneuver segmented the data regarding the trafficway division (TRFCWAY.PED.DT) into two groups. If the trafficway was two-way, not divided, with a continuous left-turn lane (TWLTL), two-way, not divided (UNDIV), or trafficway division is unknown (UNKN), the percentage of fatality crash and severe injury crash was 25.2% and 24.3%, respectively. Whereas, if the trafficway was divided highway without a traffic barrier (DIV NO), two-way, divided, unprotected (painted > 4 feet) median (DIV PNT), divided highway with traffic barrier (DIV BAR), divided highway median with a barrier (DIV MBR), one-way traffic (OW), parking lot or private property (PL/PP), or entrance/exit ramp (RAMP), the percentage of fatality crash and severe injury crash was 61.4% and 27.3%, respectively. Additionally, in the fourth level, the (GO STR, OTHR) group of controlled maneuvers, led to another split among the group of pedestrians using safety equipment: if the pedestrian was wearing reflective clothing/backpack, or if the equipment used was unknown (UNKN), the percentage predicted for fatality crash severe injury crashes is 3.2% and 9.7%. Whereas, if the pedestrian was using other equipment (OTHR) or not using safety equipment (NONE), the percentage of fatality crash and severe injury crash rise to 9.4% and 62.5%, respectively.

Table 6-41: Dependent and Independent Variables Including Bicyclist-Related Variable Used to Create the Model Using DT4000 Dataset

Dependent Variable (Abbreviation)	**Attributes**
Crash injury severity (INJSVR)	Fatality (K)
	Severe Injury (A)
	Evident and Possible (B+C)
Independent Variables	
ROADHOR [1,2]	Horizontal road terrain
ROADVERT [1,2]	Vertical road terrain
RDCOND [A,B,C]	Road surface condition
TRFCWAY [1,2]	Trafficway description
RLTNRDWY	Location of a first harmful event
LOCTYPE	Crash location type
TRFCCNTL [1,2]	The traffic control device (TCD) in effect
SURFTYPE [1,2]	Road surface type
TOTLANES [1,2]	Total number of lanes
RLTNTRWY	Crash location with respect to trafficway
INTTYPE	Intersection type where the crash occurred
TRFCINOP [1,2]	Status of the TCD
WTCOND [A, B]	Prevailing atmospheric conditions
LGTCOND	Light conditions
RDWYPC [A, B, C]	Apparent factors of the road/ highway
ENVPC [A,B,C]	Contributing environmental conditions
DRVRPC [1,2] [A,B,C,D]	Driver contributing actions/circumstances
DRVRDOIN [1,2]	Controlled Maneuver by The Driver
NMTSFQ [1,2] [A,B]	Safety Equipment Used by the Driver
RACE [1,2]	Driver race
TEENDRVR	Teen driver
NMTACT [1,2] [A,B]	Pedestrian actions/circumstances contributing to the crash
NMTLOC [1,2]	Pedestrian location with respect to the roadway
NMTSFQ [1,2] [A,B]	Safety equipment used by the pedestrian
DNMFTR [1,2] [A,B]	Individual condition relevant to the crash
NMTPRIOR [1,2]	Pedestrian actions immediately prior to the crash
PEDAGE	The age of the non-motorists involved in a crash in years.
PEDSEX	The sex of the non-motorists involved in a crash.
MOSTHARM [1,2]	Events resulting in the most severe injury
SPEEDFLAG	Vehicle speeding status

HITRUN	Hit and run
VEHTYPE [1,2]	Vehicle type involved in the crash
LGTCOND_LITE-TRFCWAY_DIV_NO	Streetlight is available at time of the crash in a divided trafficway without a traffic barrier
LGTCOND_LITE-TRFCWAY_UNDIV	Streetlight is available at time of the crash in undivided trafficway
LGTCOND_DARK-WTCOND_CLDY	No light (dark) is available at time of crash under cloudy weather
LGTCOND_DARK-WTCOND_CLEAR	No light (dark) is available at time of crash under clear weather
ROADHOR_LT_RT_CU-ROADVERT_CST_UP_DN_SAG	Curve (not straight) and hill (not level) road terrain
LGTCOND_LITE-DRVRDOIN_GO_STR	Streetlight is available at time of crash and driver going straight
DRVRDOIN_GO_STR- NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	A driver going straight and pedestrian located in the roadway
LGTCOND_DARK-DRVRDOIN_GO_STR	No light (dark) is available at time of crash and driver going straight
LGTCOND_DARK- NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	No light (dark) is available at time of crash and pedestrian located in the roadway
WTCOND_CLEAR- NMTLOC_ATI_NX-ATI_UL-NAI_MX-NAI_NX-NAI_UN-PK_LN-BIKE_LN-SHLDR	The crash occurred under clear weather and pedestrian located in the roadway
DRVRDOIN_GO_STR- NMTACT_NF_TRFC	A driver going straight and pedestrian walking not facing traffic
DRVRDOIN_GO_STR-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC	A driver going straight and pedestrian acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.
NMTLOC_ATI_MX-NAI_MX-NMTACT_OTHER_NF_TRFC-DISREG-SUDDEN-DK_CLTH-FC_TRFC	Pedestrian located in the roadway and acting other than disregarding signal, walking not facing traffic, wearing dark clothes, and darting into the roadway.

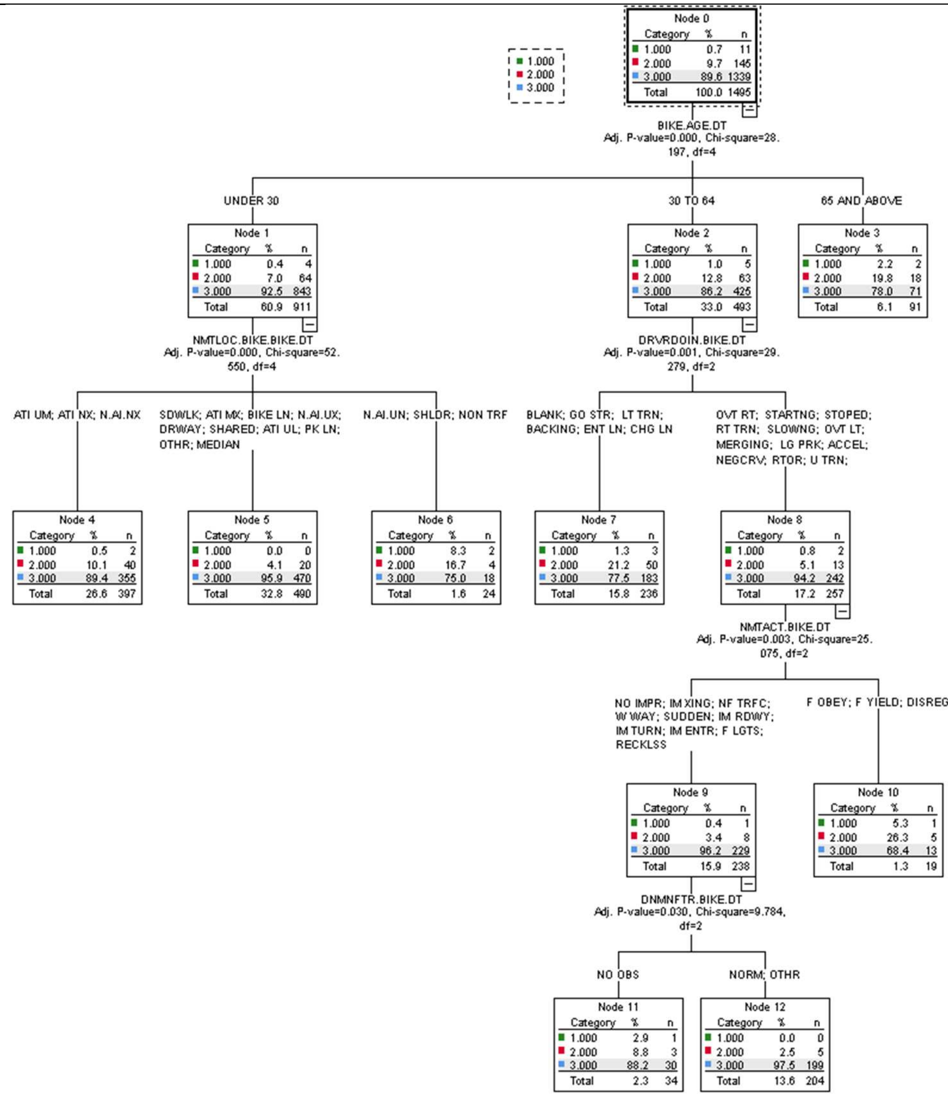


Figure 6-5: CHAID Analysis to Determine Variables that Affect Bicyclist Crash Severity Level Using DT4000 Dataset

The CHAID decision tree for bicyclist crashes is displayed in **Figure 6-5** Bicyclist crashes resulting in fatal (K), severe injury (A), and evident and possible injury (B+C) were divided into 13 nodes, and 8 terminal nodes. The decision tree structure involves the following four splitting variables:

- The age of the bicyclist involved in a crash (BIKE.AGE.DT)
- Vehicle controlled maneuver prior to the beginning of the sequence of events (DRVRDOIN.BIKE.DT)
- Bicyclist location (NMTLOC.BIKE.BIKE.DT)
- Bicyclist action (NMTACT.BIKE.DT)
- Bicyclist condition relevant to the crash (DNMFTR.BIKE.DT)

The first and top node in the CHAID decision tree output in **Figure 6-5** is "BIKE.AGE.DT", and based on the model, if the bicyclist is younger than 30 years (UNDER), the tree predicts 0.4% of fatality crashes and 4.0% of severe injury crashes; if the bicyclist age is between 30 and 64 (30 TO 64), the tree predicts a higher percentage of fatality crashes (1.1%) and severe injury crashes (10.9%); if the bicyclist is older than 65 years, the tree predicts the highest percentage of fatal and severe injury crashes among the other two groups, 2.2%, and 19.8%, respectively.

In the second level of the decision tree, the group including bicyclists age between 30 and 64 (30 TO 64), directs to another split in the tree based on vehicle-controlled maneuver prior to the beginning of the sequence of events (DRVRDOIN.BIKE.DT). if the driver was going straight (GO STR), making a left turn (LT TRN), backing (BACKING), changing lanes (CHG LN), or entering traffic lane (ENT LN), the tree predicts 1.3% and 21.2% of fatal and severe injury crashes. Whereas, if the driver was negotiating curve (NEGCRV), overtaking right (OVT RT), overtaking left (OVT LT), making a right turn (RT TRN), making U-turn (U TRN), slowing/stopping

(SLOWING), legally parked (LG PRK), stopped in traffic (STOPED), turning on red (RTOR), merging (MERGING), accelerating in the road (ACCEL), or starting in the road (STARTING), the tree predicts fewer percentages; 0.8% of fatal crashes and 5.1% of severe injury crashes. In the same level (level two) of the tree, the (UNDER 30) group divide the data into three groups based on the bicyclist's location with respect to the roadway: the first includes bicyclists who are located at intersection-unmarked / unknown if marked crosswalk (ATI UM), at intersection-not in the crosswalk (ATI NX), and not at the intersection-on roadway, not in a marked crosswalk (NAI NX); the second includes bicyclists located at intersection-in marked crosswalk (ATI MX), at the intersection-unknown location (ATI UL), not at intersection-in marked crosswalk (NAI MX), parking lane/zone (PK LN), shared-use path (SHARED), bicycle lane (BIKE LN), on a sidewalk (SDWLK), median/crossing island (MEDIAN), in driveway access (DRWAY), or other locations (OTHR); and the third includes bicyclists located in a non-traffic area (NON TRF), shoulder/roadside (SHLDR), and not at the intersection-on roadway, crosswalk availability unknown (NAI UN). For fatal crashes, the first group (ATI UM, ATI NX, N.AI.NX), the second (ATI MX, ATI UL, NAI MX, PK LN, SHARED, BIKE LN, SDWLK, DRWAY, OTHR), and the third (NON TRF, SHLDR, NAI UN) predict 0.5%, 0.00%, and 8.3%, respectively. Whereas, first, the second and third group predicts 10.1%, 4.1%, and 16.7% of the severe injury crashes, respectively.

In the third level of the tree, for the (NEGCRV, OVT RT, OVT LT, RT TRN, U TRN, SLOWING, LG PRK, STOPED, RTOR, MERGING, ACCEL, STARTING) group of drivers, bicyclist action divides the data into two subgroups: in case the bicyclist disregarded signal (DISREG), failed to yield the ROW (F YIELD), or failed to obey traffic signs/signals/officer (F OBEY), the percentage of fatality crash and severe injury crash was 5.3% and 26.3%, respectively.

Whereas in the case of bicyclists who improperly cross the roadway/jaywalking (IM XING), in roadway improperly standing/playing (IM RDWY), improper turn/merge (IM TURN), wrong-way riding (W WAY), failing to have lights on when bicycling (F LGTS), sudden movement into traffic (SUDDEN), operating in other erratic, reckless or careless manner (RECKLSS), cycling not facing traffic (NF TRFC), making improper entry to or exit from trafficway (IM ENTR), or did not act improperly (NO IMPR), the percentage of fatality crash and severe injury crash was 0.4% and 3.4%, respectively.

In the fourth level of the tree, bicyclists' actions (IM XING, IM RDWY, IM TURN, W WAY, F LGTS, SUDDEN, RECKLSS, NF TRFC, IM ENTR, NO IMPR) directed to another split in the tree based on the bicyclist condition. If the condition includes appeared normal (NORM) or involved other conditions (OTHR) such as being sick or physically impaired (PHY IMP), the percentage of severe injury crashes was to 2.5%. Whereas, if there is no observation regarding the bicyclists' condition (NO OBS), the percentage of fatal and severe injury crashes was 2.9% and 8.8%, respectively.

6.6 Crash Variable Importance Ranking Using Random Forests

Random forests (RF) are a scheme proposed by Leo Breiman in the 2000s for building a predictor ensemble with a set of decision trees that grow in randomly selected subspaces of data (Biau 2012). This technique has been commonly used to label important variables in splitting response variables such as (Jiang et al. 2016; Hong Han, Xiaoling Guo, and Hua Yu 2016; J. Lee, Abdel-Aty, and Shah 2018; Dai 2020). In general, a random forest model combines a set of unpruned decision trees (DT) (i.e., CART). Readers who are interested in understanding the CART procedure may refer to Das et al. (A. Das, Abdel-Aty, and Pande 2009) and Hossain and

Muromachi (Hossain and Muromachi 2013) as they present an extensive description of the CART algorithm. Classification trees are used to classify observations thru recursively partitioning the predictor space. As the number of trees in an RF increases, the misclassification rate converges to a limit. Hence, RF models with too many trees are free of the overfitting problem. As the forest building improves, RF uses an internal mechanism that achieves the unbiased generalization error estimate.

Through the forest building process, two datasets are generated from the complete dataset; a training dataset and a test dataset. On average, about one-third of observations are in the test dataset, which is named Out-of-Bag (OOB) samples by Breiman (Breiman 2001) which are used to estimate the RF classifier generalization error. The OOB is an RF measurement method for prediction error. The rest forms the training dataset, where each tree is trained on bootstrap samples of this dataset. The RF model performance may be enhanced by decreasing the bias of each tree by growing each tree to the maximum depth. Also, by decreasing the correlations between trees through applying two sources of randomization in each tree: a) Each tree is grown on a bootstrap sample of the training dataset (randomly drawn, with replacement) b) At each node of a tree, a certain number of variables "mtry" are randomly selected from the complete explanatory variables to compete for the best split. "mtry" is the number of input variables randomly chosen at each split, and it can be tuned by increasing or decreasing from an initial value until the minimized error rate is obtained (Jiang et al. 2016).

In this study, the two datasets (MV4000 and DT4000) with a total of 90 predictor variables were imported for variable importance analysis. The important explanatory variables in the crash model were determined by an RF model, starting with fitting an RF to the data. At that point, the OOB is recorded for each data point. This error is then averaged over the forest. To measure the importance score of the importance of a variable after training, the values are permuted among the training dataset and the OOB error is recorded. Then, the difference in before and after permutation OOB error is averaged among all trees, showing the importance score of the variable. Afterward, the standard deviation of the difference values is used to normalize the importance score.

Variables with higher importance score values are ranked as more important than other variables (Breiman 2001). In this study, the RF technique is constructed in the RStudio (V 1.2.1335) "randomForest" package, and "mtry" is used as a tuning model parameter. Regarding the number of trees in the forest (ntree), 500 trees were run for each model to obtain relatively consistent variable importance measures. Concerning the importance, the OOB error was used, but for variable impurity, two indices were used: the Mean Decrease Accuracy (MDA) and the Mean Decrease Gini (MDG) indices.

The two indices; MDA and MDG are used to evaluate the importance of each variable since the Gini index is suitable for classification, both indices are default output of the RF procedure, and using both indices is more robust than using one index (Hong Han, Xiaoling Guo, and Hua Yu 2016). As the MDA value gets larger, the variable importance increase. Whereas, MDG shows the total decrease in node impurities averaged for all trees.

Figure 6-6 to Error! Reference source not found. shows RF variable importance ranking for pedestrian and bicyclist-related variables using MV4000 and DT4000 datasets. The importance score of variables in the prediction of pedestrian and bicyclist's injury severity was carried out using the Random Forests method for each dataset; MV4000 and DT4000 crash forms. The method was implemented with 500 trees, using a training dataset of 70% of the crash observations, and using "mtry" of ($\sqrt{p}$) where p is the number of studied variables. Also, the newly created variables from the DT4000 dataset were included in the importance ranking process for variables adopted from the DT4000 crash form dataset. Note that common pedestrian and bicyclist variables in MV4000 and DT4000 datasets are used as well as the newly created variables showed in **Table 6-38** to

Table *6-41* for a more consistent comparison.

6.6.1 MV4000 Pedestrian/Bicyclist-Vehicle Crash Variables

Generally, many variables showed strong effects on injury severity of pedestrians and bicyclists involved in vehicle crashes. The most important variables are highlighted as follows: The newly created variable (DRVRDOIN.GO.STR.NMTLOC2) which refers to crashes involving the driver going straight while the pedestrian located in the roadway, light condition at the time of the crash (LGTCOND.PED.MV), vehicle-controlled maneuver prior to the event leading to the crash (DRVRDO.PED.MV), and the traffic control device (TCD) in effect at the crash location (TRFCNTL.PED.MV), were important for injury severity in pedestrian-vehicles crashes. Other variables showed a level of importance, such as the newly created variable (LGTCOND.LIGT.DRVRDOIN.GO.STR) which refers to dark/streetlight crash location with the driver going straight, roadway surface condition (ROADCOND.PED.MV), pedestrian location concerning the roadway were also important factors affecting pedestrian injury severity in pedestrian-vehicle crashes (NMTLOC.PED.MV).

As for bicyclist injury severity, three variables showed a strong effect; the newly created variable (LGTCOND.LIGT.TRFCWAY.ND) which refers to the crashes occurred in the dark but with streetlights available on an undivided roadway, light condition at the crash location (LGTCOND.BIKE.MV), and the trafficway type and level of division if available

(TRFCWAY.BIKE.MV). Bicyclist's location concerning the roadway (NMTLOC.BIKE.MV) had a significant influence on bicyclist injury severity, but not among the most important variables.

It is observed from the results shown in **Figure 6-6** and **Figure 6-7**, that light condition at the time of the crash (LGTCOND.PED.MV and LGTCOND.BIKE.MV), and the non-motorist location concerning the roadway (NMTLOC.PED.MV and NMTLOC.BIKE.MV) had a common significant effect on pedestrian and bicyclist injury severity. As for pedestrian and bicyclist injury severity, the light condition variable had an equal importance ranking for both non-motorists involved in vehicle crashes. It is noted that trafficway division type and level had a higher importance ranking for bicyclist crashes than pedestrian crashes. This finding is because a bicycle is considered a vehicle and mainly follows traffic rules and have an interaction with the roadway geometry more than a pedestrian.

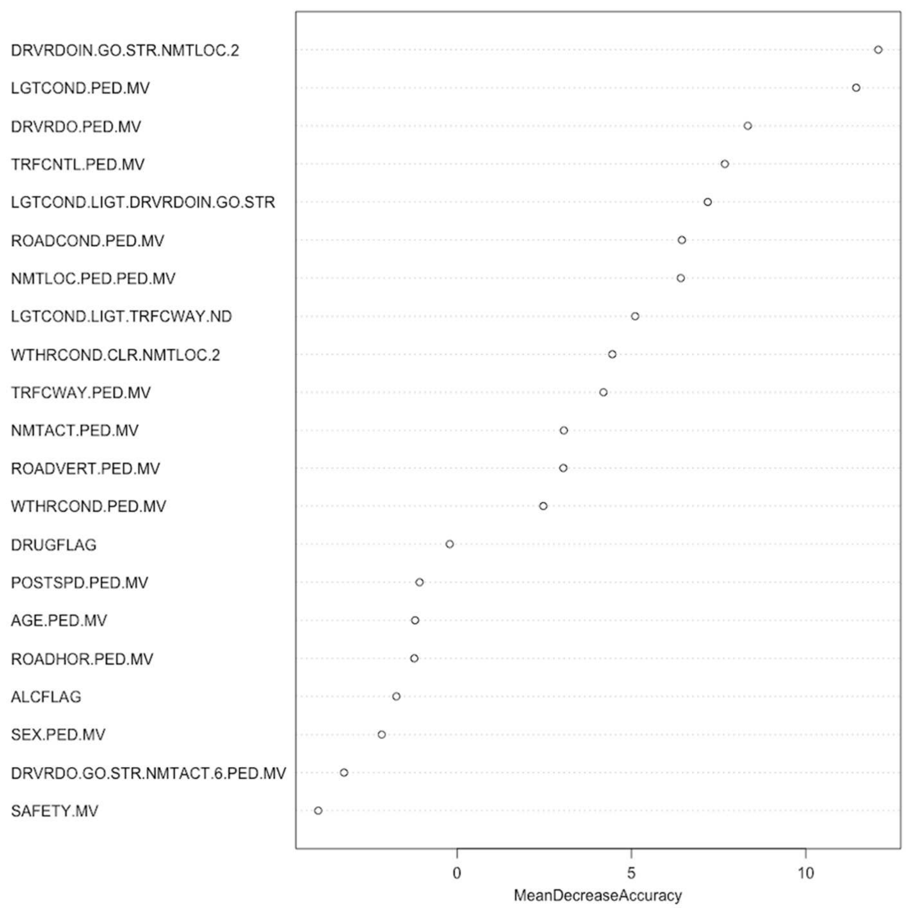


Figure 6-6: RF for MV4000 Crash Form Variable Importance Ranking for Pedestrian Crashes

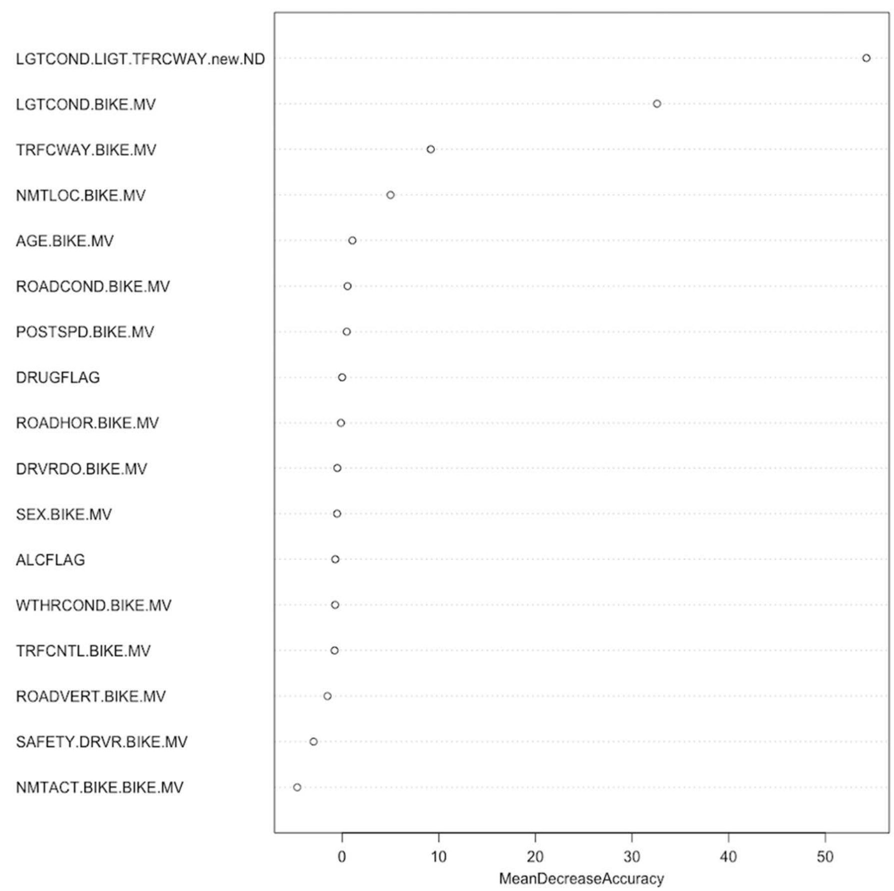


Figure 6-7: RF for MV4000 Crash Form Variable Importance Ranking for Bicyclist Crashes

6.6.2 DT4000 Pedestrian/Bicyclist-Vehicle Crash Variables

More variables showed strong effects on injury severity of pedestrians and bicyclists involved in vehicle crashes using the DT4000 dataset. This conclusion is obtained from the recategorization process that was applied to a various number of variables, as well as the adding new variables which enhance the information gathered from the crash reports. The most important variables are highlighted as follows:

Error! Reference source not found. displayed that pedestrian condition at the time of the crash (DNMFTR.PED.DT), pedestrian location with respect to the roadway (NMTLOC.PED.PED.DT), whether speed was a factor in a crash (SPEEDFLAG), and the type of traffic control device (TCD) available at the crash location (TRFCCNTL.PED.DT), were the most important factors for injury severity in pedestrian-vehicles crashes.

Other variables showed a level of importance, such as the newly created variables; (LGTCOND.DARK.DRVRDOIN.GO.STR) which refers to dark/unlit crash location with the driver going straight, (LGTCOND.LIGT.DRVRDOIN.GO.STR) which refers to dark/streetlight crash location with the driver going straight, light condition at the crash location (LGTCOND.PED.DT), driver actions that may have contributed to the crash (DRVRPC.PED.DT), and the safety equipment in use by the pedestrian at the time of the crash (NMTSFQ.PED.PED.DT).

Whereas, for bicyclist injury severity, it is observed in **Error! Reference source not found.** hat two variables showed a strong effect. First, the bicyclist age (BIKE.AGE.DT). This relates to the fact that bicyclist's age plays a role in the comprehension level of traffic rules and understanding the ROW. Second, the vehicle-controlled maneuver prior to the beginning of the sequence of events (DRVRDOIN.BIKE.DT), which relates to the consistent interaction between the driver and the bicyclist, implying that drivers' actions greatly affect the bicyclist's riding experience.

Additional variables also showed the importance in affecting the bicyclist injury severity, such as the vertical terrain of the roadway (ROADVERT.BIKE.DT), bicyclist condition at the time of the crash (DNMFTR.BIKE.DT), the safety equipment used by the bicyclist (NMTSFQ.BIKE.DT), bicyclist action that may have contributed to the crash (NMTACT.BIKE.DT), the horizontal road terrain (ROADHOR.BIKE.DT), and the bicyclist location with respect to the roadway (NMTLOC.BIKE.DT). Trafficway division type and level (TRFCWAY.BIKE.DT) had a significant influence on bicyclist injury severity, but not among the most important variables.

It is observed from the results shown in **Error! Reference source not found.** and **Error! ference source not found.**, that the non-motorists condition (DNMFTR.PED.DT and DNMFTR.BIKE.DT), and the non-motorist location (NMTLOC.PED.DT and NMTLOC.BIKE.DT) are two common important variables that are related to pedestrian and bicyclist crash severity.

Even though the RF method is capable to detect variable's importance score, obtaining the knowledge about whether a change in the value or category of the specific variable will increase or decrease the pedestrian or bicyclist's injury severity is deemed challenging.

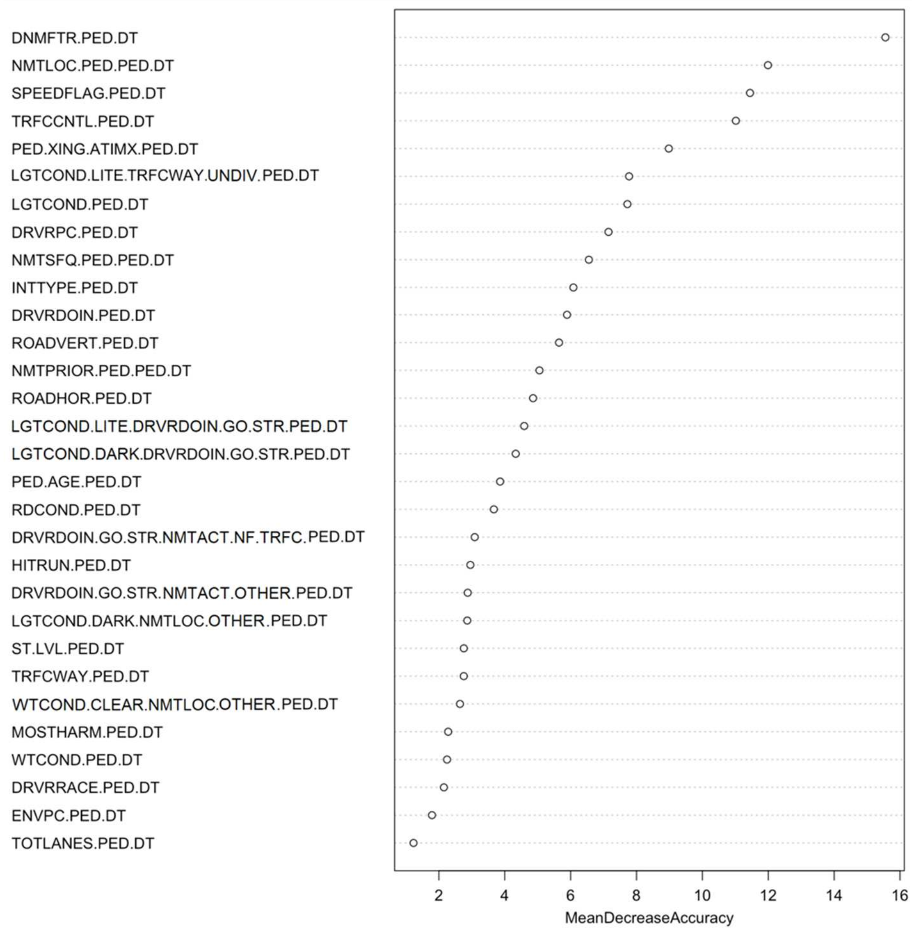


Figure 6-8: RF for DT4000 Crash Form Variable Importance Ranking for Pedestrian Crashes

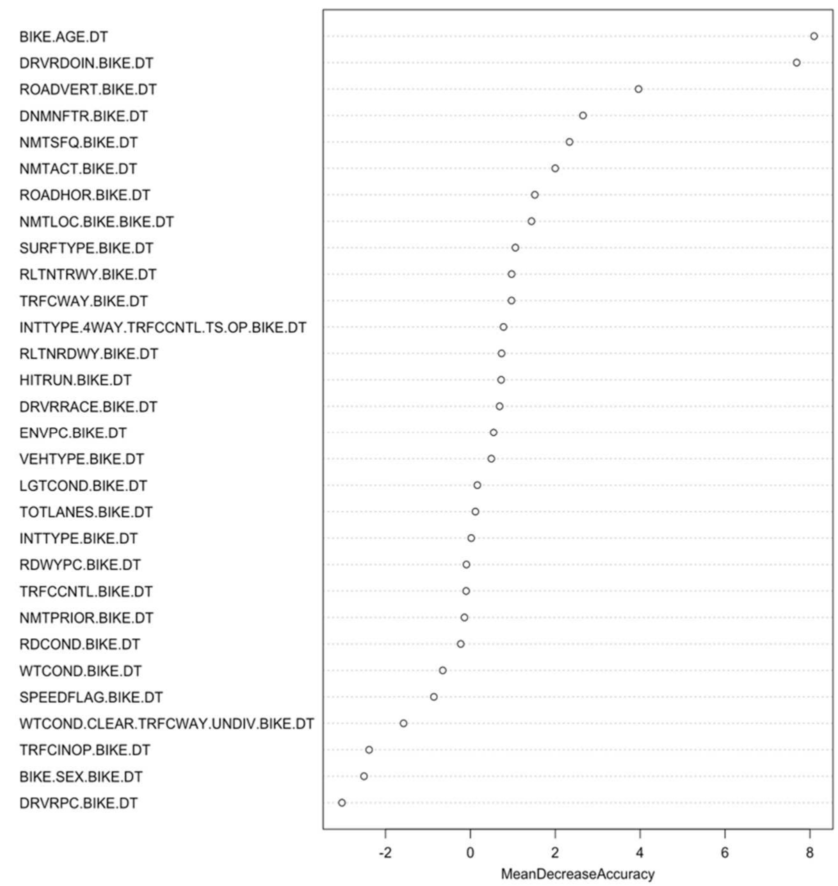


Figure 6-9: RF for DT4000 Crash Form Variable Importance Ranking for Bicyclist Crashes

6.6.3 Discussion of Variable Selection and Ranking

The principal objective of the pedestrian and bicyclist-crash variable selection and ranking was to set apart the main predictors of pedestrian and bicyclist crash severity in vehicle traffic crashes via a data mining approach. Results from the CHAID decision tree and random forests analysis in line with previous studies ((Sullivan and Flannagan 2001; Gårder 2004; Fu, Miranda-Moreno, and Saunier 2016)) revealed that for pedestrian-vehicle crashes, light condition of the roadway at the time of the crash, pedestrian location and speed were the most important predictors of the severity of pedestrian crashes. Results revealed that the majority of the fatalities and severely injured crashes occurred within the nighttime or with pedestrians located at midblock with no marked crosswalks and at intersections but not in crosswalks. Lower visibility level and drivers' failure to yield could be an explanation for this specific type of crashes ((Sullivan and Flannagan 2001)). For bicyclist-vehicle crashes, vehicles controlled maneuver, bicyclists' age, and location with respect to the roadway were the most important predictors of the severity of bicyclist crashes. These results in line with previous research (Johnson et al. 2010; Boufous et al. 2011).

The most common variables influencing pedestrian and bicyclist injury severity in vehicle crashes are pedestrian and bicyclist location with respect to the roadway, and their condition and circumstances that might have contributed to the crash. Previous research presented that crashes involving vulnerable road users (VRUs) at signalized intersections are less severe than crashes occurred elsewhere (Zegeer et al. 2010; Rifaat, Tay, and de Barros 2011; Haleem, Alluri, and Gan 2015).

However, the decision tree distinguished between crash severity predictors that are associated with specific groups of VRUs. For pedestrians, the second level of the decision tree revealed that the predicted percentage of fatal and severe injury crashes was higher for pedestrians if the crash occurs at a non-intersection location and not during daylight. This finding was in line with previous research (Mohamed et al. 2013; Hezaveh, AzadDisfany, and Cherry 2017) which pointed out the lower severity of pedestrian crossing accidents that occur at intersections.

Whereas, for bicyclists, it has been displayed in the first and third levels of the decision tree that the bicyclist age, drivers controlled maneuver, and bicyclist action at the time of the crash is significant influencing severity rate predictors. Older age bicyclists had more severe injuries, driver's controlled maneuvers including accelerating, taking right turn, and bicyclists' actions such as cycling on the wrong way/side of the road which was in line with previous studies (Kaplan and Prato 2013; Cripton et al. 2015; Behnood and Mannering 2017).

Overall, the results of the employed CHAID decision tree were steady and found to be consistent with previous research that used other statistical techniques. This implies that the use of this technique in crash severity analysis is valid. Unlike regression models, the personal judgment does not influence the model specification, which is the advantage of using the CHAID technique. Additionally, the technique is not confined to binary splits, which yields a wider decision tree in comparison to the other decision tree and helps to show the non-linear relation between dependent variables and crashes (Hezaveh, AzadDisfany, and Cherry 2017).

CHAID representation is easy to comprehend, and able to distinguish between a complex structure of many severity factors. Therefore, it is beneficial to be implemented in studying pedestrian and bicyclist crash severity factors. A drawback of CHAID is the instability issue; the random procedure of choosing training and test samples which depends on the seed number, produce different trees. Yet, in this study the tree variation was trivial, and the common important predictors presented resulted were presented. A summary of variables providing enhanced information from the DT4000 dataset over the MV4000 dataset is shown in **Table 6-42**.

Table 6-42: A summary of variables providing enhanced information from the DT4000 dataset over the MV4000 dataset

Variable	Pedestrian crashes	Bicyclist crashes	
MV4000	Crashes involving the driver going straight while the pedestrian located in the roadway (DRVRDOIN.GO.STR.NMTLOC.2)	crashes occurred in the dark but with streetlights available on an undivided roadway (LGTCOND.LIGT.TRFCWAY.new.ND)	
	Light condition at the time of the crash (LGTCOND.PED.MV)	light condition at the crash location (LGTCOND.BIKE.MV)	
	vehicle-controlled maneuver prior to the event leading to the crash (DRVRDO.PED.MV)	Trafficway type a level of division if available (TRFCWAY.BIKE.MV)	
	The traffic control device (TCD) in effect at the crash location (TRFCNTL.PED.MV)	Bicyclist's location with respect to the roadway (NMTLOC.BIKE.MV)	
	Dark/streetlight crash location with the driver going straight (LGTCOND.LIGT.DRVRDOIN.GO.STR)	Bicyclist age (AGE.BIKE.MV)	
DT4000	Pedestrian condition at the time of the crash (DNMFTR.PED.DT)	Bicyclist age (BIKE.AGE.DT)	
	Pedestrian location with respect to the roadway (NMTLOC.PED.PED.DT)	Vehicle-controlled maneuver prior to the begging of the sequence of events (DRVRDOIN.BIKE.DT)	
	Speed factor in a crash (SPEEDFLAG.PED.DT)	Vertical terrain of the roadway (ROADVERT.BIKE)	
	Type of traffic control device (TCD) available at the crash location (TRFCCNTL.PED.DT)	Bicyclist condition at the time of the crash (DNMFTR.BIKE.DT)	
	dark/unlit crash location with the driver going straight (LGTCOND.DARK.DRVRDOIN.GO.STR)	Safety equipment used by the bicyclist (NMTSFQ.BIKE.DT)	

6.7 Statistical Analysis Using the MNL Model

The MNL model was implemented using the "mlogit" package in RStudio. Also, the variable correlation was tested by the "GoodmanKruskal" package in R (Pearson R. 2016) using the "GKtauMatrix" function. Many variables from the MV4000 and DT4000 datasets are selected in the (MNL) model development. However, based on the results obtained by implementing the MNL using RStudio, the p-values of some of the variables are larger than 0.1, which means that these variables are found to be insignificant and hence are removed from the list of significant variables.

Table 6-43 and **Table 6-45** show the estimated coefficients of each variable involved in the MNL model using the MV4000 dataset for pedestrian and bicyclist crashes, respectively. The marginal effect analysis could help evaluate how the significant variables estimated in the MNL model impact the pedestrian injury outcome probabilities (Long and Freese 2001). The marginal effects of each significant factor on the likelihood of each injury-severity class are reported in **Table 6-44** and

Table 6-46 for the pedestrian and bicyclist crash models, respectively.

Table 6-43: Estimated Coefficients of Variables Included in the Pedestrian Injury Severity Model

Variable	Code	Severe Injury (A) Crash		Evident and Possible Injury (B+C) Crash	
		Coef.	P-value	Coef.	P-value
Intercept		1.25	0.00	2.70	0.00
Light condition (dark)	LGTCOND.DARK.PED.MV	-1.00	0.00	-1.65	0.00
Pedestrian location (in crosswalk)	NMTLOC.PED.PED.MV	0.79	0.00	1.03	0.00
Driver action (going straight)	DRVRDO.GO.STR.PED.MV	---	---	-1.44	0.00
Traffic control (none)	TRFCNTL.NONE.PED.MV	---	---	-0.58	0.04
Driver action (going straight) and pedestrian location (in the roadway)	DRVRDO.GO.STR.NMTLOC.2	0.97	0.00	1.27	0.00

Trafficway (undivided)	TRFCWAY.ND.PED.MV	0.64	0.06	0.66	0.04
Light condition (light) and driver action (going straight)	LGTCOND.LIGT.DRVRDO.GO.STR	---	---	1.44	0.00
Pedestrian age (above 65)	PED.AGE.MV	1.18	0.08	-0.98	0.09

Table 6-44: Marginal Effects Results for Pedestrian Crash Variables Using MV4000 Dataset for Pedestrian Crash Variables Using MV4000 Dataset

Variable	**P (Fatal (K) Crash)**	**P (Severe Injury (A) Crash)**	**P (Evident and Possible Injury (B+C) Crash)**
Light condition (dark)	0.0156	0.0722	-0.0879
Pedestrian location (in crosswalk)	-0.0124	-0.0337	0.0461
Driver action (going straight)	0.0132	0.1038	-0.1170
Traffic control (none)	0.0056	0.0163	-0.0220
Driver action (going straight) and pedestrian location (in roadway)	-0.0087	0.0087	-0.0066
Trafficway (undivided)	0.0100	0.0268	-0.0368
Light condition (light) and driver action (going straight)	-0.0129	-0.1298	0.1428
Pedestrian age (65 and above)	0.0101	-0.0243	0.0141

The analysis of the MV4000 dataset for pedestrian crashes in **Table 6-43** showed that eight variables tested by the MNL model based on their significance value and the model's overall Akaike information criterion (AIC), are selected for pedestrian severity rate prediction.

i) Three are roadway-related; light condition (dark), trafficway (undivided), and traffic control (none). Regarding the light conditions, three light conditions categories (ranged from daylight to dark) are considered in studying pedestrian crash severity with the MV4000 dataset. Pedestrians are more likely to be involved in severe injury crashes if the crash occurred in dark/unlit roadways, compared to crashes occurred in the daylight or dark/lit roadways. Additionally, the marginal effect results in **Table 6-44** indicates that the probability of severe injury crashes will increase when the pedestrian is involved in a crash in dark/unlit roadways. Many researchers highlighted the effect of wearing reflective clothes in decreasing the likelihood of being involved in a vehicle crash (i.e., (Shinar 1985; Tyrrell et al. 2016)). Four trafficway areas designed for motor vehicle operation categories (not physically divided, divided with a traffic barrier, divided without traffic barrier, and one-way traffic) are considered in this study. **Table 6-43** showed that pedestrians are more likely to be involved in severe crashes if the crash occurred in undivided roadways. The marginal effect results in **Table 6-44** indicates that undivided trafficways increases the likelihood of severe crashes. This result is in line with previous research (i.e., (LaValley et al. 2003)). With respect to traffic control in effect at the time of the crash, no traffic control availability at the crash location decrease evident and possible injury severity crashes. **Table 6-44** also indicates that no traffic control in effect at the crash location decreases the likelihood of evident and possible injury severity crashes.

ii) One is driver-related; driver action (going straight). **Table 6-43** showed that crashes involving the driver going straight do not significantly affect severe injury pedestrian crashes but are less likely to produce fatal injury severity crashes. **Table 6-44** also indicates that when the driver involved in the crash is going straight, the likelihood of evident and possible injury decreases.

iii) Two are pedestrian-related; pedestrian location (in crosswalk), and pedestrian age (above 65). For the pedestrian location, pedestrians located in a crosswalk (no information about the crosswalk marking in MV4000 dataset) showed a significant effect on decreasing the likelihood of fatal severity crashes. Regarding pedestrian age, three age groups are used in this study as previous studies divided ages (i.e., (Chakravarthy, Lotfipour, and Vaca 2007)); (pedestrians under 30 years old, pedestrians between 30-64 years old, and age 65 and above). For the pedestrian age 65 and above group, the estimation results in **Table 6-43** and the marginal effect result in **Table 6-44** both indicate that this age group increases the likelihood of severe crashes compared with younger pedestrians. This finding is supported by findings from previous research ((Chakravarthy, Lotfipour, and Vaca 2007; J.-K. Kim et al. 2008; Tay et al. 2011; Chen and Fan 2018)) and can be supported by the fact that Older adult pedestrians, those 65 years and older, have their own limitations that make them susceptible to collisions. As adults age, gradual losses in hearing, vision, and flexibility put them at a higher risk, in addition to their need for longer reaction times while in the roadway. Furthermore, once the older adult pedestrian is struck, their co-morbid conditions and limited physical reserves contribute to a higher percentage of death and disability when compared to other pedestrian age groups (Chakravarthy, Lotfipour, and Vaca 2007).

Finally, iv) two are driver action-related; “driver action (going straight) and pedestrian location (in the roadway)”, and “light condition (light) and driver action (going straight)”. These are newly created variables that were found to have a significant effect on pedestrian injury severity. For crashes involving a driver going straight while the pedestrian is located in the roadway (no information about the crosswalk marking in MV4000 dataset), the results showed a significant effect on severe injuries. **Table 6-43** and the marginal effect results in **Table 6-44** showed that pedestrians are more likely to be involved in severe crashes if the crash involved a driver going straight while the pedestrian is located in the roadway. Studying the controlled maneuver (going straight) separately, and with the non-motorist location showed that the later has a greater effect on severe injury crashes. Regarding crashes occurred in dark/streetlight roadways while the driver is going straight did not show a significant effect on severe injury crashes. However, results in **Table 6-43 and Table 6-44** both show that pedestrians are more likely to sustain an evident and possible injury if involved in crashes of such conditions.

Table 6-45: Estimated Coefficients of Variables Included in the Bicyclist Injury Severity Model for Bicyclist Crash Variables Using MV4000 Dataset

Variable	Code	Severe Injury (A) crash	Evident and Possible Injury (B+C) Crash

		Coef.	P-value	Coef.	P-value
Intercept		2.76	0.00	5.63	0.00
Bicyclist location (in the roadway)	NMTLOC.BIKE.MV	-1.74	0.02	-1.75	0.01
Light condition (light)	LGTCOND.LIGT.BIKE.MV	-5.77	0.00	-1.85	0.03
Vertical terrain (hill)	ROADVERT.BIKE.MV	1.64	0.025	1.42	0.03
Bicyclist age (< 30 years)	BIKE.AGE.MV	-2.33	0.02	-2.53	0.00
Light condition (light) and trafficway division (undivided)	LGTCOND.LIGT.TRFCWAY.ND	6.53	0.00	---	---
Driver action (turning right)	DRVRDO.BIKE.MV	-1.29	0.06	---	---
Traffic control in effect (stop sign and no TCD)	TRFCNTL.BIKE.MV	1.41	0.07	1.58	0.03

Table 6-46: Marginal Effects Results for Bicyclist Crash Variables Using MV4000 Dataset

Variable	**P (Fatal (K) Crash)**	**P (Severe Injury (A) Crash)**	**P (Evident and Possible Injury (B+C) Crash)**
Bicyclist location (in roadway)	0.0281	-0.0016	-0.0265
Light condition (light)	0.0347	-0.2777	0.2425
Vertical terrain (hill)	0.0206	0.0094	-0.0300
Bicyclist age (above 65)	0.0406	0.0112	-0.0518
Light condition (light) and trafficway division (undivided)	-0.0311	0.3513	-0.3202
Driver action (turning right)	0.0174	-0.0231	0.0057
Traffic control in effect (stop sign and no TCD)	0.0252	0.0093	-0.0345

The analysis of the MV4000 dataset for bicyclist crashes in **Table 6-45** showed that seven variables are selected for bicyclist severity rate prediction.

i) Four are roadway-related; light condition (dark), traffic control in effect (stop sign and no TCD), and vertical terrain (hill). Regarding the light conditions, three light conditions categories (ranged from daylight to dark) are considered in studying bicyclist crash severity with the MV4000 dataset. Bicyclists are less likely to be involved in severe injury crashes if the crash occurred in nighttime/streetlight roadways, compared to crashes occurred in nighttime/unlit roadways. Additionally, the marginal effect results in

Table 6-46 indicates that the probability of severe injury crashes will decrease when the bicyclist is involved in a crash in nighttime/streetlight roadways. Previous research is in line with this finding; bicyclist injury severity level could be elevated by specific crash patterns and risk factors including night without streetlight (Yan et al. 2011).

For traffic control in effect at the time of the crash, **Table 6-45** and

Table 6-46 both show that bicyclists are more likely to sustain a severe injury if involved in crashes at stop sign-controlled/or no TCD intersections. (Rash-ha Wahi et al. 2018) concludes that most bicycle–motor vehicle (BMV) crashes occurred on wet road surfaces were associated with an increased cyclist injury severity at Stop/Give-way intersections. Concerning the hill road terrain, **Table 6-45** and

Table 6-46 both show that bicyclists are more likely to sustain a severe injury if involved in crashes on roadways which are hilly. Also, researchers agreed that bottoms of hills can result in severe injury crashes since both drivers and bicyclists tend to speed on such locations. For instance, (Haworth and Debnath 2013) agreed that most BMV crashes occurred in city areas which are generally hilly, resulting in poor visibility because drivers and cyclists cannot be sure whether or not there is an oncoming vehicle hidden beyond the rise. Also, (Robartes and Chen 2017) concluded that when the automobile driver has his/her vision obscured by hill crests and is in a crash with a bicyclist, the bicyclist's severe injury and fatality risks increase by (63.4%) and (25.5%), respectively.

Finally, for bicyclists involved in crashes occurred at nighttime with streetlights in an undivided trafficway, **Table 6-45** show that bicyclists are more likely to sustain a severe injury if involved in crashes under these conditions, but has shown an insignificant effect on evident and possible injury severity. The marginal effect results in

Table 6-46 indicates that crashes occurred at nighttime with streetlights in an undivided trafficway increase the likelihood of severe crashes sustained by the bicyclist. This finding is in line with previous research. Yan and colleagues concluded that among the geometric characteristics, the presence of median and/or division was found to significantly reduce the probability of severe injuries to bicyclists on road segments (Yan et al. 2011). (J.-K. Kim et al. 2007) also found that medians can help to reduce bicyclist injury severity.

ii) One is driver-related; driver action (turning right). **Table 6-43** showed that crashes involving the driver turning right compared to drivers turning left were found to be insignificant in the prediction of evident and possible injury severity crashes, however, they were found less likely to produce severe injury crashes. **Table 6-44** also indicates that when the driver involved in the crash is turning right, the likelihood of severe injury decreases. This finding is in line with previous research (i.e., (Abdel-Aty and Keller 2005). Most common BMV collisions results when the driver looks left for oncoming vehicles when they should also be looking right for cyclists. This situation creates a lack of driver expectations about bicyclists' location and behavior (Räsänen and Summala 1998).

iii) Two are bicyclist-related; bicyclists located in the roadway (no information about the crosswalk marking in MV4000 dataset), and bicyclist age (above 65). Regarding bicyclist age, three age groups are used in this study as used for pedestrian age groups; (bicyclists under 30 years old, bicyclists between 30-64 years old, and age 65 and above). For bicyclists under 30 years old group, the estimation results in **Table 6-45** and

Table 6-46, both showed that young bicyclists (<30 years) are less likely to sustain a severe injury if involved in motor-vehicle crashes. Older pedestrians and cyclists are over-involved in serious injury and fatal crashes and under-represented in crashes of minor severity, compared to younger adult pedestrians and cyclists (Oxley et al. 2004). Additionally, numerous studies have cited bicyclists of old age as a risk factor (Yan et al. 2011; Moore et al. 2011; Eluru, Bhat, and Hensher 2008). J.-K. Kim et al. 2007, concluded that bicyclists' age group of over 55 years could double the risk of a fatality. Concerning bicyclist location at the time of the crash, the estimation results in **Table 6-45** and

Table 6-46, both showed that bicyclists located at intersections in the roadway (no specific information about the specific location of the non-motorists) are less likely to sustain severe injury when involved in motor-vehicle crashes.

Table 6-47 and **Table 6-49** show the estimated coefficients of each variable involved in the MNL model using the DT4000 dataset for pedestrian and bicyclist crashes, respectively. The marginal effects of each significant factor on the likelihood of each injury-severity class are reported in **Table 6-48** and

Table 6-50 for the pedestrian and bicyclist crash models, respectively.

Table 6-47: Estimated Coefficients of Variables Included in the Bicyclist Injury Severity Model

Variable	**Code**	**P (Severe Injury (A) crash)**		**P (Evident and Possible Injury (B+C) Crash)**	
		Coef.	P-Value	Coef.	P-Value
Intercept		1.39	0.00	2.31	0.00
Pedestrian condition (appeared normal)	DNMFTR.PED.DT	---	---	-1.52	0.00

Pedestrian location (not at intersection/on roadway-not in marked crosswalk)	NMTLOC.PED.DT	0.72	0.00	1.01	0.00
Trafficway division (divided with traffic barrier)	TRFCWAY.PED.DT	-0.91	0.09	-1.23	0.02
Trafficway division (divided without traffic barrier)	TRFCWAY.PED.DT	-1.33	0.00	1.05	0.08
Pedestrian usage of safety equipment	NMTSFQ.PED.DT	-0.75	0.00	-1.95	0.00
Pedestrian location (not at intersection/on roadway-not in marked crosswalk) and vehicles controlled maneuver (going straight)	NMTLOC.NAI.NX. DRVRDOIN.GO.STR.PED.DT	0.97	0.00	---	---
Type/level of light (dark/lighted)	LGTCOND.PED.DT	-0.85	0.00	-1.45	0.00
Driver's controlled maneuver (left turn)	DRVRDOIN.PED.DT	---	---	1.15	0.03
Pedestrian action (improper crossing of the roadway/jaywalking	NMTACT.PED.DT	2.53	0.00	3.21	0.00
Pedestrian location (shoulder/roadside)	NMTLOC.PED.DT	-1.51	0.00	-1.41	0.00
Pedestrian condition (under the influence of medication/drugs/alcohol)	DNMFTR.PED.DT	1.44	0.00	1.24	0.01
Pedestrian age (≥ 65 Years)	PED.AGE.DT	-1.30	0.00	-1.64	0.00
Vehicle type (passenger car)	VEHTYPE.PED.DT	-1.31	0.00	-1.24	0.00
Type of traffic control device –TCD (other than a stop sign, or a traffic signal).	TRFCCNTL.PED.DT	1.25	0.00	1.72	0.00

Table 6-48: Marginal Effects Results for Pedestrian Crash Variables Using DT4000 Dataset

Variable	**P (Fatal (K) Crash)**	**P (Severe Injury (A) Crash)**	**P (Evident and Possible Injury (B+C) Crash)**
Pedestrian condition (appeared normal)	-0.0081	-0.0261	0.0343
Pedestrian location (not at intersection/on roadway-not in marked crosswalk)	0.0091	-0.0438	0.0346

Trafficway division (divided with traffic barrier)	0.0884	-0.0089	-0.0795
Trafficway division (divided without traffic barrier)	0.0554	-0.0115	-0.0439
Pedestrian usage of a safety equipment (e.g., reflective lighting)	-0.0121	-0.1009	0.1131
Pedestrian location (not at intersection/on roadway-not in marked crosswalk) and vehicle's-controlled maneuver (going straight)	-0.0044	0.1500	-0.1456
Type/level of light (dark/lighted)	0.1368	-0.0095	-0.1273
Driver's controlled maneuver (left turn)	-0.0069	-0.0990	0.1060
Pedestrian action (improper crossing of the roadway/jaywalking	-0.0232	0.1686	-0.1454
Pedestrian location (shoulder/roadside)	0.0106	-0.0534	0.0428
Pedestrian condition (under the influence of medication/drugs/alcohol)	-0.0099	0.0578	-0.0478
Pedestrian age (≥ 65 Years)	0.011	-0.0895	0.0775
Vehicle type (passenger car)	-7.0417	-0.0619	0.0689
Type of traffic control device –TCD (other than a stop sign, or a traffic signal).	0.01195	0.2640	-0.2759

The analysis of the DT4000 dataset for pedestrian crashes in **Table 6-47** showed that fourteen variables are selected for pedestrian severity rate prediction.

i) Four are roadway-related; trafficway division (divided with traffic barrier), trafficway division (divided without traffic barrier), type/level of light (dark and unlit), type of the TCD (other than a stop sign, or a traffic signal). Regarding trafficway division, results in **Table 6-47** suggest that pedestrians are less likely to be involved in severe injury crashes if the crash occurred in trafficways divided with a traffic barrier. Also, the marginal effect results in **Table 6-48** indicates that the probability of severe injury crashes will decrease when the pedestrian is involved in a crash in a trafficway divided with a traffic barrier. Similarly, for trafficways divided without traffic barriers, **Table 6-47** and **Table 6-48** both show that pedestrians are less likely to sustain a severe injury if involved in crashes that occurred in trafficways divided without traffic barriers.

Regarding the light conditions, six-light conditions categories (daylight-DAY-, dawn-DAWN-, dusk-DUSK, dark/lighted-LITE-, dark/unlit-DARK-, dark-unknown lighting-DK/UN) are considered in studying pedestrian crash severity with the DT4000 dataset. The categories of light condition tested are; lite, dark and dawn, and daylight, dusk, dark-unknown lighting) as proposed by the CHAID tree in **Figure 6-4**. Results in **Table 6-47** and **Table 6-48** both show that pedestrians are less likely to sustain a severe injury if involved in crashes occurred in dark/lighted trafficways (LITE). Concerning the type of TCD applicable to the involved motor-vehicle at the time of the crash, results found in **Table 6-47** and **Table 6-48** both show that pedestrians are less likely to sustain a severe injury if involved in crashes occurred at intersections with a stop sign or a traffic signal. For instance, no traffic control, traffic signal/stop sign with flashing, and traffic control person. **Table 6-47** and **Table 6-48** both show that pedestrians are more likely to sustain a severe injury if involved in crashes occurred at intersections without traffic control, with traffic signal/stop sign with flashing, or a traffic control person.

ii) Seven are pedestrian-related; pedestrian condition (appeared normal), pedestrian location (not at intersection/on roadway-not in marked crosswalk), pedestrian usage of safety equipment, pedestrian action (improper crossing of the roadway/jaywalking, pedestrian location (shoulder/roadside), pedestrian condition (under the influence of medication/drugs/alcohol), pedestrian age ($\geq$ 65 years). **Table 6-47** and **Table 6-48** both show that pedestrians are less likely to sustain a severe injury if involved in crashes occurred while the pedestrian is using safety equipment compared to not using any safety equipment; the pedestrian is located on the roadside compared to being located in an intersection with/without a marked crosswalk, or located in a midblock (not at an intersection) with/without marked crosswalk; and if the pedestrian age group is $\geq$ 65 years compared to being a young pedestrian or among pedestrians aged 30-64.

For pedestrian action/circumstances at the time of the crash, pedestrian under the influence of medication/drugs/alcohol, and pedestrians who appeared normal, showed opposing results in **Table 6-47** and **Table 6-48**. The results display that pedestrians are more likely to sustain a severe injury if they are under the influence of medication/drugs/alcohol. Whereas, if a pedestrian appeared to be normal, he/she is less likely to sustain an evident and possible injury (was insignificant in affecting severe crashes). Lastly, results of pedestrian improper crossing of the roadway/jaywalking showed in **Table 6-47** and **Table 6-48**, revealed that a pedestrian is more likely to sustain a severe injury if he/she was crossing the roadway improperly.

iii) One is vehicle-related; Vehicle type (passenger car). For vehicle-related variables, **Table 6-47** and **Table 6-48** both show that pedestrians are less likely to sustain a severe injury if involved in crashes where the vehicle is a passenger car, compared to other groups of a vehicle type that are used in this study; i.e., passenger van, utility truck, sport utility vehicle, etc.

Finally, iv) two are driver-related; driver action (going straight). For the vehicle's controlled maneuver (left turn), results presented in **Table 6-44** indicate that when the driver involved in the crash is making a left turn, the likelihood of severe injury decreases. Moreover, results displayed in **Table 6-47** and **Table 6-48**, both revealed that a pedestrian is more likely to sustain a severe injury if he/she was located not at intersection/on roadway-not in marked crosswalk) while the driver was going straight at the time of the crash.

Table 6-49: Estimated Coefficients of Each Variable Involved in the Bicyclist Injury Severity Model

Variable	Code	P (Severe injury (A) Crash)		P (Evident and Possible Injury (B+C) Crash)	
		Coef.	P-value	Coef.	P-value
Intercept		3.05	0.07	5.57	0.00
Intersection type (4-way)	INTTYPE.BIKE.DT	---	---	2.47	0.00

Intersection type (T)	INTTYPE.BIKE.DT	-1.76	0.09	-2.38	0.02
Horizontal road terrain (straight)	ROADHOR.BIKE.DT	-1.83	0.04	-2.01	0.02
Bicyclist Action (Sudden Movement Into Traffic)	NMTACT.BIKE.DT	2.15	0.01	-2.58	0.00
Bicyclist Age (< 30 Years)	BIKE.AGE.DT	---	---	1.80	0.01
Vehicle's controlled maneuver (going straight)	DRVRDOIN.BIKE.DT	-1.34	0.07	-2.69	0.00
Vehicle's controlled maneuver (going straight) and bicyclist's action Immediately Prior to the Crash (walking facing traffic)	DRVRDOIN.GO.STR. NMTPRIOR.BIKE.DT	2.56	0.00	2.06	0.02
Vertical road terrain (level)	ROADVERT.BIKE.DT	---	---	-1.92	0.08
Bicyclist Condition (at an intersection in marked crosswalk)	DNMFTR.BIKE.DT	---	---	-1.55	0.04
Safety Equipment Used By the Bicyclist (none)	NMTSFQ.BIKE.DT	---	---	2.38	0.03

Table 6-50: Marginal Effects Results for Bicyclist Crash Variables Using DT4000 Dataset

Variable	**P (Fatal (K) Crash)**	**P (Severe Injury (A) Crash)**	**P (Evident and Possible Injury (B+C) Crash)**
Intersection type (4-way)	-0.3586	0.0086	0.3500
Intersection type (T)	0.4352	-0.1707	-0.2644
Horizontal road terrain (straight)	0.4149	-0.2283	-0.1865
Bicyclist Action (Sudden Movement Into Traffic)	0.5054	-0.2444	-0.2609
Bicyclist Age (< 30 Years)	-0.2651	-0.0132	0.2783
Vehicle's controlled maneuver (going straight)	0.4045	-0.0352	-0.3693
Vehicle's controlled maneuver (going straight) and bicyclist's action Immediately Prior to the Crash (walking facing traffic)	0.4002	-0.5172	0.1170
Vertical road terrain (level)	0.1917	0.1492	-0.3410
Bicyclist Condition (at intersection in marked crosswalk)	0.1832	-0.2534	0.0701
Safety Equipment Used By the Bicyclist (none)	0.0272	0.3287	-0.3560

The analysis of the DT4000 dataset for bicyclist crashes in **Table 6-49** showed that ten variables are selected for bicyclist severity rate prediction.

i) Four are roadway-related; intersection type (4-way), intersection type (T), horizontal road terrain (straight), vertical road terrain (level). Regarding the intersection type; two types showed significant influence on bicyclist's severity rate; 4-way intersection and T intersection among six intersection type categories (4-way intersection, T intersection, Y intersection, L intersection, five-point intersection, and roundabouts). Opposing results were displayed in **Table 6-49** and

Table 6-50 for the two intersection types; crashes occurred at 4-way intersection showed that the bicyclist is more likely to sustain a severe injury, while bicyclists struck at T-intersections are less likely to sustain a severe injury.

Vertical and horizontal road terrain, both appeared to be significant to the evident and possible injury severity level. However, level road terrain appeared insignificant in affecting severe crashes. Consistent results were displayed in

Table 6-50, which provides that crashes that occurred on straight and level graded roadways showed that the bicyclist is less likely to sustain severe injury.

ii) Two are driver-related; vehicle-controlled maneuver (going straight), and vehicles controlled maneuver (going straight) while bicyclist's actions immediately prior to the crash (walking facing traffic). For drivers going straight, the estimation and marginal effect results showed in **Table 6-49** and

Table 6-50, reveal that when the driver is going straight, he/she is less likely to severely injure a bicyclist compared to making other maneuvers such as taking a left/right turn. Whereas, considering the case when the driver is going straight while the bicyclist walking facing traffic immediately prior to the crash (a newly created variable), the bicyclists are more likely to sustain a severe injury.

Finally, iii) four are bicyclist-related; bicyclist action (sudden movement into traffic), bicyclist age ($<$ 30 years), bicyclist condition (at an intersection in marked crosswalk), safety equipment used by the bicyclist (none). Regarding bicyclist's location with respect to the roadway, results displayed in **Table 6-49** and

Table 6-50, reveal that the bicyclist is less likely to sustain a severe injury if he/she is located at an intersection and in a marked crosswalk. For the safety equipment used by the bicyclist, the results showed in **Table 6-49** and

Table 6-50, both imply that if the bicyclist has no type of safety equipment at the time of crash then he/she is more like to sustain a severe injury when struck by the motor vehicle. Bicyclist age has always been considered a risk factor in injury severity studies ((Kaplan and Giacomo Prato 2015; Behnood and Mannering 2017; S. Das et al. 2019), the results showed in **Table 6-49** implies that bicyclists of age group under 30 years, showed an insignificant effect on fatal injury severity, but the marginal effect results in

Table 6-50 indicates that bicyclists under 30 years are less likely to sustain a severe injury, compared to the other age groups of bicyclists between 30-64 years and bicyclists older than 65 years. For the bicyclist action that might have contributed to the crash, the results showed in **Table 6-49** and

Table 6-50, both imply that if the bicyclist suddenly moved/darted into traffic, then he/she is more likely to sustain a severe injury.

6.8 References

Abdel-Aty, Mohamed, and Joanne Keller. 2005. "Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections." *Accident Analysis & Prevention* 37 (3): 417–25. https://doi.org/10.1016/j.aap.2004.11.002.

Behnood, Ali, and Fred Mannering. 2017. "Determinants of Bicyclist Injury Severities in Bicycle-Vehicle Crashes: A Random Parameters Approach with Heterogeneity in Means and Variances." *Analytic Methods in Accident Research* 16 (December): 35–47. https://doi.org/10.1016/j.amar.2017.08.001.

Bertulis, Tomas, and Daniel M. Dulaski. 2014. "Driver Approach Speed and Its Impact on Driver Yielding to Pedestrian Behavior at Unsignalized Crosswalks." *Transportation*

Research Record: Journal of the Transportation Research Board 2464 (1): 46–51. https://doi.org/10.3141/2464-06.
Biau, Gerard. 2012. “Analysis of a Random Forests Model,” 33.
Boufous, Soufiane, Liz De Rome, Teresa Senserrick, and Rebecca Ivers. 2011. “Cycling Crashes in Children, Adolescents, and Adults—A Comparative Analysis.” *Traffic Injury Prevention* 12 (3): 244–50. https://doi.org/10.1080/15389588.2011.563333.
Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.
Brewer, Marcus A., Kay Fitzpatrick, Jeffrey A. Whitacre, and Dominique Lord. 2006. “Exploration of Pedestrian Gap-Acceptance Behavior at Selected Locations.” *Transportation Research Record: Journal of the Transportation Research Board* 1982 (1): 132–40. https://doi.org/10.1177/0361198106198200117.
Chakravarthy, Bharath, Shahram Lotfipour, and Federico E. Vaca. 2007. “Pedestrian Injuries: Emergency Care Considerations.” *The California Journal of Emergency Medicine* 8 (1): 15–21.
Chen, Zhen, and Wei (David) Fan. 2018. “A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity in North Carolina.” *International Journal of Transportation Science and Technology*, October. https://doi.org/10.1016/j.ijtst.2018.10.001.
Cripton, Peter A., Hui Shen, Jeff R. Brubacher, Mary Chipman, Steven M. Friedman, M. Anne Harris, Meghan Winters, et al. 2015. “Severity of Urban Cycling Injuries and the Relationship with Personal, Trip, Route and Crash Characteristics: Analyses Using Four Severity Metrics.” *BMJ Open* 5 (1): e006654. https://doi.org/10.1136/bmjopen-2014-006654.
Dai, Shenghong. 2020. “Prioritize Winter Crash Severity Influencing Factors in US Midwestern for Autonomous Vehicle,” 11.
Das, Abhishek, Mohamed Abdel-Aty, and Anurag Pande. 2009. “Using Conditional Inference Forests to Identify the Factors Affecting Crash Severity on Arterial Corridors.” *Journal of Safety Research* 40 (4): 317–27. https://doi.org/10.1016/j.jsr.2009.05.003.
Das, Subasish, Kartikeya Jha, Kay Fitzpatrick, Marcus Brewer, and Tahmida Hossain Shimu. 2019. “Pattern Identification from Older Bicyclist Fatal Crashes.” *Transportation Research Record: Journal of the Transportation Research Board* 2673 (6): 638–49. https://doi.org/10.1177/0361198119841863.
Dong, Chunjiao, Asad J. Khattak, Chunfu Shao, and Kun Xie. 2019. “Exploring the Factors Contribute to the Injury Severities of Vulnerable Roadway User Involved Crashes.” *International Journal of Injury Control and Safety Promotion* 26 (3): 302–14. https://doi.org/10.1080/17457300.2019.1595665.
Eluru, Naveen, Chandra R. Bhat, and David A. Hensher. 2008. “A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes.” *Accident Analysis & Prevention* 40 (3): 1033–54. https://doi.org/10.1016/j.aap.2007.11.010.
Fisher, Donald, and Lisandra Garay-Vega. 2012. “Advance Yield Markings and Drivers’ Performance in Response to Multiple-Threat Scenarios at Mid-Block Crosswalks.” *Accident Analysis & Prevention* 44 (1): 35–41. https://doi.org/10.1016/j.aap.2010.11.030.
Fountas, Grigorios, Achille Fonzone, Niaz Gharavi, and Tom Rye. 2020. “The Joint Effect of Weather and Lighting Conditions on Injury Severities of Single-Vehicle Accidents.”

Analytic Methods in Accident Research, February, 100124. https://doi.org/10.1016/j.amar.2020.100124.
Fu, Ting, Luis Miranda-Moreno, and Nicolas Saunier. 2016. "Pedestrian Crosswalk Safety at Nonsignalized Crossings During Nighttime: Use of Thermal Video Data and Surrogate Safety Measures." *Transportation Research Record: Journal of the Transportation Research Board* 2586 (1): 90–99. https://doi.org/10.3141/2586-10.
Gårder, Per E. 2004. "The Impact of Speed and Other Variables on Pedestrian Safety in Maine." *Accident Analysis & Prevention* 36 (4): 533–42. https://doi.org/10.1016/S0001-4575(03)00059-9.
Gershon, Pnina, Kellienne R. Sita, Chunming Zhu, Johnathon P. Ehsani, Sheila G. Klauer, Tom A. Dingus, and Bruce G. Simons-Morton. 2019. "Distracted Driving, Visual Inattention, and Crash Risk Among Teenage Drivers." *American Journal of Preventive Medicine* 56 (4): 494–500. https://doi.org/10.1016/j.amepre.2018.11.024.
Haleem, Kirolos, Priyanka Alluri, and Albert Gan. 2015. "Analyzing Pedestrian Crash Injury Severity at Signalized and Non-Signalized Locations." *Accident Analysis & Prevention* 81 (August): 14–23. https://doi.org/10.1016/j.aap.2015.04.025.
Haworth, Narelle, and Ashim Kumar Debnath. 2013. "How Similar Are Two-Unit Bicycle and Motorcycle Crashes?" *Accident Analysis & Prevention* 58 (September): 15–25. https://doi.org/10.1016/j.aap.2013.04.014.
Hezaveh, Amin Mohamadi, Mojdeh AzadDisfany, and Christopher R Cherry. 2017. "PEDESTRIAN CRASHES IN TENNESSEE; A DATA MINING APPROACH," 15.
Hong Han, Xiaoling Guo, and Hua Yu. 2016. "Variable Selection Using Mean Decrease Accuracy and Mean Decrease Gini Based on Random Forest." In *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 219–24. Beijing, China: IEEE. https://doi.org/10.1109/ICSESS.2016.7883053.
Hossain, Moinul, and Yasunori Muromachi. 2013. "Understanding Crash Mechanism on Urban Expressways Using High-Resolution Traffic Data." *Accident Analysis & Prevention* 57 (August): 17–29. https://doi.org/10.1016/j.aap.2013.03.024.
Hunter, William W., Raghavan Srinivasan, and Carol A. Martell. 2012. "Evaluation of Rectangular Rapid Flash Beacon at Pinellas Trail Crossing in Saint Petersburg, Florida." *Transportation Research Record: Journal of the Transportation Research Board* 2314 (1): 7–13. https://doi.org/10.3141/2314-02.
Jiang, Ximiao, Mohamed Abdel-Aty, Jia Hu, and Jaeyoung Lee. 2016. "Investigating Macro-Level Hotzone Identification and Variable Importance Using Big Data: A Random Forest Models Approach." *Neurocomputing* 181 (March): 53–63. https://doi.org/10.1016/j.neucom.2015.08.097.
Johnson, Marilyn, Judith Charlton, Jennifer Oxley, and Stuart Newstead. 2010. "Naturalistic Cycling Study: Identifying Risk Factors for On-Road Commuter Cyclists." *Annals of Advances in Automotive Medicine / Annual Scientific Conference* 54 (January): 275–83.
Kaplan, Sigal, and Carlo Giacomo Prato. 2015. "A Spatial Analysis of Land Use and Network Effects on Frequency and Severity of Cyclist–Motorist Crashes in the Copenhagen Region." *Traffic Injury Prevention* 16 (7): 724–31. https://doi.org/10.1080/15389588.2014.1003818.
Kaplan, Sigal, and Carlo Giacomo Prato. 2013. "Cyclist–Motorist Crash Patterns in Denmark: A Latent Class Clustering Approach." *Traffic Injury Prevention* 14 (7): 725–33. https://doi.org/10.1080/15389588.2012.759654.

Katz, A., D. Zaidel, and A. Elgrishi. 1975. “An Experimental Study of Driver and Pedestrian Interaction during the Crossing Conflict.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 17 (5): 514–27. https://doi.org/10.1177/001872087501700510.

Kemnitzer, Caitlyn R., Caitlin N. Pope, Ann Nwosu, Songzhu Zhao, Lai Wei, and Motao Zhu. 2019. “An Investigation of Driver, Pedestrian, and Environmental Characteristics and Resulting Pedestrian Injury.” *Traffic Injury Prevention* 20 (5): 510–14. https://doi.org/10.1080/15389588.2019.1612886.

Kim, Dohyung. 2019. “The Transportation Safety of Elderly Pedestrians: Modeling Contributing Factors to Elderly Pedestrian Collisions.” *Accident Analysis & Prevention* 131 (October): 268–74. https://doi.org/10.1016/j.aap.2019.07.009.

Kim, Joon-Ki, Sungyop Kim, Gudmundur F. Ulfarsson, and Luis A. Porrello. 2007. “Bicyclist Injury Severities in Bicycle–Motor Vehicle Accidents.” *Accident Analysis & Prevention* 39 (2): 238–51. https://doi.org/10.1016/j.aap.2006.07.002.

Kim, Joon-Ki, Gudmundur F. Ulfarsson, Venkataraman N. Shankar, and Sungyop Kim. 2008. “Age and Pedestrian Injury Severity in Motor-Vehicle Crashes: A Heteroskedastic Logit Analysis.” *Accident Analysis & Prevention* 40 (5): 1695–1702. https://doi.org/10.1016/j.aap.2008.06.005.

Kordani, Ali Abdi, Bahram Shirini, and Mirbahador Yazdani. 2019. “The Effect of Road Shoulder and Weather Conditions on the Occurrence of Rollover Crashes in Two-Lane Highways.” *Periodica Polytechnica Civil Engineering*, October. https://doi.org/10.3311/PPci.12588.

Kutela, Boniphace, and Hualiang Teng. 2019. “Prediction of Drivers and Pedestrians’ Behaviors at Signalized Mid-Block Danish Offset Crosswalks Using Bayesian Networks.” *Journal of Safety Research* 69 (June): 75–83. https://doi.org/10.1016/j.jsr.2019.02.008.

LaValley, Jonathon, Cameron S. Crandall, Laura Banks, David P. Sklar, and Leverson Boodlal. 2003. “Rural and Urban Fatal Pedestrian Crashes Among United States American Indians and Alaskan Natives.” *Annual Proceedings / Association for the Advancement of Automotive Medicine* 47: 127–43.

Lee, Chris, and Mohamed Abdel-Aty. 2005. “Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida.” *Accident Analysis & Prevention* 37 (4): 775–86. https://doi.org/10.1016/j.aap.2005.03.019.

Lee, Jaeyoung, Mohamed Abdel-Aty, and Imran Shah. 2018. “Evaluation of Surrogate Measures for Pedestrian Trips at Intersections and Crash Modeling.” *Accident Analysis & Prevention*, May. https://doi.org/10.1016/j.aap.2018.05.015.

Lin, Ciyun, Dayong Wu, Hongchao Liu, Xueting Xia, and Nischal Bhattarai. 2020. “Factor Identification and Prediction for Teen Driver Crash Severity Using Machine Learning: A Case Study.” *Applied Sciences* 10 (5): 1675. https://doi.org/10.3390/app10051675.

Long, J. Scott, and Jeremy Freese. 2001. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, Tex: Stata Press.

Mitman, Meghan Fehlig, David R. Ragland, and Charles V. Zegeer. 2008. “Marked-Crosswalk Dilemma: Uncovering Some Missing Links in a 35-Year Debate.” *Transportation Research Record: Journal of the Transportation Research Board* 2073 (1): 86–93. https://doi.org/10.3141/2073-10.

Mohamed, Mohamed Gomaa, Nicolas Saunier, Luis F. Miranda-Moreno, and Satish V. Ukkusuri. 2013. “A Clustering Regression Approach: A Comprehensive Injury Severity

Analysis of Pedestrian–Vehicle Crashes in New York, US and Montreal, Canada." *Safety Science* 54 (April): 27–37. https://doi.org/10.1016/j.ssci.2012.11.001.

Moore, Darren N., William H. Schneider, Peter T. Savolainen, and Mohamadreza Farzaneh. 2011. "Mixed Logit Analysis of Bicyclist Injury Severity Resulting from Motor Vehicle Crashes at Intersection and Non-Intersection Locations." *Accident Analysis & Prevention* 43 (3): 621–30. https://doi.org/10.1016/j.aap.2010.09.015.

Morrison, Christopher N., Jason Thompson, Michelle C. Kondo, and Ben Beck. 2019. "On-Road Bicycle Lane Types, Roadway Characteristics, and Risks for Bicycle Crashes." *Accident Analysis & Prevention* 123 (February): 123–31. https://doi.org/10.1016/j.aap.2018.11.017.

National Highway Traffic Safety Administration. 2017. "MMUCC." Text. NHTSA. August 24, 2017. https://www.nhtsa.gov/mmucc-1.

Obeid, Hassan, Hoseb Abkarian, Maya Abou-Zeid, and Isam Kaysi. 2017. "Analyzing Driver-Pedestrian Interaction in a Mixed-Street Environment Using a Driving Simulator." *Accident Analysis & Prevention* 108 (November): 56–65. https://doi.org/10.1016/j.aap.2017.08.005.

Oxley, Jennifer, Bruce Corben, Brian Fildes, Mary O'Hare, and Talib Rothengatter. 2004. "Older Vulnerable Road Users – Measures to Reduce Crash and Injury Risk," 182.

Pearson R. 2016. "GoodmanKruskal.Pdf." 2016. https://cran.r-project.org/web/packages/GoodmanKruskal/GoodmanKruskal.pdf.

Pelé, Deneubourg, and Sueur. 2019. "Decision-Making Processes Underlying Pedestrian Behaviors at Signalized Crossings: Part 2. Do Pedestrians Show Cultural Herding Behavior?" *Safety* 5 (4): 82. https://doi.org/10.3390/safety5040082.

Räsänen, Mikko, and Heikki Summala. 1998. "Attention and Expectation Problems in Bicycle–Car Collisions: An in-Depth Study." *Accident Analysis & Prevention* 30 (September): 657–66. https://doi.org/10.1016/s0001-4575(98)00007-4.

Rash-ha Wahi, Rabbani, Narelle Haworth, Ashim Kumar Debnath, and Mark King. 2018. "Influence of Type of Traffic Control on Injury Severity in Bicycle–Motor Vehicle Crashes at Intersections." *Transportation Research Record* 2672 (38): 199–209. https://doi.org/10.1177/0361198118773576.

Rifaat, Shakil Mohammad, Richard Tay, and Alexandre de Barros. 2011. "Effect of Street Pattern on the Severity of Crashes Involving Vulnerable Road Users." *Accident Analysis & Prevention* 43 (1): 276–83. https://doi.org/10.1016/j.aap.2010.08.024.

Robartes, Erin, and T. Donna Chen. 2017. "The Effect of Crash Characteristics on Cyclist Injuries: An Analysis of Virginia Automobile-Bicycle Crash Data." *Accident Analysis & Prevention* 104 (July): 165–73. https://doi.org/10.1016/j.aap.2017.04.020.

Schneider, Robert J., Offer Grembek, and Matthew Braughton. 2013. "Pedestrian Crash Risk on Boundary Roadways: University Campus Case Study." *Transportation Research Record: Journal of the Transportation Research Board* 2393 (1): 164–73. https://doi.org/10.3141/2393-19.

Sheykhfard, Abbas, and Farshidreza Haghighi. 2019. "Performance Analysis of Urban Drivers Encountering Pedestrian." *Transportation Research Part F: Traffic Psychology and Behaviour* 62 (April): 160–74. https://doi.org/10.1016/j.trf.2018.12.019.

Shi, Jiangang, Yanyan Chen, Futian Ren, and Jian Rong. 2007. "Research on Pedestrian Behavior and Traffic Characteristics at Unsignalized Midblock Crosswalk: Case Study in

Beijing." *Transportation Research Record: Journal of the Transportation Research Board* 2038 (1): 23–33. https://doi.org/10.3141/2038-04.

Shinar, David. 1985. "The Effects of Expectancy, Clothing Reflectance, and Detection Criterion on Nighttime Pedestrian Visibility." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 27 (June): 327–33. https://doi.org/10.1177/001872088502700308.

Simons-Morton, Bruce, Neil Lerner, and Jeremiah Singer. 2005. "The Observed Effects of Teenage Passengers on the Risky Driving Behavior of Teenage Drivers." *Accident Analysis & Prevention* 37 (6): 973–82. https://doi.org/10.1016/j.aap.2005.04.014.

Sullivan, John M, and Michael J Flannagan. 2001. "CHARACTERISTICS OF PEDESTRIAN RISK IN DARKNESS," 27.

Sun, Ming, Xiaoduan Sun, and Donghui Shan. 2019. "Pedestrian Crash Analysis with Latent Class Clustering Method." *Accident Analysis & Prevention* 124 (March): 50–57. https://doi.org/10.1016/j.aap.2018.12.016.

Tay, Richard, Jaisung Choi, Lina Kattan, and Amjad Khan. 2011. "A Multinomial Logit Model of Pedestrian–Vehicle Crash Severity." *International Journal of Sustainable Transportation* 5 (4): 233–49. https://doi.org/10.1080/15568318.2010.497547.

Tyrrell, Richard A., Joanne M. Wood, D. Alfred Owens, Stephanie Whetsel Borzendowski, and Ashley Stafford Sewall. 2016. "The Conspicuity of Pedestrians at Night: A Review." *Clinical and Experimental Optometry* 99 (5): 425–34. https://doi.org/10.1111/cxo.12447.

"VDOT_Pedestrians_Crash_Assessment_2014-2018.Pdf." n.d. Accessed April 26, 2020. http://www.virginiadot.org/business/resources/HSIP/VDOT_Pedestrians_Crash_Assessment_2014-2018.pdf.

"Wisconsin DMV Official Government Site - Final Year-End Crash Statistics." n.d. Accessed April 18, 2020. https://wisconsindot.gov/Pages/about-wisdot/newsroom/statistics/final.aspx.

Wisconsin State,. 2020. "2017−18 Wisconsin Statutes." 2020. http://docs.legis.wisconsin.gov/statutes/statutes/346.pdf.

Wootton, Isaac A, and Lisa K Spainhour. 2004. "Improved Methods for Reviewing Florida Pedestrian Fatal Crash Data," 11.

Yan, Xinping, Ming Ma, Helai Huang, Mohamed Abdel-Aty, and Chaozhong Wu. 2011. "Motor Vehicle–Bicycle Crashes in Beijing: Irregular Maneuvers, Crash Patterns, and Injury Severity." *Accident Analysis & Prevention* 43 (5): 1751–58. https://doi.org/10.1016/j.aap.2011.04.006.

Yue, Lishengsa, Mohamed Abdel-Aty, Yina Wu, Ou Zheng, and Jinghui Yuan. 2020. "In-Depth Approach for Identifying Crash Causation Patterns and Its Implications for Pedestrian Crash Prevention." *Journal of Safety Research* 73 (June): 119–32. https://doi.org/10.1016/j.jsr.2020.02.020.

Zegeer, Charlie, William Hunter, Loren Staplin, Fran Bents, Richard Huey, and Janet Barlow. 2010. "Safer Vulnerable Road Users: Pedestrians, Bicyclists, Motorcyclists, and Older Users," 59.

Chapter 7 Conclusions and Future Work Recommendations

7.1 Research Conclusions

In this research, exploration, and evaluation of a variety of factors significantly affecting motor vehicle crashes involving pedestrians and bicyclists, commonly defined as vulnerable road users (VRUs), is applied.

The comprehensive safety analysis in this study covered the following main points:

- Discover possible underlying factor structure connecting exogenous variables and crashes involving pedestrians based on the corridor level, using the SEM technique.
- Developing crash count models and responsible party choice models to respectively address factors relating to roles in a crash by pedestrians and drivers.

- Identify, categorize, and quantify the factors contributing to crashes where a pedestrian is responsible, a driver is responsible, or both parties are responsible.
- specify the crash party at-fault, which provides evidence about whether pedestrians, bicyclists or drivers are more likely to be involved in severe crashes, and to identify the contributing factors that affect the fault of a specific road user group.
- An extensive investigation of the available information regarding the crash (i.e., issued citations, actions/circumstances that may have played a role in the crash occurrence, and crash scenario completed by the police officer) are considered.
- evaluate the enhancements of crash report form for its effectiveness of reporting VRU involved motor vehicle crashes.
- Thoroughly study the modified and new data fields, and their associated attribute values.
- Comprehensively evaluate the effectiveness of improved data collection in terms of a better understanding of factors associated with and contributing to VRU crashes.

7.1.1 Conclusions of the Corridor Crash Analysis Study

The study applies SEM to develop a conjectured structure that provides a clear portrait between a large number of highway corridor specific variables and VRU crashes. The structure is featured by the relationship between three exogenous latent variables representing bicyclist and pedestrian-oriented roadway design, exposure and surrounding social status, and one endogenous variable representing a single value VRU safety quantification for both pedestrian and bicyclist crashes. The relationship between latent and observed variables can also be conveniently established by using the measurement model. Combining both the structural and measurement models in a single modeling process enables the effective distinction between direct, indirect, and synergic effects between variables and thus more accurately captures the physical underpinning for VRU crashes. Hence, the notable findings from this highway corridor based VRU study are as follows:

- The model suggests that bicycle/pedestrian-oriented roadway, exposure, and low social status are strongly related to VRUs' crash frequency.
- SEM helps to explain the potentially conflicting information such that an observed variable may affect more than one latent variable in different ways (i.e., High_Spd_Limit), and the results show that high-speed limit positively influences pedestrian and bicyclist's exposure to traffic leading to an increase in crash frequency.
- It is noted that some significant variables in the models were not significant in previous research.

7.1.2 Conclusions of Party At-Fault Assignment Study

The importance of determining the at-fault status of each of the parties emerged from the transportation professional's need to understand the injury severity of traffic crashes where the driver is at-fault or not-at-fault. This knowledge may be used to educate at-fault drivers and at-fault VRUs about the possible risk produced to other not-at-fault drivers and VRUs. Besides, comparing the injury severity of the VRUs at-fault party with injury severity of the not-at-fault party allows the identification of the major factors affecting both parties. In this study, fault investigation included three different outputs: driver at-fault, VRU at-fault, and unknown party at-fault, as shown in **Figure 5-2.** Police crash reports from 2017-2018 were reviewed in the DT4000 form. Moreover, a case review was performed for every particular traffic crash by using all possible data sources, namely police narrative construction including eyewitness statement/testimony, issued VRU and driver traffic citations including violated pedestrian/bicycle laws, toxic driver and VRU behavior/actions and any contributing circumstances to crash occurrence. Hence, the notable findings from this fault party study are as follows:

- The data sources of information served as the foundation to help us assign the final score that will contribute to the fault assignment of each party. Conditions/circumstances are useful in terms of proposing preventive actions.
- This fault assignment guideline is designed for the Milwaukee area since results of previous research on fault assignment concluded that the fault assignment results are not necessarily uniform with distant geographic locations (Ulfarsson, Kim, and Booth 2010), hence a comprehensive and unique guideline to assign the fault to pedestrians and drivers was developed.
- The analysis using Z-test and XGboost as statistical modeling techniques helps to rank variables' importance in terms of predicting driver at-fault and VRU at-fault crashes. Subsequently, the MNL models quantified the effect of the variables on injury severity prediction.
- The manual review of the crash narrative is time-consuming and labor-intensive, which may be subject to the reviewer's experience and judgment. Some efforts have been made to automate manual review through text-mining techniques and the results are described in detail in **Appendix B**.

7.1.3 Conclusions of the Data Quality Evaluation Study

Regarding the investigation on the level of completion of the DT4000 crash form data fields, major upgrades were applied to the DT4000 data fields and their completeness throughout the crash report. For instance, horizontal road terrain (ROADHOR [1,2]), vertical road terrain (ROADVERT [1,2]), road surface condition (RDCOND [A,B,C]), controlled maneuver by the driver (DRVRDOIN [1,2]), trafficway description (TRFCWAY [1, 2]), apparent factors of the road/highway (RDWYPC [A, B, C]), driver contributing actions/circumstances (DRVRPC [1,2] [A,B,C,D]), non-motorists actions/circumstances contributing to the crash (NMTACT [1,2] [A,B]), and non-motorist location with respect to the roadway (NMTLOC [1,2]), showed significant detailing in the DT4000 dataset compared to the MV4000 dataset. The formerly mentioned data fields also showed prominent contributions based on their sample size -this reflects the data field completion level-. Not only these data fields were more detailed, but also provided extra information and enhanced the information regarding the crash circumstances and boosted the understanding of the police narrative scenario. Additionally, several data fields were added to the DT4000 dataset and supplemented information missing from the MV4000 dataset such as a total number of lanes (TOTLANES [1,2]), the status of the TCD (TRFCINOP [1,2]), individual condition relevant to the crash (DNMFTR [1,2] [A, B]), and non-motorist actions immediately prior to the crash (NMTPRIOR [1,2]). After assessing the changes in the attributes of new and common data fields, conclusions concerning the examination of the new and recategorized and enhancement of the injury severity model 's accuracy. The importance of DT4000 over MV4000 is stated as follows:

- Regarding pedestrian-vehicle injury severity prediction, pedestrian condition at the time of the crash (DNMFTR.PED.DT), pedestrian location with respect to the roadway (NMTLOC.PED.PED.DT), whether speed was a factor in a crash (SPEEDFLAG), vehicle type (VEHTYPE), the type of traffic control device (TCD) available at the crash location (TRFCCNTL.PED.DT), driver actions that may have contributed to the crash (DRVRPC.PED.DT), and the safety equipment in use by the pedestrian at the time of the crash (NMTSFQ.PED.PED.DT) were the most important and significant factors for pedestrian-vehicle crash injury severity. In addition to the following newly created variables; (LGTCOND.DARK.DRVRDOIN.GO.STR) which refers to dark/unlit crash location with the driver going straight, (LGTCOND.LIGT.DRVRDOIN.GO.STR) which refers to dark/streetlight crash location with the driver going straight.

- Whereas, for bicyclist-vehicle injury severity prediction, the DT4000 dataset provided better attributes -in terms of the variable estimated coefficient- and variables that enhance the prediction of the injury severity compared to the MV4000 dataset through the inclusion of the new data fields and added attributes. Intersection type (INTTYPE), safety equipment usage by the bicyclist (NMTSFQ.BIKE.DT) provided more insights, specifically to the severe injury level of crash severity, which was not well studied using only the MV4000 crash form data fields. Furthermore, vehicle-controlled maneuver prior to the beginning of the sequence of events (DRVRDOIN.BIKE.DT), the vertical terrain of the roadway (ROADVERT.BIKE.DT), bicyclist condition at the time of the crash (DNMFTR.BIKE.DT), bicyclist action that may have contributed to the crash (NMTACT.BIKE.DT), the horizontal road terrain (ROADHOR.BIKE.DT), and the bicyclist location with respect to the roadway (NMTLOC.BIKE.DT). furthermore, trafficway division level and type (TRFCWAY.BIKE.DT) had a significant influence on bicyclist injury severity but were not considered among the most important variables. Hence, DT4000 crash form data fields provided additional insights beyond the MV4000 crash form data fields.

- Few variables showed common effects on both pedestrian and bicyclist crash severity prediction, i.e., the non-motorists condition (DNMFTR.PED.DT and DNMFTR.BIKE.DT), and the non-motorist location (NMTLOC.PED.DT and NMTLOC.BIKE.DT).

7.2 Future Work Recommendations

Based on the findings of the study in hand, future work is recommended as follows:

- The SEM study was limited by not having direct pedestrian and bicyclist volume counts; there may be other land-use variables beyond those considered in this study that contributes to increased pedestrian and bicycle exposure. Future studies should try to use more refined pedestrian and bicyclist exposure data. Pedestrian and bicyclist counts will be helpful to improve the accuracy of latent variable exposure.
- Additionally, crashes were not analyzed if they were not reported to police, or not geocoded in the crash database. The ten-year time period provides more crash data for analysis, but it also increases the chance that a particular corridor had different characteristics when the earliest crashes occurred. The database contained crashes with motor vehicles only, as they appeared to be the most severe, but they have been found to represent only a fraction of total pedestrian and bicycle crashes.
- Although the sample size of 200 corridors is adequate, more sites are desirable to improve the model fit and significance of the input variables. Further work should also

use behavioral data to ensure that these factors are well studied and reduce the potential for omitted variable bias.

- To improve the results of the fault party study, there is room to enhance the text mining classification results for efficiency and accuracy. For example, other features could be engineered, other than the “textlength” generalizable feature which performed well in this study.
- Also, trying different combinations of n-grams might increase the classification accuracy, such as testing trigrams with 4-grams, bigrams with trigrams, bi-grams, and 4-grams, or maybe testing bi grams separately not combined with uni-grams as in this study.
- Concerning the classification models, the RF algorithm was used and performed better than the CART model. RF model is easy to tune and is considered a good general-purpose algorithm. However, other models such as boosted decision trees, or support vector machines (SVMs) may perform better and can be tested.
- For the data quality study, using a larger sample size may provide a chance to study other newly added DT4000 crash form data fields, that did not show a significance to the crash severity models. Also, combining attributes that serve for the same meaning in a future study, is believed to add more value to such new data fields. As an example, the data field describing the relevant condition of the individual (DNMFTR) contains (11) attributes that may distract the police officer while filling the right description of the individual condition. For instance, the attribute “appeared normal” may be confused with “under the influence of medication/drugs/alcohol” especially in cases where the consumption is minimal and does not show clear symptoms yet. Another

example is the data field describing actions/circumstances of the non-motorists that may have contributed to the crash (NMTACT). The attribute “sudden movement into traffic” and “improper passing”, also the attributes “dark clothing” and “not visible (dark clothing, no lighting, etc.)” may be mis-filled through the crash report because of the similarity these attributes carry to each other.

Appendices

Appendix A: Citations for drivers and non-motorists

STATNM [1, 2] [A, B, C, D]Code	Citation Indication	Driver Violation	VRU Violation
341.04 (3)	*Penalty for operating the unregistered or improperly registered vehicle.*	✓	
343.05	*Operators to be licensed.*	✓	
344.62 (1)	*Motor vehicle liability insurance required*	✓	
346.03	Applicability of rules of the road to authorized emergency vehicles	✓	
346.04	Obedience to traffic officers, signs, and signals; fleeing from an officer.	✓	✓
346.04 (2), 346.04 (3)	No operator of a vehicle shall disobey the instructions of any official traffic sign or signal unless otherwise directed by a traffic officer.	✓	
346.05	Vehicles to be driven on the right side of the roadway.	✓	
346.06	Operators of vehicles proceeding in opposite directions shall pass each other to the right, and upon roadways having width for not more than one line of traffic in each direction each operator shall give to the other at least one-half of the main traveled portion of the roadway as nearly as possible.	✓	✓
346.07	Overtaking and passing on the left.	✓	
346.072	Passing stopped emergency or roadside service vehicles.	✓	✓
346.075	Overtaking certain vehicles and devices.	✓	✓
346.13	Driving on roadways laned for traffic.	✓	
346.14	Distance between vehicles.	✓	
346.15	Driving on a divided highway. Whenever any highway has been divided into 2 roadways by an intervening unpaved or otherwise clearly indicated dividing space or by a physical barrier constructed to substantially impede crossing by vehicular traffic, the operator of a vehicle shall drive only to the right of the space or barrier and no operator of a vehicle shall drive over, across, or within the space or barrier except through an opening or at a crossover or intersection established by the authority in charge of the maintenance of the highway, except that the operator of a vehicle when making a left turn to or from a private driveway, alley, or highway or making a U-turn may drive across a paved	✓	

	dividing space or a physical barrier not constructed to impede crossing by vehicular traffic unless the crossing is prohibited by signs erected by the authority in charge of the maintenance of the highway.		
346.18 (2)	General rules of right-of-way. Turning left or making a U-turn at the intersection.	✓	✓
346.19 (1)	General rules of right-of-way. Upon the approach of any authorized emergency vehicle giving an audible signal by siren, the operator of a vehicle shall yield the ROW.	✓	✓
346.23 (1)	Crossing controlled intersection or crosswalk.		✓
346.24	Crossing at uncontrolled intersection or crosswalk.		✓
346.24 (3)	Whenever any vehicle is stopped at an intersection or crosswalk to permit a pedestrian, personal delivery device, bicyclist, or rider of an electric scooter or an electric personal assistive mobility device to cross the roadway, the operator of any other vehicle approaching from the rear may not overtake and pass the stopped vehicle.	✓	
346.25	Every pedestrian, bicyclist, or rider of an electric scooter or an electric personal assistive mobility device crossing a roadway at any point other than within a marked or unmarked crosswalk shall yield the right-of-way to all vehicles upon the roadway.		✓
346.26	Blind pedestrian on highway.		✓
346.27	The operator of a vehicle shall yield the right-of-way to persons engaged in maintenance or construction work on a highway whenever the operator is notified of their presence by flagmen or warning signs.	✓	
346.31	Required position and method of turning at intersections.	✓	
346.32	Required position for turning into a private road or driveway.	✓	
346.35	Method of giving signals on turning and stopping.	✓	✓
346.37 (1)	Vehicular traffic facing a green signal may proceed straight through, make a U-turn, or turn right or left unless a sign at such place prohibits the turning maneuver, but vehicular traffic shall yield the right-of-way to other vehicles and pedestrians lawfully within the intersection or an adjacent crosswalk at the time the signal is exhibited.	✓	
347.48	Safety belts and child safety restraint.	✓	
346.78	Play vehicles not to be used on the roadway. No person riding upon any play vehicle.		✓
346.80 (3)(b)	Persons riding bicycles upon a roadway may not ride more than 2 abreast except upon any path, trail, lane or other way		✓

	set aside for the exclusive use of bicycles, electric scooters, and electric personal assistive mobility devices.		
346.804	Riding a bicycle on the sidewalk.		✓
346.87	Limitations on backing. The operator of a vehicle shall not back the same unless such movement can be made with reasonable safety.	✓	
346.89(1)	Inattentive driving. No person while driving a motor vehicle may be engaged or occupied with an activity, other than driving the vehicle, that interferes or reasonably appears to interfere with the person's ability to drive the vehicle safely.	✓	
346.94 (2)	RACING. No operator of a motor vehicle shall participate in any race or speed or endurance contest upon any highway.	✓	
347.489(1)	A bicycle, motor bicycle, personal delivery device, electric scooter, or electric personal assistive mobility device shall also be equipped with a red reflector that has a diameter of at least 2 inches.		✓
347.489 (2)	No person may operate a bicycle, motor bicycle, electric scooter, or electric personal assistive mobility device upon a highway, bicycle lane, or bicycle way unless it is equipped with a brake in good working condition, adequate to control the movement of and to stop the bicycle, motor bicycle, electric scooter, or electric personal assistive mobility device whenever necessary.		✓
940.09	Homicide by intoxicated use of vehicle or firearm.	✓	

Note: Citations showed in *italic* refer to violations related to vehicle registration and license availability which were excluded from being considered since they do not indicate the party is at-fault.

Appendix B: Police Narrative Text Mining Process

Since a manual review of the crash narrative is a time-consuming and labor-intensive process, text mining techniques will be exploited as an alternative means of enhancing the efficiency of the case review. The reason is that crash narratives in crash reports are unstructured and can be fed directly into machine learning algorithms to perform data mining tasks; and the aim is to test if the algorithms can automatically discover information from the 3,641 crash scenarios in the dataset.

Text mining of the police narratives

Police narrative includes dates, numbers, and many unnecessary link words that need to be processed as an input text data first, and then patterns, rules, and information are extracted from the text to later be interpreted and used as a judgment for initial fault assignment. Zhang et al. (2019) evaluated various methods of processing crash narratives and compared several classification models to determine an optimal approach for secondary crash identification. **Figure B1** illustrates the process of mining police narrative keywords. As illustrated in **Figure B1**, the process consists of five steps. These steps were implemented in several RStudio packages (i.e., qdap, dplyr, tm, wordcloud, dendextend, RWeka, caret, irlba, lsa, and quanteda).

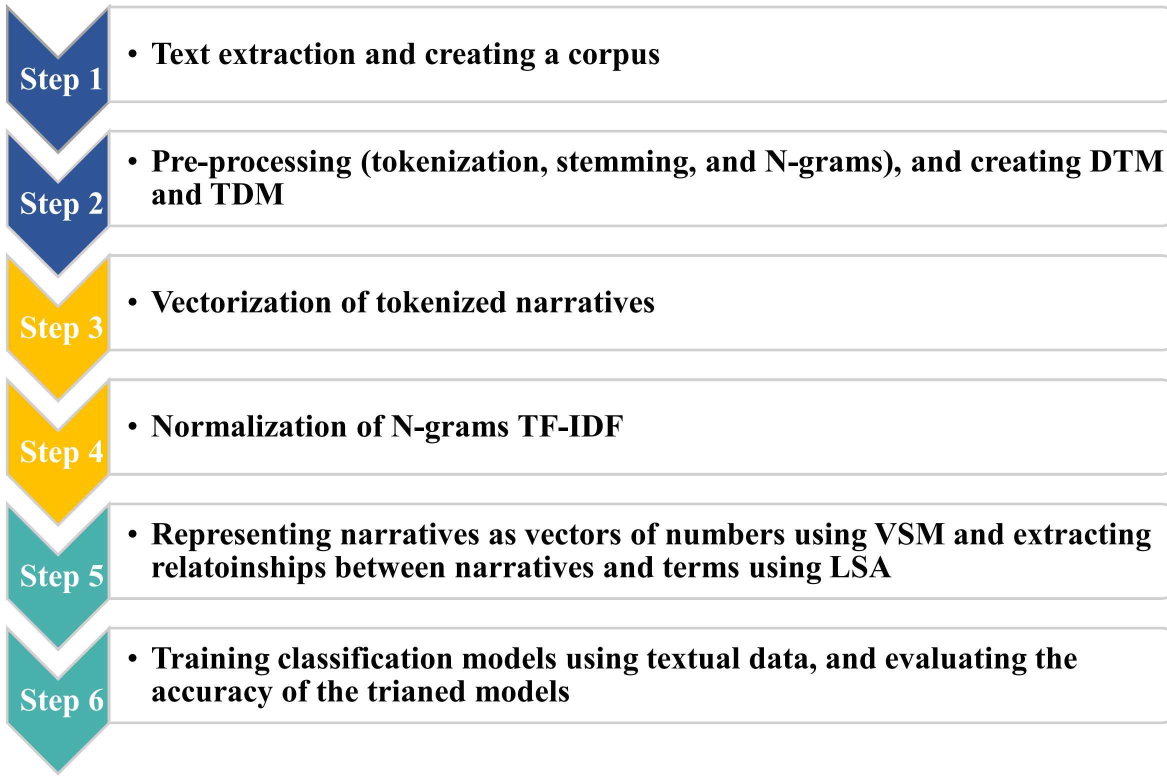


Figure B1: The process generated for text mining the police narratives

Step 1: Text extraction and creating a corpus. Using "tm" package, the column where the narrative is located in Excel is converted into "Corpus"; a collection of text documents.

Step 2: Pre-processing (tokenization, stemming, and N-grams), and creating DTM and TDM. The objective of text mining using the "bag-of-words" approach, is to convert the text to a data frame that includes the words and their frequencies; referred to as document term matrix (DTM) and term-document matrix (TDM).

This step deals with removing sparsity which is related to terms frequency. In DTM, terms form columns, and several columns may represent the same term. In text mining, it is recommended to treat sparsity. In feature extraction, the goal is to use the corpus as a predictor to predict whether the narrative deems a specific party to be at-fault. Next, is tokenization.

This process includes breaking each narrative text into individual terms, represented by n-consecutive words, leading to converting the text into a numerical vector where each word is assigned a special index. The process is also referred to as “bag-of-words”. When tokenization is complete, each row represents a document, each column represents a definite token, and each cell provides the token count for a document (G. Lee 2017; Nayak, Piyatrapoomi, and Weligamage 2010).

Even though this step takes some time; however, it is very important to make sure that DTM and TDM are clean to achieve valuable results. A general scan of the narrative identifies the potential elements that need to be eliminated from the text-called “noise”, and the main cause of this noise is because of the inconsistency in the narrative writing which generated from the fact that different police officers are assigned to write this piece of text in every police report.

Also, common phrases such as “right of way” and “roundabout” are concatenated to appear as the same word and not being counted as a new word. This ensures that only root words appear in the DTM and TDM. Since the corpus is pre-processed, the next step is creating the DTM and TDM. DTM refers to the mathematical matrix which describes terms frequencies, in which rows correspond to documents and columns correspond to terms. Whereas, TDM is the transpose of DTM, and is used for language analysis. Next, exploratory analysis can be performed, and frequently used words may be viewed from the TDM.

Step 3: Vectorization of tokenized narratives. Vectorization is applied after the narratives are tokenized (where each narrative is broken into tokens). This process involves assigning a unique integer index to each unique term, where narratives are represented by a vector where each column represents the frequency of a specific term in a narrative. There are several methods to do vectorization, such as dot product which is implemented as a proxy for correlation between two vectors in this study.

Step 4: Normalization of N-grams TF-IDF. The constructed distributional vectors are evaluated using the oldest method; term-frequency-inverse document frequency (TF-IDF).

TF-IDF is one of the oldest approaches to measure term significance in the document from a corpus of documents -the corpus of narratives in this study-, developed by (Gerard Salton and Buckley 1988). Terms are given different weights than simple frequencies: TF-IDF measures of relevance or TF-IDF scores, depending on term frequency in the document (the more gives a higher score of TF), and in the corpus (the more it appears in various documents the less relevant became IDF) (Vrbanec and Meštrović 2020). TF is calculated as:

$$TF(t,\, d) = \frac{freq\ (t,d)}{\sum_i^n freq\ (t_{i,d})} \qquad \textbf{Eq. B-1}$$

Where:

- TF(t, d): the proportion of the count of the term of interest t in narrative d;
- $freq$ (t, d): count of instances of the term of interest t in narrative d;
- n: number of distinct terms in narrative d.

IDF measures if the term is common or rare across all the narratives (corpus) and is calculated as:

$$\text{IDF } (t) = \log\left(\frac{N}{\text{count } (t)}\right) \qquad \textbf{Eq. B-2}$$

Where:

- N: count distinct narratives in the corpus;
- $\text{count } (t)$: count of narratives in the corpus in which the term if interest t is present.

Then, TF and IDF are combined to enhance the document-term frequency matrix as:

$$\text{TF-IDF }(t, d)\text{: TF}(t, \text{d}) * IDF\ (t) \qquad \textbf{Eq. B-3}$$

The TF-IDF analysis results in one TF-IDF value for each word in each narrative, then the TF-IDF values are normalized using the "weightTfIdf" function.

Step 5: Representing narratives as vectors of numbers using VSM and extracting relationships between narratives and terms using LSA. One of the first approaches in measuring semantic similarity/paraphrase detection between documents namely the vector space model (VSM), which was originally proposed by (G. Salton, Wong, and Yang 1975) for information retrieval is examined. The purpose of examining the VSM is to represent each entity in the collection (letters in words, words in sentences, sentences in documents, documents in the corpus of documents) as a point in n-dimensional space; i.e., as a vector in VSM (Turney and Pantel 2010). The closer the points in this space are, the more semantically similar they are and vice versa. In VSM, each dimension corresponds to one term or word from the narrative set. Weights may be determined by using various weighting schemes; TF-IDF is commonly used in VSM. The main drawbacks of the similarity model are high dimensionality, sparseness, and vocabulary problems. Therefore, there are various modifications and generalizations of this classical version of the VSM. (Vrbanec and Meštrović 2020).

Another approach considered for examination is the latent semantic analysis (LSA) model proposed by (Landauer, Foltz, and Laham 1998). LSA model is carried out to generate word vectors leveraging indirect cooccurrence statistics since similar words are likely to appear in the same context.

For instance, the word "suddenly" is more likely to cooccur with the word "in the roadway" than with the word "in crosswalk". Through the literature, LSA has been successfully implemented for developing several applications in natural language processing (NLP) such as (Turney and

Pantel 2010; Rao and Kak 2011; Šarić et al. 2012; Patra et al. 2020). It also utilizes the vector space model but uses a dimension reduction technique identified by singular value decomposition (SVD) of the initial matrix. LSA is known to overcome the high dimensionality and sparseness of the standard VSM model. **Table B1** shows the chosen text models for comparison and their corresponding similarity metrics. Among the deterministic approaches, VSM would be considered to be the simplest way to represent documents for information retrieval. The deterministic models use the cosine similarity measure on the normalized representation of the documents and the queries (Ruthven and Lalmas 2003).

Table B1: Deterministic models used in the comparative evaluation

Model	Representation	Similarity
VSM	Frequency vector	Cosine similarity
LSA	K dimensional vector in the eigenspace	

The similarity between two narratives is calculated as a cosine similarity as in the following equation:

$$\text{Sim}\ (\vec{A}, \vec{B}) = \frac{\vec{A}\ W\ \vec{B}}{\|\vec{A}\|\|\vec{B}\|} = cos\ \theta = \frac{A\ .B}{\|A\|_2\ \|B\|_2} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \qquad \textbf{Eq. B-4}$$

Where:

- $\vec{A}\ and\ \vec{B}$ are binary vectors that represent each narrative;
- W is the semantic similarity matrix which comprises the similarity of each word pairs;
- θ: Angle between vector A and vector B;
- $A\ .B$: dot product of A and B;

- $\parallel A \parallel_2 \parallel B \parallel_2$: length of A times length of B.

Step 6: Training classification models using textual data and evaluating the accuracy of the trained models. Steps 1 through 5 are considered a preparation for building the classification model for the crash party at fault.

Text Mining Results and Discussion

In this section, the classification algorithms used are single decision trees (i.e., CART), and the random forest (RF) models. Random forest (RF) is an ensemble learning method that consists of several decision tree classifiers. Each tree in the random forest uses only a subset of the selected features for classification; to make sure that the results of different trees are uncorrelated. RF is known to generate classification results based on the majority rule. The standard performance metric for the RF is the out-of-bag (OOB) estimate of the error (Franklin 2005). The RF method is recognized for reducing overfitting and accurate prediction compared to single decision trees (i.e., classification and regression trees (CART)). To achieve optimum performance, the main hyperparameters should also be tuned. In this study, the hyperparameters comprise of: i) the number of decision trees in the forest, ii) the number of influencing features considered in each split, and iii) the minimum number of samples required for internal and leaf nodes.

The Classification and Regression Trees (CART) method is a non-parametric statistical algorithm developed by Leo Breiman et al. (Breiman 1984). It builds both classifications and regression trees. CART is based on a binary splitting of the attributes. CART methodology comprises three main stages: growing or splitting decision trees, pruning, and selection of the optimal tree as follows: i) splitting; the process of tree building starting with splitting the root node into two child nodes, ii) pruning; develops an optimal tree via shedding off the large tree's branches, and iii) optimal tree selection; a tree with the smallest prediction error for new samples -test data-. In this study, the prediction error is measured by using cross-validation.

First, missing cases were checked via the “complete.cases” function, and then completed before data exploration. The most obvious observation about the collected crash narratives is the large variance in narratives’ text length. Therefore, a new feature of our data frame is created and named “text length”, where the length of characters for every narrative is calculated using the “nchar” function. The illustration of the text length summary for our data is shown in **Figure B2**, where we see a lot of variance in the data. The summary showed that the max number of characters is (2266) characters, and a minimum of (2) characters-such narratives were later excluded since they are incomplete. The median value or the 50th percentile; the value at which half of the narrative text is shorter and longer than this value is (353) characters. The 3rd quartile value showed that 75% of the narrative text is (590) characters or less.

While processing the calculations of classification models over the corpus through the text mining process, it became obvious that the models’ results could not be highly accurate and informative while using the four at-fault party categories; driver at-fault (DRVR), VRU at-fault (VRU), both parties at-fault (BOTH), no or unknown party at-fault (NO/UNKN), as intended while planning the experiments. It was necessary to determine the best combination of fault categories for every model and corpus. Fault categories considered for the text mining analysis were decided from running the classification models using the training part of the dataset and evaluated on the testing parts through the 3-cross validation method because there are three categories in it; driver at-fault “DRVR”, VRU at-fault “VRU”, no party at-fault “NONE”.

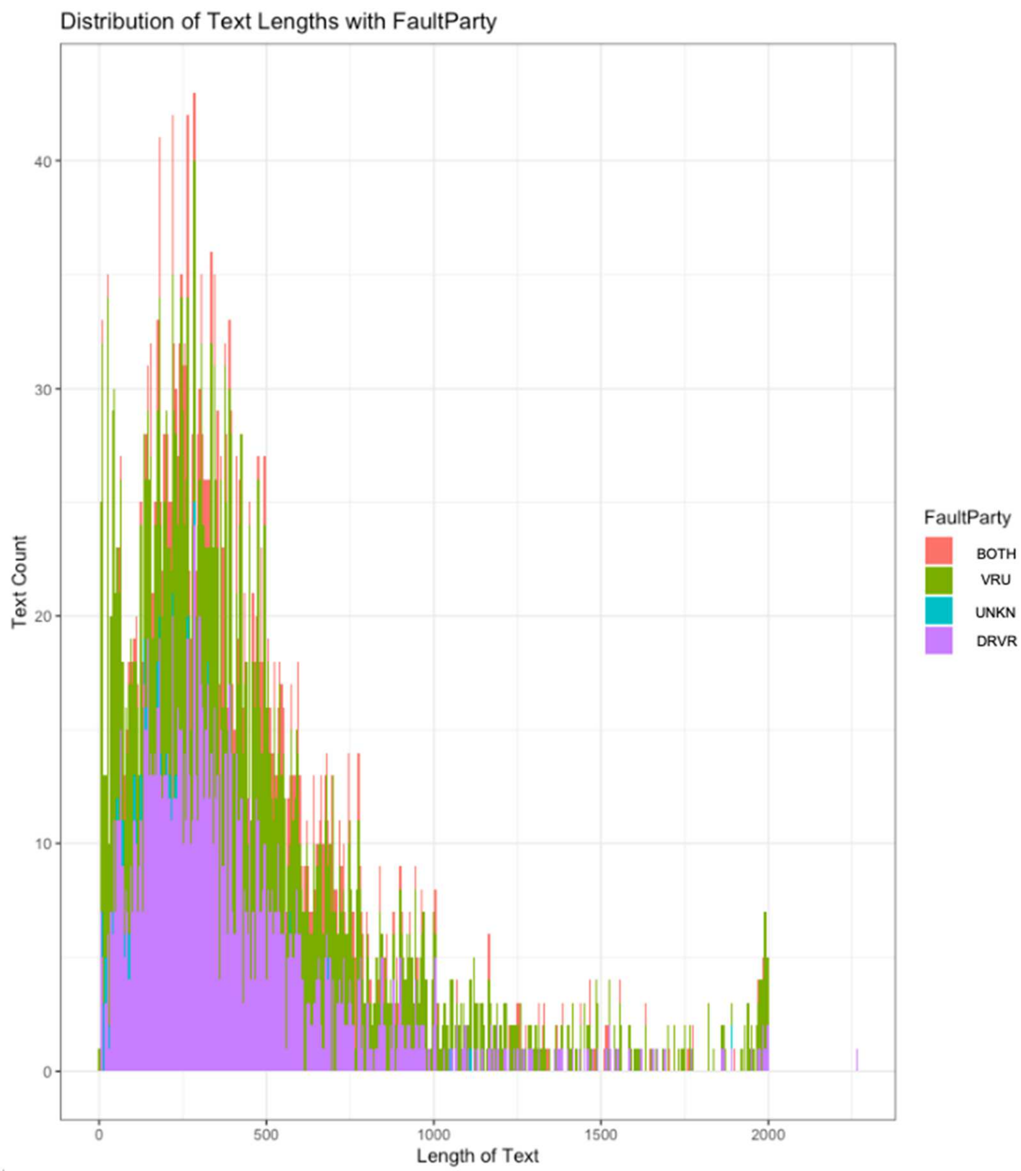


Figure B2: An illustration of the propensity of the four at-fault categories based on the "textlength" feature engineered in the initial stages of the narrative mining process.

In machine learning (ML), the goal is to create models that can generalize and perform well on new data, called "generalization". Hence, the dataset is split between training and test datasets with a ratio of (3:1) using the "createDataPartition" function from the "caret" package to make a stratified split. This function randomly samples the data and ensures that the split is stratified meaning that the proportions are maintained across the splits. Narratives are then tokenized, meaning that their text is decomposed into distinct pieces.

Next, narratives were pre-processed; punctuation (such as |*~’”`#$%^_=-;/<:.{}) which appeared to deliver no information and therefore was omitted and replaced by spaces. Stop words (such as “I”, “they”, “was”, “having”, “isn’t”, “don’t”, “but”, “each”, “very”) and broken words which generated from removing the punctuation and adding spaces instead, are both removed to obtain more accurate results from mining the remaining text. The “quanteda” package’s built-in stop-word list for English was used after being inspected to make sure it applies to the mining problem tackled in this study as every stop-ward list is different among different packages. Some words are added to the set of stop words predefined by R, are words that are appeared differently throughout the records, such as Unit1, U1, Unit 1, unit 1. Additionally, words that are differently abbreviated, those with spelling mistakes, and words with an upper case were converted to lower case using “tokens_tolower” function. The text is then ready for the stemming process, meaning that inflected/derived words are reduced to their word stem using the “tokens_wordstem” function. For instance, “turned”, “turning”, “turns” are reduced to “turn”.

Next, the document-frequency matrix (DFM) is constructed where each row represents a specific narrative using the “dfm” function; each column represents a distinct token; and each cell contains the count of that token for a particular narrative. Since word order is not preserved, the “bag-of-words” approach is used throughout this police narrative mining process. It is noted that after constructing the DFM, the matrix suffers from the dimensionality or sparsity problem, and feature engineering is proposed to contribute to solving this issue.

Cross-validation is performed after pre-processing is applied to the narratives, to create the first bag-of-words model. However, "make. names" function is used before training the model to clean column names.

Using the "caret" package, stratified folds are created for 10-fold cross-validation (k=10) repeated 3 times, meaning that 30 random stratified samples were created to get more accurate estimates through the "createMultiFolds" function. The first trained model is a single decision tree model; classification and regression tree (CART) model. Hyperparameter tuning is used via the "rpart" package by trying 7 different configurations (tuneLength=7) of the CART algorithm. Accuracy was used to select the optimal model which resulted in 81.75% at a complexity parameter (CP) value -the tuning parameter for "rpart"- of (0.0411449).

However, longer narratives were observed to have higher term counts, and terms that appear more frequently across the corpus do not necessarily reflect their significant importance. To improve the accuracy result of the CART model, a TF-IDF is used to adjust the DFM and normalize the term counts across the narrative based on the text length. Matrix adjustment is needed to accommodate these two issues. Hence, TF-IDF is calculated. TF -which is document-centric meaning rows are normalized- is calculated for each narrative using "term.frequency" function, and IDF -which is corpus centric meaning columns are normalized - is calculated using "inverse.doc.freq" function. Then, TF and IDF are combined into TF-IDF. Using the TF-IDF data, the CART model is applied to check for any improvement in the model accuracy. Results showed that accuracy has improved slightly to reach 82.84%.

The original data representation was the document term frequency matrix DFM, each row represented a document and each column was a term as a result of data preprocessing (tokenization, stop-word removal, lower casing, removing certain numbers, and stemming). Then, the data is transformed using the TF-IDF. Next, the N-grams are used to augment the representation which involves single terms; called unigrams or 1-grams and bigrams.

N-grams allows the extension of the bag-of-words model to include word order to provide more signal to the machine learning model. Tri grams are added at this stage of analysis to enrich the current feature matrix using the "tokens_ngrams" function provided by the "quanteda" package. However, we've more than doubled the total size of the matrix (dimensionality problem). To check if adding tri grams improve the model effectiveness, the CART model was re-run, with stratified folds created for 10-fold cross-validation (k=10) repeated 3 times. The results showed that the model's accuracy declined, hence, the addition of trigrams appeared to negatively impact the single decision tree model; CART.

However, issues with document-term matrix (i.e., a wide number of columns), document-term matrices sparsity (i.e., sparse features), and the scalability issue (i.e., huge computation amount), lead to the dimensionality problem which resulted from the aforementioned issues. So, to solve these issues, VSM and LSA may be utilized. The vector space model (VSM) is utilized which allows us to address these problems.

VSM represents documents (i.e., narratives) as vectors of numbers, allowing working with document geometrically "i.e., dot product". Also, other tools may be implemented (e.g., latent semantic analysis (LSA)) which extracts relationships between the narratives via the matrix factorization technique; singular value decomposition (SVD).

SVD allows implementing feature reduction/extraction to improve the data representation while simultaneously allowing for more robust models (i.e., random forest) to improve the classification accuracy. However, the reduced factorization matrices are approximations, and new data need to be projected into the matrix, but the information loss is negligible compared to a loss of information from. Modeling results of the CART and RF models are reported in **Table B2**.

Table B2: modeling results throughout the narrative mining process N=3,642

Model	**Samples**	**CP**	**Accuracy**
1st CART model, with simple document term frequency matrix	2551	0.0411449	0.8174607=81.75%
2nd CART with TF-IDF	2551	0.04472272	0.8284448=82.84%
4th CART with SVD	2551	0.04472272	0.8955932=89.56%
Model	**Samples**	**mtry**	**Accuracy**
RF model with LSA SVD	2551	2	0.9938598=99.38%
RF model results with (text length) feature aggregation	2551	251	0.9942504=99.42%

Note: accuracy was used to select the optimal model using the largest value.

The process of projecting new data consists of the following: i) using "term. frequency" function to normalize the document vector; ii) using "tf.idf" function to complete TF-IDF projection; and iii) applying SVD projection to the document vector.

Using the "irlba" package, which uses truncated SVD to assign a certain number of the most important extracted features. The CART model is implemented after using the LSA technique, and the results showed a significant increase in models' accuracy (89.56%). From the perspective of single decision trees, we gained accuracy by adding bigrams, and then by adding SVD to bigrams, the increase is more than (7%). To approve that the LSA technique increases the information density of each feature, the random forest (RF) model is considered.

Results showed that using the RF model increased the model's accuracy to (%). The "mtry" parameter controls how much data is used in building the individual tree, and by default

RF in R builds 500 trees. The results also showed that the best RF model was built using approximately half of the data being available for each tree of the RF (250), and this is (9.8%) accuracy higher than the accuracy of a single decision tree model.

Hence, the RF performed better and showed more ability in making use of the data preprocessing pipeline. In particular, it can make better use of the LSA SVD matrix factorization than single decision trees. More informative results were acquired by using the “confusion matrix” function from the “caret” package. Obtained accuracy is slightly higher as the model worked on all of the data as opposed to cross-validation where the testing data was not included.

The “confusionMatrix” function also provides several performance metrics (i.e., sensitivity, specificity, Pos Pred Value, Neg Pred Value) that are useful in interpreting the results. While accuracy is considered intuitive and common in deciding the optimal model, it is not the only metric to be used.

Hence, using other metrics such as sensitivity and specificity provides help in building the most effective machine learning model. The results showed that the model is better at predicting VRUs at-fault than drivers at-fault. Sensitivity (0.9979) and specificity (0.9885). In this study, the preference is to be able to correctly predict both, VRUs and drivers’ fault through police narratives. Therefore, engineered features may provide an enhancement to the model’s accuracy and that is discussed next.

Revisiting the (text length) feature; count of the number of characters in narratives, this feature appeared to be predictive through inspection provided earlier. Hence, the (text length) feature was aggregated into the training data and the RF model was examined. Results of the RF model with the aggregation of the (text length) feature, showed that accuracy has increased slightly (0.05%) from 99.37% to 99.42%, and that sensitivity and specificity increased reaching (0.9958) and (0.9902), respectively. These results denote a reduction in DRVR errors 13 wrong to 11 wrong. These results indicate that the (text length) is a promising feature, therefore it is next tested in the production of unseen data (test data).

Given that there is some room for enhancing the model's performance, cosine similarity was implemented to engineer a new feature using the "lsa" package. The similarity in vector space is using the cosine between document vectors and is considered an improvement over the dot product and works well in high dimensional spaces (which suffers from high dimensionality). Thus, the cosine similarity was examined next. The results showed an increase in accuracy and sensitivity; whereas, specificity has decreased.

The results showed that both the DRVR and VRU's similarities are important features, and this may be indicative of overfitting since it didn't increase both measures simultaneously. When (text length) feature raised both sensitivity and specificity at the same time, it was considered a good feature. Hence, the similarity feature is opted out.

Finally, the test data is transformed into vector space, and the RF model is tested with all the (300) features engineered/extracted using the SVD LSA including the (text length) feature. Preprocessing was applied to the test data (i.e., tokenizing, converting to a lower case, removing stop words, stemming, adding bigrams, and converting to document term frequency matrix).

The results showed that the accuracy has been decreased slightly by (0.34%) using the test data (from 99.33% to 98.99%) than using the training dataset, and sensitivity reached a 100% value (1) for the DRVR denoting that the model predicts all DRVR at-fault correctly. Nevertheless, specificity was slightly reduced from 0.9902 to 0.9770, meaning that 97.70% of VRU were correctly predicted, so this is a tradeoff and is acceptable. Also, note that with the training data (measured by cross-validation which is an estimate of generalization) the reported accuracy was a little over 0.3% better than with test data. Note that stratified samples were created for both training and test datasets, so we know that the proportion of DRVR, NONE/UNKN, and VRU in the test dataset mimics that of the training dataset.

Table B3: Confusion matrix of Rf on test data without similarity

Confusion matrix of the Rf model using test data and excluding the similarity feature			
	Reference		
Prediction	DRVR	NONE/UNKN	VRU
DRVR	613	0	11
NONE/UNKN	0	7	0
VRU	6	0	460
Statistics			
Class	DRVR	NONE/UNKN	VRU
Accuracy	0.9899		
Sensitivity	1.0000	1.0000	0.9766
Specificity	0.9770	1.0000	1.0000

Formerly, **Figure B2** showed the propensity of the three at-fault categories based on the "textlength" engineered feature which highlighted a significant number of VRU at-fault crashes that were associated with a high number of characters in the police narratives.

This feature was tested statistically through the RF model and showed an improvement in the model's accuracy when included with other features extracted from the text of the narratives. For instance, the following narrative shows that the police officer used a large number of words to describe a VRU at-fault crash: "*unit 1 driver identified by xxx. unit 2 operator identified by xxx. unit 1 driver was driving north on xxx in the city of xxx. unit 2 was riding his bicycle north on the right side of xxx unit 2 driver stated he was listening to music with earbud headphones. unit 2* *driver turned across traffic way to reach a path on the opposite side of the road. unit 2 driver stated he did not see unit 1. this caused unit 2 to strike unit 1. witness verified this account of course of events. minor injuries reported by hospital staff on operator of unit 2. unit 1 occupants uninjured. there was prior damage to unit 1 front passenger area".*

Regarding false positive classifications, they can result from the presence of the formerly mentioned keywords in the crash narrative, which then dominate the narrative vector and result in a higher probability of the crash being falsely classified as driver or VRU at-fault crash. Two crash narratives are provided below showing incorrectly classified crashes as a driver at-fault crash and VRU at-fault crash clarifies this point. The location and identifiable information in the following narratives were redacted for information protection purposes.

- Narrative example of false-positive driver at-fault crash: *"unit # 1 was* ***backing*** *from being parked when unit #2 the pedestrian entered the roadway not at a crosswalk to cross the street westbound. unit #2 was jaywalking".*
- Narrative example of false-positive VRU at-fault crash: *"unit 1 was traveling SB on xxx ave then proceeded to turn left or EB on xxx ave/hwy. a pedestrian was walking southbound in crosswalk. it is unknown if the pedestrian walk button was showing walk or dont walk at the time of the collision. The driver of unit 1 stated that he could not see the pedestrian. note that it was dark and rainy, and visibility was fairly poor. the pedestrian was wearing* ***dark clothing****. pedestrian claimed possible injury. nothing further".*

Moreover, some keywords are not related to a specific party being at-fault appeared less frequently, maybe assigned high TF-IDF values. This intensifies the influence of such keywords on the classification result; for instance: *"on the above date and time I responded to the* *intersection of xxx ave at xxx dr for a crash. upon arrival I made contact with unit 1 who explained to me that prior to the collision she was on south bound xxx dr in the right turn lane to turn onto west bound xxx ave. unit 1 explained that when she approached the right turn lane she had a red traffic signal. she stopped and did not see any pedestrian traffic in the cross walk. unit 1 began to look to her left to watch for traffic on west bound xxx ave. when she saw traffic was clear she began to make the right turn. as unit 1* ***accelerated*** *she collided with unit 2 who was east bound on the sidewalk and entered the intersection on his bicycle. unit 1 told me that she was approximately 10 feet south of the crosswalk where she collided with unit 2. I made contact with unit 2 who explained to me that prior to the collision he was east bound on the sidewalk off of xxx ave. as he approached xxx dr he was going to continue east through the intersection. unit 2 told me that he had a signal for pedestrians to proceed through the cross walk and as he entered the intersection unit 1 accelerated to make a right turn onto xxx ave and they collided. A witness".*

The keyword *"accelerated"* occasionally appeared in the corpus of crash narratives. Hence, its TF-IDF value was relatively high (0.09985466), and it was assigned as the fourth most important keyword for classifying driver at-fault crashes, after "fled_scene", "not_see", and "backing". Conversely, "speeding" was among the least important words for identifying driver at-fault crashes, with a low TF-IDF value of (0.001891378).

This designates that the bag-of-word representation is incapable to consider the exact word meaning. Word normalization must be improved in further analysis to lessen this issue. None of the important keywords appeared in false-negative narratives. For instance, in driver at-fault crash false negative narratives, no presence of the defined important keywords. The following narrative explains this point and shows an example that was not classified as driver at-fault crash: "*unit #2 was crossing n. xxx st on his bicycle and heading eastbound on w. xxx. unit #2 had the right of way* *to cross at the cross walk on n. xxx. unit #1 failed to yield for the bicyclist to cross the street fully. unit #1 made a right turn from w. xxx on to n. xxx and crashed into unit #2 crossing the street in the crosswalk. unit #1 did not stay on scene after crashing into unit #2. NFA city of xxx police department*".

Another observed result concerning the false-negative classification of driver at-fault crashes is that even with using a stratified split to split the training and test datasets, some keywords that appear in the test dataset may not appear in the training dataset leading to the model's inability to detect such keywords. This point is illustrated in the following narrative: *"unit #1 struck unit #2 after an argument and fled the scene n/b on n. xxx st. unit #2 suffered a broken leg and several abrasions. unit #2 was transported by med #5 to xxx hospital and admitted at 1832. operator of unit #1 was located and arrested for recklessly endangering safety. MPD case # xxx".* This scenario described that the driver was recklessly driving, making it a driver at-fault crash. Despite this, "reckless" did not appear in the training set, hence, was not detected by the model. Extra narratives can enhance the model's ability and lessen the effect of this issue.

Through the manual narrative revision, some scenarios were found to be confusing due to the lack of a specific definition of a driver and VRU at-fault crash, especially when a narrative involves an eyewitness testimony that leans towards showing that the driver is at-fault even if not. Hence, a specific standard was followed in determining the fault party manually to determine a crash was a result of the fault of a specific party. For instance, a driver violating a specific traffic rule that was not included in the field attributes mentioned in **Table 5-7** (i.e., failure to yield the ROW), was considered a driver at-fault crash. Otherwise, if the narrative did not involve such violation even if a citation related to the crash was issued, it will not be considered a driver at-fault crash since the citation is considered in the fault assignment as a sperate data source.

The following narratives do not indicate any of the two fault parties were responsible for the crash, and therefore may be debated not to be deemed as a driver at-fault crash and VRU at-fault crash, respectively.

- *"unit 2 was riding her bike in the crosswalk on xxx st at w. xxx ave. unit 1 was traveling eastbound on x and struck unit 2. driver of unit 1 said he was unable to stop in enough time"*.
- *"operator of unit two is a bicycle. operator does not remember which direction he was riding his bike prior to the accident. operator does not remember which direction the striking vehicle was coming from. based on the location of the bicycle and operator after the crash the vehicle could possibly have been traveling northbound on n xxx St. it is still unknown which direction the bicyclist was riding. Milwaukee fire department"*.

Such scenarios express how challenging is the task of correctly classifying the crash's responsible fault party crashes for a human being, let alone machines. Even so, the fact that the model classified these crashes reveals its sensitivity.

Appendix C: Definition and Attribute of Data Fields

Variable and Attribute Codes	Variable and attribute Code Indication	
	Roadway Level	
Horizontal Road Terrain		
ROADHOR-MV	The horizontal road terrain at the point of impact. The options for this field are either straight or curve. The field will only be filled in on this summary if curve (C) was indicated.	
ROADHOR [1,2]-DT	The curvature of the roadway in the direction of travel for the vehicle.	
	ST	Straight
	LT	Curve Left
	RT	Curve Right
	CU	Curve-Unknown Direction
	UNKN	Unknown
Vertical Road Terrain		
ROADVERT-MV	The vertical road terrain at the point of impact. The options for this field are either flat or hill. The field will only be filed in on this summary if hill H was indicated.	
ROADVERT[1,2]-DT	The grade of the roadway in the direction of travel for this vehicle.	
	LVL	Level
	CST	Hillcrest
	UP	Uphill
	DN	Downhill
	SAG	Sag (Bottom)
	UNKN	Unknown
Road Surface Condition		
ROADCOND-MV	The surface condition of the road at the point of origin for the unit apparently most at fault. If blank the road condition is DRY.	
RDCOND [A,B,C]-DT	The roadway surface condition at the time and place of a crash.	
	DRY	Dry
	WET	Wet
	SNOW	Snow
	SLUSH	Slush
	ICE	Ice
	WATER	Water (Standing/Moving)
	SAND	Sand
	MUD	Mud/Dirt
	GRAVL	Gravel
	OIL	Oil
	UNKN	Unknown
Trafficway Description		
TRFCWAY-MV	Text describing areas designed for motor vehicle operation.	
	BLNK	Blank
	ND	Not physically divided
	D/WO	Divided highway without traffic barrier
	D/B	Divided highway with a traffic barrier
	OW	One-way traffic
	OTHR	Parking lot or private property
TRFCWAY[1,2]-DT	Indication of whether or not the trafficway for this vehicle is divided and whether it serves one-way or two-way traffic.	
	UNDIV	Two-Way-Not Divided
	TWLTL	Two-Way, Not Divided, With A Continuous Left Turn Lane
	DIV NO	Divided Hwy W/O Traffic Barrier
	DIV PNT	Two-Way, Divided, Unprotected (Painted > 4 Feet) Median
	DIV BAR	Divided Hwy W/Traffic Barrier
	DIV MBR	Divided Hwy Median W/Barrier
	OW	One-Way Traffic
	PL/PP	Parking Lot or Private Property
	RAMP	Entrance/Exit Ramp

	UNKN	Unknown
Total Number of Lands		
TOTLANES[1,2]-DT	The total number of lanes in the roadway on which this motor vehicle was traveling. For undivided highways - total through lanes in both directions, excluding designated turn lanes. For divided highways - total through lanes for roadway the motor vehicle under consideration was traveling. <u>This is a new variable suggested in the DT4000 crash form.</u>	
Location of First Harmful Event		
RLTNRDWY-MV	Location of the first harmful event in relation to a roadway.	
	GORE	Gore
	LTSH	Outside should-left
	MED	Median
	OFF	Off roadway - location unknown
	ON	On roadway
	PLOT	Private lot or private prop
	RAMP	On ramp
	RTSH	Outside shoulder-right
	SHLD	Shoulder
RLTNRDWY-DT	The location of the first harmful event as it relates to its position within or outside the trafficway	
	ON	On Roadway
	LTSH	Shoulder Left
	RTSH	Shoulder Right
	MED B	Median Barrier
	R SIDE	Roadside
	GORE	Gore
	SEP	Separator
	PARK	In Parking Lane or Zone
	OFF	Off Roadway, Location Unknown
	O ROW	Outside Right-Of-Way (Trafficway)
	CTLT	Continuous Left Turn Lane
Crash Location with Respect to Trafficway		
RLTNTRWY-DT	Identifies the location of a crash with respect it's relation to a trafficway. <u>This is a new variable included in the DT4000 crash form.</u>	
	ON	Trafficway - On Road
	OFF	Trafficway - Not On Road
	P LOT	Non-Trafficway - Parking Lot
	OTHR	Non Trafficway-Other
Crash Location Type		
ACCDLOC-MV	The type of location at which a crash occurred. Types I and N are public roadway crashes.	
	I	Intersection related
	N	Non intersection related
	PL	Parking lot
	PP	Private property
LOCTYPE-DT	The location type of a crash.	
	I	Intersection (public roadway),
	N	Non-intersection (public roadway)
	PL	Parking lot
	PP	Private Property
Intersection Type		
INTTYPE-DT	The type of intersection in which a crash occurred. An intersection consists of two or more roadways that intersect at the same level. <u>This is a new variable included in the DT4000 crash form.</u>	
	NA	Not At Intersection
	4 WAY	Four-Way Intersection,
	T	T-Intersection
	L	L-Intersection
	RAB	Roundabout
	5	Five-Point or More
Status of the TCD		

TRFCINOP[1,2]-DT	Indicates whether a traffic control device was inoperable or missing at the time of the crash (Y/N/UNKN). This is a new variable included in the DT4000 crash form.	
Crash Occurrence Within an Interchange Area		
RLTNJNIC-DT	The coding of this data element is based on the location of the first harmful event of the crash. It identifies the crash's location with respect to presence in a junction or proximity to components typically in junction or interchange areas. This field identifies if a crash occurred within the Interchange area. (Y/N/UNKN). This is a new variable included in the DT4000 crash form.	
Environmental Level		
Prevailing Atmospheric Conditions		
WTHRCOND-MV	A code which identifies the weather condition at the time of a crash.	
	BLNK	Blank
	CLR	Clear
	CLDY	Cloudy
	RAIN	Rain
	RAIN	Rain
	SNOW	Snow
	FOG	Fog / smog / smoke
	SLET	Sleet / hail
	WIND	Blowing sand / dirt / snow
	XWIND	Severe crosswinds
WTCOND[A,B]-DT	The prevailing atmospheric conditions that existed at the time of the crash.	
	CLEAR	Clear
	CLDY	Cloudy
	RAIN	Rain
	SNOW	Snow
	SLEET	Sleet/Hail
	WIND	Severe Winds
	FRZ RN	Freezing Rain or Freezing Drizzle
	FOG	Fog
	B SNOW	Blowing Snow
	SMOG	Smog/Smoke
	B DIRT	Blowing Sand, Soil, Dirt
Light Conditions		
LGTCOND-MV	Light condition at time of crash. If blank the light condition is DAY.	
	DARK	Nighttime-Unlit
	LIGT	Nighttime–Street Lights
LGTCOND-DT	The type/level of light that existed at the time of the motor vehicle crash.	
	DAY	Daylight
	DAWN	Dawn
	DUSK	Dusk
	LITE	Dark/Lighted
	DARK	Dark/Unlit
	DK/UN	Dark-Unknown Lighting
Contributing environmental Conditions		
ENVPC[A,B,C]-DT	Apparent environmental conditions which may have contributed to the crash. This is a new variable included in the DT4000 crash form.	
	NONE	None
	WTHR	Weather Conditions
	OBSTR	Visual Obstruction(s)
	GLARE	Glare
	ANML	Animal(s) In Roadway
Driver Level		
The Driver Condition Relevant to the Crash		
DNMFTR [1,2] [A,B]-DRVR-DT	Any relevant condition of the individual (motorist or non-motorist) that is directly related to the crash.	
	NORM	Appeared Normal
	PHY IMP	Physically Impaired
	EMO	Emotional (Depressed, Angry, Disturbed, Etc.)
	SICK	Ill (Sick)- Fainted

	SLEEP	Asleep or Fatigued
	UI MDA	Under the Influence of Medication/Drugs/Alcohol
	WCHAIR	Paraplegic or Restricted to Wheelchair
	CONF	Confused or Disoriented (Non-Lucid)
	BLIND	Blind
	CANE	Using Cane or Crutches
	NO OBS	Not Observed
Distraction/Inattentive Driving		
DISTFLAG-DT	Flag indicating whether a crash involved distracting or inattentive driving.	
Driver Contributing Actions/Circumstances		
DRVRPC[1,2]-MV	Lists the possible driver contributing circumstances (driver factors) in a collision.	
	DC	Driver condition
	DIS	Physically disabled
	DTC	Disregard traffic control
	FTC	Following too close
	FTY	Failure to yield
	FVC	Failure to keep vehicle under control
	IC	In conflict
	ID	Inattentive driving
	IO	Improper overtake
	IT	Improper turn
	LOC	Left of center
	OTR	Other
	SPD	Exceed speed limit
	TFC	Too fast for conditions
	UB	Unsafe backing
DRVRPC[1,2][A,B,C,D]-DT	The actions by the driver that may have contributed to the crash, based on the judgment of the law enforcement officer investigating the crash.	
	SPD	Exceed Speed Limit
	TFC	Speed Too Fast/Cond
	FTY	Failed To Yield Right-Of-Way
	FTC	Following Too Close
	IT	Improper Turn
	UB	Unsafe Backing
	FVC	Failure To Control
	ROR	Ran Off Roadway
	DRED	Disregarded Red Light
	DSS	Disregarded Stop Sign
	DTC	Disregarded Other Traffic Control
	DRM	Disregarded Other Road Markings
	IOR	Improper Overtaking / Passing Right
	IOL	Improper Overtaking / Passing Left
	WW	Wrong Side or Wrong Way
	FDL	Failed To Keep In Designated Lane
	AR	Operated Motor Vehicle In Aggressive/Reckless Manner
	ID	Operated Motor Vehicle In Inattentive, Careless or Erratic Manner
	IC	Swerved or Avoided Due To Wind, Slippery Surface, Motor Vehicle, Object, Non-Motorist In Roadway, etc.
	OVR	Over-Correcting/Over-Steering
	RAC	Racing
	NO	No Contributing Action
	NOT SEE	Looked But Did Not See
Controlled Maneuver by the Driver		
DRVRDO[1,2]-MV	What the driver of unit was doing at the time of the crash.	
	BACKING	Backing up
	CHG LN	Changing lanes
	GO STR	Going straight
	IL PRK	Illegally parked
	LG PRK	Legally parked

	LT TRN	Making left turn
	MERGING	Merging into traffic
	NEGCRV	Negotiating curve
	NPASZN	Violate no pass zone
	OVT LT	Overtaking on the left
	OVT RT	Overtaking on right
	PARKNG	Parking maneuver
	RT TRN	Right turn
	RTOR	Right turn on red
	SL/ST	Slowing or stopped
	STOPED	Stopped in traffic
	UTURN	U turn
DRVRDOIN[1,2]-DT	The controlled maneuver for this motor vehicle prior to the beginning of the sequence of events.	
	GO STR	Going Straight
	NEGCRV	Negotiating Curve
	BACKING	Backing
	CHG LN	Changing Lanes
	OVT RT	Overtake Right
	OVT LT	Overtake Left
	RT TRN	Right Turn
	LT TRN	Left Turn
	UTRN	U Turn
	LVG LN	Leaving Traffic Lane
	ENT LN	Entering Traffic Lane
	SLOWNG	Slow/Stopping
	LG PRK	Legally Parked
	STOPED	Stop in Traffic
	NO PASS	Viol No Pass Zn
	PARKNG	Park Maneuver
	RTOR	Turn on Red
	MERGING	Merging
	ACCEL	Accelerating in Road
	STARTNG	Starting in Road
Safety Equipment Used by the Driver		
SAFETY[1,2]-DR-MV	The type of safety equipment, if any, that was used by a driver, bicyclist or pedestrian involved in a crash.	
	SH/LP	Shoulder & lap belt
	LAP	Lap belt only
	SHLD	Shoulder belt only
	CHILD	Child safety seat
	HT/EY	Helmet & eye protection
	EYE	No helmet / eye protection only
	NA	Not applicable-non-motorist
	HLMT	Helmet
SFTYWQP [1, 2]-DR-DT	The restraint equipment in use at the time of the crash (excluding motorcyclists).	
	SH/LP	Shoulder & Lap Belt
	LAP	Lap Belt Only
	SHLD	Shoulder Belt Only
	UNTYPE	Restraint Used - Type Unknown
	CH/FF	Child Restraint System - Forward Facing
	CH/RF	Child Restraint System - Rear Facing
	BOOST	Booster Seat
	CH/UN	Child Restraint - Type Unknown
Driver Race		
RACE[1,2]-DT	The race of the driver per the Wisconsin Uniform Traffic Citation. <u>This is a new variable included in the DT4000 crash form.</u>	
	A	Asian
	B	Black
	I	Indian

	H	Hispanic
	W	White
Teen Drivers		
TEENDRVR-DT	Flag indicating whether a crash involved a driver between the age of 16 and 19. This is a new variable included in the DT4000 crash form.	
	Pedestrian Level	
The Pedestrian Condition Relevant to the Crash		
DNMFTR[1,2][A,B]-PED-DT	Any relevant condition of the individual (motorist or non-motorist) that is directly related to the crash.	
	NORM	Appeared Normal
	PHY IMP	Physically Impaired
	EMO	Emotional (Depressed, Angry, Disturbed, Etc.)
	SICK - Ill	Ill (Sick), Fainted
	SLEEP	Asleep or Fatigued
	UI MDA	Under the Influence of Medication/Drugs/Alcohol
	CONF	Confused or Disoriented (Non-Lucid)
	WCHAIR	Paraplegic or Restricted to Wheelchair
	BLIND	Blind
	CANE	Using Cane or Crutches
	NO OBS	Not Observed
Pedestrian Actions/Circumstances Contributing to the Crash		
NMTACT[1,2][A,B]-PED-MV	This data field was retrieved from "NMTACT[1,2][A,B]" in the DT4000 crash from using the SAS code translation Excel file provided through the WisTransportal website. Attribute "6" was created to combine other actions and was named "OTHR".	
	0	BLANK
	1	WALKING NOT FACING TRAFFIC
	2	DISREGARDED SIGNAL
	3	DARTING INTO ROAD
	4	DARK CLOTHING
	5	WALKING FACING TRAFFIC
NMTACT[1,2][A,B]-PED-DT	The actions/circumstances of the non-motorist that may have contributed to the crash, based on the judgement of the law enforcement officer investigating the crash.	
	NF TRFC	Walking Not Facing Traffic
	DISREG	Disregarded Signal
	SUDDEN	Sudden, Movement into Traffic
	DK CLTH	Dark Clothing
	FC TRFC	Walking Facing Traffic
	NO IMPR	No Improper Action
	IM XING	Improper Crossing of Roadway (Jaywalking)
	F YIELD	Failure to Yield Right-Of-Way
	F OBEY	Failure to Obey Traffic Signs, Signals, or Officer
	IM RDWY	In Roadway Improperly (Standing, Lying, Working, Playing)
	DISABLD	Disabled Vehicle Related (Working On, Pushing, Leaving/Approaching)
	STOPPED	Entering/Exiting Parked/Standing Vehicle
	INATTV	Inattentive (Talking, Eating, Etc.)
	NOT VIS	Not Visible (Dark Clothing, No Lighting, Etc.)
	IM TURN	Improper Turn/Merge
	IM PASS	Improper Passing
	W WAY	Wrong-Way Riding or Walking
	F LGTS	Failing to Have Lights on When Required (Bicycling)
	NO EQIP	Operation Without Required Equipment (Bicycle Reflectors)
	IM CHNG	Improper or Erratic Lane Changing
	F LANE	Failure to Keep in Proper Lane or Running Off Road
	IM ENTR	Making Improper Entry to or Exit from Trafficway
	RECKLSS	Operating in Other Erratic, Reckless or Careless Manner
	PASSNG	Passing with Insufficient Distance or Inadequate Visibility or Failing to Yield to Overtaking Vehicle
Pedestrian Actions Immediately Prior to the Crash		

NMTPRIOR[1,2]-PED-DT	The action of a non-motorist immediately prior to a crash. No such data field in MV4000 crash form. This is a new variable included in the DT4000 crash form.	
	XING	Crossing Roadway
	WAITING	Waiting to Cross Roadway
	W TRFC	Walking / Cycling Along Roadway with Traffic (In or Adjacent to Travel Lane)
	A TRFC	Walking / Cycling Along Roadway Against Traffic (In or Adjacent to Travel Lane)
	SIDE WK	Walking / Cycling on Sidewalk
	RDWY OT	In Roadway - Other
	ADJACNT	Adjacent to Roadway (E.G., Shoulder, Median)
	NONE	None
	JOGGING	Jogging / Running
	STOPPED	Entering/Exiting Parked or Stopped Motor Vehicle
	DISABLD	Disabled Vehicle Related
Pedestrian Location with Respect to the Roadway		
NMTLOC [1,2]-PED-MV	This data field was retrieved from "NMTLOC [1,2]-PED-DT" in the DT4000 crash from using the SAS code translation Excel file provided through the WisTransportal website.	
	0	BLANK
	1	IN CROSSWALK
	2	IN ROADWAY
	3	NOT IN ROADWAY
	4	ON SIDEWALK
NMTLOC[1,2]-PED-DT	The location of the non-motorist with respect to the roadway at the time of the crash.	
	ATI MX	At Intersection-In Marked Crosswalk
	ATI UM	At Intersection-Unmarked / Unknown If Marked Crosswalk
	ATI NX	At Intersection-Not in Crosswalk
	ATI UL	At Intersection-Unknown Location
	NAI MX	Not at Intersection-In Marked Crosswalk
	NAI NX	Not at Intersection-On Roadway, Not in Marked Crosswalk
	NAI UN	Not at Intersection-On Roadway, Crosswalk Availability Unknown
	PK LN	Parking Lane/Zone
	BIKE LN	Bicycle Lane
	SHLDR	Shoulder / Roadside
	SDWLK	Sidewalk
	MEDIAN	Median / Crossing Island
	DRWAY	Driveway Access
	SHARED	Shared-Use Path
	NON TRF	Non-Trafficway Area
	NOT RPT	Not Reported
Safety Equipment Used by the Pedestrian		
NMTSFQ[1,2][A,B]-PED-DT	The safety equipment in use by the operator non-motorist at the time of the crash (excluding motorcyclists). Note that the SAFETY [1, 2] data field in the MV4000 crash dataset indicates that the field shows the type of safety equipment that was used by a driver, bicyclist or pedestrian involved in the crash, while the data. did not show that this field was filled for pedestrians nor bicyclists. Hence, the NMTSFQ data field is a new filed included in the DT4000 crash form.	
	NONE	None
	HLMT	Helmet
	PADS	Protective Pads Used (Elbow, Knees, Shin, etc.)
	REFL	Reflective Clothing (Jacket, Backpack, etc.)
	LTNG	Lighting
Bicyclist Level		
The bicyclist condition relevant to the crash		
DNMFTR [1,2] [A,B]-BIKE-DT	Any relevant condition of the individual (motorist or non-motorist) that is directly related to the crash.	
	NORM	Appeared Normal
	PHY IMP	Physically Impaired
	EMO	Emotional (Depressed, Angry, Disturbed, Etc.)

	SICK	Ill (Sick), Fainted
	SLEEP	Asleep or Fatigued
	UI MDA	Under the Influence of Medication/Drugs/Alcohol
	CONF	Confused or Disoriented (Non-Lucid)
	WCHAIR	Paraplegic or Restricted to Wheelchair
	BLIND	Blind
	CANE	Using Cane or Crutches
	NO OBS	Not Observed
Bicyclist Actions/Circumstances Contributing to the Crash		
NMTACT [1,2] [A,B]-BIKE-MV	This data field was retrieved from “NMTACT[1,2][A,B]” in the DT4000 crash from using the SAS code translation Excel file provided through the WisTransportal website. Attribute “6” was created to combine other actions and was named “OTHR”.	
	0	BLANK
	1	WALKING NOT FACING TRAFFIC
	2	DISREGARDED SIGNAL
	3	DARTING INTO ROAD
	4	DARK CLOTHING
	5	WALKING FACING TRAFFIC
NMTACT [1,2] [A,B]-BIKE-DT	The actions/circumstances of the non-motorist that may have contributed to the crash, based on the judgement of the law enforcement officer investigating the crash.	
	NF TRFC	Walking Not Facing Traffic
	DISREG	Disregarded Signal
	SUDDEN	Sudden, Movement into Traffic
	DK CLTH	Dark Clothing
	FC TRFC	Walking Facing Traffic
	NO IMPR	No Improper Action
	IM XING	Improper Crossing of Roadway (Jaywalking)
	F YIELD	Failure to Yield Right-Of-Way
	F OBEY	Failure to Obey Traffic Signs, Signals, or Officer
	IM RDWY	In Roadway Improperly (Standing, Lying, Working, Playing)
	DISABLD	Disabled Vehicle Related (Working On, Pushing, Leaving/Approaching)
	STOPPED	Entering/Exiting Parked/Standing Vehicle
	INATTV	Inattentive (Talking, Eating, Etc.)
	NOT VIS	Not Visible (Dark Clothing, No Lighting, Etc.)
	IM TURN	Improper Turn/Merge
	IM PASS	Improper Passing
	W WAY	Wrong-Way Riding or Walking
	F LGTS	Failing to Have Lights on When Required (Bicycling)
	NO EQIP	Operation Without Required Equipment (Bicycle Reflectors)
	IM CHNG	Improper or Erratic Lane Changing
	F LANE	Failure to Keep in Proper Lane or Running Off Road
	IM ENTR	Making Improper Entry to or Exit from Trafficway
	RECKLSS	Operating in Other Erratic, Reckless or Careless Manner
	PASSNG	Passing with Insufficient Distance or Inadequate Visibility or Failing to Yield to Overtaking Vehicle
Bicyclist Actions Immediately Prior to the Crash		
NMTPRIOR[1,2]-BIKE-DT	The action of a non-motorist immediately prior to a crash. No such data field in MV4000 crash form. This is a new variable included in the DT4000 crash form.	
	XING	Crossing Roadway
	WAITING	Waiting to Cross Roadway
	W TRFC	Walking/Cycling Along Roadway with Traffic (In or Adjacent to Travel Lane)
	A TRFC	Walking/Cycling Along Roadway Against Traffic (In or Adjacent to Travel Lane)
	SIDE WK	Walking/Cycling on Sidewalk
	RDWY OT	In Roadway - Other
	ADJACNT	Adjacent to Roadway (E.G., Shoulder, Median)
	WORKING	Working in Trafficway (Incident Response)
	NONE	None
	JOGGING	Jogging/Running

	STOPPED	Entering/Exiting Parked or Stopped Motor Vehicle
	DISABLD	Disabled Vehicle Related
Bicyclist Location with Respect to the Roadway		
NMTLOC [1,2]-BIKE-MV	This data field was retrieved from "NMTLOC[1,2]- BIKE -DT" in the DT4000 crash from using the SAS code translation Excel file provided through the WisTransportal website.	
	0	BLANK
	1	IN CROSSWALK
	2	IN ROADWAY
	3	NOT IN ROADWAY
	4	ON SIDEWALK
NMTLOC[1,2]- BIKE -DT	The location of the non-motorist with respect to the roadway at the time of the crash.	
	ATI MX	At Intersection-In Marked Crosswalk
	ATI UM	At Intersection-Unmarked / Unknown If Marked Crosswalk
	ATI NX	At Intersection-Not in Crosswalk
	ATI UL	At Intersection-Unknown Location
	NAI MX	Not at Intersection-In Marked Crosswalk
	NAI NX	Not at Intersection-On Roadway, Not in Marked Crosswalk
	NAI UN	Not at Intersection-On Roadway, Crosswalk Availability Unknown
	PK LN	Parking Lane/Zone
	BIKE LN	Bicycle Lane
	SHLDR	Shoulder / Roadside
	SDWLK	Sidewalk
	MEDIAN	Median / Crossing Island
	DRWAY	Driveway Access
	SHARED	Shared-Use Path
	NON TRF	Non-Trafficway Area
	NOT RPT	Not Reported
	Safety Equipment Used by the Bicyclist	
NMTSFQ[1,2][A,B]-BIKE-DT	The safety equipment in use by the operator non-motorist at the time of the crash (excluding motorcyclists). <u>Note that the SAFETY [1, 2] data field in the MV4000 crash dataset indicates that the field shows the type of safety equipment that was used by a driver, bicyclist or pedestrian involved in the crash, while the data. did not show that this field was filled for pedestrians nor bicyclists. Hence, the NMTSFQ data field is a new filed included in the DT4000 crash form.</u>	
	NONE	None
	HLMT	Helmet
	PADS	Protective Pads Used (Elbow, Knees, Shin, etc.)
	REFL	Reflective Clothing (Jacket, Backpack, etc.)
	LTNG	Lighting
Crash Level		
Events Resulting in the Most Severe Injury		
ACCDTYPE-MV	Description of type of crash based on the first harmful event. ***MVIT** - Motor Vehicle in Transit involves moving vehicles. This field appears blank.	
	ATTEN	Impact attenuator
	BIKE	Bicycle
	BRP AR	Bridge parapet
	BRPIER	Bridge/pier/abutment
	BRRAIL	Bridge rail
	CULVRT	Culvert
	CURB	Curb
	DEER	Deer
	DITCH	Ditch
	EMBKMT	Embankment
	FENCE	Fence
	FIRE	Fire / Explosion
	GR END	Guardrail end
	GR FAC	Guardrail face
	IMMER	Immersion
	JKNIF	Jackknife

	LTPOLE	Lum light support
	MAILBOX	Mailbox
	MED B	Median barrier
	MVIT*	Vehicle in transit
	OBNFX	Object not fixed
	SIGN	Overhead signpost
	OTH FX	Other object fixed
	OTH NC	Other non-collision
	OT ANL	Other animal
	OT RDY	Veh trans other rdwy
	OT PST	Other post
	OVRTRN	Overturned vehicle
	PED	Pedestrian
	PKVEH	Parked vehicle
	TFSIGN	Traffic sign
	TF SIG	Traffic signal
	TRAIN	Train
	TREE	Tree
	UT PL	Utility Pole
MOSTHARM[1,2]-DT	Event that resulted in the most severe injury or, if no injury, the greatest property damage involving this motor vehicle.	
	MVIT	Motor Vehicle in Transport
	PKVEH	Parked Motor Vehicle
	BIKE	Pedal cycle
	PED	Pedestrian
	TRAIN	Railway Vehicle (Train, Engine)
Vehicle Level		
Vehicle Type Involved in the Crash		
VEHTYPE [1,2]-MV	The type of vehicle that was involved in a crash.	
	ATV	Snowmobile
	ATV, BIKE	Bicycle
	BLNK	Blank
	BUS	Passenger bus
	CAR	Passenger car
	CYCLE	Motorcycle
	EM AMB	Ambulance on emergency
	EM FIRE	Fire truck / fire fighter on emergency
	EM POL	Police on emergency
	FARM	Farm tractor / self-propelled
	HOME	Motor home
	HRSDRWN*	Horse drawn implement (carriage, wagon, buggy)
	MISC	Miscellaneous
	MOPED	Moped
	OTHR	Other working machine
	PED	Pedestrian
	PLOW	Snowplow
	SBS	School bus / pupil transport
	TRAIN	Railway train
	TRK DB	Truck tractor (double bottom)
	TRK NA	Truck tractor (not attached)
	TRK SA	Truck tractor (semi attached)
	TRK ST	Straight truck (insert truck)
	TRK UT	Utility truck
VEHTYPE [1,2]-DT	Specific category for the type of vehicle which was involved in a crash.	
	CAR	Passenger Car
	SUV	(Sport) Utility Vehicle
	P VAN	Passenger Van
	C VAN	Cargo Van (10,000 Lbs. or Less)
	UT TRK	Utility Truck/Pickup Truck

	HOME	Motor Home
	S BUS	School Bus
	PT BUS	Pupil Transportation School Bus
	T BUS	Passenger Bus/Transit Bus
	COACH	Motor Coach
	OT BUS	Other Bus
	CYCLE	Motorcycle
	MOPED	Moped
	LSPD	Low Speed Vehicle
	GOLF	Golf Cart
	ATV	ATV/UTV (Utility Terrain Vehicle)
	SNOW	Snowmobile
	EM POL	Police on Emergency
	ST TRK	Straight Truck
	TRK NA	Truck Tractor (Trailer Not Attached)
	TRK TA	Truck Tractor (Trailer Attached)
	TRK DB	Truck Tractor (More Than One Trailer)
	AMB EM	Ambulance on Emergency
	FIRE EM	Fire Truck on Emergency
	FARM	Farm Tractor/Self Propelled
	AGCMV	AgCMV (Ag Commercial Motor Vehicle)
	OTHR	Other Working Machine
	TRAIN	Railway Train
	PLOW	Snowplow
	MISC	Miscellaneous
	BIKE	Bicycle
	FIREF EM	Fire Fighter on Emergency
	TRAILER	Trailer
	HRSDRWN	Horse and Buggy
	MINI	Minibike/Dirt Bike
	ACYCLE	Autocycle
	ATV	ATV
	UTV	UTV (Utility Terrain Vehicle)
Extent of Vehicle Damage		
VEHDMG [1,2]-MV	The extent of vehicle damage.	
	BLNK	Blank
	V MNR	Very Minor
	MNR	Minor
	MOD	Moderate
	SVR	Severe
	V SVR	Very Severe
VEHDMG [1,2]-DT	Identifies the extent to which the damage affects the vehicles operability rather than the cost to repair.	
	NO	No Damage
	MINOR	Minor Damage
	FUNC	Functional Damage
	DISABL	Disabling Damage
	NAS	Not at Scene

Curriculum Vitae

Objective

A Hard-working, highly motivated Transportation Engineer focusing on the areas of highway safety, statistical data analysis, and data mining, and GIS applications in Transportation. Interested in fundamental research in Transportation data analysis with a focus on Transportation safety. Experienced in data visualization, predictive modeling, causal analysis, and quantitative analysis.

Experience

January 2016-present |Graduate Research and Teaching Assistant
University of Wisconsin-Milwaukee | Milwaukee-Wisconsin, USA

- Prepared deliverables (reports, presentations, research papers, and grant proposals)
- Worked on WisDOT funded projects:
- Identifying Highly Correlated Variables Relating to the Potential Causes of Reportable Wisconsin Traffic Crashes.
- Comprehensive Evaluation of DT4000 Data Quality for Pedestrian and Bicycle Crashes
- Worked on Driver Yield data collection in Milwaukee project: Collected field driver yielding data.
- Worked on evaluating data quality for pedestrian and bicyclist crashes project: Collected, analyzed, and summarized crash data from the MV4000 and DT4000 crash report forms.
- Taught the following courses: CIV ENG 201 Statics, CIV ENG 202 Dynamics, CIV ENG 280 Computer-Based Engineering Analysis-Statistics.

June 2018-August 2019 | Civil Engineering Intern
City of Milwaukee | Milwaukee, Wisconsin, USA

- Assist engineers with plan reviews, field inspections, and stormwater management charge calculations.
- Designed permeable parking lots.
- Prepared sketches for sewer rehabilitation and maintenance projects.
- Prepare engineering designs and plans using the MicroStation drawing tool.
- Reviewed and prepared responses to Stormwater Management Plans and Building Permits,
- ensured submittals are compliant with environmental, safety, and other governmental regulations.
- Inspectedprojectsitestomonitorprogressandensureconformancetodesignspecifications.
- Estimated quantities and cost of materials, equipment, and labor and prepared cost estimates.

March 2012-October 2014 | Design Engineer and Quantity Surveyor
Ministry Of Public Works & Housing | Amman, Jordan

- Road design using AutoCAD Civil3D for Irbid Ring Road project, Irbid, Jordan
- Provided training for a group of 20 employees in the Geographic Information Department in the ministry.

- Highway engineer and quantity surveyor for the study: “A monitoring study of an underground quarry loaded by fill materials for the construction of airport street”.

Education

January 2016-present | Ph.D. in Civil Engineering (GPA 3.70)

University of Wisconsin- Milwaukee, Milwaukee-Wisconsin, USA

Major: Transportation Engineering.
Minor: Urban Planning University of Wisconsin-Milwaukee, Milwaukee, Wisconsin
Dissertation Topic: Pedestrian and Bicycle Safety Analysis.

- Served as the president for the Wisconsin chapter of the Arab American Association of Engineers and Architects (AAAEA) at UW-Milwaukee, 2017-2019.
- Course work: CIV ENG 592 Traffic Control | CIV ENG 794 Traffic Planning and Operations | CIV ENG 700 Graduation Seminar-Technical Writing | IND ENG 455 Operations Research I | CIV ENG 596 Transportation Facilities Design | CIV ENG 792 Methods of Transportation Analysis |URBPLAN 793 Applies Projects in Urban GIS | URBPLAN 772 Pedestrian/Bicyclist Transport |IND ENG 716 Engineering Statistical Analysis.

January 2013-April 2015 | M.Sc. in Civil Engineering (GPA 3.70)

University of Jordan | Amman, Jordan

Major: Transportation Engineering
Thesis Topic: Traveler’s revealed and stated preference analysis for the proposed bus rapid transit service in Jordan.

September 2008-August 2012 | B.Sc. in Civil Engineering

Al-Balqa’ Applied University | Amman, Jordan

Major: Highway and Bridges Engineering
Senior Design/Graduation Project Topic: Green and Sustainable Restaurant Design

Skills

- Traffic and Planning: VISSIM, Synchro, ArcMap GIS
- Structural Analysis: PROKON, ETABS, STAAD.Pro
- Programming: C++
- Statistical Software Tools: R, RStudio, SAS, Biogeme, SPSS, STATA, Minitab, MATLAB
- Stormwater Quality and Quantity Analysis Tools: HydroCAD, WinSLAMM
- Drawing Software: MicroStation, AutoCAD Civil 3D
- Technical writing and presentation skills

Publications

Journal papers:

1. Farah Al-Mahameed, Xiao Qin, Robert Schneider, Mohammad Razaur Shaon, "Analyzing Pedestrian And Bicyclist Crashes At The Corridor Level: A Structural Equation Modeling Approach" Transportation Research Record: Journal of the Transportation Research Board (IF=1.029), May 2019, Vol. 2673, No.7, pp. 308-318. https://doi.org/10.1177/0361198119845353
2. Farah Al-Mahameed, Xiao Qin, Robert Schneider, "A Comprehensive Review and Evaluation of the DT4000 Crash Form Data Quality for Pedestrian and Bicycle Crash Analysis", to be Submitted
3. Farah Al-Mahameed, Xiao Qin, Robert Schneider, "A Guideline for Determining Road User Fault Status - An Application to Pedestrian/Bicyclist-Vehicle Crashes in Wisconsin", to be Submitted

Conference Presentations:

1) *Analyzing Pedestrian and Bicyclist Crashes at the Corridor Level: A Structural Equation Modeling Approach (19-03034),* Transportation Research Board (TRB) annual meeting 2019, January 13-17, 2019 Washington DC, USA
2) *Identifying Vulnerable Road Users Safety Issues Along Street Corridors*, Lifesavers National Conference on Highway Safety Priorities 2019, March 31-April 2, 2019 Louisville, Kentucky, USA.
3) *Integrating Exploratory Factor Analysis And Confirmatory Factor Analysis To Find Robust Predictors Of Pedestrian/Bicyclist Crashes,* Association of Transportation Safety Information Professionals' (ATSIP) 2018 Traffic Records Forum, September 6, 2018, Milwaukee, WI, USA

Awards

- Nominated among the finalists in the Three-Minute Thesis Competition during the TRB Annual Meeting held in Washington, D.C., January 2019.
- Received A ($1000) scholarship from the Lifesavers Traffic Safety Scholars (TSS) Program to attend and present a research paper during the Lifesavers conference held in Louisville, KY, March 31-April 2, 2019 and presented my university as a traffic safety scholar.
- Received the Chancellor's Graduate Student Award, University of Wisconsin-Milwaukee, Spring 2018.
-

Certificates

- ArcGIS 1 & 2 (16 Hours) | September 1, 2013 – September 5, 2013
  InfoGraph | Amman, Jordan
- Primavera V.3 (21 Hours) February | 12, 2012 – February 23, 2012
  Jordan Engineers Association | Amman, Jordan

- Application of Software Analysis & Design of High-rise Buildings-ETABS (21 Hours) | September 4, 2011 – September 22, 2011
  Jordan Engineers Association | Amman, Jordan
- Design of Concrete Structures (21 Hours) | June 15, 2011 – July 6, 2011
  Jordan Engineers Association | Amman, Jordan

Leadership involvements

- TRB Bicycle and Pedestrian Data Subcommittee (ABJ35)
- TRB Transportation Data and Information Systems Committee (ABJ20)
- Arab American Association of Engineers and Architects (AAAEA) Wisconsin chapter president, 2017-2019
- Jordan Green Building Council (JGBC) active member, 2013-2015
- Jordan Engineers Association (JEA) member, 2012-present

Languages

English: IELTS overall score of 7.0

Arabic: Native or bilingual proficiency