

**ANALYSIS OF CONNECTED AND AUTOMATED VEHICLES TO ADVANCE
MULTIMODAL SAFETY**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Antora Mohsena Haque
December 2022

Copyright © 2022 by Antora Mohsena Haque
All rights reserved.

This dissertation is dedicated to my parents, my husband, my sister, my friends, and family who inspired me every day to achieve my dreams and complete my dissertation.

ACKNOWLEDGMENT

First, I want to thank almighty Allah for his blessings upon me in every aspect of life. I believe I have come this far and now that I can achieve the doctoral degree are possible for his continuous blessings and miracles. I truly believe he has made all my hard times easier and difficult targets achievable.

I want to thank everyone I have come across in life who encouraged and supported my endeavors to pursue higher studies abroad. I want to convey my most profound gratitude to Dr. Asad J. Khattak. Without his support, my dream of earning a doctoral degree in Civil Engineering would be impossible. I express my appreciation to him for his knowledge, mentorship, and guidance in completing my research work. Dr. Khattak has always given me the freedom and flexibility to do research at a comfortable pace using my strongest skills. Words would fall short of expressing how grateful I am to him. It has been my honor to work under his supervision and learn from him.

A debt of gratitude is extended to my dissertation committee members, Dr. Christopher Cherry, Dr. Anahita Khojandi, and Dr. Russel Zaretski, for their tremendous guidance and support. Their insights have helped me broaden my thought process and improvise my dissertation. A special thanks to Dr. Christopher Cherry, who supported me during the most challenging time in my PhD life. I want to acknowledge my co-authors, Dr. Asad J. Khattak, Dr. Candace Brakewood, Dr. Christopher Cherry, Dr. Anahita Khojandi, Dr. Ramin Arvin, Dr. Iman Mahdinia, Shahrbanoo Rezaei, Jing Guo, A Latif Patwary, and Meredith King for their valuable knowledge and insight into my research. I wish to thank the Department of Civil and Environmental Engineering faculty and staff and the Collaborative Sciences Center for Road Safety (CSCRS) for their assistance and support.

Thank you to my lab mates, especially Mojdeh Azad, Dr. Ali Boggs, Dr. Steve Lee, Dr. Iman Mahdinia, Dr. Numan Ahmed, Dr. Amin Mohammadnazar, A Latif Patwary, Dr. Nitesh Raj Shah, Kinzee Clark, Dr. Abubakr Salah, Dr. Yi Wen, Yangsong Gu, Jing Guo, and Sameer Aryal whose friendship and support made my PhD life memorable, especially inside our office - John D. Tickle Building Room 311.

I want to thank my friends from the Bangladesh office, Emmy, Touhid, and Mehedi, who encouraged me to pursue my career goals. I want to thank my undergraduate life friends, Dr. Taslima Akter Panna, Dr. Md. Tanvir Hossain Shubho, and Rifat Hossain, who gave strategic advice whenever I stumbled in the dark. I am grateful to my dear friend Ashfia, Putul, and Sadya for their sympathetic ear to me.

I want to convey a special gratitude to Dr. William Hahn from the Student Counseling Center, who helped me to hold my sanity and stay focused on my research. I want to thank my uncle Shantu, who encouraged me to apply to UTK for the doctoral degree and live in Knoxville, Tennessee. I appreciate my friends in Knoxville for their time and care.

My acknowledgment would be incomplete without thanking my biggest supporters, my husband - Dr. Md. Badruddoja Majumder Sabuj, and my father, mother, and sister. Their love, prayers, and sacrifices for me have been enormous.

ABSTRACT

Transportation Engineers and Planners are working towards improving transportation safety by researching Connected and Automated vehicles (CAV), which are anticipated to address human error factors in crash studies. This dissertation investigates various safety features of CAVs such as Enhanced Pedestrian in Crossing Warning (EPCW), Enhanced Vehicle Turning Right in Front of Bus Warning (EVTRW), Advanced Driver-Assistance Systems (ADAS), and Autopilot feature to understand their performance at this early stage of deployment which can be insightful for CAV researchers. The two research questions addressed in this dissertation are 1) how can the data from CAVs be utilized to increase transit, pedestrian, motor vehicle, and infrastructure safety; and 2) how can the data from CAVs be utilized to understand their readiness and effectiveness. Due to the limitation of publicly available CAV data, knowledge about their safety performances is limited. This dissertation attempts to address a segment of this gap. First, the safety features of connected buses are studied to determine the association between pedestrian and bus-based alerts with crashes that can aid the safety of vulnerable road users (VRUs), transit buses, and intersections. A literature review of Automated shuttles follows this to identify the potential locations of AV shuttle deployment at an early stage. Later, the safety features of AVs are studied by comparing lower-level automated vehicles to understand which AVs perform better in terms of safety at different road types, weather and lighting conditions, and driver behaviors. Various data science, spatial, and frequentist techniques including panel poisson regression model to address heterogeneity, ordered logistic regression model, binary logistic regression model, text analysis, hazardous location identification method in GIS, and novel safety surrogate measures (SSM), are utilized to conduct the analysis. Finally, research implications and future works are discussed to understand these emerging vehicles comprehensively. Long term scientific value of this dissertation is that it harnesses big data from cutting-edge intelligent transportation systems and infrastructures and integrates them with conventional crash data to perform analysis that can aid in advancing multimodal safety. The dissertation will be helpful for city planners, policymakers, civil engineers, designers, and researchers working towards achieving Vision Zero goals.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION	1
CHAPTER 2. ENHANCING THE SAFETY OF TRANSIT USERS AND PEDESTRIANS IN URBAN AREAS: THE ROLE OF CONNECTED VEHICLES AND SMART INFRASTRUCTURE TO PROACTIVELY IDENTIFY HIGH-RISK HAZARDOUS LOCATIONS	7
Abstract.....	8
Introduction	8
Literature Review	9
Prior Research on Basic Safety Messages	9
Prior Research on Transit Bus-involved crashes	11
Prior Research on Hazardous locations	11
Methodology	12
Data	12
Data Processing.....	16
Data Analysis	21
Results.....	24
Hazardous locations for transit bus-involved crashes using ArcGIS.....	24
Grids with ≥ 1 alert	26
Discussion	30
Comparison of hazardous and non-hazardous areas based on Alert/Grid and Crash/Grid...	30
Bus-involved crash reduction by EPCW	36
Limitations	36
Conclusions	39
Acknowledgments.....	40
CHAPTER 3. EVALUATING PEDESTRIAN SAFETY IN BUS-PEDESTRIAN INTERACTIONS WITH THE APPLICATION OF ENHANCED PEDESTRIAN COLLISION WARNING SYSTEMS: ANALYSIS OF LARGE-SCALE BASIC SAFETY MESSAGE DATA	41
Abstract.....	42
Introduction	42

Literature Review.....	43
Methodology	45
Data	45
Data preparation.....	46
Contingency tables.....	48
Alert Zone identification.....	48
Mean Time Difference to Intersection.....	48
Study Area	49
Results.....	49
Frequency Distribution of Alerts	49
Comparison of three intersections	50
Identification of Alert Zones for buses	56
Discussion	57
Performance of EPCW.....	57
Land Use	57
Temporal Distribution of EPCW	58
Limitations	58
Conclusion	66
Acknowledgments.....	68
CHAPTER 4. A SYNTHESIS AND COMPARISON OF AMERICAN AUTOMATED SHUTTLE PILOT PROJECTS	69
Abstract.....	70
Introduction	70
Prior Research.....	71
European research on automated shuttles	71
American research on automated shuttles.....	71
Gaps in research on automated shuttles	72
Objectives.....	72
Methodology	72

Results.....	74
Identification of pilot projects.....	74
Selection of cases.....	74
Summary of the selected cases	74
Synthesis of the state of the practice.....	76
Areas for improvement and future research.....	79
Conclusions	80
CHAPTER 5. INVESTIGATING CRASHES INVOLVING ADVANCED DRIVER ASSISTANCE SYSTEMS (ADAS) EQUIPPED AUTOMATED VEHICLES - THE ROLE OF AUTOMATION	86
Abstract.....	87
Introduction	87
Literature Review.....	88
Methods	89
Data.....	89
Ordered Logistic Regression Model	92
Marginal Effect	93
Results.....	93
Frequency Distribution of variables.....	93
Ordered Logistic Regression Model Results	94
Discussion	95
Factors Associated with AV Crash Severity.....	95
AV Crashes at Different Road Types	101
Limitations	106
Conclusions	107
Acknowledgments.....	108
CHAPTER 6. UNDERSTANDING THE SURVIVABILITY OF DRIVERS FROM MEDIA-REPORTED AUTOMATED VEHICLE-INVOLVED FATAL CRASHES: A CASE STUDY OF TESLA-INVOLVED FATAL CRASHES	109
Abstract.....	110

Introduction	110
Literature Review.....	111
Data.....	111
Methods	114
Human Crash Contributing Factors	114
Correlation Analysis: Somers' D Test	114
Binary Logistic Regression.....	115
Text Mining Analysis	115
Results and Discussion	115
Descriptive Statistics.....	115
Cross Tabulation: Tesla Driver Survivability versus Key Factors	118
Correlation analysis through Somers' D Tests	124
Binary Logistic Regression Model on Tesla Drivers' Fatality	125
Text Analysis on Crash Reports	132
Comparison of Tesla of Luxury Electric Vehicles	132
Limitations	132
Conclusions	134
Acknowledgments.....	135
CHAPTER 7. CONCLUSION.....	136
Implications	138
Future Research	138
LIST OF REFERENCES	140
VITA	154

LIST OF TABLES

Table 2.1 Number of observations in grids with ≥ 1 alert-based datasets	31
Table 2.2. Descriptive Statistics of grids with ≥ 1 alert.	32
Table 2.3. Descriptive Frequencies of Key Variables.	33
Table 2.4. Regression Models for alerts on grids with ≥ 1 alert.	34
Table 3.1. Frequency of Alerts by “EventCodeType” and “AppMessageType”	53
Table 3.2. Contingency Table of Alerts under all “AppMessageType,” only EPCW Application, and Pedestrian Based “EventCodeType”	55
Table 4.1. Timeline of automated shuttle deployments	77
Table 4.2. Summary of six case studies	82
Table 4.3. Typology based on locations	83
Table 4.4. Typical service characteristics	84
Table 5.1. Frequency Distribution of Important Variables	96
Table 5.2. Frequency Distribution of Other Important Variables	98
Table 5.3. Ordered Logistic Regression Model	99
Table 6.1. Summary of the Literature Review	116
Table 6.2. Driving Errors and Violations by Tesla Driver in Single-Vehicle Crashes	119
Table 6.3 Driving Errors and Violations by Tesla Driver in Multi-Vehicle Crashes	120
Table 6.4. Distribution of Injury Severity in Single-Vehicle Crashes	121
Table 6.5 Distribution of Injury Severity in Multi-Vehicle Crashes	121
Table 6.6. Cross Tabulation of Fatality of Tesla Driver in Single-Vehicle Crashes Versus Human Factors	126
Table 6.7 Cross Tabulation of Fatality of Tesla Driver in Multi-Vehicle Crashes Versus Human Factors	127
Table 6.8 Comparison of Fire-involved Fatal Crashes between Drivers of All Vehicles and Tesla in the US	128
Table 6.9. Results of Correlational Analysis: Tesla Driver Survivability versus Key Factors in Single-Vehicle Crashes	129
Table 6.10 Results of Correlational Analysis: Tesla Driver Survivability versus Key Factors in Multi-Vehicle Crashes	130
Table 6.11 Results of the Binary Logistic Regression Model	131

LIST OF FIGURES

Figure 1.1. Conceptual Framework of the Study	5
Figure 1.2. Outline of the Dissertation.....	6
Figure 2.1. Technology packages relevant to transit buses.	13
Figure 2.2. Methodological Framework.	14
Figure 2.3 Onboard Subsystem.....	17
Figure 2.4 Infrastructure Subsystem-Roadside Units (RSU).....	17
Figure 2.5. EPCW WarnAlert to Bus Driver	18
Figure 2.6 EVTRW Inform Alert to Bus Driver.....	18
Figure 2.7. Percentage of usable alerts after merging Alert and BSM datasets.....	20
Figure 2.8. Map of hazardous locations for bus-involved crashes in Cuyahoga County	25
Figure 2.9. Hazardous locations vs. grids with ≥ 1 alert (Both EPCW & EVTRW).....	27
Figure 2.10 Crash/Grid in grids that include ≥ 1 alert using Both EPCW & EVTRW Dataset ...	37
Figure 2.11 Alert/Grid in grids that include ≥ 1 alert using Both EPCW & EVTRW Dataset	37
Figure 2.12. Bus-involved crashes from 1999 to 2019 in the US	38
Figure 2.13. Reduction of bus-involved crashes by EPCW.....	38
Figure 3.1. On-board Subsystem (top), Infrastructure Subsystem (middle), and EPCW “WarnAlert” (bottom)	47
Figure 3.2. Percentage of alerts after merging Alert and BSM datasets.....	51
Figure 3.3. Three selected intersections in Cleveland, Ohio	52
Figure 3.4. Percentage of “Info, “InformAlert” and “WarnAlert” at (a) all three intersections, (b) Signalized, (c) Unsignalized, and (d) Midblock crossing	54
Figure 3.5. Categories of bus stops based on their curbside location	59
Figure 3.6. (a) All, (b) “FarsidePed,” (c) “NearsidePed,” (d) “LeftsidePed,” and (e) “RightsidePed” alerts at Signalized Intersection	60
Figure 3.7. (a) All, (b) “FarsidePed,” (c) “NearsidePed,” (d) “LeftsidePed,” and (e) “RightsidePed” alerts at Unsignalized Intersection	61
Figure 3.8. Pedestrian based alerts at Midblock Crossing.....	62
Figure 3.9. Trendline of MTDIs at intersections	63
Figure 3.10. Land Use around the three intersections of the study area	64
Figure 3.11. Timeseries heatmap of Alerts at intersections.....	65
Figure 4.1. Methodology.....	73

Figure 4.2. Three key dimensions of automated shuttle deployments.....	81
Figure 4.3. Typical stakeholders in automated shuttle projects.....	85
Figure 5.1. ADAS-equipped AV crashes in the US (2021-2022).....	90
Figure 5.2 Margins plot of the probability of crashes-based on (A) Vehicle Make, (B) Vehicle Weight, (C) Speeding of Subject Vehicle, (D) Pre-crash maneuver of Crash Partner Vehicle, and (E) Collision Type.....	102
Figure 5.3 (A) Vehicle Make, (B) AV Crash Severity, and (C) Crash With at Different Roadway Types.....	103
Figure 5.4 (A) Collision Type and (B) Pre-crash maneuver of Crash Partner Vehicle at Different Roadway Types.....	104
Figure 5.5 Crash Rates by AV Type.....	105
Figure 6.1. Conceptual Framework	112
Figure 6.2. State wise Distribution of Single-Vehicle Fatal Crashes involving Tesla Vehicle in the US	122
Figure 6.3. State wise Distribution of Multi-Vehicle Fatal Crashes involving Tesla Vehicle in the US	123
Figure 6.4 Word Cloud for Media Reported Tesla-Involved (a) Single-Vehicle and (b) Multi-Vehicle Fatal Crashes	133
Figure 6.5 Year to Date Sales of Luxury Electric Vehicles.....	133

CHAPTER 1. INTRODUCTION

People need to travel for various needs almost daily, and these inevitable trips can become fatal due to human error, vehicle error, faulty infrastructure, the environment, and bad weather. In these hazardous conditions, along with the drivers and passengers of the vehicles, vulnerable road users like pedestrians are also at high risk. While some issues are unavoidable, scientists are working to eradicate human error by developing connectivity and automation technology in the big picture. Connected Vehicles (CV) use vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) technology to exchange Basic Safety Messages (BSM) between each other and infrastructures to alert drivers to ensure the safety of vehicles and road users. On the other hand, automated vehicles use Lidar, Radar, and cameras to acquire images of their surroundings and use their artificial intelligence-based algorithms to enable various automation features to ensure safety. The emergence of these technologies has brought a ray of hope for carefree and safe traveling opportunities. But before fully deploying connected and automated vehicles (CAV), these technologies need to be evaluated and improved to yield the best possible benefit they can offer.

Although numerous CAV pilot projects are occurring, almost all focus on automobiles. Simultaneously, there are very few applications of connectivity and automation in alternative modes like public transit buses. Whereas 11.5% of workers in the largest cities of US metro areas and 5% of all workers in the US choose transit to commute to work every day (1), transit is not a vehicle that can be ignored in any safety analysis. Existing literature lacks investigation of V2V and V2I technologies' utility in transit buses. Automation has not been applied to large transit buses, but it is being tested in some automated shuttles, which are smaller versions of these large transit buses. Therefore, the application of connectivity and automation should also be studied for transit buses to ensure safety. For now, automation features in automobile needs more rigorous analysis, which can later help in its application in other modes like transit buses. Hence, this dissertation focuses on large-scale data on connectivity and automation technology in transit buses and automobiles. Through rigorous analytical methods, this dissertation demonstrates how V2V, V2I, and various automation safety features can be utilized to advance multimodal safety.

Chapter 2 utilizes data from the “Enhanced Transit Safety Retrofit Package (E-TRP)” project, which contains big data on Basic Safety Message (BSM) and Alerts generated in 24 Connected transit buses in Cuyahoga County, Cleveland, Ohio. This publicly available data is combined with five years of transit-involved crashes to explore two safety features of connectivity technology: Enhanced Pedestrian in Crossing Warning (EPCW) and Enhanced Vehicle Turning Right in Front of Bus Warning (EVTRW). Association between pedestrian and bus-based alerts and historic crash locations is analyzed only for areas where at least one alert is generated using Panel Poisson Regression and hotspot analysis. The Panel model helps to control unobserved variables and understand individual heterogeneity. This study contributes a framework for harnessing and utilizing unique data from Connected Vehicles to gain critical knowledge on pedestrians and transit safety that provides long-term scientific value. The results and implications are discussed to illustrate the potential of basic safety messages and alerts to predict crash-prone hazardous locations.

Chapter 3 uses only the Enhanced Pedestrian in Crossing Warning (EPCW) portion of the “Enhanced Transit Safety Retrofit Package (E-TRP)” project that have been generated at three different intersections of Cuyahoga County, Cleveland, Ohio. The study observes the pattern of different types of EPCW alerts at a Signalized, an unsignalized, and a midblock intersection. Safety surrogate measures and the radius of the alerts from each intersection are measured. Land use and timeseries heatmaps are produced to understand EPCW's pattern better. The results and

implications are discussed to realize the potential of EPCW alerts to improve pedestrian safety at intersections.

Chapter 4 synthesizes the state of the practice of automated shuttle pilot projects in the US. The case study method identifies shuttle deployment patterns by studying ongoing and completed projects open to the broader public since 2016, as the deployment of automated shuttle pilot projects began in 2016 in the United States. The pilot projects are identified manually from online news articles and websites. The case studies are synthesized along three dimensions: locations, service characteristics, and stakeholders. The results will assist cities that are considering future deployments of automated shuttles in their locality.

Chapter 5 uses Crash Data of ADAS-equipped AVs in the US from the National Highway Traffic Safety Administration (NHTSA) database. The Advanced Driver Assistance Systems (ADAS) studied in this paper include Adaptive Cruise Control (ACC) and Lane Keep Assist (LKA) systems. This study provides an in-depth understanding of driver behavior, roadway, and vehicle features' role in the crash severity of ADAS-equipped AVs. An Ordered Logistic Regression model, which uses a proportional odds assumption, is applied to observe their relationship. Next, AV crashes on highway/freeways and non-highway locations are compared to get valuable insights.

Chapter 6 uses Fatal crashes of Tesla cars that have happened in the US since its deployment. One of the safety features of Tesla's automation technology, Autopilot Engagement, and human behaviors are studied for these fatal crashes. The study provides a deeper understanding of the role of driver behavior, roadway and vehicular features, and automation features in automated cars, considering Tesla as an example. Somers' D statistics and a Binary Logistic Regression model are used to investigate the association between Tesla driver survivability and other independent variables.

In summary, the dissertation's main contribution is exploration of the association of various Advanced Driver Assistance Systems (ADAS) in terms of safety features at different levels in Connected and Automated Vehicles by harnessing big data generated under-real world conditions. The research integrates unstructured raw data on V2V and V2I communication technology from connected vehicles with geometric features of the roads, driver behavior, and crash data to develop unique and insightful datasets about transit, pedestrian, and automobile safety. The analyses show that real-world crashes, crash severity, safety warnings, and performance of safety features can help predict hazardous locations and conditions that can assist drivers from potential crashes. The research is conducted at three levels: The microscopic level (probability of crash severity), mesoscopic (intersection), and macro level (crashes at road network). The research shows how alerts like EPCW and EVTRW exchanged via V2V and V2I can be utilized for the safety improvement of transit buses and pedestrians. The research observes and identifies locations where the automated shuttles can be deployed immediately. The research also examines the ADAS and autopilot features in lower-level automated vehicles to understand the association between automation features and crashes.

It is expected that under this dissertation, five chapters in reputed transportation journals be published:

Haque, A.M., Arvin, R., Khattak, A.J. (2022) Enhancing the Safety of Transit Users and Pedestrians in Urban Areas: The Role of Connected Vehicles and Smart Infrastructure to Proactively Identify High-Risk Hazardous Locations.

- Under preparation for submitting to Accident Analysis and Prevention Journal

Haque, A.M., Khattak, A.J. (2022). Evaluating Pedestrian Safety in Bus-Pedestrian Interactions with the Application of Enhanced Pedestrian Collision Warning Systems: Analysis of Large-Scale Basic Safety Message Data.

- Under review in Transportation Research Record
- Peer-reviewed conference paper: Presented in the Transportation Research Board 101st Annual Meeting in 2022

Haque, A.M., Brakewood, C. A Synthesis and Comparison of American Automated Shuttle Pilot Projects

- Published journal article: Case Studies on Transport Policy
- Peer-reviewed conference paper: Presented in the Transportation Research Board 98th Annual Meeting in 2019

Haque, A.M., Mahdinia, I., Safaitaherkhani, M., Khattak, A.J. Investigating Crashes Involving Advanced Driver Assistance Systems (ADAS) equipped Automated Vehicles - The Role of Automation

- Under preparation for submitting to Transportation Research Part - C

Haque, A.M., Ahmad, N., King, M., Khattak, A.J. Understanding the Survivability from Media-Reported Automated Vehicle-Involved Fatal Crashes: A Case Study of Tesla-Involved Fatal Crashes

- Under preparation for submitting to the Journal of Transportation Safety and Security

This research is important because it will assist the development of transportation systems in smart cities. Knowledge gained from this dissertation can assist policymakers in decision-making regarding deploying these emerging technologies and improving the safety of vulnerable road users. Methodological frameworks and a novel safety surrogate measure introduced in this dissertation would expand the knowledge body of CAVs, assisting researchers investigating these technologies. Furthermore, the results of this study can help agencies and manufacturers of these technologies in improving the safety features' performances. The overall framework of the study and the outline of the proposal are shown in **Figure 1.1** and **Figure 1.2**, respectively.

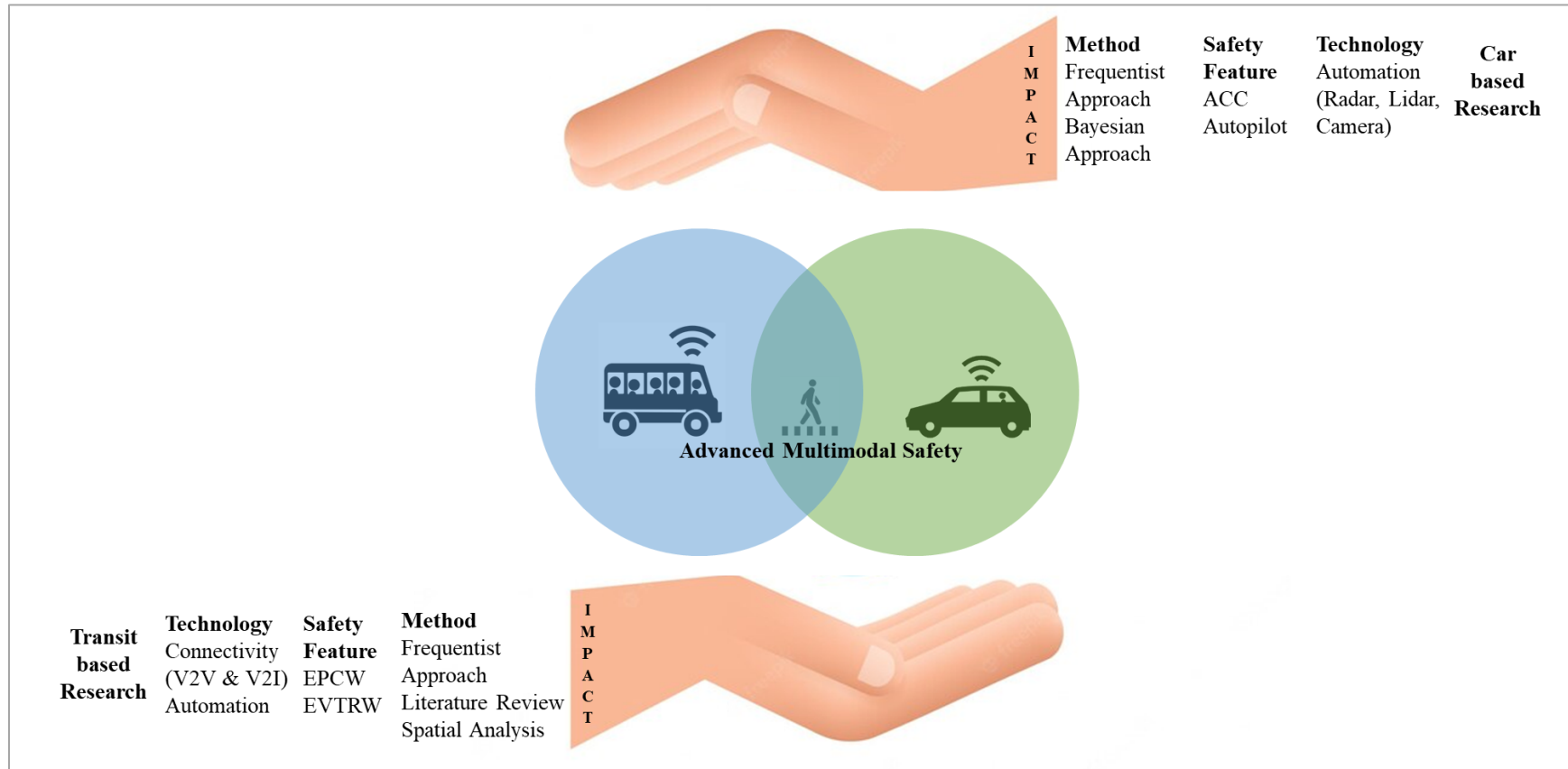


Figure 1.1. Conceptual Framework of the Study

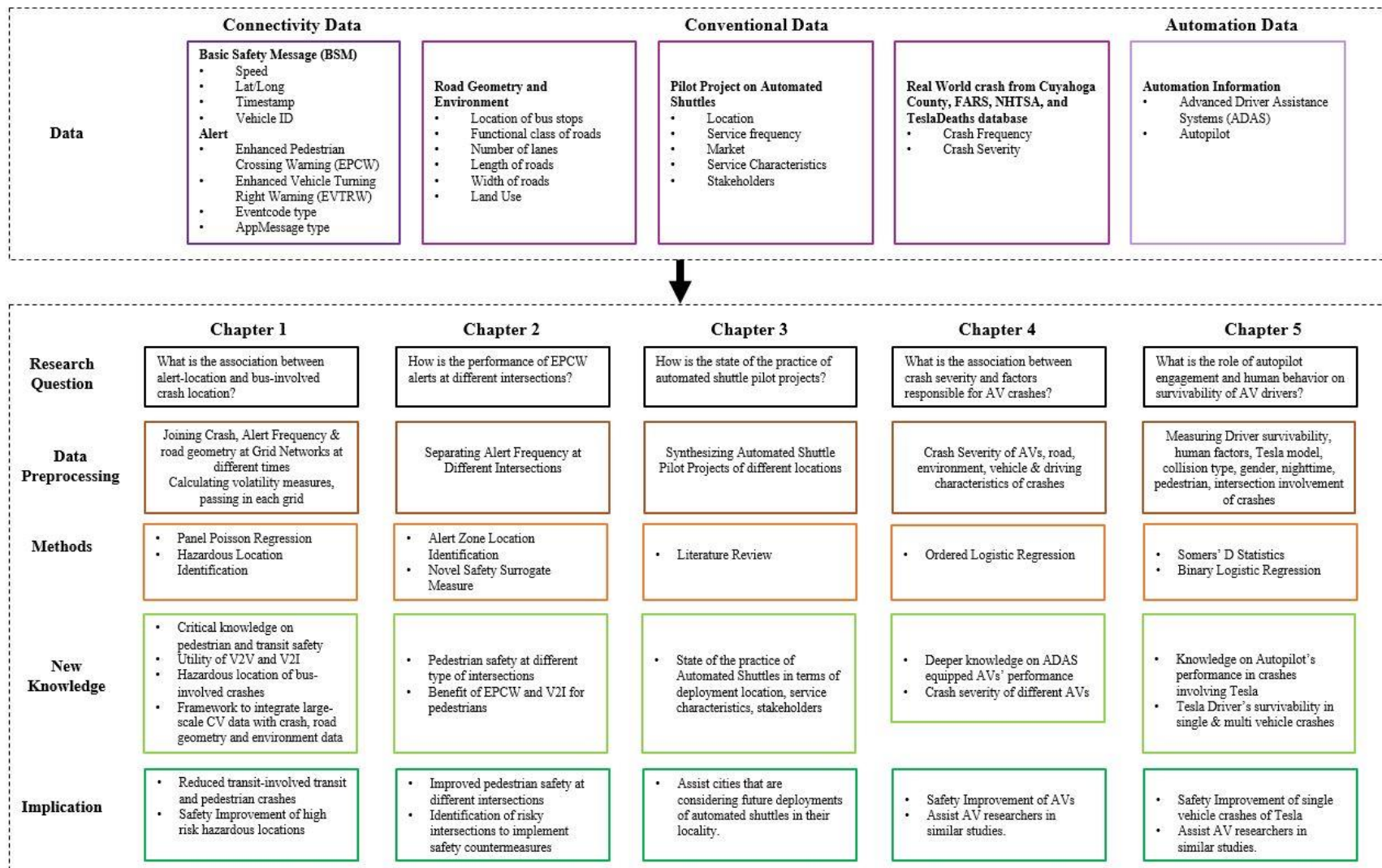


Figure 1.2. Outline of the Dissertation

**CHAPTER 2. ENHANCING THE SAFETY OF TRANSIT USERS AND
PEDESTRIANS IN URBAN AREAS: THE ROLE OF CONNECTED
VEHICLES AND SMART INFRASTRUCTURE TO PROACTIVELY
IDENTIFY HIGH-RISK HAZARDOUS LOCATIONS**

Abstract

With automation and connectivity technologies, new possibilities are emerging for improving the safety of all road users, including the safety of alternative travel modes and vulnerable road users, in line with vision zero plans. This study explores the safety of transit users, especially focusing on interactions and crashes with vulnerable pedestrians. This study uses large-scale smart infrastructure and vehicle communication data for novel applications in personal and transit vehicles from recently deployed connected buses in the US. Smart vehicle and infrastructure technologies can improve public transit safety, e.g., by detecting pedestrians to enhance the safety of all road users in the future. The study analyzes Basic Safety Message (BSM) and alerts to help bus drivers avoid crashes, including crashes with pedestrians. This study aims to enhance transit safety by exploring two Advanced Driver Assistance Systems: Enhanced Pedestrian in Crossing Warning (EPCW) and Vehicle Turning Right in Front of Bus Warning (EVTRW). A publicly available five months of BSM data and alerts received by 24 connected buses, local crash statistics, and geometric features of the roads in Cuyahoga County, Ohio, are used for the analysis. The study area is divided into a grid network. Four speed volatility measures that quantify variations in driving behavior are calculated. Datasets are developed based on EPCW and EVTRW alerts. Panel Poisson regression models suggest that a 1% increase in EPCW alerts is associated with a 0.239% reduction in bus-involved crashes and a 1% increase in bus-driving volatility correlates with a 0.041 unit increase in the average number of bus-involved crashes if all else remains equal. Hazardous location identification maps show that alerts help identify high-risk hazardous locations for bus-involved crashes. EPCW systems can clearly reduce bus crashes with pedestrians. The analysis provides support for policymaker decisions about the value of smart infrastructure, transit-based CV technologies, and vulnerable road user safety.

Introduction

Automation and connectivity technologies are increasingly diffusing through the personal vehicle market and are potentially substantial contributors to the vision zero goal. However, their use for alternative modes and vulnerable road users is limited. New possibilities are emerging for improving the safety of transit users and vulnerable road users, including pedestrians. This study aims to enhance transit safety by exploring the role of two advanced driver assistance systems (ADAS) - the Enhanced Pedestrian in Crossing Warning (EPCW) and Vehicle Turning Right in Front of Bus Warning (EVTRW). These are based on smart vehicle and infrastructure technologies that can benefit public transit and detect pedestrians to enhance safety for all road users in the future.

Connected Vehicle's (CV) Basic Safety Message (BSM) data are state-of-the-art technology that many American cities are recently testing to examine their potential for helping avoid collisions or crashes. According to Federal Highway Administration (FHWA), BSMs are the "minimum transmitted data package exchanged between CVs which include information about the vehicle's current location, speed, heading, acceleration, braking status, and vehicle size" (2). CV's BSM data can be used to understand contributing factors in crashes that can directly impact safety.

Recent research on road safety and contributing factors to crashes has started using BSMs and alert messages received by CVs in various pilot projects. Due to the high confidentiality of CV projects, these data are not disclosed publicly and the research on them is limited. Among the few conducted studies, the focus have been on all vehicles or cars, but not on public transport like buses.

Generally, the number of cars is much higher than the number of buses on any route in US cities. However, the number of passengers carried by bus is much higher than the number of passengers carried by car. Hence, fatal bus-involved crashes cause more casualties than fatal car crashes. Therefore, measures for bus crash reduction are vital.

Deployment of connected buses allows researchers to investigate the potential of BSMs and alerts to help develop bus crash countermeasures. This research hypothesizes that alerts and BSM data can help identify high-risk hazardous locations for bus-involved crashes. This study aims to enhance transit safety based on which the objectives are formulated. The objectives of this study are: 1) To develop a framework for extracting useful information from connected bus pilot projects (BSMs and alerts) and bus-involved crash data; 2) To investigate associations between alert locations and bus-involved crash locations in a road network; 3) To develop a random-effect Panel Poisson Regression model to consider spatial and temporal effect in the CV alerts and crash data.

Initially, hazardous locations for bus-involved crashes are identified using ArcGIS. Five years of bus crash statistics are used to create a layer of hazardous and non-hazardous locations. This layer is overlaid with the location of alerts received by CVs to see whether the alerts also appear in those crash-prone locations or not. Later, a Panel Poisson regression model is used to identify if the number of bus-involved crashes changes with changes in alerts at different times for specific locations.

The intellectual merit and contribution of this study are broad. The study harnesses publicly available big data on connected transit buses to explore the safety benefits of two ADAS, EPCW and EVTRW. These are emerging technologies that have the potential to be adopted in the future. The study develops a framework for combining big data with data from other sources (four different sources) to develop a master database for safety studies that future researchers can follow. BSMs, EPCW alerts, and EVTRW alerts are unique, and they are structured to gain critical knowledge on pedestrians and transit safety that provides long-term scientific value. Historic crash statistics are used to identify real-world hazardous locations for bus-involved crashes. The study highlights the impact of CV technology on transit and pedestrian safety.

Literature Review

This literature review provides an overview of recent studies on three topics relevant to transit bus safety: 1) Analysis of Basic Safety Message data as a tool to reduce bus-involved crashes; 2) Frequency and types of transit bus-involved crashes; 3) Analysis of hazardous locations for transit buses. By synthesizing these three separate areas of study, this review will highlight opportunities for new research to improve transit bus safety.

Prior Research on Basic Safety Messages

Road users are considered the main contributors to all types of crashes. Transportation experts have started taking the initiative to control human factors indirectly by introducing connectivity and automation in vehicles and roadway infrastructure. Recently, transportation agencies have been testing automation and connectivity in various environmental conditions. Connectivity is part of the Intelligent Transportation System (ITS), where BSMs are shared between vehicles and infrastructures to avoid potential crashes. Instantaneous driving behavior, maneuvers, trajectory, individual origin and destination, and traffic data can be acquired from the connected vehicles (3).

A handful of studies have been conducted to explore the relationship between crash frequencies and drivers' volatile behavior or volatility (4-7).

Because untreated or raw BSM data are complex to understand, data analytics methods should be used to extract critical information from them (4). When longitudinal accelerations are used to find drivers' volatility's effect on crash frequency, braking shows substantial volatility (5). One study showed that with a one percent increase in volatility at intersections, crash frequencies could increase by 1.25%, 0.25%, and 0.35% for acceleration/deceleration, passing level volatility captured through coefficient of variation of speed, and mean absolute deviance of vehicular jerk respectively (6). Although various measures derive driving volatility, none have used lateral swipe crash or crash severity to measure volatility (7). Intersections with high variations in the longitudinal movement are prone to higher rear-end crashes, and intersections with lateral volatility have a higher risk of a head-on collision (3). These studies were conducted using overall crash data. However, they did not specify how the results could be generalized for transit bus-related crashes.

In a step towards the future of public transit, many cities and transit agencies introduced the "Smart Bus." These transit buses use the Internet of Things (IoT) technology to exchange information between buses and waiting passengers' smartphones. The information may include real-time location information, arrival time, waiting time, and passenger count of buses (8; 9). The Smart Buses are equipped with GPS devices to track their positions and share this information with a central server periodically from where the location information of the transit buses is accessed and displayed on the computer screens of the bus stops and smartphone apps of the passengers (10; 11). USDOT's ITS Architecture website reports several service packages related to public transit buses that generate and use BSM data. Among those packages, the "Transit Pedestrian Indication," "Transit Vehicle at Station/Stop Warnings," and "Vehicle Turning Right in Front of a Transit Vehicle" package involves the exchange of BSMs between onboard equipment (OBE)'s of connected transit and other CVs, roadside units (RSU), and pedestrians with connected personal information devices. Besides BSMs, there is also an exchange of other information like vehicle location in three dimensions, motion information with other vehicles in the vicinity, transit vehicle's route and stops, and location of the pedestrian wearing the personal information device. Other interesting service packages involve exchanging information with the Transit Management Center, Traffic Management Center, RSUs, and other CVs for improved mobility and ridership. For instance, the "transit signal priority" package uses V2I communications to access one or a series of intersections to save time spent at the red light. The "transit vehicle tracking package" monitors transit buses' location via an automated vehicle location system to update the transit system's schedule in real-time. The "transit passenger counting package" uses sensors to count the number of passengers entering and exiting the buses, which gives an idea of ridership and passenger load at bus stops. The "transit fare collection management" package uses traveler fare cards or electronic devices like smartphones to collect transit fare passengers either from onboard or infrastructures. The "transit traveler information" package enables travelers to have access to transit information like real-time transit schedules, stop annunciation, display of arrival signs, and transit trip itineraries. Information from these packages can be used for expanding ITS-based transit safety analysis (12).

Despite all these efforts, research on transit-based ITS is limited. In 2018, FTA published the "Strategic Transit Automation Research Plan," which summarizes transit-based service packages and their use cases from the perspective of SAE automation levels. It is observed from **Figure 2.1**

that the “Transit Bus Advanced Driver Assistance System” uses automation of SAE levels 1 and 2 to increase the safety of bus operations (13).

Recently, FTA conducted another research report on a pilot project undertaken in an urban bus operational safety platform in Cuyahoga county. It was reported that over the time of six months study period, connected bus drivers’ response to Pedestrian in Crossing Warning (PCW) situations increased by 16% and average drivers’ reaction time to PCW warning situations decreased by 18% (14). This chapter will expand on the data of this research.

Prior Research on Transit Bus-involved crashes

Transit buses mostly ride on designated routes at a decent speed, yet they fall into many road accidents and crashes. Buses are typically large and have many blind spots, which increases the chances of a crash (15). According to the United States Department of Transportation’s (USDOT) Commercial Motor vehicle-related crash report, 12% of the fatal crashes in 2016 involved exclusively large trucks and buses. It also reported that 51% of fatal crashes involved at least one large truck or transit bus and 50% of the fatalities resulted from the involvement of at least one large truck or bus (16). Existing research on transit bus-involved crashes has focused on identifying factors (driver, vehicle, or environment) responsible for the crashes, examining the characteristics of bus-involved crashes, and the impacts of several traffic measures on bus-involved crashes. The studies report that transit buses can be unsafe and are more prone to sideswipe crashes than rear-end crashes (17). However, the probability of bus-involved crash frequency is lower in bus priority lanes as they are located in the slowest traffic lane (18). Various statistical regression models like Poisson, Negative Binomial, Pearson, Linear, and Logistic models have been used to understand the relationship between bus-involved crash frequency and causal factors of the crashes (19). Machine learning techniques like Neural Networking and Random Forest can also be used to acquire better results. The biggest challenge is data in this kind of analysis, and Intelligent Transportation System (ITS) is a good source of data useful for transit safety (20). Transit officials aim to reduce bus-involved crashes by addressing all aspects within their control. Transit experts cannot control drivers’ behavior. However, by identifying hazardous crash locations, they can allocate resources for improving road conditions as countermeasures for crash risks.

Prior Research on Hazardous locations

High-frequency crash locations that are deemed to be hazardous are identified using crash locations and road design information. One study has defined high-risk crash locations as locations having a higher number of crashes than other similar locations considering local risk factors (21). High-risk crash locations are also known as hazardous road locations, accident-prone locations, and black spots. Some researchers have defined them as priority investigation locations (22) and the priority is given based on crash patterns, contributing factors, and countermeasures as the first step of the highway safety management process (23). High-risk crash locations are statistically analyzed using a negative binomial regression model with the help of crash frequency data. This modeling technique has been used by many researchers (24-28). Lately, the Empirical Bayes estimate (EB) of the total crash frequency method has been found to better identify high-risk crash locations (22). Recently, depth functions in multivariate depth-based Full Bayes is found to be a superior method for such analysis compared to other analogous methods. All these studies used crash statistics based on cars, and none of them used crash statistics of transit buses. Therefore, specific research on bus-involved crashes should be undertaken to improve their safety.

This literature review identifies opportunities to improve transit bus safety. For instance, identifying hazardous locations proactively for transit buses has not been explored to the best of the authors' knowledge. Transit bus-related crash analysis is also rather limited in the literature, and it is still unknown whether connectivity is helping reduce the number of bus-involved crashes or not, specifically in hazardous locations. Hence, attention should be given to these gaps to improve transit bus safety.

Methodology

The methodological framework of this study is shown in **Figure 2.2**. Data collection involved acquiring BSMs, alerts, crashes, and geographic information about the study area. As this paper involves the analysis of big data, R studio, STATA, and ArcGIS software are used for data mining and hazardous location analysis. The steps of the research are explained sequentially in the following section.

Data

Four types of data are used for this research: Alerts received by connected buses, BSM data of those connected transit buses, crash data, and geographic information of the road network. The data are described briefly in this section.

The alerts and BSM dataset are from the “Enhanced Transit Safety Retrofit Package (E-TRP)” project and the project data package is collected from the USDOT’s ITS Joint Program Office’s Website (29). The project is from Cuyahoga County, Cleveland, Ohio. From February to August 2018, the buses collected 239,453 BSM alerts. Cutting-edge technological equipment is used in a transit environment to test the impact of connectivity on pedestrian and transit safety. There are three key technologies used for this unique dataset. First, a Dedicated Short Range Communications (DSRC) system is used for both vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication. Second, a High-precision Global Navigation Satellite System (GNSS) is used for tracking connected vehicles. Third, Forward-Looking Infrared (FLIR) cameras are used for detecting pedestrians. E-TRP features are installed on 24 transit buses and three intersections. These buses are from Greater Cleveland Regional Transit Authority (GCRTA) and they are of different makes and models. Equipment installed on the transit buses is termed onboard (OBUs) transit vehicle-based subsystems as shown in **Figure 2.3**. Equipment installed at the pedestrian crossings is termed infrastructure-based subsystems (Roadside Equipment – RSUs), as shown in **Figure 2.4** (14).

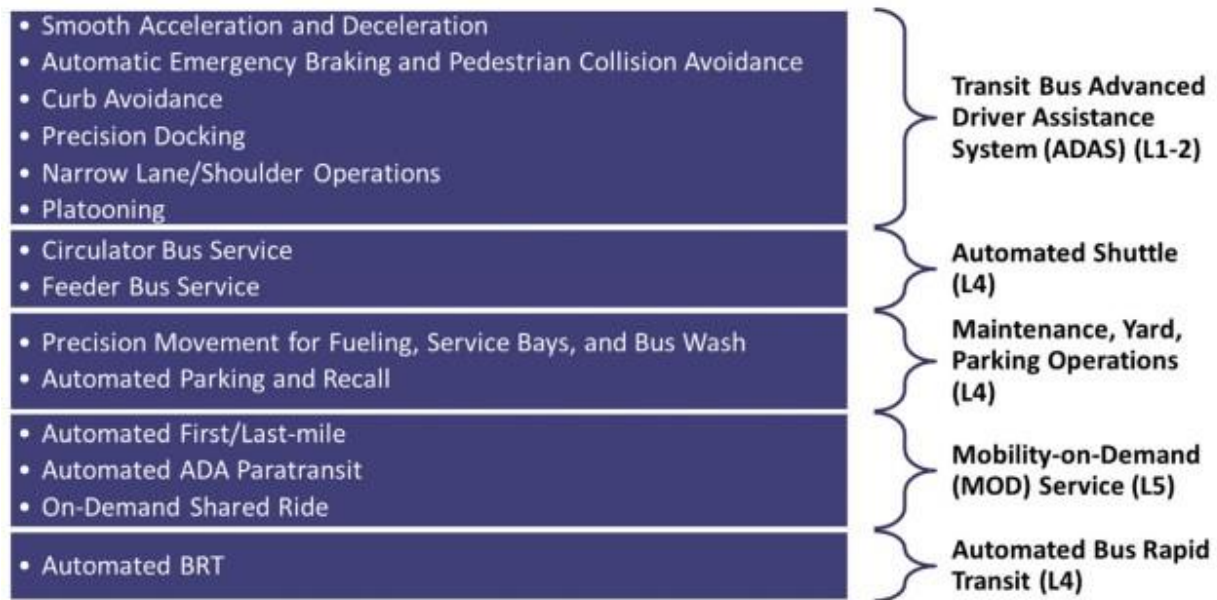


Figure 2.1. Technology packages relevant to transit buses.

Source: US Department of Transportation (13)

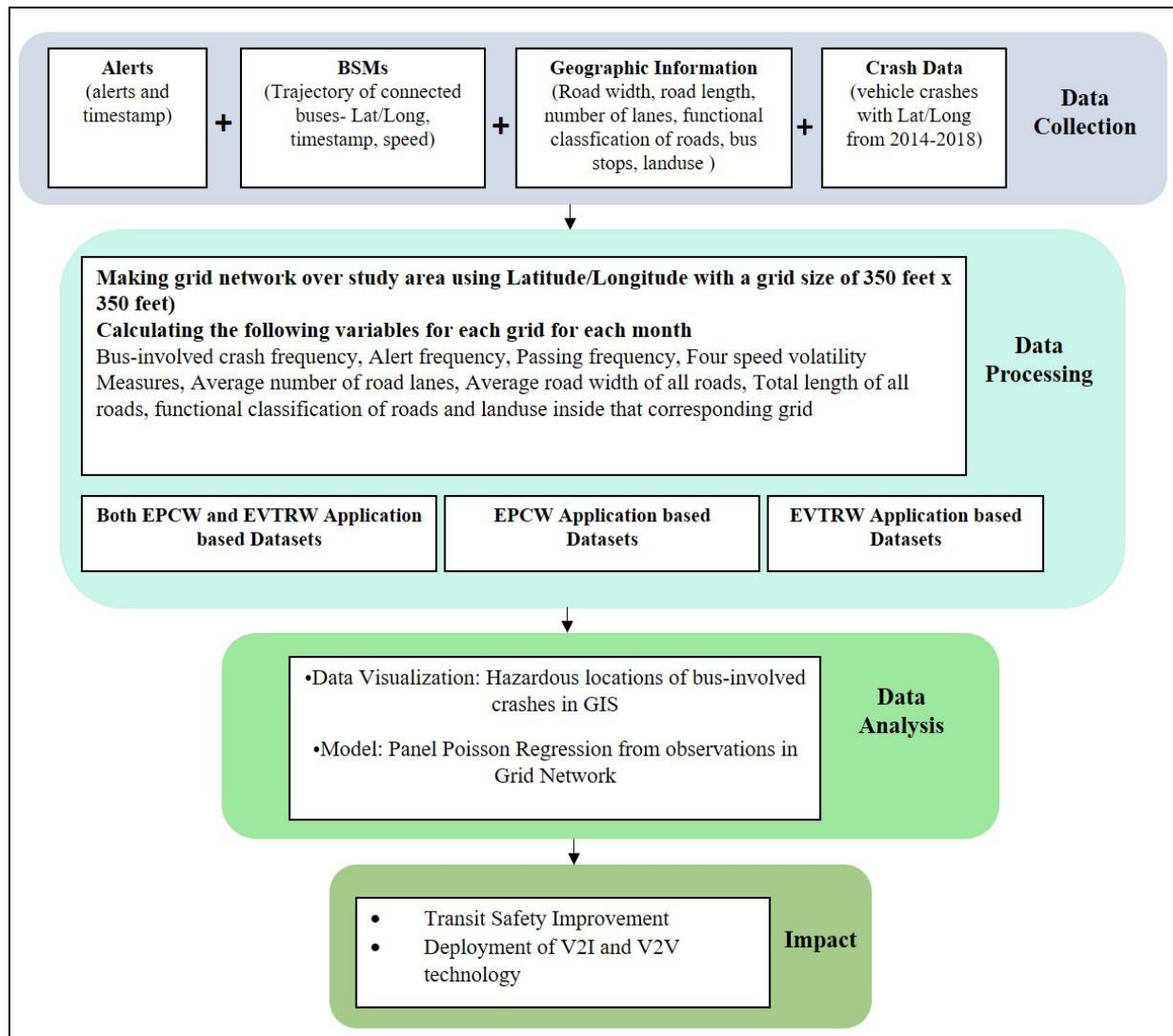


Figure 2.2. Methodological Framework.

Two CV applications are used: EPCW and EVTRW.

“EPCW is a V2I application that alerts the driver of the connected bus about pedestrians whenever a pedestrian enters or is in an E-TRP equipped crosswalk or is in a curbside of the potential path of that connected bus. This application generates three types of alerts (AppMessage Type): 1)“Info” - When the bus enters, is inside, or exits the geofenced area; 2)“InformAlert”-When a pedestrian is on the curb in the potential path of the connected bus; 3)“WarnAlert” – When pedestrians are in the crosswalk in the path of the bus. “InformAlert” and “WarnAlert” are generated only when the traffic signal red light does not protect the pedestrian. EPCW application also obtains information about the position (lat/long), timestamp, heading, speed, and elevation information of the connected buses. Geometric intersection information is received as Map Data Message (MAP). Signal Phase and Timing (SPaT Message) and status of crosswalks (pedestrians detected and zones affected) are received from the roadside infrastructure DSRC broadcast. Vehicle information like turn signal status information is also obtained.” (14)

“EVTRW is a V2V application that alerts the driver of the connected bus whenever a CV, in this case, a connected bus, tries to make an illegal right turn in front of another connected bus that is departing from a near-side bus stop. This application also generates three types of alerts. However, “Info” is generated when the door of the connected bus is open and then closed when it is at the bus stop. “InformAlert” is generated when a connected bus moves from behind to beside the connected bus; “WarnAlert” is generated when the other connected bus starts to turn right in front of the connected bus. “InformAlert” and “WarnAlert” are generated only when the doors of the connected bus open and close in a cycle and the bus is in the forward gear without the foot brake applied. EVTRW obtains similar position and time information to the EPCW application. This application also obtains vehicle information like door status, gear position, and brake status (14).” An example of an EPCW “WarnAlert” and EVTRW “InformAlert” are shown in **Figure 2.5** and **Figure 2.6**.

The E-TRP package has seven dataset files, among which only Alert and BSM files are used in this study. All the datasets used in developing the master data of this study are described below:

Alert Dataset

This dataset contains the EventCode Type, AppMessage Type, and timestamp of those alerts, alert id, device id, and vehicle id. EventCode Type indicates the type of events generated, which include no event generated, application active, application inactive, bus entered area, bus exited area, an area is detected, direction of where the pedestrians are (nearside, farside, leftside, rightside), pedestrian in a road, vehicle is ahead, vehicle is ahead on left side, vehicle is ahead on right side, vehicle is on left side, vehicle is on right side, vehicle is behind on left side, vehicle is behind on right side, vehicle is behind, event is complete, and vehicle is inside the area. The AppMessage Type indicates type of the message like Info, InformAlert, and WarnAlert. Each transit bus has a unique vehicle id, and the devices used in the buses to send or receive messages have unique device ids.

BSM Dataset

This log file contains the Global Positioning System (GPS) locations of the connected buses, timestamp, speed, and heading at a rate of 5 Hz. Speed is in kilometers per hour (KPH). February’s data is excluded from this study intentionally, as February was only the trial period to install and test the devices. August’s data is also excluded since it has data for only two days.

Crash Dataset

The crash data of Cuyahoga county, Ohio, is collected from 2014 to 2018 from the Ohio Department of Public Safety's (DPS) website (30). The Crash data is collected for five years because the number of transit bus-involved crashes is relatively small per year, and five years is deemed a reasonable period to understand the crash pattern. The dataset contains four types of files. Among those, only crash statistics and unit statistics files are merged by the local report number and timestamp to get the GPS location of the transit bus-involved crashes.

Geographic Information of the Road Network Dataset

The information about the width of the road, number of road lanes, length of the road segments, and functional classification of road information is collected in GIS shapefile format from the Northeast Ohio Areawide Coordinating Agency's website from the year 2016 (31). The GIS shapefile of Cuyahoga county's Land use information is collected from Cuyahoga County Open Data website (32). The bus stop locations are acquired from the General Transit Feed Specification (GTFS) dataset for GCRTA (33). These are then converted into a GIS shapefile to extract this information spatially.

Data Processing

Merging Crash Statistics and Unit Statistics

Crash statistics from each year are merged with unit statistics of that year based on the Local report number and timestamp to acquire the GPS locations of bus-related crashes. This is done for each of the five year's data. The files for five years are combined, and only bus-involved crashes are kept. The total number of bus-involved crashes was 1,411 in Cuyahoga County.

Cleaning the BSM Dataset

At first, the BSMs are projected and overlaid on the Cuyahoga county boundary in GIS using their GPS locations (latitude/longitude). It was clear from observation that some BSMs fell outside the county boundary (766), some of which fell into the water body (237) and some on the ground (529). The transit buses had routes extending outside the county boundary and the 766 BSMs outside the county boundaries fell into two of their routes. The data points in the water are considered outliers since GIS generated 114 million BSM data points from March to July. Therefore, the data points falling into the waterbody are trivial compared to all BSMs inside the county. However, this may still induce a question of validity about the BSM locations inside the county. The E-TRP pilot project used GNSS, satellite signals of which can sometimes be blocked for enough time to interrupt the correct positioning of BSMs or alerts if the location is an urban canyon, tree-lined street, underpass, or tunnel (34). It can be inferred that as the connected buses circulated in and outside the downtown area, some GPS locations might have been affected by multipath fading due to the proximity of the buildings of Cuyahoga county's downtown and roads below the highways. This can help explain why some BSMs fell into the waterbody (Lake Erie). The transit route goes beyond the county boundary in the west and south, which explains why some BSMs fell outside the county boundary. It is mention-worthy that the E-TRP pilot project team used the uBlox neo-7P GNSS receiver, which has a Horizontal Accuracy of 2.5 m for 50% Circular Error of Probability (CEP). Based on this fact, it is expected that at least half of the time, the coordinates are within 2.5 m of the exact location (35). The research is conducted assuming these observations as correct.

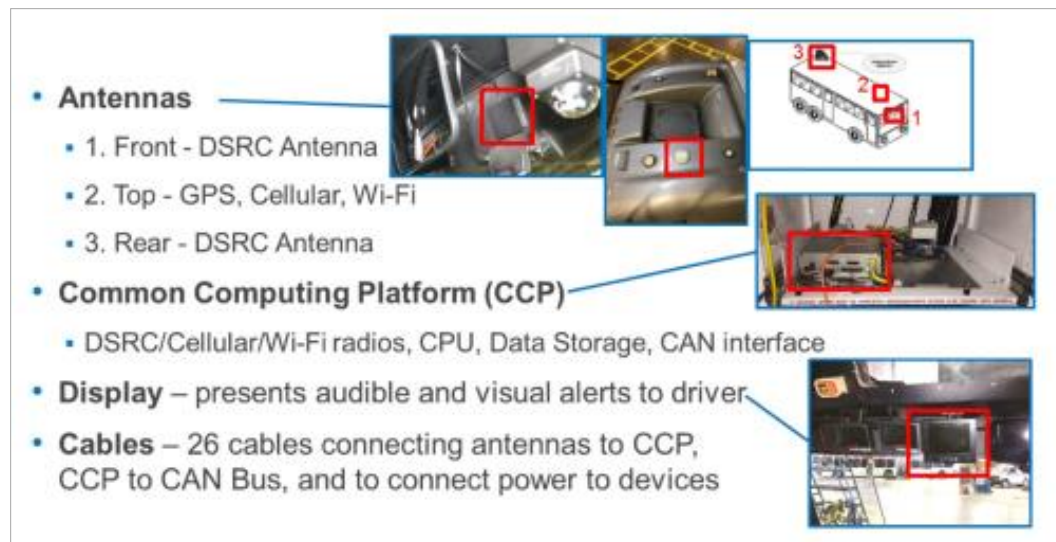


Figure 2.3 Onboard Subsystem

Source: E-TRP Pilot Project (14)



Figure 2.4 Infrastructure Subsystem-Roadside Units (RSU)

Source: E-TRP Pilot Project (14)



Figure 2.5. EPCW WarnAlert to Bus Driver

Source: E-TRP Report (14)



Figure 2.6 EVTRW Inform Alert to Bus Driver

Source: E-TRP Report (14)

Merging Alert Dataset and BSM Dataset

To get the geographic location of each alert, Alert and BSM datasets needed to be merged. Both datasets had vehicle id and timestamp as common variables, but the alert dataset's timestamp had a rounded value for seconds. The BSM dataset had a vehicle trajectory for every 0.2 seconds. Hence, it was never possible to get the exact fraction of seconds when the alerts occurred from this dataset. For the convenience of merging, the BSM dataset's values for seconds are rounded up. By doing this, every second contains five observations. While merging, from every five observations for each second from the BSM dataset, one observation is selected randomly and merged with the alert dataset's observation of the same time and vehicle id. After merging, it was discovered that the BSM dataset was missing observations for some time when the alerts occurred. Therefore, the merged file missed some alerts. The device to generate and store the alerts could have failed to operate properly at that time. The percentage of usable alerts after merging the two datasets is shown in **Figure 2.7**, which illustrates that not many alerts are lost due to this matter.

After the merging process, samples containing irrelevant applications like RCI and TSPW are removed from the dataset. Two samples having missing values under application are also removed. After the cleaning process, other information is added to the merged dataset.

Adding geographic information with the merged dataset

With the help of spatial joining in ArcGIS, road and land use information are connected to the merged dataset of alerts and BSMs.

Making a grid network of the study area

Previous studies have considered road intersections as their unit of analysis. However, this leaves out many other locations of the road network where crashes occur. A grid network consisting of grids, each grid having a dimension of 350'x 350', is developed in R. The grids are considered the location unit for the analysis to cover the entire study area. 350' is equivalent to approximately 100 meters (106.68 meters). Some prior research regarding pedestrian crashes near bus stops used 100 meters buffer area (36) for their analysis. The E-TRP data involves BSMs that convey alerts about pedestrians and other connected buses to the drivers, considering the 350'x 350' grid deemed reasonable for analysis.

In the developed grid network, there were grids where alerts were generated. There have been many more grids where alerts are not generated (0 alerts), but the BSMs of connected buses had their trajectory information. As the geographic unit of analysis in this study is 350'x350' (106.68 m x 106.68 m), most of the BSMs and Alerts falling inside the grids can be acceptable. There will remain a chance of some error for the GPS locations on the border of those grids, which is a limitation of the data preparation process. Since this study is at a macro level, based on the accuracy level of GNSS, the BSMs inside the study area are considered reasonable to use. As the connected transit buses are from GCRTA and the study area of the pilot project is Cuyahoga county, discarding the BSMs that fell outside the county boundary is logical in this study. For future works, a more rigorous pre-processing approach, such as Kalman filtering (37; 38), can be applied to filter out the outliers extensively.

For this study, grids with a minimum of one alert (Grids with ≥ 1 alert) were considered because the aim is to observe whether an association remains between crashes and alerts. In this way, 334 unique grids were identified over five months, containing a minimum of one alert.

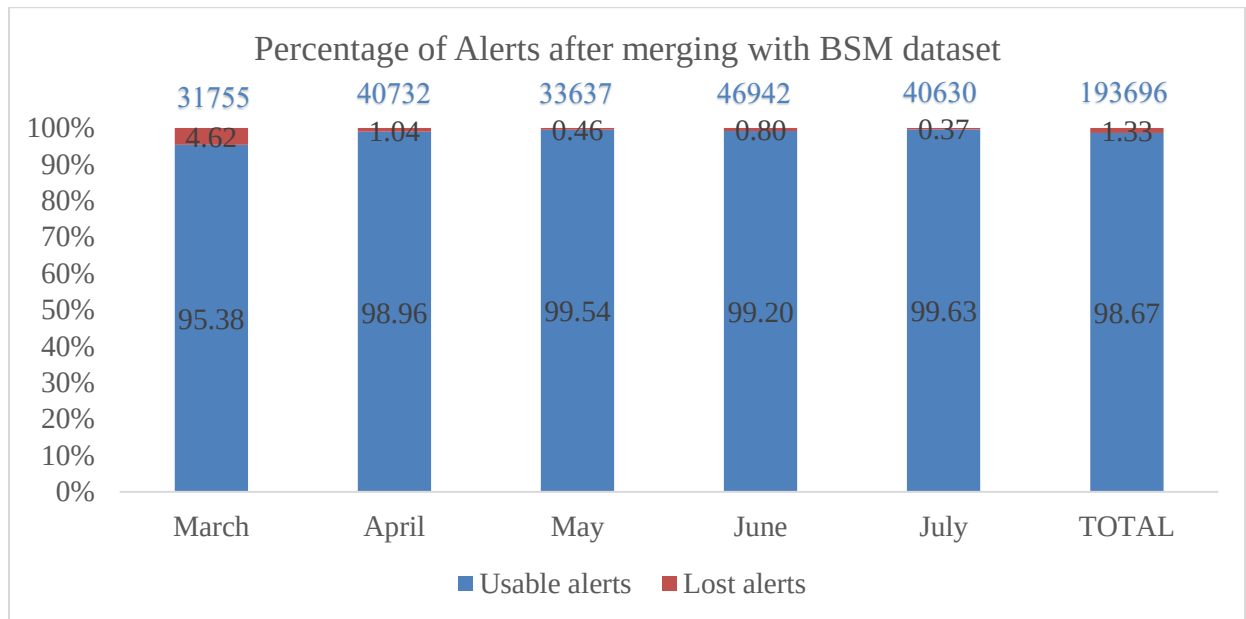


Figure 2.7. Percentage of usable alerts after merging Alert and BSM datasets.

Next, the total frequency of alerts, crashes, and passing frequency for each grid are computed using the Latitude-Longitude information. As the unit of study was shrunk down to these selected grids, the total number of usable alerts, BSMs, and bus-involved crashes also reduced in frequency. Four volatility measures, the average number of lanes, average road width, total length of all roads per grid, the number of bus stops per grid, and the number of passings by buses in each grid, are calculated. Later, a dummy variable indicating the presence of bus stops per grid is developed. The follow-up study of this research will incorporate the grids with zero alerts for further investigation.

Dividing the master dataset into three datasets based on CV Applications

After the development of the grid network and the required variables, the master dataset is divided into three datasets, as shown below:

Both EPCW & EVTRW dataset: This is the cleaned dataset that contains both EPCW and EVTRW-based alerts. From March-July, 170,065 alerts were generated in the grids.

EPCW dataset: This dataset contains only EPCW-based alerts. From March-July, 119,744 EPCW-based alerts were generated in the grids.

EVTRW dataset: This dataset contains only EVTRW-based alerts. From March-July, 50,321 EVTRW-based alerts were generated in the grids.

Data Analysis

Panel Poisson Regression Model

Panel data is longitudinal or cross-sectional time-series data that helps to observe the behavior of entities across time. For this paper, each grid is an entity. Panel data is used to control unobserved variables and helps understand individual heterogeneity (39).

General Poisson Regression Model (40) is used to predict count data when the dependent variable is discrete and there is a preponderance of zeros and small values in the dependent variable. According to Poisson distribution, the probability of n crashes in a grid i is as follows:

$$P(n_i) = \lambda_i^{n_i} \exp(-\lambda_i) / n_i! \quad (1)$$

Here, λ_i is the expected number of crashes in grid i . The Poisson parameter is laid out in the following form to fit the regression model.

$$\ln(\lambda_i) = \beta X_i \quad (2)$$

Here, X_i is the matrix independent variable and β is a vector of the coefficients.

Poisson Regression model (41) for one-way Panel data can be written as follows:

$$y_{it} \sim \text{Poisson}(u_{it}) \quad (3)$$

$$\text{Here, } u_{it} = \alpha_i \lambda_{it} = \alpha_i e^{\beta X'_{it}} \quad (4)$$

Where, $i = 1, \dots, N$; $t = 1, \dots, T$

α_i denote the individual effects. α_i need to be independent and identically distributed (iid) random variables for the model to be a random effect model. The model is estimated by assuming a distribution for α_i .

Random effects Panel Poisson Regression model is conditional of λ_{it} and parameter α_i . y_{it} is iid Poisson distributed with parameter $u_{it} = \alpha_i \lambda_{it} = \alpha_i e^{\beta X'_{it}}$, with the individual effects α_i considered

as an iid random variable. The joint density for observations in all time periods for the i th individual is as follows:

$$P[y_{i1}, \dots, y_{it} | \lambda_{i1}, \dots, \lambda_{iT}] \quad (5)$$

Density $g(\alpha)$ of α_i needs to be specified to achieve equation (5).

Let,

$$\alpha_i \sim \text{iid gamma}(\theta, \theta) \quad (6)$$

$$E(\alpha_i) = 1 \text{ and } V(\alpha_i) = \frac{1}{\theta}$$

$$g(\alpha_i) = \frac{\theta^\theta}{\Gamma(\theta)} \alpha_i^{\theta-1} e^{-\theta \alpha_i} \quad (7)$$

Location-based volatility measure calculation

Volatility measures potentially represent instantaneous driving behaviors or decisions of drivers that increase crash risk and severity while passing through a location in the study area (6). Volatility indices can identify locations where driving behavior differs from the other locations. Recently, researchers have started examining volatility indices to measure abnormal and extreme driving behaviors so that crash risks can be mitigated when a driver maneuvers in an unsafe manner (5). As speed directly influences safety outcomes and is related to the risk of being involved in a crash, this paper uses speed profiles to measure volatility. Four types of speed volatility measures (3) are identified from prior studies and their formulas are as follows:

$$\text{Standard Deviation of Speed: SpeedSD}_i = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

$$\text{Coefficient of variation of Speed: SpeedVol}_i = \frac{SD_i}{\bar{x}} * 100 \quad (9)$$

$$\text{Mean Absolute Deviation of Speed: SpeedMAD}_i = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (10)$$

$$\text{Quartile Coefficient of Variation of Speed: QuartCOV}_i = \frac{Q3_i - Q1_i}{Q3_i + Q1_i} * 100 \quad (11)$$

In equations (8) through (11), subscript i indicates the index number of the grids, j indicates the index number of BSM alerts in the grid “ i ,” n indicates the number of BSM alerts in grid “ i ,” x_i indicates the speed in the grid “ i ,” $\bar{x}$ indicates the mean speed in the grid “ i ,” SD_i indicates the standard deviation of speed in the grid “ i .” For instance, *SpeedSD* indicates the extent of deviation present in a set of speed values. In equation (8), the Mean Speed of a grid “ i ” is subtracted from all the speed values located inside grid “ i ,” and the subtracted value is squared and then summed for that grid “ i .” Hence, the *SpeedSD* of each corresponding grid is calculated.

Similarly, *SpeedVol_i* indicates the degree of dispersion of all speed values in the grid “ i ” from the mean speed in the grid “ i .” *SpeedVol* is presented as a percentage and it compares the speed volatility across different sets of speed values. *SpeedMAD_i* indicates the average distance between each speed value in the grid “ i ” and the mean of all speeds in the grid “ i .” In equation 11, $Q1_i$ and

$Q3_i$ indicate 25th and 75th percentile speed in the grid “i,” respectively. $QuartCOV_i$ indicates the relative dispersion of all speed values in the grid “i” based on the interquartile range of all speed values in the grid “i.”

All four of these speed volatility measures are unitless. Volatility measures with high numeric values indicate the volatile behavior of drivers in that grid, which can indicate incidents like near-miss situations when a driver brakes hard to avoid a collision with a pedestrian or a vehicle. Volatility measures with low numeric values indicate less volatile behavior of a driver in that grid, which can indicate incidents when a driver makes a gradual stop at a bus stop. When the buses stop at bus stops, their speed goes down to zero. In such situations, buses are maneuvered steadily by the drivers to make a gradual stop. When driving volatility is measured around bus stops, volatility would have some numeric value, but it is expected to be on the low side.

Hazardous Location Analysis in ArcGIS

An analysis tool in ArcGIS is used to proactively identify hazardous locations for buses. The tool uses a spatial clustering process that identifies statistically significant spatial clusters using Getis-Ord G_i^* statistic. Getis-Ord G_i^* (G_i^* statistic) is an indicator of local autocorrelation as it evaluates the degree to which spatial auto-correlation varies locally over the study area. In this statistical method, a neighborhood of a specified geographic distance is considered around each feature and then the degree to which features surround this feature with similarly high or low values is measured (42). A statistically significant cluster of features with high values are hazardous locations and a statistically significant cluster of features with low values are non-hazardous locations. The G_i^* statistic that ArcGIS uses is a version of G_i^* in which the G_i^* index is combined with the z-score into one index (42). However, the analysis produces a new output feature class where each location or feature has a z-score, p-value, and confidence level bin (G_i_Bin) value (43). G_i_Bin field identifies the hazardous and non-hazardous locations. If the values of the features fall in the ± 3 bins, the location is statistically significant with a 99% confidence level. If the values fall in the ± 2 bins, the location is significant with a 95% confidence level. If the feature’s values fall inside ± 1 bins, that reflects a 90% confidence level. The clusters of features in bin 0 is not statistically significant. The formula for G_i^* statistic is as follows (42):

$$G_i^* (d) = \frac{\sum_j w_{ij} (d) x_j - \bar{X} \sum_j w_{ij} (d)}{S \sqrt{\frac{n \sum_j w_{ij}^2 - (\sum_j w_{ij} (d))^2}{n-1}}} \quad (12)$$

Here,

$$\bar{X} = \frac{\sum_i x_i}{n} \quad \text{and}$$

$$S = \sqrt{\frac{\sum_j x_j^2}{n} - (\bar{X})^2}$$

In equation 12, $G_i^* (d)$ is the statistic computed for a feature i , at a distance d , standardized as a z-score. x_j denotes the attribute value of each neighbor, w_{ij} denotes the spatial weight for the target-neighbor i - j pair, and n denotes the sample size of the dataset. The geographic distance between each i - j pair is measured following the Euclidean distance method. The spatial weight is calculated from a spatial weight matrix. This matrix is an $n \times n$ table, which surrounds each feature, and all the values in the matrix are a weight representing the association between a pair of features in the dataset. The weights are assigned based on the Inverse Distance Weighted (IDW) method. IDW

method identifies the cell values that do not have any value by using a linearly weighted combination of a set of sample points. The weight is an inverse function of a distance fixed before running the operation (44). The general formula for IDW (45) is as follows:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i) \quad (13)$$

Here,

$\hat{Z}(s_0)$ denotes the value for the location s_0

N denotes the number of sample points surrounding the prediction location that will be used in the prediction

λ_i denotes the weights assigned to each measured point. This weight decreases as weight increases

$Z(s_i)$ is the observed value at the location s_i

The formula to determine the weights is as follows:

$$\lambda_i = \frac{\frac{1}{\sum_{i=1}^N d_{i0}^p}}{d_{i0}^p} \quad (14)$$

As the distance d becomes larger, the weight reduces by a factor of p . The value d_{i0} indicates the distance between the prediction location s_0 and the measured locations, s_i . The Inverse Distance Weighted (IDW) interpolation method is applied to this new output feature class to map the hazardous locations over the study area. The distance used for this research is 10,000 meters.

Results

Results from both spatial and statistical analysis are described here. At first, hazardous locations for bus-involved crashes are identified in ArcGIS. Then grids containing BSM alerts and vehicle trajectories are overlaid on the hazardous location map to visually understand whether BSM alerts can spatially indicate hazardous locations for bus-involved crashes. Next, a Random Effect Panel Poisson regression model's results for grids having ≥ 1 alert are described.

Hazardous locations for transit bus-involved crashes using ArcGIS

Figure 2.8 shows the hazardous locations for transit bus-involved crashes in Cuyahoga County. ArcGIS identifies 800 hazardous locations in Cuyahoga County based on the location of five years of bus crash data. The Gi_bin shown as red and orange circles (205/800 areas) are considered more hazardous for bus-involved crashes and they are concentrated mainly within 3 miles (approximately) radius of the downtown area. After that, there are still many locations where crashes happened. However, those locations are very dispersed and bus-involved crashes happening there are low in frequency. As a result, those locations (594/800 areas) are found as non-hazardous areas (Gi_Bin shown as green and blue circles). However, identifying the scale of hazardousness in the areas between these hazardous locations needed to be determined. IDW interpolation method is applied, which uses the Gi_Bin based areas and interpolates to distribute more and less crash-prone areas across the region. Thus, hazardous areas for bus-involved crashes in Cuyahoga County are developed, represented as the red to yellow areas in **Figure 2.8**. The non-hazardous areas are represented as green-shaded areas.

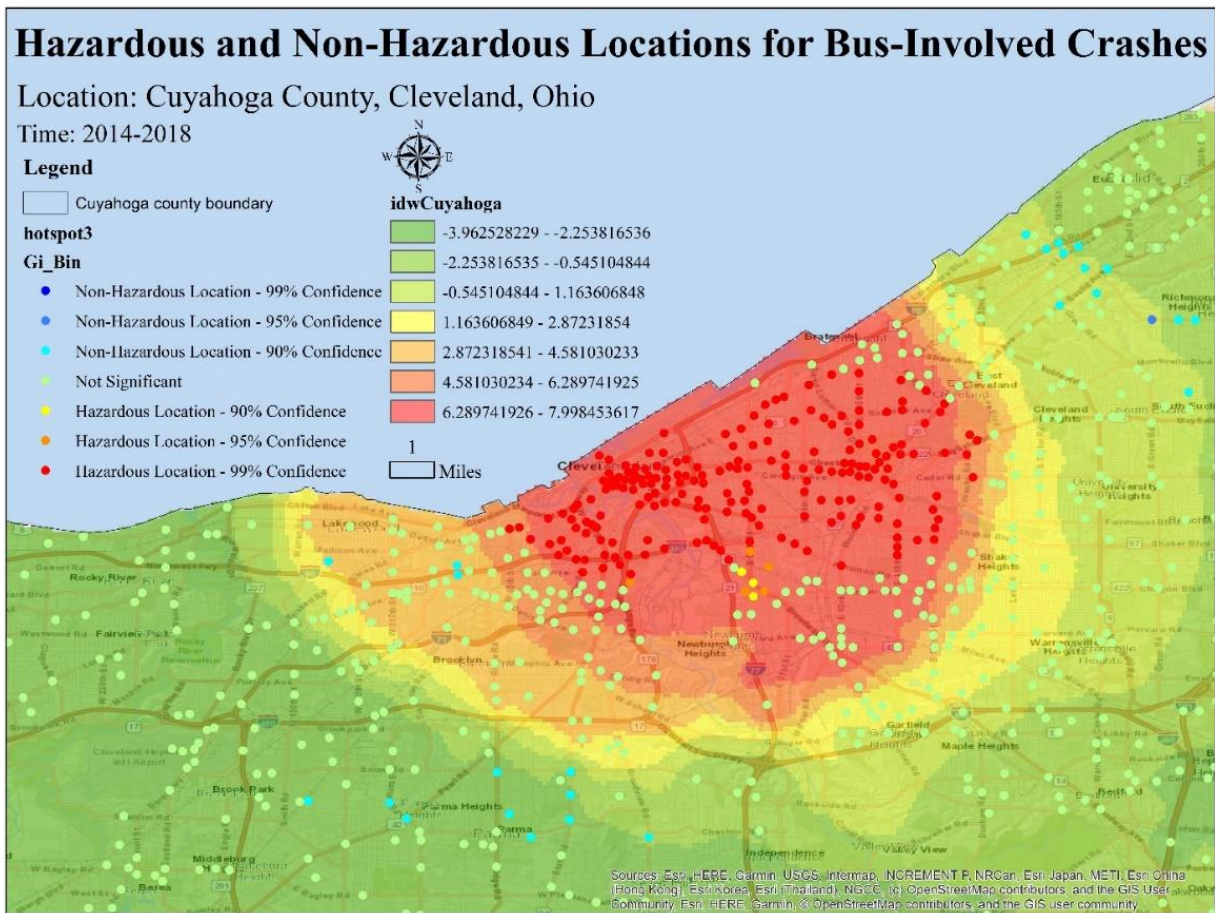


Figure 2.8. Map of hazardous locations for bus-involved crashes in Cuyahoga County

Figure 2.9 shows the locations of the 352 grids of the network (a total of 200,998 grids) that have a minimum of one alert (either EPCW or EVTRW). This map is also prepared using ArcGIS. Understandably, alerts are concentrated in the downtown and near the three selected intersections, and alerts are also seen along routes where the connected buses drove as part of the E-TRP pilot project. It is visible from **Figure 2.9** that grids with a minimum of one alert are more concentrated in the hazard-prone locations of bus-involved crashes.

Grids with ≥ 1 alert

Table 2.1 shows the number of observations or grids in the datasets that contain ≥ 1 alert. As the datasets are developed as panel data for five months, each unique grid can have up to five observations (forming a group) depending on whether the connected transit bus received any alert in that grid in those five months. It is clear from **Table 2.1** that, in the “Both EPCW & EVTRW” dataset, there are 807 observations or information about 807 grids, among which 334 are unique. The “EPCW” dataset has only 24 unique grids and 85 observations in 5 months. The “EVTRW” dataset has 326 unique grids and 782 observations in five months. In the “Both EPCW & EVTRW,” “EPCW,” and “EVTRW” datasets, the average number of observations per group is 2.4, 3.5, and 2.4, respectively. This indicates that some groups did not have observations in all five months and are unbalanced panel data.

In the “Both EPCW & EVTRW” alert-based dataset, 471 of 807 grids do not have crashes and 227 of 334 (67.96%) unique grids do not have crashes. Specifically, in the “EPCW” alert-based dataset, 42 of 85 grids do not contain any crashes. Among 24 unique grids of the “EPCW” alert-based dataset, 14 grids (58.33%) do not have any crashes. In the “EVTRW” alert-based dataset, 446 of 782 grids do not have crashes, and 219 of 326 (67.18%) unique grids do not have crashes.

Descriptive Statistics on grids with ≥ 1 alert

Table 2.2 shows the descriptive statistics of the three datasets. In the “Both EPCW & EVTRW” alert-based dataset, the dependent variable *Crash Frequency* represents a total 5-years crash in each of the 334 unique grids. It has a minimum value of zero, a maximum value of 10, and an average of 0.98. This indicates that there are grids that do not have crashes and grids that have high crash frequencies. 221 crashes happened in the 334 unique grids (794 crashes in 807 observations) in five years. Hence, crash frequency per unique grid is 0.98 in 5 years, which is very low. This is expected because the number of bus-involved crashes is generally low and Cuyahoga county is no different in this matter.

As the values of independent variables are available for five months, their descriptive statistics are measured based on all 807 grids in five months for the “Both EPCW & EVTRW” dataset. Among the independent variables, alerts are important because the primary goal of this research is to find if there exists an association between the location of crashes and alerts. *Number of Alerts provided to bus drivers* in the grids in five months has a minimum value of 1 and a maximum value of 10,449. This indicates that in the study period, the highest number of alerts a grid received in a month was 10,449. Its mean value is 220.74, which indicates that, on average, the grids received 220.74 alerts in a month. It implies that many grids have a higher frequency of alerts. The histogram of this variable is right-skewed. So the log of this variable has been developed for use. *Log of Alerts* has a minimum value of 0, maximum of 9.25, and mean of 2.58.

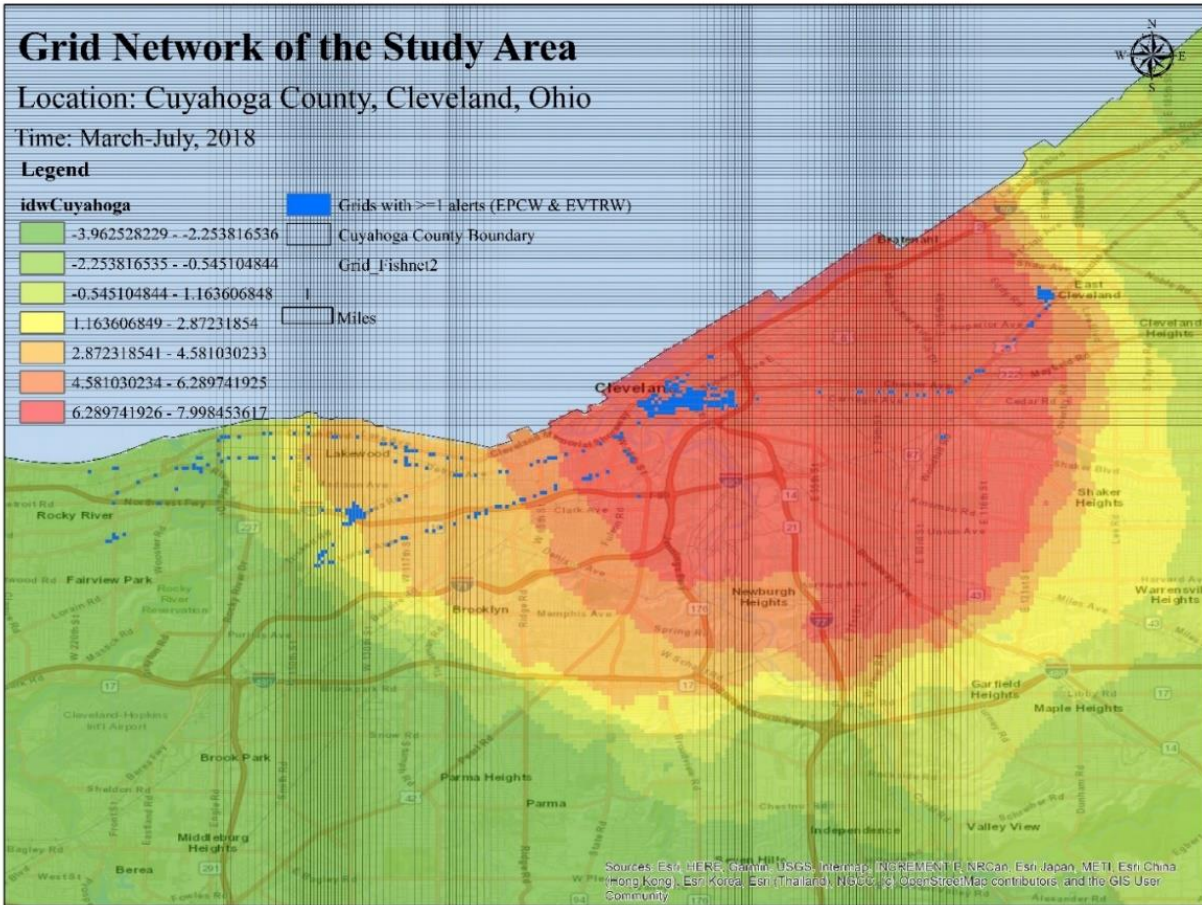


Figure 2.9. Hazardous locations vs. grids with ≥ 1 alert (Both EPCW & EVTRW).

Among the volatility measures, *SpeedVol*, *SpeedSd*, *SpeedMAD*, and *QuartCOV* have mean values of 1.43, 3.14, 2.61, and 83.69, respectively. Driver behavior is considered volatile if the volatility measures have a value higher than one. The minimum values of all four volatility measures indicate that there are locations where drivers showed very little or no volatile behavior, but some locations showed high volatile behavior. Drivers should be concerned about those locations. The variable *Number of passings* has a minimum value of seven, a maximum value of 8,036, a mean of 1,716.43, and a standard deviation of 1,400.08. This indicates that the connected buses passed these grids a minimum of seven times and 1,716.43 times in a month on average. It implies that the buses have passed many times on their routes, but alerts were sometimes generated on those journeys. There can be some places where a very high number of alerts have occurred. *Log of passing* has a minimum, maximum, and mean values of 1.95, 8.99, and 7.02. The geographic variables are constant over time. The variable *Average number of lanes* has a minimum value of one, a maximum value of eight, and a mean value of 4.39. This means this grid network has some grids with eight-lane roads, but on average, these grids have four-lane roads. The variable *Average road width (feet)* has a minimum value of 10, a maximum value of 96, and a mean value of 59.63. The *Total length of roads (feet)* variable has a mean of 446.33 feet, indicating that the roads falling inside the grids totals 446.33 feet on average. The variable *Number of bus stops* has a minimum value of zero and a mean value of 1.24. This indicates that although some grids do not have any bus stops, these grids have at least one bus stop on average.

In the “EPCW” alert-based dataset, *Crash Frequency* has minimum, maximum, and mean values of zero, 10, and 0.91, respectively. Only 24 transit crashes (77 crashes in five panels) happened in these 24 grids over the 5-years. So, it can be said that, on average, 0.91 transit crashes have occurred in five years per unique grid. Among the dependent variables, *Log of Alerts* has minimum and maximum values similar to the previous dataset but has a mean of 5.38. This indicates that the frequency of alerts is very high in the EPCW dataset’s grids. The mean values of the four volatility measures are similar to the “Both EPCW & EVTRW” alert-based dataset. The variable *Number of passings* has a minimum value of 15, a maximum value of 5,583, mean of 2,588.66. This indicates that the connected buses have passed through “EPCW” alert-based grids on an average more times than “Both EPCW & EVTRW” grids. It can be assumed that buses have visited the three intersections more frequently than the rest of the routes of this project. The *Total length of roads (feet)* variable has a mean of 446.87, similar to the previous dataset.

In the “EVTRW” alert-based dataset, the dependent variable *Crash Frequency* has almost identical minimum and maximum values as the “Both EPCW & EVTRW” dataset. This is expected because most grids of the “Both EPCW & EVTRW” dataset are from the EVTRW dataset (782 of 807 = 96.9%). Among the dependent variables, the *Log of Alerts* has a mean of 2.29. This clearly indicates that the frequency of EVTRW alerts is very low in this dataset’s grids. This dataset’s mean values of all the volatility measures are higher than those observed in the “EPCW” dataset. This indicates that drivers have been more volatile when driving in these “EVTRW” alert-based grids. The variable *Number of passings* has a mean of 1,724.93, indicating that the connected buses have passed through “EVTRW” alert-based grids fewer times than “EPCW” alert-based grids.

Table 2.3 shows the descriptive frequency of key variables of the three datasets, namely the *Presence of bus stops*, *Functional classification of roads*, and *Land use*. The *Presence of bus stops* is a categorical variable that has two attributes. In the “Both EPCW & EVTRW” dataset, *No bus stops* (0) have a frequency of 284 grids (35.19%) and ≥ 1 bus stop (1) has a frequency of 528 grids (64.81%). This indicates that most of the grids contain ≥ 1 bus stops. In the “EPCW” dataset, 58

out of 85 grids (68.24%) grids have ≥ 1 bus stops which are higher in percentage than the grids found in the previous dataset. In the “EVTRW” dataset, the variable *Presence of bus stops* indicates that along the trajectory of the connected buses over five months, 521 of 782 (66.62%) grids had bus stops. The variable *Functional classification of roads* has four levels: *Arterials*, *Collectors*, *Interstate*, and *Mixed*. The *Mixed* level represents the presence of more than one type of road in the grid. It is observed that, among the four types, *Arterial* roads are found in the majority of the grids in all three datasets, followed by *Collectors* and *Mixed* roads. The variable *Land use* has six levels: *Commercial*, *Green Space*, *Industrial*, *Institutional*, *Residential*, and *Utility*. The frequency distribution of the land use in the “EPCW” dataset indicates that most grids are *Commercial*, followed by *Green Space*, *Institutional*, and *Residential*. In the “EVTRW” dataset, most grids are *Commercial*, followed by *Residential*, *Institutional*, and *Industrial*.

Models on grids with ≥ 1 alert

Random effects Panel Poisson regression model is applied to the three types of datasets to investigate the association between *Crash Frequency* and other independent variables. The initial models (not shown in **Table 2.4**) gave some unexpected results as the association between crashes and the three volatility measures like, *SpeedSd*, *SpeedVol*, and *SpeedMAD*, were negative. These variables are omitted from the presented models, and only the variables with correct signs are kept. The models that proved that the panel estimator is better than pooled estimator are kept (with the help of the *likelihood ratio test of $\alpha = 0$*). **Table 2.4** shows the panel Poisson models on grids with a minimum of one BSM alert. *Alpha (α)* value > 0 indicates the presence of overdispersion over time in panel datasets. The panel models express the effects of time on the dependent variable. Average Marginal Effects (AME) are also presented to explain the association between crash frequency and other independent variables.

In model 1, the variable *Log of Alerts* has a coefficient of -0.090, indicating that one unit increase in log of EPCW or log of EVTRW alert is associated with a decrease of 0.090 log count of crashes. This also means that a 1% increase in either EPCW or EVTRW alert is associated with a 0.090% decrease in bus-involved crashes if all else remains constant. This implies that an increased number of alerts can lead to a decrease in crash frequency. It can be deduced that bus drivers will get more cautious and drive safely to avoid a potential crash if a location has high alert frequencies. However, this coefficient is not statistically significant in model 1 ($p > 0.1$). In this model, the variables *QuartCOV*, *Log of passing*, *Presence of bus stops*, *Industrial*, *Residential*, and *Utility* have a coefficient of 0.015, 0.329, 0.764, -2.086, -1.491, and -1.571. They are all statistically significant (p -value < 0.05). One unit increase of *QuartCOV* is associated with a 0.015 unit increase in the log count of *Crash Frequency* if all else remains constant. From the values of AME, it can be said that a 1% increase in *QuartCOV* is associated with a 0.015 unit increase in the average number of bus-involved crashes if all else remains constant. More volatile driving behavior may cause more probability of crashes. The coefficient of *Log of passing* indicates that a 1% increase in passing is associated with a 0.329% increase in bus-involved crashes if all else remains constant. The AME of the *Presence of bus stops* indicates that the grids with bus stops are associated with 0.764 units more crashes than grids without bus stops. The *Land use* variable's base is *Commercial*. The AME of *Industrial* indicates that *industrial* land use-based grids are associated with 2.086 units less crashes than commercial land use-based grids on average if all else remains constant. Likewise, *Residential* land use-based grids are associated with 1.491 units less crashes than commercial land use-based grids on average if all else remains constant. *Utility* land use-based grids are associated with 1.571 units less crashes than commercial land use-based grids on

average if all else remains constant. This model also contains an interaction term, *Presence of bus stops*Log of Alerts (1)* and the explanation of its coefficient is complex. This term's AME of 0.173 indicates that a 1% increase of the *Log of Alerts* in grids with bus stops is associated with 0.173 units more bus-involved crashes on average than the grids without bus stops. It can be deduced that the presence of bus stops increases the probability of bus-involved crashes even with the increasing number of alerts. Hence, bus drivers should be careful near bus stops.

In model 2, the variable *Log of Alerts* is statistically significant (p-value <0.01) with a coefficient of -0.239. This indicates that one unit increase in log of EPCW alert is associated with a decrease of 0.239 log count of crashes. It also means that a 1% increase in EPCW alerts is associated with a 0.239% decrease in bus-involved crashes if all else remains constant. *Log of passing* has a coefficient of 0.996 (p-value <0.1), which indicates that a 1% increase in the number of passings in a grid is associated with a 0.996% increase in bus-involved crashes if all else remains constant. From the values of AME, it can be said that a 1% increase in *QuartCOV* is associated with a 0.041 unit increase in the average number of bus-involved crashes if all else remains constant. The *Functional classification of roads* variable has *Arterials* as the base. Among all the levels of this variable, only *Mixed* has a statistically significant coefficient. The AME of *Mixed* indicates that grids with *Mixed* roads are associated with 1.127 units less crashes than grids with only *Arterial* roads on average if all else remains constant. In model 2, only *Green Space* is statistically significant among the levels of *Land use* variable. The AME of *Green Space* indicates that grids with *Green Space* land use are associated with 0.746 units (p < 0.05) less crashes than grids with *Commercial* land use if all else remains constant.

In model 3, the *Log of Alerts* has a negative coefficient of 0.054, which indicates that a 1% increase in EVTRW alert is associated with a 0.054% decrease in bus-involved crashes when all else remains constant. Nevertheless, this value is not statistically significant, similar to model 1's results. *QuartCOV*, *Log of passing*, *Presence of bus stops*, *Industrial*, *Residential*, and *Utility* have statistically significant coefficients. Between EPCW and EVTRW systems, EPCW has shown to decrease more crashes considering their coefficients in the models.

Discussion

Comparison of hazardous and non-hazardous areas based on Alert/Grid and Crash/Grid

As the hazardous and non-hazardous locations are developed based on the number and location of crashes, it is expected that number of crashes per grid will be higher in hazardous areas and lower in non-hazardous areas, as shown in **Figure 2.10**. However, it would be interesting to see whether the number of alerts per grid is also higher in hazardous areas or not. **Figure 2.11** shows that the grids falling inside the hazardous areas also have higher alerts/grid than non-hazardous areas. This implies that alerts can indicate high-risk hazardous locations for bus-involved crashes. Nevertheless, it also implies that even though the *Log of Alerts* is negatively associated with the crashes according to the Panel Poisson models, the number of alerts/grid can be high in crash-prone hazardous areas. This indicates that their locations do not entirely determine their association; there can be other factors causing this negative association that need further investigation.

Table 2.1 Number of observations in grids with ≥ 1 alert-based datasets

Dataset based on application type	Both EPCW & EVTRW	EPCW	EVTRW
Total number of grids (Observations)	807	85	782
Total number of unique grids (groups)	334	24	326
Observations per group			
Minimum	1.00	1.00	1.00
Average	2.40	3.50	2.40
Maximum	5.00	5.00	5.00

Table 2.2. Descriptive Statistics of grids with ≥ 1 alert.

Application Type	Both EPCW & EVTRW				EPCW				EVTRW			
	N (number of grids = 807)				N (number of grids = 85)				N (number of grids = 782)			
Variables	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
Dependent (5 yrs. total)												
<i>Total Crash Frequency (5 yrs.total)</i>	0.98	1.72	0.00	10.00	0.91	1.41	0.00	10.00	1.02	1.74	0.00	10.00
Independent (5 months)												
<i>Number of Alerts provided to bus drivers</i>	210.74	792.35	1.00	10449.00	1408.75	1664.97	1.00	7374.00	64.35	401.89	1.00	10449.00
<i>Log of Alerts</i>	2.58	2.07	0.00	9.25	5.38	2.99	0.00	8.91	2.29	1.63	0.00	9.25
Vehicle Kinematics												
<i>SpeedVol</i>	1.43	1.09	0.13	12.87	1.08	0.92	0.19	8.04	1.46	1.10	0.13	12.87
<i>SpeedSd</i>	3.14	1.27	0.28	11.09	3.01	0.78	1.20	4.59	3.16	1.28	0.28	11.09
<i>SpeedMad</i>	2.61	1.22	0.03	10.76	2.54	0.84	0.26	4.08	2.61	1.23	0.03	10.76
<i>QuartCOV</i>	83.69	24.62	7.85	99.92	80.72	31.98	8.58	99.46	84.91	23.12	7.85	99.72
<i>Number of passings</i>	1716.43	1400.08	7.00	8036.00	2588.66	1356.79	15.00	5583.00	1724.93	1405.94	7.00	8036.00
<i>Log of passing</i>	7.02	1.08	1.95	8.99	7.46	1.38	2.71	8.63	7.05	1.03	1.95	8.99
Geographic Information												
<i>Average number of lanes</i>	4.39	1.20	1.00	8.00	3.74	1.39	2.00	8.00	4.42	1.19	1.00	8.00
<i>Average road width (feet)</i>	59.63	12.49	10.00	96.00	58.03	17.38	30.00	96.00	59.84	14.34	10.00	96.00
<i>Total length of roads (feet)</i>	446.33	213.39	8.33	1243.27	446.87	196.29	48.16	869.44	447.09	215.63	8.33	1243.27
<i>Number of bus stops</i>	1.24	1.37	0.00	9.00	1.24	1.10	0.00	4.00	1.27	1.37	0.00	9.00
Std. dev. = Standard Deviation, Min=Minimum, Max=Maximum												

Table 2.3. Descriptive Frequencies of Key Variables.

Variable	Levels	Both EPCW & EVTRW		EPCW		EVTRW	
		N (number of grids = 807)		N (number of grids = 85)		N (number of grids = 782)	
		Freq.	Percent	Freq.	Percent	Freq.	Percent
<i>Presence of bus stop</i>	<i>≥ 1 bus stop</i>	523.00	64.81	58.00	68.24	521	66.62
	<i>No bus stops</i>	284.00	35.19	27.00	31.76	261	33.38
<i>Functional classification of roads</i>	<i>Arterials</i>	416.00	51.55	53.00	62.35	401	51.28
	<i>Collectors</i>	225.00	27.88	15.00	17.65	216	27.62
	<i>Interstate</i>	21.00	2.60	-	-	21	2.69
	<i>Mixed</i>	145.00	17.97	17.00	20.00	144	18.41
<i>Land use</i>	<i>Commercial</i>	405.00	50.19	44.00	51.76	395	50.51
	<i>Green Space</i>	33.00	4.09	15.00	17.65	33	4.22
	<i>Industrial</i>	81.00	10.04	1.00	1.18	81	10.36
	<i>Institutional</i>	105.00	13.01	15.00	17.65	95	12.15
	<i>Residential</i>	149.00	18.46	10	11.76	144	18.41
	<i>Utility</i>	34.00	4.21	-	-	34	4.35

Table 2.4. Regression Models for alerts on grids with ≥ 1 alert.

Dataset (Model)	Both EPCW & EVTRW (Model 1)			EPCW (Model 2)			EVTRW (Model 3)		
Variables	Coef.	z-stat.	AME	Coef.	z-stat.	AME	Coef.	z-stat.	AME
Dependent (5 yrs. total)	Crash Frequency			Crash Frequency			Crash Frequency		
Independent (5 months)									
Constant	-4.425***	-4.700		-9.725***	-2.300		-	-4.480	
							4.226***		
Log of alert	-0.090	-0.660	0.022	-0.239***	-5.280	-0.239	-0.054	-0.380	0.046
Vehicle Kinematics									
QuartCOV	0.015***	2.640	0.015	0.041***	3.270	0.041	0.014**	2.460	0.014
Log of Passing	0.329***	2.710	0.329	0.996*	1.950	0.996	0.309**	2.520	-0.309
Geographic Information									
Presence of bus stops	0.764**	1.950	1.211				0.772*	1.930	1.114
Presence of bus stops*Log of alert	0.173	1.230	0.173				0.149	1.010	
Functional classification of roads									
Arterials (base)									
Collectors				-28.975	-0.000	-28.975			
Mixed				-1.127***	-2.680	-1.127			
Land use									
Commercial (base)									
Green Space	-0.688	-1.190	-0.688	-0.746**	-2.170	-0.746	-0.653	-1.170	-0.653
Industrial	-2.086***	-3.300	-2.086	1.425	0.000	1.425	-	-3.430	-2.149
							2.149***		
Institutional	-0.381	-1.240	-0.381	-27.766	-0.000	-27.766	-0.361	-1.18	-0.361
Residential	-1.491***	-4.750	-1.491	0.378	0.800	0.378	-	-4.760	-1.461
							1.461***		
Utility	-1.571**	-2.080	-1.571				-1.615**	-2.150	-1.615
<i>L</i> alpha			0.735			-15.831			0.695
alpha			2.086			0.000			2.003
Wald χ^2 (Prob > $\chi^2 = 0.000$)			102.610			47.190			100.460
LR test of $\alpha = 0$: $\bar{\chi}^2(01)$			506.460			0.000			438.130
Prob $\geq \bar{\chi}^2$			0.000			0.500			0.000
Log-likelihood at constant (C)			-735.531			-60.964			-731.672
Log-likelihood at convergence (B)			-684.909			-79.799			-682.291
McFadden's ρ^2			0.069			0.236			0.067
df			12			11			12
AIC			1393.817			143.929			1388.581
BIC			1450.137			170.798			1444.524

Table 2.4 Continued.

Dataset (Model)	Both EPCW & EVTRW (Model 1)	EPCW (Model 2)	EVTRW (Model 3)
<i>N</i> (Number of observations)	807	85	782
*** <i>p</i> -value<.01, ** <i>p</i> -value<.05, * <i>p</i> -value<.1			
<i>Coef.</i> = Coefficient, <i>z</i> -stat = z-statistics, <i>AME</i> = Average Marginal Effect			

Bus-involved crash reduction by EPCW

According to bus-involved crash data provided by Federal Motor Carrier Safety Administration (FMCSA) website (46), there have been more than 60,000 bus-involved crashes in the US since 2013 (as shown in **Figure 2.12**). Most of these are Property Damage Only (PDO) crashes (approximately 50000/year on average between 2009-2019), followed by Injury crashes (approximately 14000/year on average between 2009-2019). Bus-involved fatal crashes are comparatively low (244/year on average between 2009-2019).

The association between alerts and bus-involved crashes is negative in all three models shown in **Table 2.4**, but only the association between EPCW alerts and bus-involved crashes is statistically significant. Hence, only the impact of deploying EPCW (V2I) in transit buses has been examined. Crash reduction values are calculated for hypothetical numbers of bus-involved crashes using the coefficient values from model 2. **Figure 2.13** shows the number of bus-involved crashes that EPCW can potentially reduce upon deployment. It is observed that EPCW can reduce 143 bus-involved crashes if the total number of bus-involved crashes is 60,000. This number may appear small, but it is significant if these were severe injuries or fatal crashes. Therefore, the deployment of EPCW systems could be beneficial for road users.

Limitations

There are some limitations of this E-TRP pilot project's dataset. The first issue is the unavailability of vehicle trajectory data for the many unique alerts. Although those alerts are valid, they could not be used, as their GPS locations could not be traced. The second issue is that the alert and BSM files' timestamp has different values of seconds with different decimal places. So, while merging the two datasets, many alerts were lost. The third issue is that some GPS locations fell outside the county boundary and some on the waterbody beside Cuyahoga County. Those alerts also had to be excluded from the main dataset. The fourth issue is a common issue of the absence of driver information in the BSM dataset. Researchers and CV companies can consider resolving these issues to achieve more accurate BSMs and improved CV technologies. Geometric feature data could not be acquired for some of the grids along the trajectory of the connected buses. These grids had to be dropped while conducting analysis, but their inclusion could have provided more insight into the associations. Many of the grids covered more than one land use and selecting a specific land use for such grids was complex. The land use that occupied the majority of a grid was considered the grid's land use. The road network shapefile did not contain all roads in many grids. Therefore, the total length of roads in a grid was manually calculated using ArcGIS for those grids. This made the data collection process time-consuming and tedious. Moreover, the data used in this study is a product of averaging five years of transit bus-involved crashes and five months of BSM data. A short period of volatility measures and BSM alerts are used to examine the correlation with 5-year average bus-involved crashes. However, to make the analysis as valid as possible, the authors used all available data despite the difficulty of processing such large-scale data. Another critical issue is that grids with zero alerts were not analyzed, and their frequency is enormous compared to the number of grids with ≥ 1 alert. The inclusion of those grids could have yielded more insights.

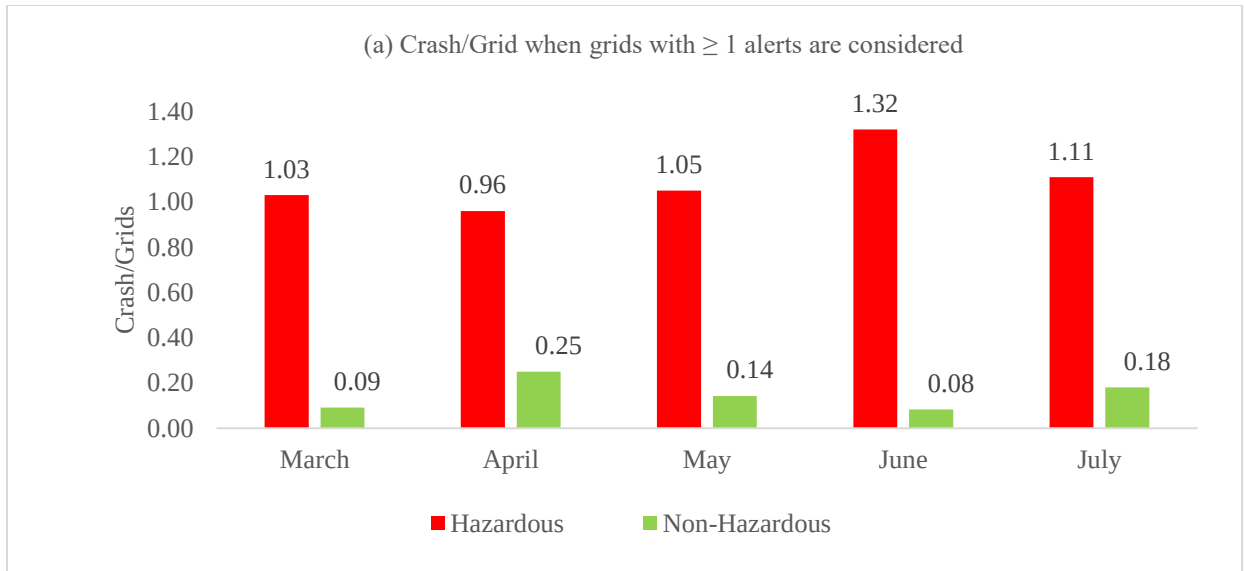


Figure 2.10 Crash/Grid in grids that include ≥ 1 alert using Both EPCW & EVTRW Dataset

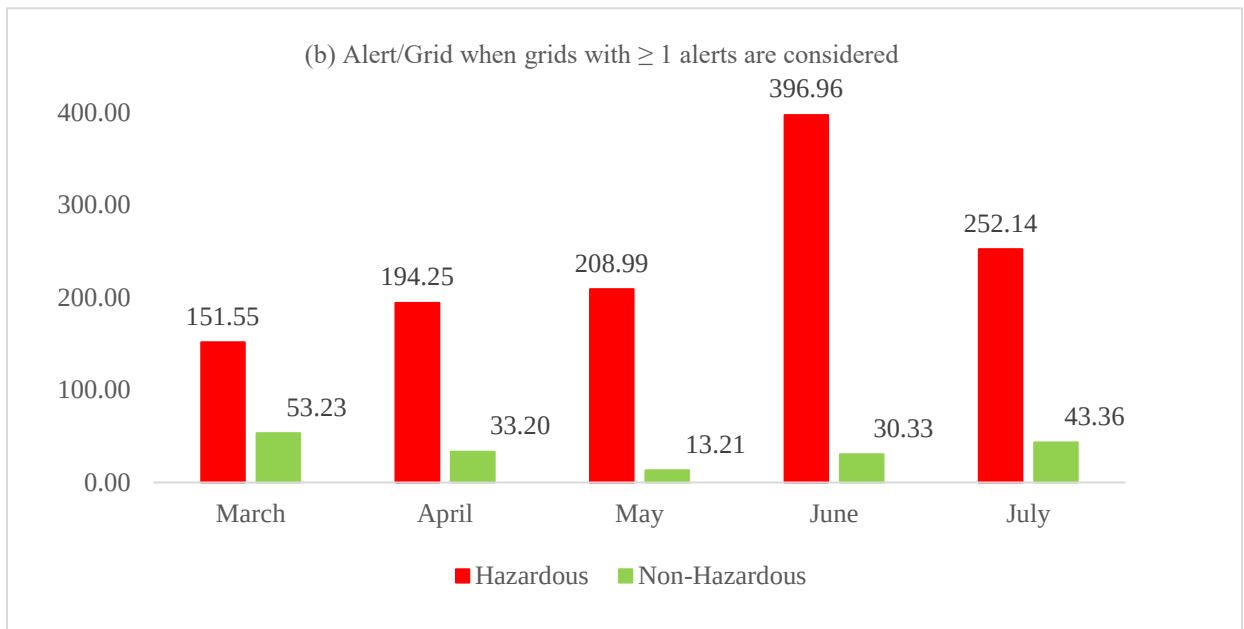


Figure 2.11 Alert/Grid in grids that include ≥ 1 alert using Both EPCW & EVTRW Dataset

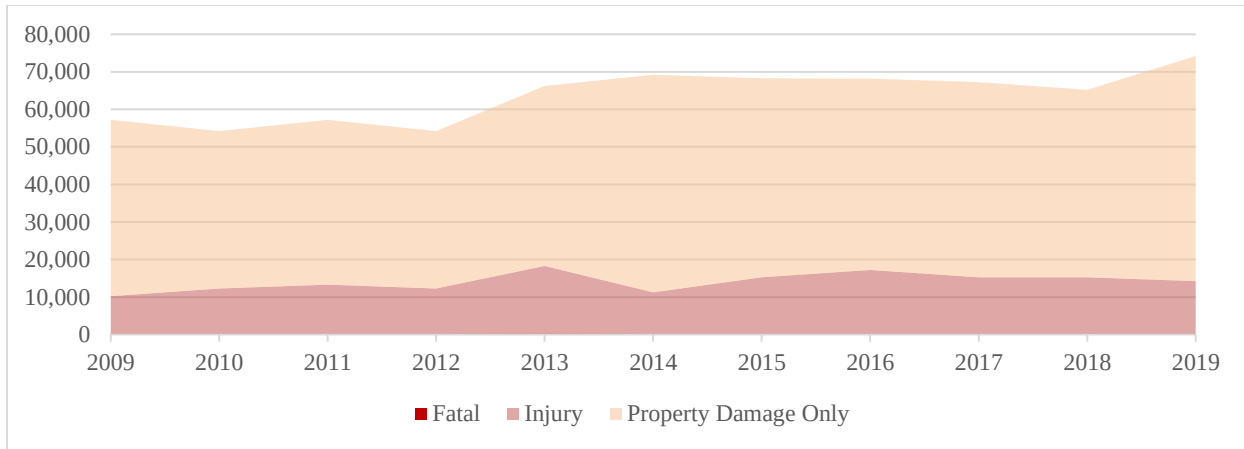


Figure 2.12. Bus-involved crashes from 1999 to 2019 in the US

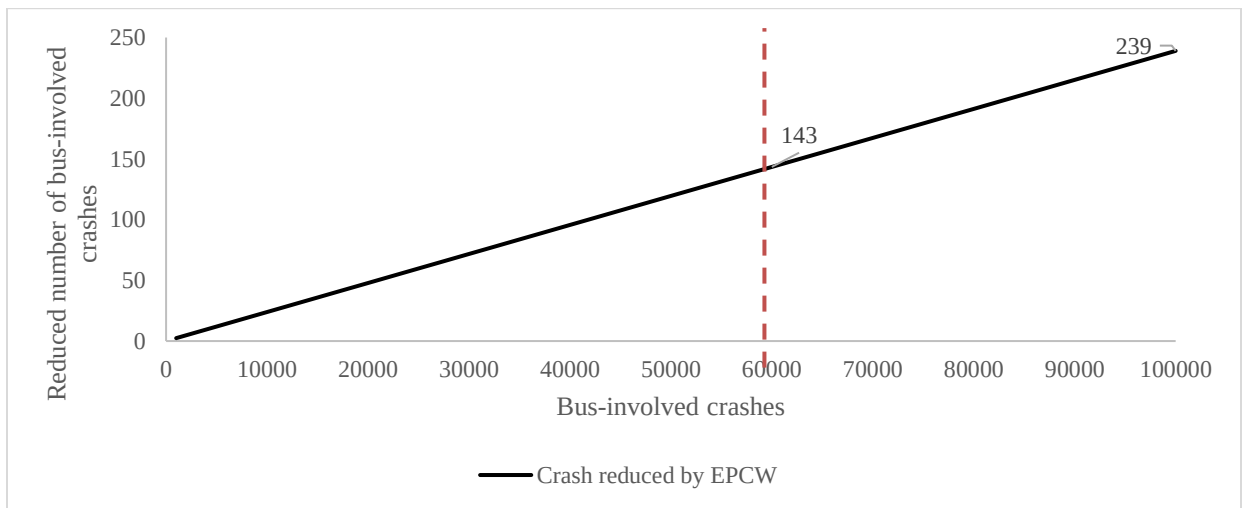


Figure 2.13. Reduction of bus-involved crashes by EPCW

Conclusions

This study has explored the role of smart vehicle and infrastructure technologies in improving public transit safety, especially focusing on detecting pedestrians to enhance safety for all road users in the future. By harnessing the big data from newly available connected buses that provide drivers with alerts, this study uses alerts and BSMs to identify high-risk bus-involved crash locations. The key findings of the paper are as follows:

- Panel Poisson regression models find that bus-involved crash locations are negatively associated with alert locations. In grids containing ≥ 1 alert, a 1% increase in EPCW alerts is associated with a 0.239% decrease in bus-involved crashes when all else remains constant. A negative association was observed between EVTRW alerts and bus-involved crashes, but it was not statistically significant.
- Bus-involved crash locations are also negatively associated with *Mixed* roads and *Green Space* land use. EPCW dataset-based model showed that grids with *Mixed* roads (grid contains more than one type of road) are associated with 1.127 units less bus-involved crashes on average than Arterial road-based grids when all else remains constant. *Green Space* land use-based grids are associated with 0.764 units less bus-involved crashes on average than *Commercial* land use-based grids when all else remains constant.
- The models also suggest that bus-involved crashes are positively associated with bus drivers' volatility, passing, and presence of bus stops. "EPCW" dataset-based model showed that a 1% increase of QuartCOV is associated with a 0.041 unit increase in the average number of bus-involved crashes if all else remains constant.
- "EVTRW" and "Both EPCW and EVTRW" dataset-based models also showed a significant positive association between *Crash Frequency* and *QuartCOV*, *Log of passing*, and *Presence of bus stops*.
- Valid Panel Poisson models indicate that these associations are affected by time along with space.
- Hazardous location identification maps show that alerts help identify high hazardous locations that need the attention of transit agencies and bus drivers.

This research contributes to the knowledge area concerned with smart infrastructure and new transit technologies that enhance safety. The models developed in this research are based on new and emerging data that can transform transit buses' increased safety. The analysis of CV technologies and volatility provides rich insights. Although the models are not generalizable as they are only for bus-involved crashes in a small county, the models can be improved by including more volatility measures and the geographic and socio-economic information of the locations. The next step of this study can be incorporating locations with BSMS but without alerts in the model and analyzing causality. Future research work can apply the model to busy metropolitan cities and rural areas to develop generalizable models for these CV technologies. The results will be valuable for transit agencies to undertake countermeasures to reduce bus-involved crashes. Future endeavors can also investigate the association of crash severity with alerts and BSMs. The validity of these associations can also be studied at different locations, such as near bus stops or at intersections. For now, the key findings of this research can help Cuyahoga county's transit authority become aware of their area's hazardous locations for bus-involved crashes. They can analyze the geographic conditions and design of the roads in these locations to find the underlying

reasons which may be causing these crashes. Based on this research, countermeasures can be deployed to increase transit safety. Before undertaking any significant infrastructural changes, bus drivers could be made aware of these hazardous locations by installing roadside alert signs at the entrances to these locations. Finally, CV technologies are in the test phase. When fully deployed, the effectiveness of BSMs and alerts in identifying high-risk hazardous locations might need to be studied again.

Acknowledgments

The authors would like to thank the Civil, Environmental, Industrial & Systems Engineering (CEE) Information Technology service providers of the University of Tennessee, Knoxville, for their constant IT support and advice.

**CHAPTER 3. EVALUATING PEDESTRIAN SAFETY IN BUS-
PEDESTRIAN INTERACTIONS WITH THE APPLICATION OF
ENHANCED PEDESTRIAN COLLISION WARNING SYSTEMS:
ANALYSIS OF LARGE-SCALE BASIC SAFETY MESSAGE DATA**

Abstract

Intersections are one of the most hazardous locations for vehicles and pedestrians due to the abundance of conflicts. Connected Vehicle (CV) technologies can potentially reduce some of these conflicts. In this regard, a lightly researched area is the use of CV technologies in public transportation. This study explores the safety benefits of an Advanced Driver Assistance Systems called Enhanced Pedestrian Collision Warning (EPCW) by harnessing a large-scale data from a pilot project of 24 connected buses that traversed three intersections in Cleveland, Ohio, US. For each intersection, a novel safety surrogate measure “Mean Time Difference to Intersection (MTDI)” and Alert zone radii are measured. MTDI, which denotes the mean time a CV would take to reach the center of the intersection from the moment of receiving “WarnAlert” traveling at the speed at which “WarnAlert” was received, showed that the EPCW alerts have potential to improve safety at the unsignalized and midblock intersections. Alert zone, which indicates the mean distance (radii) of receiving EPCW alerts near an intersection, will be helpful for drivers to identify pedestrians and drive safely at or near intersections if the radii is large. Analysis showed that the midblock crossing had the largest alert zone radii (169.78 feet), followed by the signalized (142.09 feet) and unsignalized (102.13 feet) intersections. The results are relevant for policymakers who aim to promote safe roadways by using new safety technologies for all users. The paper suggests further investigation of EPCW to improve safety at intersections.

Introduction

Pedestrians are the most vulnerable of road users as they are exposed and are difficult to observe while driving, irrespective of the time of the day (47). According to the National Highway Traffic Safety Administration (NHTSA), “6283 pedestrians were killed in traffic crashes in 2018, which is 17% of the total traffic fatalities of that year. Among those pedestrian fatalities, 17% occurred at intersections (48).” Hence, increasing the safety of pedestrians at intersections is of utmost importance. According to the American Association of State Highway and Transportation Officials (AASHTO), an intersection is a joining or crossing of two or more highways, and the highways that radiate from an intersection are called intersection legs (49). Addressing safety challenges at intersections requires an in-depth understanding of the factors contributing to the crashes like the geometry and traffic signal phasing of that intersection, road type, land use, weather, time, visibility, and vehicle, driver, and other road user errors, etc. (50). The United States Department of Transportation’s (USDOT) Federal Highway Administration (FHWA) has published a manual that provides countermeasures to improve intersection safety (51), but collisions at intersections are still prevailing.

A handful of research (50; 52-54) has been conducted to identify which type of intersection has more scope for safety improvements, but there has not been any concrete affirmation about one. Studies have reported that drivers are 33% more likely to make errors at intersections than at midblock crossings (50) and found more crashes per million conflict at signalized intersections compared to unsignalized intersections (52). Pedestrians are found to avoid most of contingent collisions in unsignalized intersections but not in the signalized ones (53). The safety of signalized intersections was positively associated with design speed, but the safety of unsignalized intersection was negatively associated with design speed (54). Therefore, for the safety of pedestrians from bus-involved pedestrian crashes, it is important to understand which intersection needs more pedestrian safety improvements and where transit bus drivers should be more careful to avoid collision with pedestrians.

With the emergence of connectivity and automation technology in vehicles, overall crash risks might be reduced (55). Connected Vehicle's (CV) Basic Safety Message (BSM) data is state-of-the-art technology that many U.S. cities are testing to examine its potential for helping avoid collisions. According to FHWA, BSMs are the "minimum transmitted data package exchanged between CVs and infrastructures (Vehicle to infrastructure –V2I), which include information about the vehicle's current location, speed, heading, acceleration, braking status, and vehicle size (2)." Some CVs also have Advanced Driver Assistance Systems (ADAS), such as Pedestrian Collision Warning (PCW) system that detects pedestrians and warns drivers about potential pedestrian-involved collisions. BSMs and PCWs from CV pilot projects can be used to study pedestrian crashes at intersections to increase pedestrian safety and contribute to the ongoing experiments of CVs. Therefore, it is important to harness the data from the pilot projects on Connected and Automated Vehicles (CAVs) and explore how these can be beneficial to the drivers and road users at intersections.

Research on pedestrian safety at intersections using BSM data has commenced, but it is hampered by the confidentiality of CV projects. BSM data is not disclosed publicly and the research on it is limited. Among the few publicly available CV pilots, big data from the "Enhanced Transit Safety Retrofit Package (E-TRP)" project is harnessed for this study. The data contains BSMs, and Enhanced Pedestrian in Crossing Warning (EPCW) alerts generated for 24 connected buses traversing three intersections in Cuyahoga County, Cleveland, Ohio, from March-August 2018. The objectives of this study are as follows: 1) Harness data from connected transit buses to understand the vulnerability of pedestrians at intersections; 2) Explore the trend of EPCW alerts at different types of intersections; 3) Identify alert zones within the intersections where bus drivers should be more careful to avoid bus-involved pedestrian crashes.

The scope and potential impact of the work is broad despite a limited size of the dataset (data on only three intersections). The work explores EPCW, an innovative technology that is likely to be adopted in the future. The microscopic BSM and EPCW alerts data are unique and rare which are structured to gain critical knowledge that provides long-term scientific value. A novel safety surrogate measure (SSM) has been developed that future studies can use. The work highlights the benefit of ADAS and Roadside Infrastructure (V2I) for pedestrians.

Literature Review

Text mining on the current literature revealed the following topics discussed. Based on the frequency of appearance, it is discovered that the content of the studies could be categorized as turning in lanes, crash severity, traffic signal, Pedestrian Collision Warning (PCW) application, sight distance and alignment, and time series analysis. A key focus of this study is the PCW system (56). This literature review covers CAVs, BSMs, and SSMs to highlight opportunities for new research on PCW systems to improve pedestrian safety.

Transportation agencies are testing automation and connectivity in various environmental conditions. Connectivity is part of the Intelligent Transportation System (ITS), where BSMs are shared between vehicles and infrastructure to avoid potential crashes. Data on traffic, individual origin and destination, instantaneous driving behavior, maneuvers, and trajectory can be acquired from the connected vehicles (3). Because untreated or raw BSM data are complex to understand, data analytics are used to extract critical information from them (4). For instance, a handful of

studies have developed frameworks to explore BSM data and advance crash investigations at roadways (57) and intersections (4-6; 58; 59) using CAV data.

Many of these CAV based studies used a safety performance measure called volatility and its different versions to represent instantaneous driving behavior (6) at intersections. One percent increase in acceleration/deceleration based volatility has been found to increase crash frequency by 1.25%, while a vehicular jerk based volatility measure has been found to increase crash frequency by 0.35%. (6). Intersections with high variations in the longitudinal movement showed tendency to have higher rear-end crashes, and intersections with lateral volatility have been identified to possess a higher risk of a head-on collision (3). Higher crash frequencies are also found to be associated with time-varying stochastic volatility of vehicle speeds when vehicles pass through intersections (59). Therefore, countermeasures that can improve safety of intersections are necessary and deployment of CAV technologies can be one of them. It was observed that with 50% CAV penetration rate, the number of conflicts at both signalized and unsignalized intersections can decrease substantially (60). In terms of their benefits, CAV technologies like Forward Collision Warning, Cooperative Adaptive Cruise Control, and Cooperative Intersection Collision Avoidance Systems could save \$76 billion per year in the US (61).

Since CAV pilot projects generate BSM data, researchers are developing and applying various safety surrogate measures (SSMs) to evaluate the safety performances of CAVs. SSMs are used to analyze near crash events, which is also effective in safety improvement initiatives as historic crash data are limited. Some of the widely used SSM in CAV studies are time-to-collision (TTC), time exposed time-to-collision (TET), time integral time-to-collision (TIT), rear-end collision risk index (RCRI), time to accident (TA), deceleration rate to avoid the crash (DRAC), maximum speed, and standard deviation of speed (62). There are some other relevant SSMs that can be used in CAV based studies. Such SSMs include time difference to the point of intersection (TDPI), distance between stop position and pedestrian (DSPP), and vehicle-pedestrian speed-distance profile (63). Researchers are developing SSMs for evaluating driver's compliance to various ADAS. For instance, Deceleration-based SSM (DSSM), which reflects driver's perception of collision risks in the car-following situation, showed the potential to enhance the driver's compliance to the ADAS (64).

Notably, a vital topic in the ITS community (65) is pedestrian protection. Since its innovation, the ADAS has been playing an important role in the ITS's mission of protecting pedestrians, and for this reason, it is researched often (66; 67). A noteworthy ADAS is the PCW system. This technology is initially being tested either in driving simulators or test environments like parking lots. Tests in real world settings found that the PCW system can improve the reaction time of transit bus drivers by 26% (68) and can reduce vehicle-pedestrian crash risks by 70% (69). Tests in driving simulator-based scenarios found that PCWS has the potential to avoid a pedestrian collision (70) and has the potential to increase drivers' levels of perception, vigilance, and anticipation of pedestrians if PCW system is combined with Augmented Reality (AR) that contains Head-UP Display (HUD) interfaces (56). Some limitations of the system include possessing the risk of rear-end collision in cases where drivers brake aggressively (70), producing false alarms when there is no potential risk of collision (71) and failing to generate alerts in some risky scenarios. The risky scenarios include situations where pedestrians enter from two sides of the vision field; pedestrians cross in the right-hand half of the driver's vision when the subject vehicle is about to finish a left-turning maneuver; pedestrians are obstructed by buildings or vehicles; and when the mean speed of the subject vehicle is below 6 km/h, between 18 km/h and 24 km/h and

beyond 30 km/h (72). As the PCW system is still in the development stage it has scope for improvements. Besides testing the technology in regular cars, researchers are also working concurrently to improve pedestrian collision prediction algorithms in mobile phones, autonomous vehicles (73), and connected vehicles. For example, a study on Smartphone-based Distracted Pedestrian Localization developed a system utilizing low energy (BLE) Bluetooth beacons to track distracted pedestrians within 26.25 feet (eight meters) of the main beacon as alert zone near intersections(74).

While many studies cover CAV data and performance, no study (to the authors' knowledge) has attempted to identify vehicle alert zones based on the location of receiving PCW alerts to prevent collisions. Recently, FTA researched a pilot project undertaken in an urban bus operational safety platform in Cuyahoga county (14). This paper will expand on the data of this research. Since transit bus and pedestrian conflicts often occur at intersections, and it is unknown which type of intersection needs more safety improvement to prevent potential collision, it is worthwhile to examine intersections microscopically to understand them at a deeper level. Existing literature has not utilized BSM and alerts data for studying diverse intersection users. Hence, this approach is unique and is expected to reveal new information that may help enhance pedestrian safety.

Methodology

The framework of this study is developed based on the objectives. RStudio and ArcGIS software are used to conduct all the analyses.

Data

The data is from the “Enhanced Transit Safety Retrofit Package (E-TRP)” project and it is collected from the USDOT’s ITE Joint Program Office’s Website (29). The project is from Cuyahoga County, Cleveland, Ohio. From February to August 2018, 24 buses of Greater Cleveland Regional Transit Authority (GCRTA) collected 239453 BSM alerts. Cutting-edge technological equipment is used in a transit environment to test the impact of connectivity on pedestrian and transit safety. There are three key technologies used for this unique dataset. First, a Dedicated Short-Range Communications (DSRC) system is used for both vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication. Second, a High-precision Global Navigation Satellite System (GNSS) is used for tracking the connected vehicles. Third, Forward-Looking Infrared (FLIR) cameras are used for detecting pedestrians. E-TRP features are installed on 24 transit buses and three intersections. The installations on the transit buses are called on-board transit vehicle-based subsystems, and the installations at the pedestrian crossings are called infrastructure-based subsystems (14).

Two CV applications are used: Enhanced Pedestrian in Crossing Warning (EPCW) and Vehicle Turning Right in Front of Bus Warning (EVTRW). As this study is related to pedestrian safety and bus-pedestrian collisions, only EPCW application-based alerts will be applied. EPCW is a V2I application that alerts the driver of the connected bus about pedestrians. This application generates three app messages: 1)“Info” - When the bus enters, is inside, or exits the geo-fenced area. 2)“InformAlert”-When a pedestrian is on the curb in the potential path of the connected bus. 3)“WarnAlert” – When pedestrians are in the crosswalk in the path of the bus.

EPCW application also obtains position (latitude/longitude), timestamp, heading, speed, and elevation information of the connected buses. (14).” The subsystems and EPCW “WarnAlert” are shown in **Figure 3.1**.

Alerts are also classified under event types and are stored under the “EventCodeType” variable. An event is an incident for which an alert is either stored or informed to the driver. The events under AppMessage Type “Info” are: “NOEVENTID,” “ApplicationActive,” “ApplicationInactive,” “EnteredArea,” “ExitedArea,” “DetectedArea,” “EventComplete,” and “InArea.” The EventCodeType under “InformAlert” and “WarnAlert” are: “NearsidePed,” “FarsidePed,” “LeftsidePed,” and “RightsidePed.” EventCodeType indicates the type of events generated, which include no event generated, application active, application inactive, bus entered area, bus exited area, an area is detected, direction of where the pedestrians are (nearside, farside, leftside, rightside), pedestrian in a road, vehicle is ahead, vehicle is ahead on left side, vehicle is ahead on right side, vehicle is on left side, vehicle is on right side, vehicle is behind on left side, vehicle is behind on right side, vehicle is behind, event is complete, and vehicle is inside the area. The AppMessage Type indicates the type of message like Info, InformAlert, and WarnAlert.

The package has seven types of datasets, among which only alerts and BSM datasets are used in this study.

Alert Dataset

This dataset contains the EventCode Type, AppMessage Type, and timestamp of those alerts, alert id, device id, and vehicle id. Each transit bus has a unique vehicle id. The devices used in the buses to send or receive messages have unique device ids.

BSM Dataset

This log file contains the Global Positioning System (GPS) locations of the connected buses, timestamp, speed, and heading at a rate of 5 Hz. Speed is in kilometers per hour (KPH), later converted to miles per hour (mph). February’s data is excluded from this paper intentionally, as this is only the trial period to install and test the devices.

Data preparation

Alert and BSM datasets needed to be merged to get the geographic location of each alert. Both datasets had vehicle id and timestamp as common variables, but the alert dataset’s timestamp had a rounded value for seconds. The BSM dataset had a vehicle trajectory every 0.2 seconds. Hence, it was never possible to get the exact fraction of seconds when the alerts occurred from this dataset. For the convenience of merging, the BSM dataset’s values for seconds were rounded up. By doing this, every second had five observations. While merging, from every five observations for each second from the BSM dataset, one observation was selected randomly and merged with the alert dataset’s observation of the same time and vehicle id. After merging, it was discovered that the BSM dataset was missing observations for some times when the alerts occurred. Therefore, the merged file missed some alerts.

The device to generate and store the alerts could have failed to operate properly at that time. The number of alerts after merging the two datasets is shown in **Figure 3.2**, which illustrates that not many alerts were lost due to this matter.

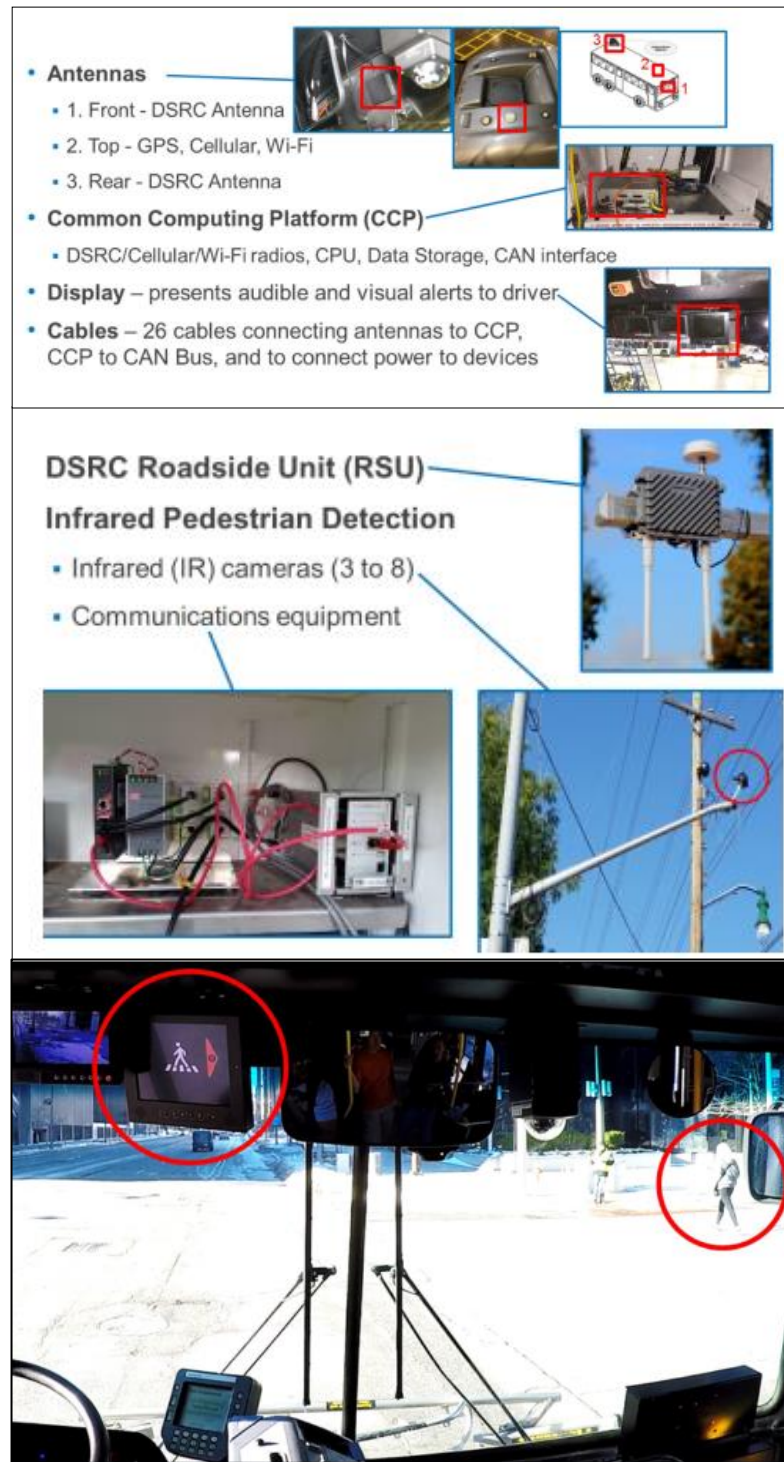


Figure 3.1. On-board Subsystem (top), Infrastructure Subsystem (middle), and EPCW “WarnAlert” (bottom)

Source: (14)

The successfully merged alerts are taken as the final dataset. After error-checking through descriptive statistics, the data seemed reasonable. No outliers are found. As the data came from credible sources such as USDOT and Ohio DPS websites, the data is valid and reliable.

Contingency tables

A contingency table shows the counts from various combinations of categories for a cross-classification of two or more categorical variables. If one categorical variable is denoted as X and another one is denoted as Y, and their categories are denoted as I and J, respectively, then the rectangular table of I rows and J columns is a contingency table, which has cells that display the IJ possible combination of outcomes (75; 76). For this study, two categorical variables are considered: type of EventCodes and type of intersections, which are denoted as X and Y, respectively. The number of categories in X and Y are 14 and 3, respectively. The contingency table for these two categorical variables will be 14 rows by 3 columns.

If the null hypothesis (H_0) considers certain fixed values (Π_{ij}) as cell probabilities, it will mean that the categorical variables are not associated. When the sample size is n, and the cell count is n_{ij} , the expected frequencies will be $\mu_{ij} = n\Pi_{ij}$. To test the null hypothesis, the Pearson chi-square statistic (χ^2) is calculated, and its formula is as follows:

$$\chi^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \quad (1)$$

Here, larger χ^2 indicates more contradiction to H_0 . The P-value represents the probability that the chi-squared right tail value is above the observed χ^2 (75).

Alert Zone identification

One limitation of GPS data is that the reported locations are not always accurate. Using the alerts that fell outside the road boundary to calculate the average distance of alerts from the center of intersections would have given erroneous results. This matter is resolved by shifting these alerts inside of the road. The “Near” tool from GIS is used to relocate the alerts. A spatial join tool is used to measure the distance of each type of pedestrian-based alert from the center of the intersection. The projection system of the Intersection and Alerts used in GIS is “NAD 1983 (2011) StatePlane Ohio North FIPS 3401 (Meters),” which gave the measured distances in meters. Later the distances are converted from meters to feet for convenience. The average distance of all types of pedestrian alerts is considered the radius of the alert zones.

Mean Time Difference to Intersection

A unique safety surrogate measure (SSM) is developed to examine the performance of the EPCW system. This SSM is named “Mean Time Difference to Intersection” (MTDI) which denotes the mean time a connected bus would take to reach the intersection from the moment of receiving “WarnAlert” at an intersection, considering that the bus would be moving at the speed at which “WarnAlert” was received. Only “WarnAlerts” are considered, as these are generated when pedestrians are at the crosswalks. The equation to calculate MTDI is as follows:

$$\text{MTDI} = \frac{T}{F} \quad (2)$$

Here,

T = time to reach intersection from the moment of receiving the “WarnAlert”

F = frequency of “WarnAlerts” in the intersection

Shorter values of MTDI will be more dangerous and unsafe for pedestrians than longer MTDIs. If the time taken for buses to reach the intersection increases, pedestrians will have more time to cross the intersection.

Study Area

The study area consists of three intersections in downtown Cleveland, Ohio. **Figure 3.3** shows the location of the intersections in Google Maps, along with their geometric features. Alerts and BSMs have been collected for only 100-meter (328.08 feet) buffer zones surrounding these intersections. All three intersections are located approximately within a half-mile distance from each other.

The signalized intersection is an intersection of Superior Avenue, which is a six-lane principal arterial, and East 12th St, which is a two-lane major collector. The width of Superior Avenue is 76 ft., its crosswalks are marked with white lines, and left turns are restricted. It has a bus-only lane going in the east-west direction. The East 12th St has lanes for left turns and red-painted crosswalks.

The unsignalized intersection is an intersection of East Roadway and Rockwell Avenue, both of which are 2-lane principal arterials. Left turns are not restricted on both roads. Both are one-way roads, but the East roadway has a bus-only lane. The crosswalks are marked with white hatched lines.

The mid-block crossing is in the middle of the block of Euclid Road, which is a minor arterial of four lanes, and East. 19th Street. Euclid Avenue is a two-way road, and it has two bus-only lanes going in opposite directions on the south side of the median.

Results

Alerts located inside a 100-meter (328.08 feet) buffer area around the three intersections are considered first (124,060 alerts). Next, among all types of applications, only alerts that fell under the EPCW application (122,260) are included in this study.

Frequency Distribution of Alerts

Categorical variables used for the analysis are types of intersections, event codes, and app messages. The frequency distribution of alerts under these categories is shown in **Table 3.1**. Most alerts are generated at the signalized intersection, followed by the unsignalized intersection and the midblock crossing. This could be because more people were crossing roads at the signalized intersections compared to the other two intersections. Only pedestrian-related alerts are “FarsidePed,” “LeftsidePed,” “NearsidePed,” and “RightsidePed.” Most pedestrian-related alerts were “WarnAlert,” followed by “InformAlert,” and none were “Info” alerts. Conversely, only information type alerts were “EnteredArea,” “EventComplete,” “ExitedArea,” and “InArea.” These alerts did not have any “InformAlert” or “WarnAlert” type alerts.

Figure 3.4 (a) shows that in the entire dataset, 76% of alerts are “Info” alerts, which are irrelevant to the research. Only 15% of the total alerts are “WarnAlerts” and valuable for the analysis. **Figure 3.4** also shows the percentage of alerts under different “AppMessageTypes.” It is clear that “Info”

message types are dominant in all three intersections. “WarnAlerts” are the least dominant and are the lowest at the Midblock crossing (3% of all alerts are at Midblock). Signalized intersection received the highest percentage (20%) of “WarnAlert.” The ratio of “WarnAlert” vs. the other two alerts is lowest at the Midblock and highest at the signalized intersection.

It is noteworthy that the percentages of “WarnAlert” are very similar between the signalized (20%) and unsignalized intersection (19%). The ratios of “WarnAlert” vs. other alerts are also close for these two intersections (25.57% and 23.39%). The percentage of “WarnAlert” is low in the midblock crossing (3%) and hence, the ratio of “WarnAlert” vs. other alerts (2.81%) is also low for this midblock crossing. The differences in exposure of bus activity and pedestrian activity likely is responsible for differences in alerts-as would be geometry, surrounding land use and presence of other vehicular traffic. The three intersections represent three unique situations, and they are not comparable. Therefore, each of them is examined with equal importance in this study.

The dataset contained other relevant variables that are considered for analysis. For instance, the speed at the time of the EPCW alert generation (N=29813) ranged from 0 to 10.48 mph. This indicates that some buses were stopped when they received the alert. Lower speed values indicate that the buses were either decelerating or accelerating at that moment. The distance between the bus and intersections ranged from 0.36 to 327 feet, indicating that alerts were received from 327 feet away from the intersection.

Comparison of three intersections

Table 3.2 shows that among all pedestrian-based alerts (29813 alerts) at the three intersections, “NearsidePed” alerts are mostly observed, followed by “LeftsidePed,” “FarsidePed,” and “RightsidePed.” The signalized intersection received the highest number of pedestrian-based alerts among the three intersections. It mostly received “NearsidePed” alerts, followed by “LeftsidePed,” “FarsidePed,” and “RightsidePed.” The unsignalized intersection received the second-highest number of alerts. Like the signalized intersection, it also received mostly the “NearsidePed” alerts. The unsignalized intersection received the lowest “FarsidePed” alert among the three intersections. The Midblock crossing received the lowest number of alerts among the three intersections. It is noticeable that Midblock did not receive any “LeftsidePed” or “RightsidePed” alerts in the entire study period, whereas the bus stop is in the middle lane of the road and pedestrians come to the bus stop from both sides. It can be deduced that pedestrians were not crossing the road from the left or right when the connected buses came to the midblock crossing. However, some “FarsidePed” and “NearsidePed” alerts were received here.

It is also critical to know if there is any statistically significant association between the types of intersections and types of event codes. Data for only the EPCW application type were considered to evaluate the association between categories of intersections and event codes, as shown in **Table 3.2**. The null hypothesis was that there is no correlation between the type of intersections and event codes. The calculated Chi-square value for this dataset was 6672, which is larger than the critical Chi-square value for degrees of freedom (df) = 6. The p-value is also less than 0.01. Hence, the null hypothesis was rejected, and it was concluded that there is a significant relation between “EventCodeTypes” and intersections.

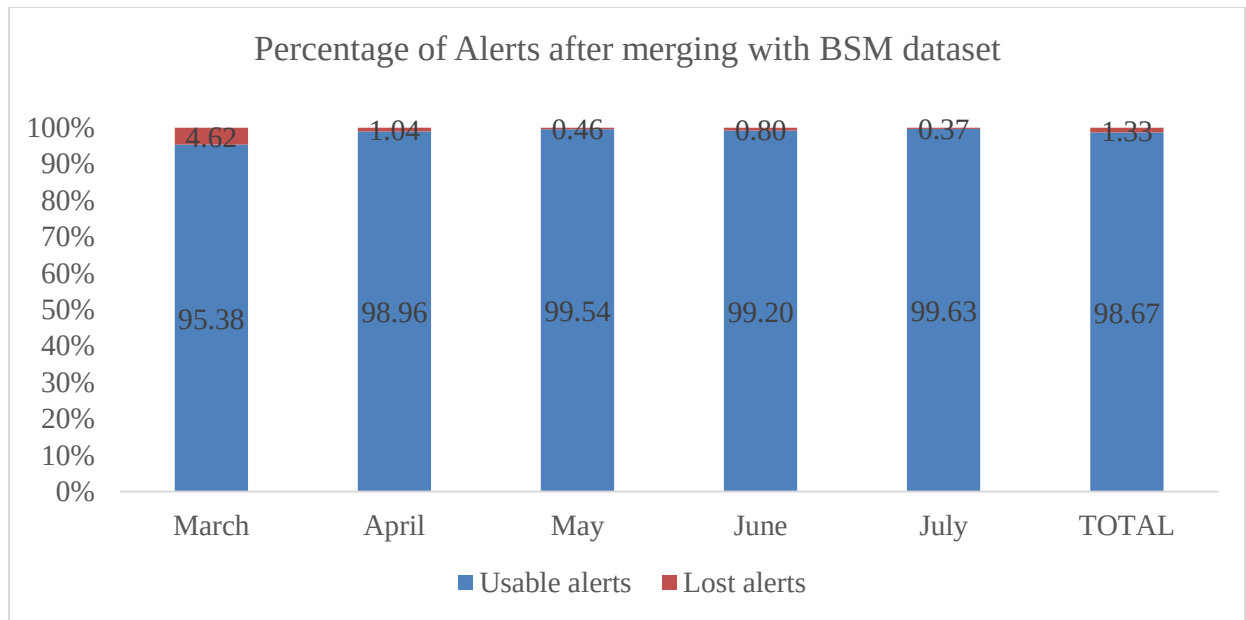


Figure 3.2. Percentage of alerts after merging Alert and BSM datasets

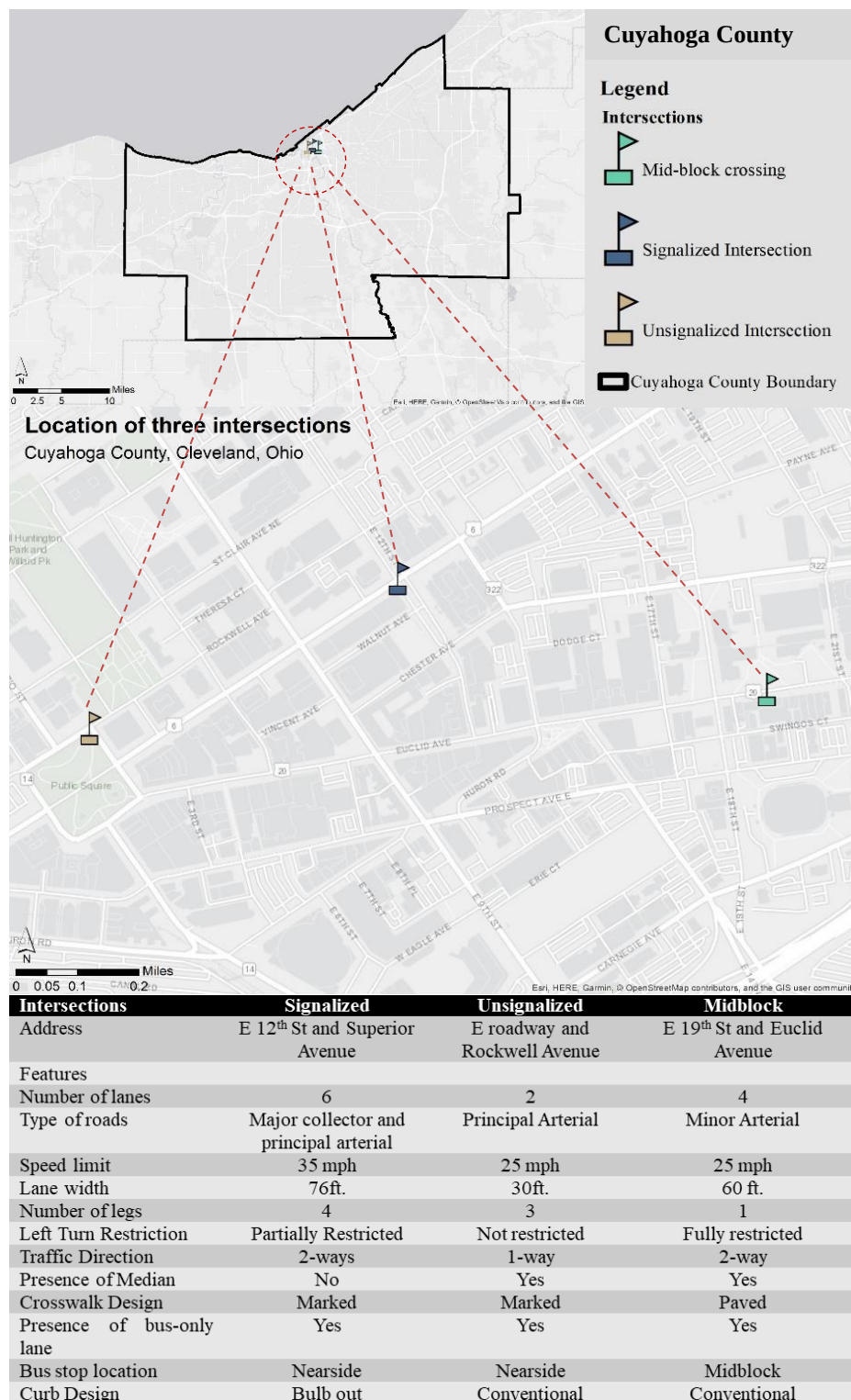


Figure 3.3. Three selected intersections in Cleveland, Ohio

Source: Google Maps and (77)

Table 3.1. Frequency of Alerts by “EventCodeType” and “AppMessageType”

EventCodeType	AppMessage	Intersection			Total Alerts
		Signalized	Unsignalized	Midblock	
EnteredArea	Info	5737 (11%)	5104 (14%)	9624 (27%)	20465 (17%)
	InformAlert	0	0	0	0
	WarnAlert	0	0	0	0
EventComplete	Info	17409 (34%)	11281 (31%)	3734 (10%)	32424 (27%)
	InformAlert	0	0	0	0
	WarnAlert	0	0	0	0
ExitedArea	Info	5930 (12%)	5097 (14%)	9706 (27%)	20733 (17%)
	InformAlert	0	0	0	0
	WarnAlert	0	0	0	0
InArea	Info	5506 (11%)	4160 (12%)	9159 (26%)	18825 (15%)
	InformAlert	0	0	0	0
	WarnAlert	0	0	0	0
FarsidePed	Info	0	0	0	0
	InformAlert	1385 (3%)	33 (0%)	1283 (4%)	2701 (2%)
	WarnAlert	1431 (3%)	44 (0%)	381 (1%)	1856 (2%)
LeftsidePed	Info	0	0	0	0
	InformAlert	716 (1%)	1982 (6%)	0 (0%)	2698 (2%)
	WarnAlert	2851 (6%)	2503 (7%)	0 (0%)	5354 (4%)
NearsidePed	Info	0	0	0	0
	InformAlert	2844 (6%)	1165 (3%)	1238 (3%)	5247 (4%)
	WarnAlert	4059 (8%)	3432 (10%)	597 (2%)	8088 (7%)
RightsidePed	Info	0	0	0	0
	InformAlert	723 (1%)	347 (1%)	0 (0%)	1070 (1%)
	WarnAlert	1956 (4%)	843 (2%)	0 (0%)	2799 (2%)
Total		50547 (100%)	35991 (100%)	35722 (100%)	122260 (100%)

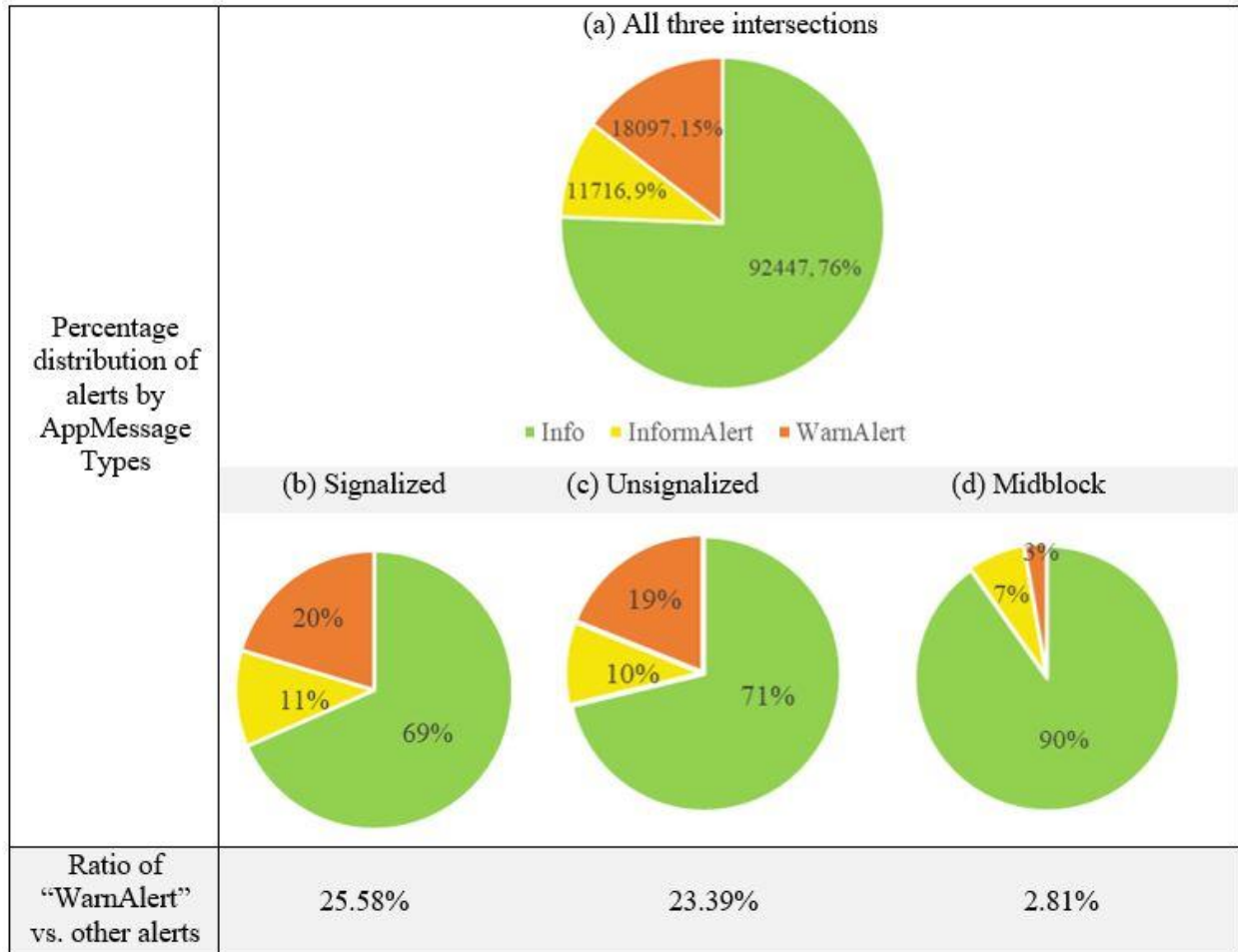


Figure 3.4. Percentage of “Info, “InformAlert” and “WarnAlert” at (a) all three intersections, (b) Signalized, (c) Unsignalized, and (d) Midblock crossing

Table 3.2. Contingency Table of Alerts under all “AppMessageType,” only EPCW Application, and Pedestrian Based “EventCodeType”

			Intersection			Total
			Signalized	Unsignalized	Midblock	
EventCodeType	FarsidePed	Count	2816	77	1664	4557
		% Within Intersection	18%	1%	48%	15%
	LeftsidePed	Count	3567	4485	0	8052
		% Within Intersection	22%	43%	0%	27%
	NearsidePed	Count	6903	4597	1835	13335
		% Within Intersection	43%	44%	52%	45%
	RightsidePed	Count	2679	1190	0	3869
		% Within Intersection	17%	12%	0%	13%
	Total	Count	15965	10349	3499	29813
		% Within Intersection	100%	100%	100%	100%
Pearson Chi-Square: 6672 ^a , df = 6, Asymptotic Significance (2-sided): 0.000						
Likelihood Ratio: 8148.446, df = 6, Asymptotic Significance (2-sided): 0.000						
N= 29813						
a. an expected count of <5 was seen in 0 cells (0.0%). 454.08 is the expected minimum count for cells						

Identification of Alert Zones for buses

According to the Transit Capacity and Quality of Service Manual (TCQSM 3rd Ed.), bus stops are categorized into three types: farside, midblock, and nearside, as shown by **Figure 3.5**. The three intersections selected for this study and the alerts received by connected buses are discussed further in the next section.

Alerts at the Signalized Intersection

The connected buses received all pedestrian-based alerts at the signalized intersection, as shown in **Figure 3.6**. This was the most complex intersection. It had three nearside bus stops, traffic running both ways, and it was a junction of four north-south direction lanes and two east-west direction lanes. All the alerts were generated in all four directions. Most of the alerts were along Superior Avenue (east-west). The “NearsidePed” alerts were generated when pedestrians were on the crosswalks of nearside stops. Logically, the number of these alerts was the highest at this intersection. The “LeftsidePed” alerts were more prominent at the southwest corner of this intersection. This might be because there were more pedestrians along East 12th Street and on the crosswalk across East 12th Street going from southwest to southeast. Another reason for getting this “LeftsidePed” alert could be the left turn lane, which exposes pedestrians to left-turning vehicles. The “RightsidePed” alert was the lowest in number at this intersection. Probably, not many pedestrians were walking across East 12th Street from the east to the west side through the crosswalk. The average distance between the CenterPoint of this intersection and the “FarsidePed” alerts, the “NearsidePed” alerts, the “Leftsideped” alerts, and the “RightsidePed” alerts was 135.66 feet, 168.57 feet, 101.94 feet, and 134.19 feet, respectively. So, it can be said that the alert zone for this signalized intersection is situated within a radius of 142.09 feet from the center of this intersection.

Alerts at the Unsignalized Intersection

Similar to the signalized intersection, the connected buses also received all the pedestrian-based alerts at the unsignalized intersection, as shown in **Figure 3.7**. This intersection was a T-intersection with one-way roads and a nearside bus stop. “FarsidePed” alerts were generated when the buses were passing through Rockwell Avenue from east to west, but these alerts were not high in number. The majority of the “NearsidePed” alerts were generated along the East roadway when the buses were approaching the intersection. Many pedestrians may have been on the crosswalk of Rockwell Avenue that connects the public square and Key Tower.

The “Leftsideped” alerts were likely generated when pedestrians were along the west side of the East roadway or on the crosswalk of the East roadway connecting the public square and the courthouse. The “RightsidePed” alerts were likely generated when pedestrians were along the east side of the East roadway or on the crosswalk of the East roadway connecting the public square and the courthouse. The most numerous alerts were “NearsidePed” and “LeftsidePed.” So, from the magnitude of the frequency, it can be inferred that at this unsignalized intersection, pedestrians on the near side, left side, and right side are more at risk from upcoming buses than those on the far side. The average distance between the center point of this intersection and “FarsidePed” alerts, the “NearsidePed” alerts, the “LeftsidePed” alerts, and the “RightsidePed” alerts was 97.90 feet, 122.64 feet, 84.48 feet, and 89.86 feet, respectively. So, it can be said that the alert zone for this unsignalized intersection is situated within a radius of 102.13 feet from the center of this intersection.

Alerts at the Midblock Crossing

The midblock crossing had one midblock bus stop. Only “FarsidePed” and “NearsidePed” alerts were received by the connected buses at this midblock crossing, as shown in **Figure 3.8**. The “FarsidePed” alerts were received more towards the east as the buses were approaching the far side of the following block. The “NearsidePed” alerts were received more towards the west as buses were approaching another nearside bus stop. It appears that in this midblock intersection, there were more pedestrians in the farside and nearside areas for whom these alerts were generated compared to the left and right sides. It can be inferred from this data that when a bus stop is in the middle of the block, pedestrians at the crosswalks or curbs of nearside or farside bus stops would be at the risk of collision. The average distance between the midblock crossing and “FarsidePed” alerts was 85.20 feet, and between the midblock crossing and “NearsidePed,” it was 246.49 feet. So, it can be said that the alert zone for this midblock crossing is situated within a radius of 169.78 feet from the center of this intersection.

As this study does not have a large enough sample size to conclude these claims in general for all signalized, unsignalized, or midblock intersections, further study is needed.

Discussion

Performance of EPCW

To understand the performance of EPCW, pedestrian crash frequencies were observed. But the frequency of pedestrian crashes for three intersections was negligible from 2014-2018. The research report on the E-TRP pilot project found that over the six-month study period, connected bus drivers’ response to PCW alerts increased by 16%, and average drivers’ reaction time to the PCW warning situation decreased by 18% (14). These SSMs indicated that connectivity can help bus drivers to react more quickly over time and increase pedestrian safety. In this study, a unique SSM called MTDI was developed to quantify the efficacy of EPCWs at different intersections. **Figure 3.9** shows the MTDI values for three intersections for six months and their trendlines. The trendline for signalized intersection is downward sloping, indicating that MTDI decreased over the period. It can be inferred that EPCW alerts were not as effective as expected at this intersection. MTDI at the unsignalized intersection showed an upward sloping trendline, indicating that the mean time difference of connected buses to reach this intersection increased with time. Hence, EPCW alerts were effective at the unsignalized intersection. MTDI for midblock crossing excludes August, as it had only three “WarnAlerts.” From March to July, MTDI at midblock crossing also had an increasing slope, which indicated that EPCW was also effective at the midblock intersection. Considering all SSMs and the dataset, it was evident that EPCW has the potential to improve pedestrian safety at intersections. With connectivity in transit, the unsignalized intersection and the midblock crossing intersection of the study area had higher potential for improvement, based on analysis of the data. However, this is the observation from the data, which can change based on further evidence. The area needs further investigation to understand which intersection has more potential for improvement than the others with CV technologies.

Land Use

Land use around the three intersections is shown in **Figure 3.10**. All three intersections are located downtown, which is mainly a commercial area. The area also has some plots that are used for green space, industrial, institutional, residential, and utility purposes. The signalized intersection is

surrounded by three commercial buildings (various restaurants, offices, and banks) and one residential building (commercial spaces and some apartments). The unsignalized intersection is surrounded by a green space (Public Square), one institutional building (Court House), and a commercial building. The midblock crossing is surrounded by institutional buildings (Marshall College of Law) and commercial buildings. Apparently, commercial areas remain busy during office hours on weekdays and remain quiet during the weekends. It is predictable that there will be more pedestrians at these intersections during the rush hours and office hours on the weekdays compared to the other times of the days and weekends. Therefore, the frequency of alerts at these intersections is also expected to follow a peaking pattern.

Temporal Distribution of EPCW

The time series analysis has been conducted to see underlying patterns of how the alerts are distributed over time, as shown in **Figure 3.11**. For all three intersections, higher frequency (100-200+) of alerts is mainly observed from 10 am to 11 pm and some lower frequency (<100) of alerts are observed between 12 am to 9 am. The signalized intersection has more alerts between 12 am – 3 am compared to the other two intersections. Overall, alerts are negligible to low on weekends and moderate to high on weekdays for all intersections in all months. Some of the highest frequency of alerts (around 200s) has been observed on the Fridays of June and 1st Monday and last Wednesday of July at the signalized intersection. Overall, the temporal distribution of the alerts resonates with the busy hours of the surrounding land uses.

Notably, the signalized intersection has no alerts on most days in March and April because the project's connected buses did not take that route on those days. Similarly, the connected buses also had fewer operational days in March and April compared to the other three months along the route of the unsignalized intersection. Although the alert frequency is lesser in the midblock crossing compared to the other two intersections, the midblock intersection has a similar pattern of alert distribution for all five months.

Limitations

This study could only analyze the alerts through descriptive statistics and some spatial analyses because of the few sample intersections. Robust statistical models could not be run to show the associations between types of alerts and characteristics of intersections. As the number of bus-involved pedestrian crashes was not significantly high in the study period, it was irrational to draw any conclusion from them. Utilizing GPS data for spatial analysis at a microscopic level held the possibility of error in the calculation as GPS locations do not always fall in the correct position.

On-street Bus Stop Locations

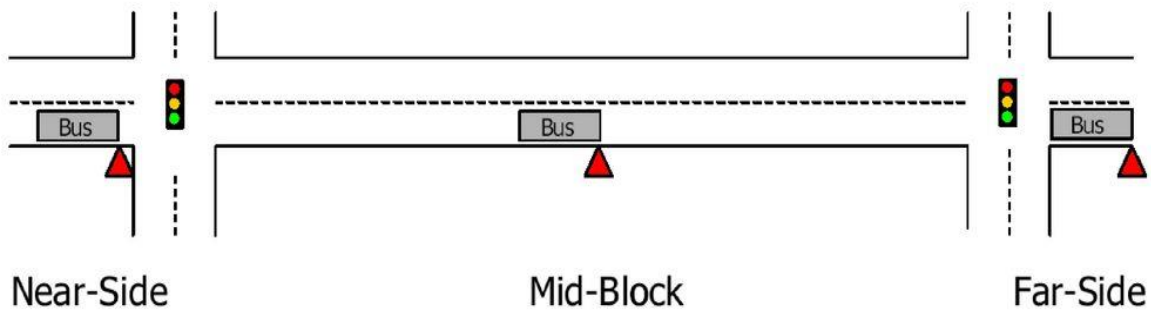


Figure 3.5. Categories of bus stops based on their curbside location

Source: TCQSM 3rd Ed. (78)

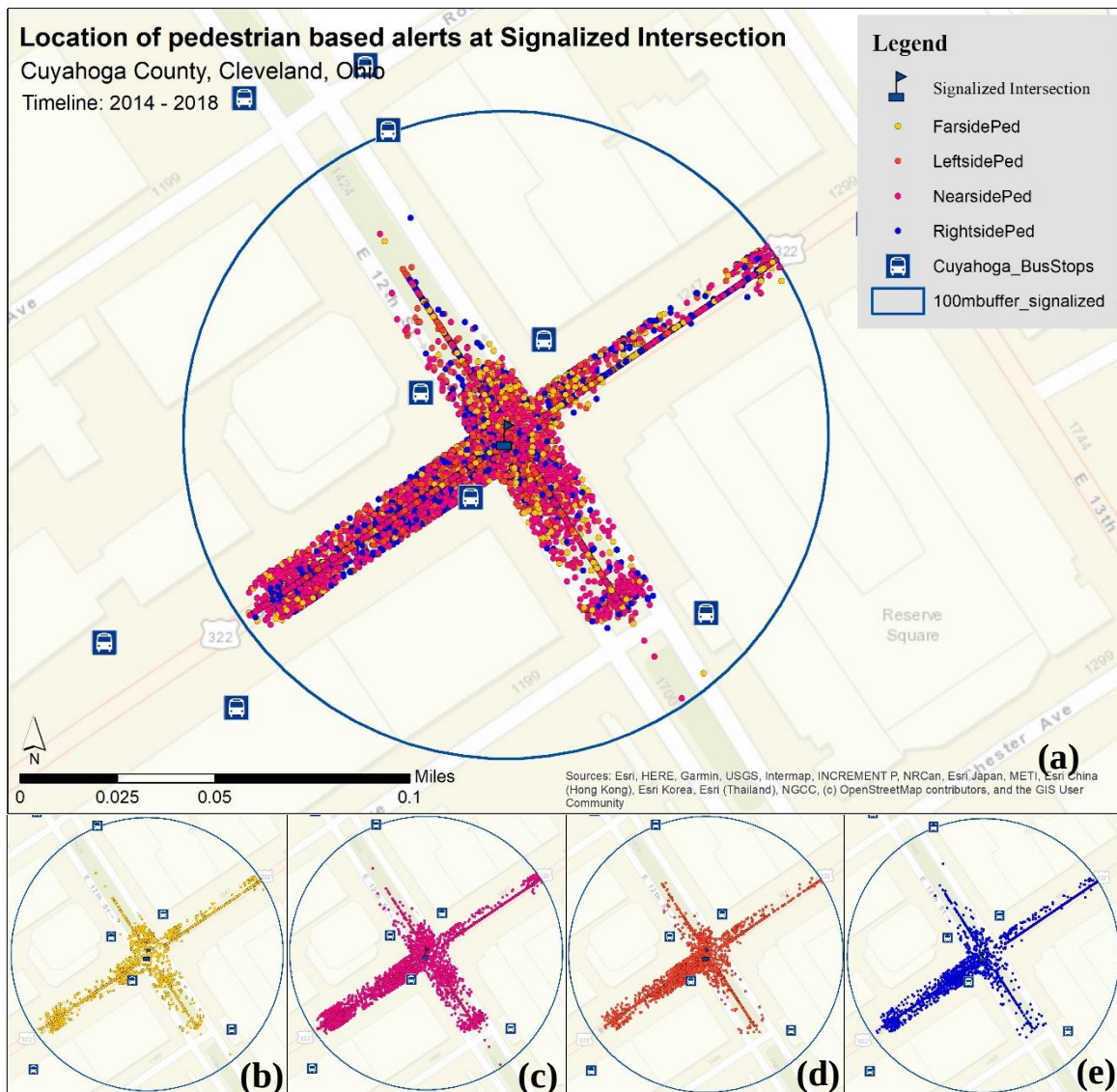


Figure 3.6. (a) All, (b) “FarsidePed,” (c) “NearsidePed,” (d) “LeftsidePed,” and (e) “RightsidePed” alerts at Signalized Intersection

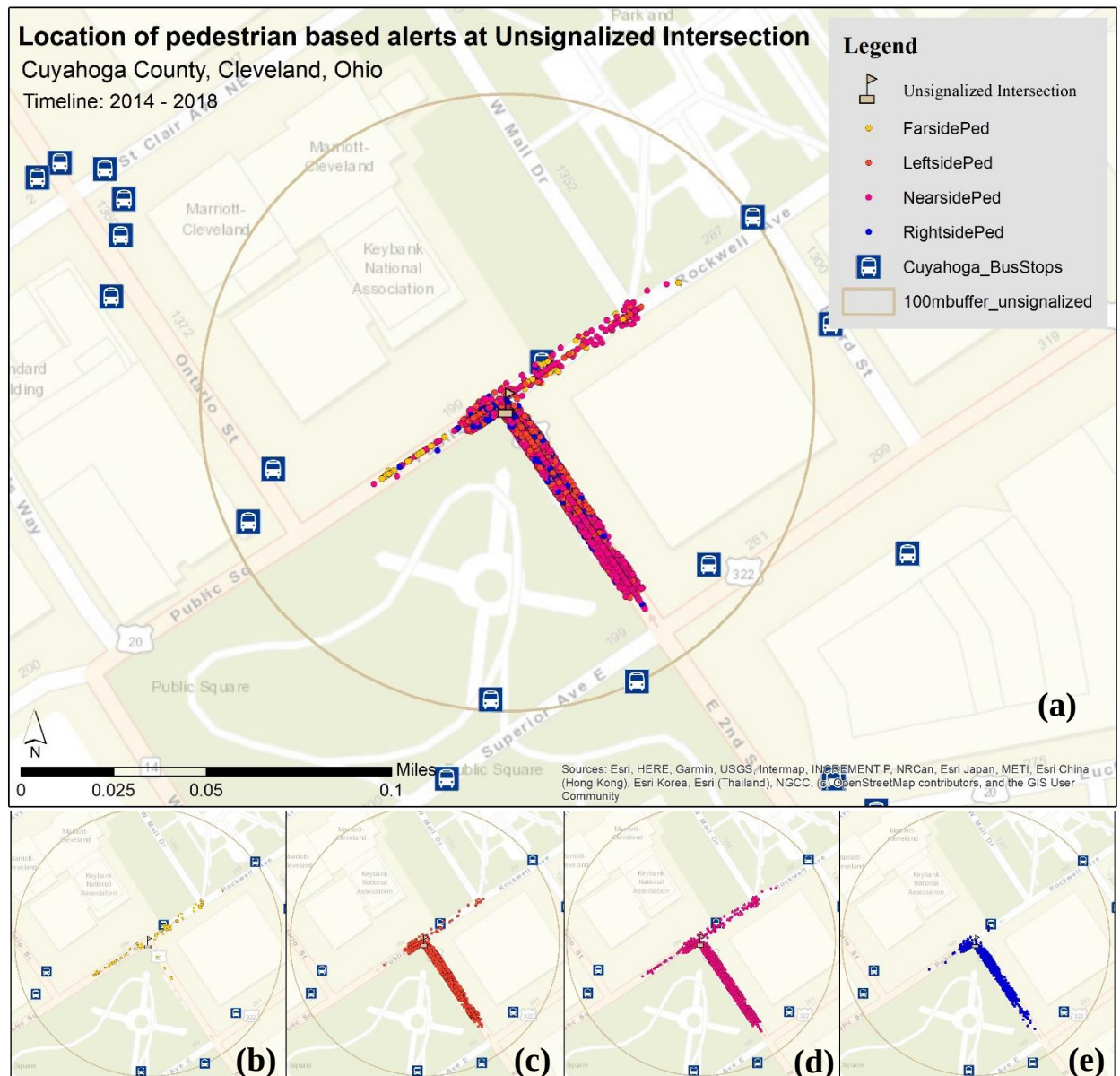


Figure 3.7. (a) All, (b) “FarsidePed,” (c) “NearsidePed,” (d) “LeftsidePed,” and (e) “RightsidePed” alerts at Unsignalized Intersection

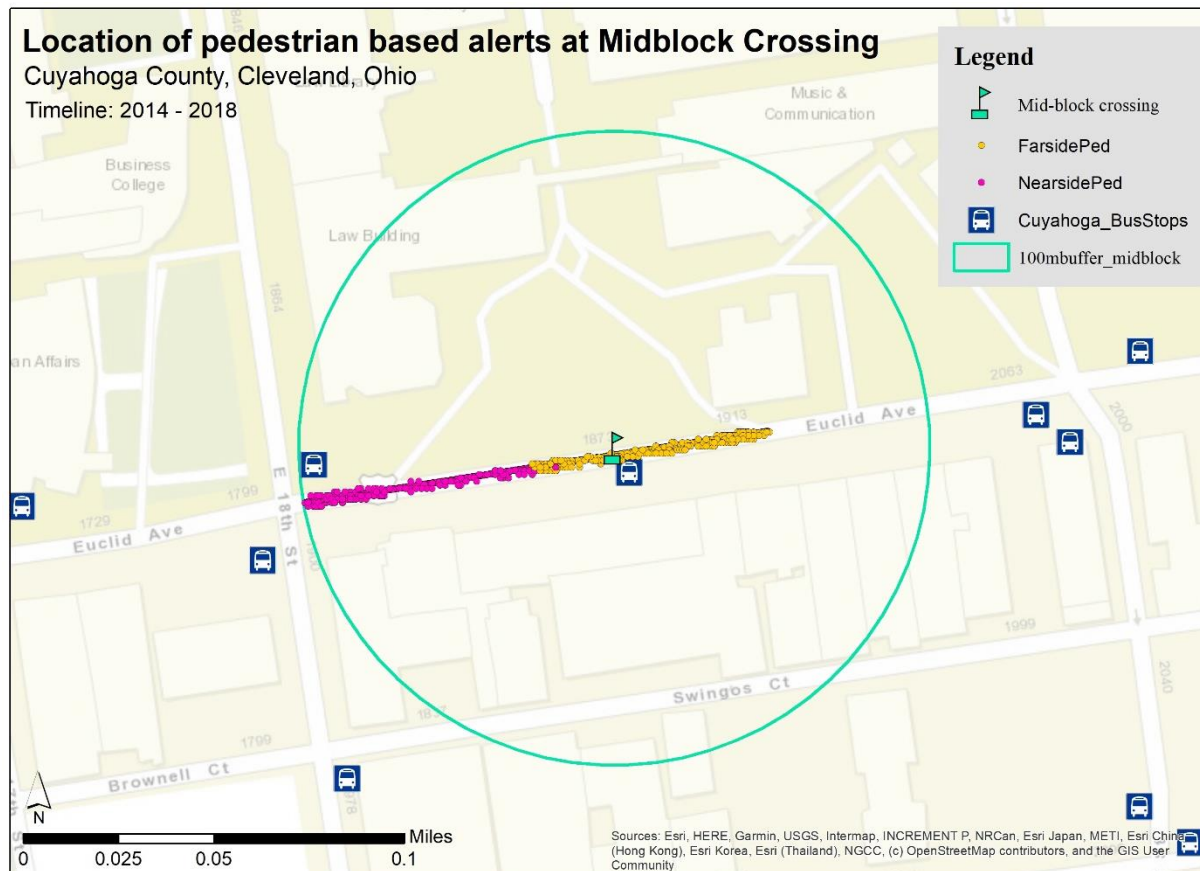


Figure 3.8. Pedestrian based alerts at Midblock Crossing

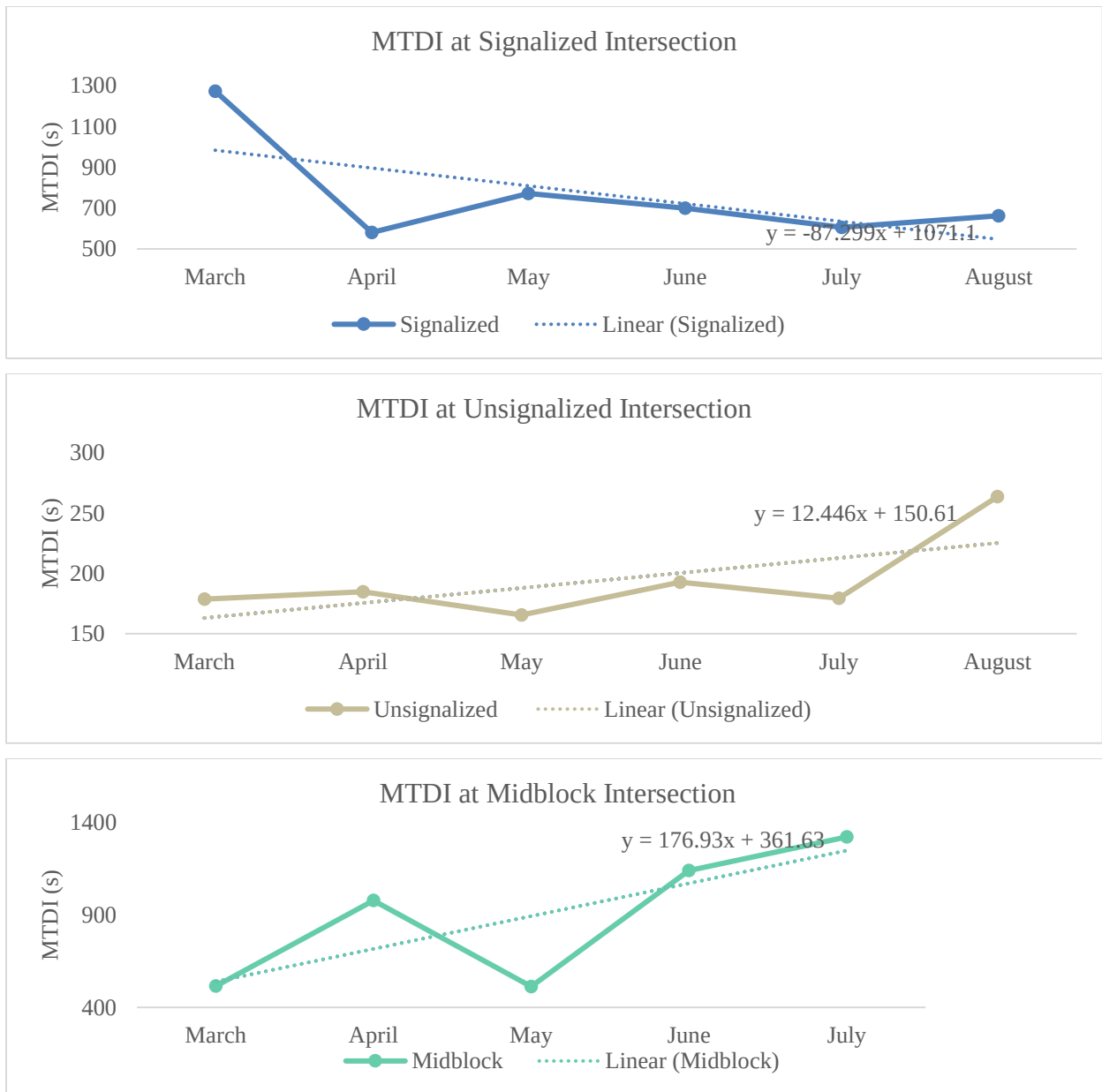


Figure 3.9. Trendline of MTDI at intersections

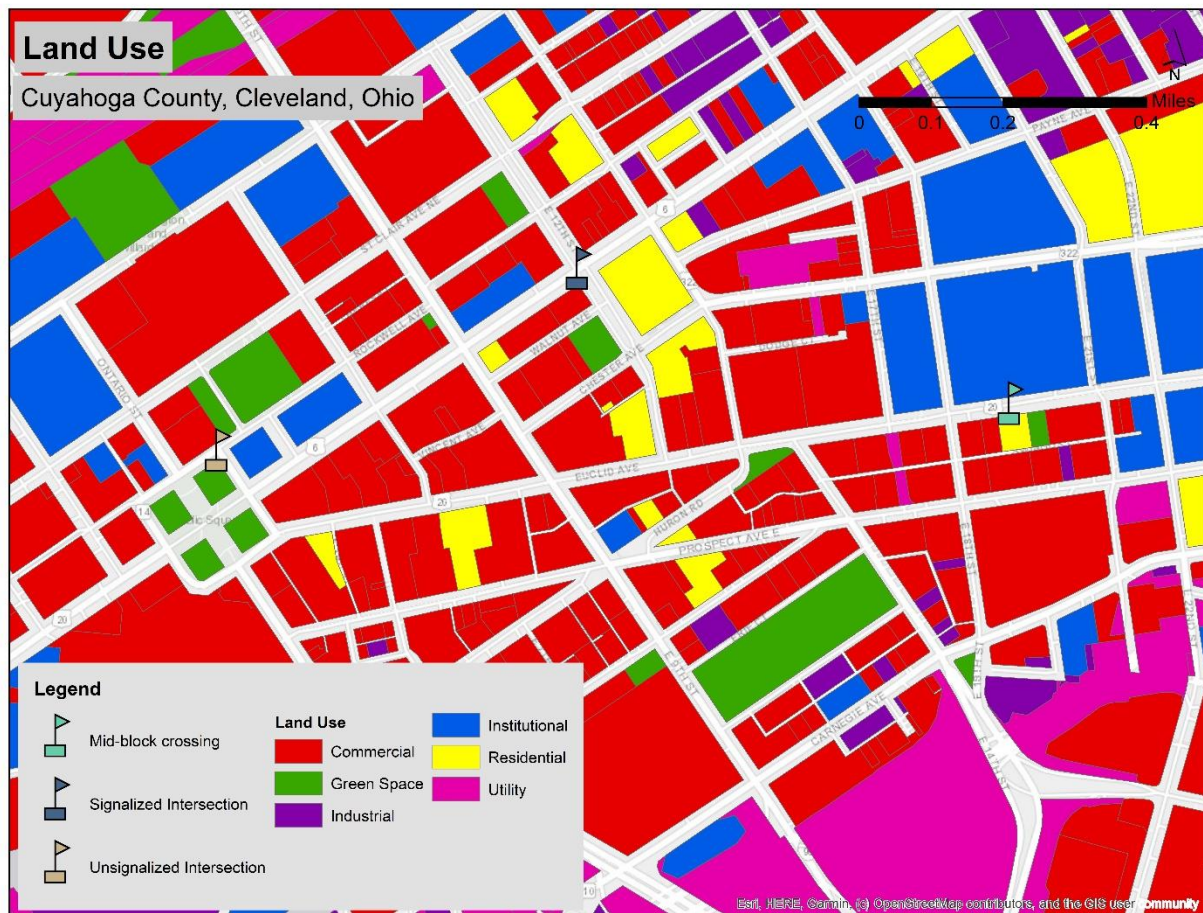


Figure 3.10. Land Use around the three intersections of the study area

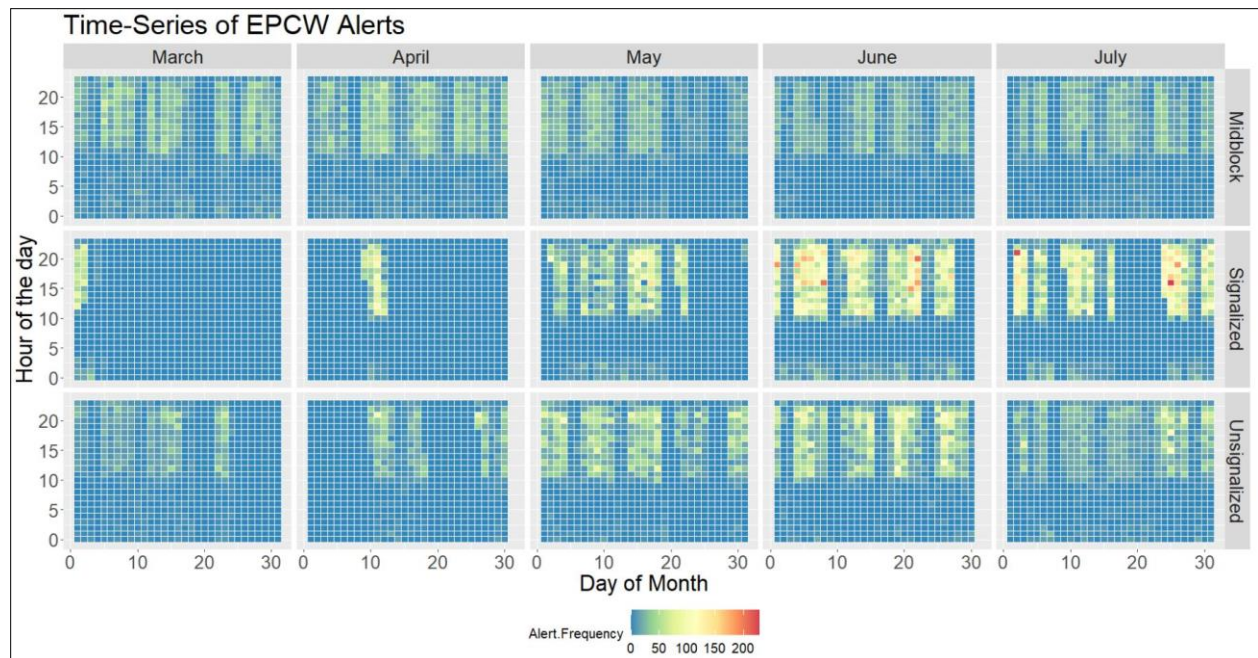


Figure 3.11. Timeseries heatmap of Alerts at intersections

This study developed a framework for exploiting BSM data to identify alert zones for transit buses and pedestrians. But the results might not be generalizable for other cities because of differences in pedestrian and vehicle density, geometric design of the intersections, and weather. Another issue was the unavailability of vehicle trajectory data for the many unique alerts. Those alerts could not be used as their GPS locations could not be traced.

Lastly, false alerts in the data could not be identified and removed because of the unavailability of video footage of the test runs and no mention of false alerts in the published dataset. The research report published by FTA indicates that along with false alerts, incorrect alerts were present in this dataset. False alerts were those alerts that the detector generated when no pedestrian was present at the bus stop. On the other hand, incorrect alerts were generated when the detector identified an incorrect location of the pedestrian. Specifically, false alerts occurred due to lighting and environmental conditions, and incorrect alerts occurred due to the proximity of the zones in this project. To detect the rate of acceptable alerts, The E-TRP project team used 15% of the EPCW alerts as samples. Among the samples analyzed, 81% were correct, 10% were incorrect, and 9% were false alerts (14). As the majority of the alerts of the sample were correct, the authors of this study have decided to proceed with the analysis of the data.

Conclusion

Addressing safety challenges at intersections can be accomplished through the use of innovative technologies and the application of data science. The emergence of CAVs has introduced a new branch of unique datasets that can be harnessed for such safety studies. Examining data from the CAV pilot projects at a microscopic level can yield benefits for drivers and pedestrians at intersections. This study aims to understand the nature of pedestrian-based alerts received by connected buses at three different types of intersections. Of interest was the type and frequency of pedestrian-based alerts and how they differed across different intersections. Another matter of interest was to see if these alerts can assist in identifying alert zones where bus drivers should be careful to avoid potential pedestrian bus collisions. Alert zones were identified spatially in GIS. Since the data was a unique BSM dataset, collected from credible sources and error-checked through descriptive statistics, the data seems valid and contributed to the findings.

Key findings are worth mentioning. First, the location of receiving pedestrian-based alerts might depend on the type of bus stop location. The midblock crossing received only “FarsidePed” and “NearsidePed” alerts, whereas the other two intersections received all four types of pedestrian-based alerts. Second, based on the data collected in this pilot project, the signalized intersections received the most EPCW alerts, followed by the unsignalized and the midblock intersections. However, this finding does not confirm which intersection has more potential for improvement, and further studies are needed in this area. With the full deployment of CV transit, researchers may recognize which intersections have greater potential for pedestrian safety improvements. Third, “NearsidePed” was the most generated alert type in all three intersections. The next most generated alert was “LeftsidePed,” although midblock didn’t have any “LeftsidePed” alert. Fourth, Alert zones were identified within a radius of 142.09 feet, 102.13 feet, and 169.78 feet from the center of the signalized, unsignalized, and midblock crossing intersections, respectively. Presumably, if transit drivers are more vigilant when entering and driving through these alert zones, then they may be able to avoid collisions with pedestrians. Finally, based on the data in hand, the novel SSM MTDI metric showed that the unsignalized and midblock crossing intersections in the study area

had more time to reach the middle of the intersection than the signalized intersections for pedestrians after applying connectivity.

This study highlights the importance of connectivity technology by traversing the usefulness of BSM alerts in enhancing pedestrian safety. As the findings are based on the unique CV transit bus data, they contribute to the analysis of ongoing experiments on connected vehicles. During the test, alerts were successfully provided to drivers. This shows that the system can successfully sense pedestrians and provide alerts to bus drivers. Overall, this study provides early evidence about the performance of EPCW technologies, which is relevant to policy makers who aim to promote safe roads for all users. Notably, EPCW deployment can be considered as an option in other contexts, given their successful implementation at intersections in this case study. The paper will also be useful for transportation analysts as it contributes a methodology that harnesses BSM data generated by new sensors to improve the safety of transit and pedestrians. The methods used in this paper to identify alert zones and calculate MTDI can be applied immediately. Transit agencies might install these connectivity systems in their buses and intersections along their route to collect BSMs and analyze the BSMs to identify alert zones for each intersection. Transit agencies can share this information with the transit drivers so that they can be more cautious about the presence of pedestrians and drive more safely in and around intersections. By checking the MTDI of the intersections over time, transit managers can identify dangerous intersections and implement safety countermeasures.

Conversely, it is evident from the FTA report (14) that although GCRTA highly valued the EPCW system and expressed their keen interest to deploy the system in the future, they did not continue to use the technology (OBUs and DSRC RSUs) after the pilot ended. This is because the transit agency's current budgets fell short to cover the full deployment of the EPCW infrastructure across their transit network. Such outcomes are a common phenomenon in similar pilot projects. To accelerate the diffusion and deployment of promising technologies, agencies can undertake a phased deployment of successful transit technology and install the systems at their most vulnerable intersections initially, where the technology is likely to have a substantial and high impact. In this instance, GCRTA could continue investing in the technology instead of removing it from the network's study area and later expand deployment to other vulnerable areas. Similarly, other transit agencies that conduct such tests should consider continued investment from testing to technology deployment (if sufficient benefits were found) rather than removing the transit technology at the conclusion of the technology pilot. The issue of maintenance cost of the technology would need to be addressed by the transit agency as well.

Finally, the success of EPCW in this case study provides the motivation for further investigation of the technology. Future studies can investigate the creation and impacts of different types of alerts and warnings temporally and in diverse spatial settings. Investigating the frequency of alerts, precisely "WarnAlert," at unique intersections such as midblock crossings and roundabouts would be an interesting topic for future research. By including more intersections, factors contributing to different types of alerts at different intersections can be identified through robust statistical models. Future efforts should focus on associating BSM alerts with variables such as crashes, road geometry, traffic signal design, exposure data like pedestrian and vehicle flow at intersections, bus driver characteristics, speed, acceleration, reaction time, stopping sight distance of CVs, socio-economic characteristics of the area, and surrounding land uses to further the understanding on bus-involved pedestrian crash studies.

Acknowledgments

The Collaborative Sciences Center for Road Safety, CSCRS, a U.S. DOT-funded National Transportation Center, partially supported this study under Grant No. 69A3551747113. The Civil, Environmental, Industrial & Systems Engineering Information Technology service providers of the University of Tennessee, Knoxville are acknowledged for their IT support.

CHAPTER 4. A SYNTHESIS AND COMPARISON OF AMERICAN AUTOMATED SHUTTLE PILOT PROJECTS

Abstract

Automated shuttles are an emerging mode of transportation in many American cities, but it is not yet well understood where they should be deployed. Therefore, the study aims to synthesize the state of the practice of automated shuttle pilot projects to identify patterns of shuttle deployments in the United States. The case study approach is used as the method. Nineteen ongoing or completed automated shuttle pilot projects are initially identified. Among them, automated shuttle projects in Ann Arbor, Michigan; Arlington, Texas; San Ramon, California; Denver, Colorado; Las Vegas, Nevada; and Minneapolis, Minnesota are chosen to study in detail to identify patterns. The key findings are used to develop a synthesis of the state of the practice, which is formulated along three dimensions: deployment locations, service characteristics, and stakeholders. The dimensions suggest that university campuses, stadiums, office parks, transit stations, entertainment areas, and downtowns are well-suited locations for automated shuttle operations. Common service characteristics are low speeds (10-15 mph), vehicle capacities of 6-12 passengers, short routes (less than a mile) with few stops, and free fares. Stakeholders are divided into four groups: public organizations, partner organizations, shuttle manufacturers, and private operators. Deployment locations influence the types of service characteristics and stakeholder combinations of each deployment project. These findings will assist cities that are considering future deployments of automated shuttles in their locality.

Introduction

The introduction of automated vehicles (AV) is expected to bring about momentous change in the transportation industry. As part of this movement toward AVs, automated shuttles have recently been deployed in many American cities. According to the Federal Transit Administration's (FTA) Strategic Transit Automation Research Plan of 2018, automated shuttles are small, low speed (generally less than 25 mph) vehicles that "do not require a human operator, although early demonstrations all have included an on-board human attendant to observe passengers, record data, answer questions, and serve as a safety operator if needed" (79). These shuttles typically meet the Society of Automotive Engineers (SAE)'s definition of Level 4 automation, which is a high automation mode in which the system uses an automated driving system that involves the driver's assistance in some situations. Additionally, the shuttles are typically "limited to operating in certain environments such as parking lots, busways, campuses, downtown districts, and retirement communities" (79).

To prepare for the future of the transportation system, the United States Department of Transportation (USDOT) has outlined guiding principles for automated vehicles (80); however, this policy document does not have specific guidance for automated shuttles. To guide the transit industry, particularly in the case of automated shuttles, the FTA has initiated a five-year-long research plan to communicate deployment decisions; this is currently underway and will continue through 2022 (79). As this initiative is still underway, it would be helpful to provide planners and policymakers an opportunity to learn about the initial similarities, differences, and key findings of early automated shuttle projects. Therefore, this paper aims to create an initial synthesis and comparison of the state of the practice of automated shuttle pilot projects, which can be insightful for American cities that are interested in deploying automated shuttles.

The chapter is organized as follows: first, prior research on automated shuttles is reviewed, and the objectives of this research are set forth. Then, the case study method is used; this includes

identifying ongoing or successfully completed deployments in the United States and selecting six unique cases to study in detail. The case study results are then used to formulate the state of the practice along three key dimensions: deployment locations, service characteristics of the shuttles, and stakeholders who are involved in the project. This is followed by areas for future research and the conclusions.

Prior Research

Research on automated shuttles is being undertaken by many countries. European research is arguably the most advanced. Thus, this brief literature review will focus on European research as well as American research. For a more comprehensive literature review, readers are referred to (81).

European research on automated shuttles

Many deployment projects have taken place in the form of demonstrations, pilots and showcases in Europe, among which Citymobil2, WEpod, and CarPostal are noteworthy (82-85). The Citymobil2 project analyzed twelve automated shuttle deployments and classified their deployment locations into four types: city centers, within a “major facility,” from a public transport node to a major facility, and from a public transport node to a residential area.

Other studies have analyzed automated shuttles from the perspective of users, including travel behavior, demographics, and user perceptions of safety (86-88). One study conducted a stated preference survey to understand the user perspective of shared AVs as a potential solution for last mile trips from transit stations and found that first class train travelers are more likely to choose shared AVs as their egress mode (89). Another study surveyed actual users of automated mini buses and found that regular users tended to be younger people; additionally, passengers generally did not have concerns about safety or security onboard (88). A third study conducted surveys of the passenger experience after travelling in a driverless shuttle bus and concluded that passengers perceived driverless shuttle bus safety to be better than that of conventional buses, whereas personal in-vehicle security was weaker than that of conventional buses (87).

The economics and cost effectiveness of automated shuttles have been discussed by many authors (90-92). One early study used simulation techniques to develop automated demand responsive transport service (ADRTS) scenarios and found that operational costs are likely to be similar to current bus systems but less than traditional demand responsive services, primarily because ADRTS eliminates driver costs and scheduling constraints (93). Another noteworthy study concluded that deployment of automated shuttles will be competitive if demand is sufficient when deployed in dense urban areas (94).

American research on automated shuttles

The United States is arguably lagging behind European countries in terms of automated shuttle deployment and research, even though there has been some background research (81). Most American studies on the user perspective were done prior to automated shuttle deployment (95; 96). Two of these studies have revealed some skepticism about shuttle safety. For example, one study reported that parents may be unwilling to send their children to school in automated shuttles (96). Another study of transit riders found that their willingness to ride in a driverless shuttle increases significantly with the presence of onboard safety staff (95). However, it is likely that

attitudes will change after more Americans have experienced riding in an automated shuttle in the future. A limited amount of research has been undertaken on the economic aspects of the automated shuttle deployments in the United States; one noteworthy study showed the cost efficiency of automated shuttles over paratransit services (97).

The next segment of prior work includes the studies that were undertaken since American cities started deploying automated shuttle pilot projects. To date very few relevant academic works have been published on automated shuttles because the shuttles are still in their test phases. The Eno Center for Transportation published a report titled “Beyond Speculation 2.0: Automated Vehicles and Public Policy,” which contains policy recommendations for testing automated vehicles on public roads (98); however, the report is for automated vehicles, it is not specific for automated shuttles. The most relevant reference for this study is a report by Volpe National Transportation Systems Center of the United States Department of Transportation that analyzed the state of the practice of automated shuttle projects in 2018. The report analyzed a few national and international pilot projects, and it summarized the vehicle and service characteristics of the automated shuttles, identified the parties involved in development, elaborated on the issues related to the shuttles’ deployment, and provided mitigation strategies for resolving those issues (99). This synthesis is build upon this report.

Gaps in research on automated shuttles

The prior research mentioned in the previous sections has studied automated shuttle deployments up until 2018. However, this is a rapidly changing area and many new pilot projects have launched after 2018, particularly in the United States. Patterns of pilot project likely exist, but there has been limited effort to document them. This synthesis study seeks to address this research gap.

Objectives

Given the rapid advancements in automated shuttle pilots in the United States, the objective of this study is to synthesize the current state of the practice of automated shuttle pilot projects in the United States along three dimensions: locations, service characteristics, and stakeholders. The first dimension, locations, examines areas where shuttles have been deployed. The second dimension refers to the service characteristics such as fares, the presence of onboard staff, right-of-way, route length, speed, stops, and vehicle capacity. The third dimension considers the stakeholder organizations that are involved in the deployment, operation, and manufacturing of these automated shuttles. The methodology used for this study is discussed next.

Methodology

The case study method was chosen to conduct this research for two reasons: (1) deployment of automated shuttle is a contemporary event and (2) the researchers do not have any control over the events relevant to the research. According to Yin, the case study method is appropriate for this kind of research (100). To conduct the case study analysis, a four-step process was used in this study as shown in **Figure 4.1**.

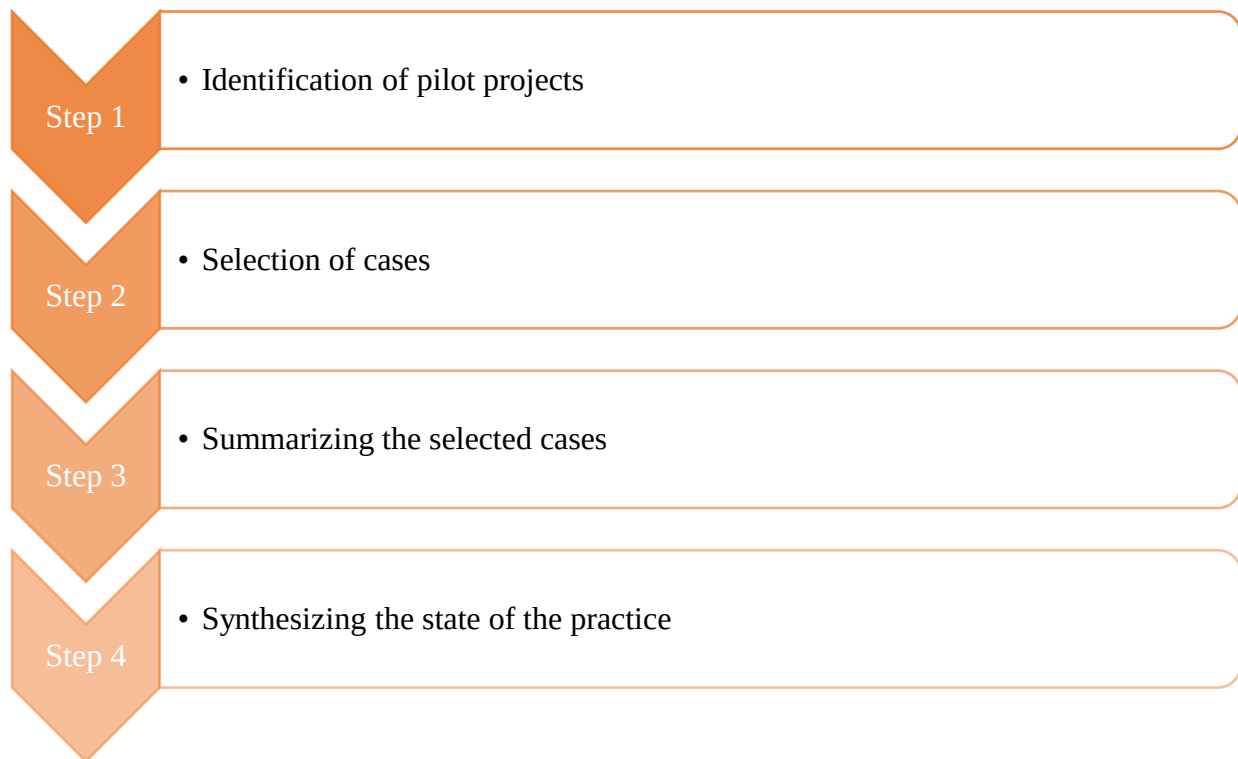


Figure 4.1. Methodology

The first step was to identify as many automated shuttle pilot projects as possible in the United States to make a comprehensive list. The pilots were identified by searching online. This study considered only active or completed project; any planned, postponed, or cancelled projects were excluded from the study. It should also be noted that only pilot projects open to the broader public were considered; exclusive/internal projects like one for the United States Army at Fort Bragg were excluded from this study (101).

In the second step, a smaller number of pilot projects from the comprehensive list were selected for further study. This selection was based on (1) different location types and (2) the amount of publicly available information. Each of these cases also involved an automated vehicle with a capacity of six to twelve passengers serving the general public along a fixed route.

In the third step, each case was summarized. The summary of each case included the project's deployment location, launch time, speed, route length, number of stops, vehicle capacity, ridership market, the presence of onboard staff, project's aim, project completion time, and key findings if it was a completed project.

In the last step, a synthesis of the key findings of the case studies was created along three dimensions: type of location, service characteristics, and key stakeholders.

Results

Identification of pilot projects

Deployment of automated shuttle pilot projects began in 2016 in the United States. Initial pilots were typically deployed in research centers to test shuttle technology. Gradually, deployment locations were expanded to include different types of roads and environmental conditions in different cities. This list contains all of the automated shuttle cases that were deployed up until 1st June, 2019, to the best of authors' knowledge. As can be seen in **Table 4.1**, there were six pilot cases on university campuses, three cases in entertainment areas, three cases near transit stations, two cases in the vicinity of stadiums, one case in an office park area, and one in a downtown area. At least three cases serve multiple locations: Detroit, Columbus, and Providence.

Selection of cases

Six unique cases were selected from **Table 4.1**. These case studies are intended to be representative of their general location type. Each of these cases were either ongoing or completed and served the general public. The selected cases were the automated shuttle projects in Ann Arbor, Michigan (MI); Arlington, Texas (TX); San Ramon, California (CA); Denver, Colorado (CO); Las Vegas, Nevada (NV); and Minneapolis, Minnesota (MN). A brief summary of each case is provided in the next section and **Table 4.2**.

Summary of the selected cases

Case Study 1: Ann Arbor, Michigan

The University of Michigan (U-M) ran two automated shuttles on Mcity's test ground from June 2018 to December 2019. The shuttles ran on U-M's North Campus research complex's Plymouth Road. The route was one mile long and had two stops. Speeds were approximately 12 mph, and the shuttles had a capacity of 11 passengers. The primary riders of these shuttles were U-M's students, faculty and staff and their escorted guests. An onboard staff person stayed inside the

shuttles to handle emergency situations. Through this pilot project, Mcity was observing the reaction of other vehicles, pedestrians, and riders with the shuttle's exterior and interior cameras. The project aimed to understand user acceptability of automated shuttles (102-104).

Case Study 2: Arlington, Texas

In the one-year pilot project "Milo", the City of Arlington tested Easymile's automated shuttle on an off-street trail from August 2017 to August 2018. Milo had three routes inside Arlington's entertainment district: the AT&T Stadium Events route (0.44 mile), the Globe Life Park Events route (0.285 mile), and the Public Demo route (0.47 mile). Each route had two stops. The shuttles ran at a speed of approximately 15 mph. Vehicle capacity of the shuttles were 12 passengers, or 10 passengers when a one wheelchair was inside. The shuttle had an onboard staff member for safety reasons. Project Milo held 17 public demos and 18 special interest group demos, in addition to serving 78 events such as football games, baseball games, and concerts. It gave free rides to visitors on event days, such as sports games (105; 106).

Case Study 3: San Ramon, California

In San Ramon's pilot project, the Contra Costa Transportation Authority (CCTA) led a test of manufacturer EasyMile's automated shuttle. The first phase occurred in 2017 inside Bishop Ranch's office park; in the next phase, the project has plans to test the shuttles on a public street that bisects Bishop Ranch. The shuttles ran at approximately 12 mph, and the vehicle capacity was 12 passengers. Initially, the shuttles were being ridden only by shuttle operators, but newer phases will allow selected employees of Bishop Ranch to ride. An onboard staff person was always present to stop the shuttle in emergency situations. The ultimate plan is to extend the shuttles' service route to a Bay Area Rapid Transit station and San Ramon city center (107-112).

Case Study 4: Denver, Colorado

Denver's automated shuttle pilot project – Route 61AV – ran from January to August 2019 and was led by the local transit agency, the Regional Transit District (RTD), and the City of Denver. The shuttle served a one-mile circulatory loop that started at the 61st and Peña commuter rail station and continued past the companies Panasonic and EasyMile's offices. The shuttle ran at an average speed of 12-15 mph, and it stopped at four stops every 15 minutes. The shuttle had a vehicle capacity of 12 passengers. An onboard staff person was present to answer questions and ensure the safety of passengers. The project ended after giving free rides to almost 600 people (113-115).

Case Study 5: Las Vegas, Nevada

In the automated shuttle pilot in Las Vegas, the American Automobile Association (AAA) led a pilot of manufacturer Navya's shuttle from November 2017 to October 2018. The shuttle ran on Fremont Street in mixed traffic, which was the first pilot project where an automated shuttle operated in mixed traffic in the United States. Despite having a minor crash on its launch day, the project continued. The shuttle ran at approximately 15 mph on a 0.6-mile-long loop that had three stops. This shuttle had a vehicle capacity of eight people. It gave free rides to 32,827 people, who were typically tourists, residents, and local business people. An onboard staff person was present who could take action if an emergency occurred. AAA surveyed riders and found that 30% of shuttle riders agreed that they have more positive attitudes towards AV technology after riding on the automated shuttle. Similarly, 98% of riders said they would recommend a shuttle ride to their friends and family (116; 117).

Case Study 6: Minneapolis, Minnesota

In the initial phase of an automated shuttle demonstration in downtown Minneapolis, the Minnesota Department of Transportation (MnDOT) led a three-day demonstration of an automated shuttle from January 26th to January 28th 2018 during the Super Bowl weekend. A street in front of the Minneapolis Central Library called Nicolett Mall between 3rd and 4th Street South was blocked off from other vehicles and pedestrians to make way for the demonstration. The shuttle ran at approximately 12-15 mph, had two stops, and had a vehicle capacity of 12 passengers. Onboard staff members were present inside the shuttle to talk to the riders and survey them during the demonstration (118). Additional demonstrations were held that spring in other locations, including one on the University of Minnesota's campus. Key characteristics of these cases are shown in **Figure 4.2**.

Synthesis of the state of the practice

The synthesis of the state of the practice was developed along three dimensions, which are interrelated. The following sections will describe the dimensions and their relationship with each other (**Figure 4.2**).

Dimension 1: Locations

Six types of deployment locations were identified from the initial list of existing pilots shown in **Table 4.1**, which were university campuses, stadiums, office parks, transit stations, entertainment areas, and downtowns. Based on the six in-depth case studies, it can be concluded that all of these location types were in populated areas within a metropolitan region that requires good transportation facilities to reduce congestion and/or increase accessibility. However, the six case studies revealed that the different location types had unique characteristics pertaining to circulation patterns, service frequencies, and ridership markets, as described briefly below.

The university campus shuttle circulated in a loop between the different buildings of the campus, whereas the stadium shuttle operated on a linear route between nearby parking lots and the stadium. The office park shuttle circulated between parking lots and office buildings, while the transit station shuttle served transit stations and nearby destinations. The entertainment area shuttle circulated in a tourist area, whereas the downtown shuttle different locations in a dense downtown area. The results of the six case studies also revealed that service frequencies varied, which may have been influenced by the type of deployment location. For example, the stadium shuttle ran more frequently on game days, whereas the university campus shuttle likely operated most frequently during the academic semester. The ridership market appears to have varied from location to location. For example, the university campus shuttle was intended to transport students, faculty members, staff, and campus visitors. The stadium shuttle primarily transported fans and visitors on game days, whereas the office park shuttle presumably served employees coming from nearby parking lots. The transit station shuttle aimed to address the first mile-last mile problem of transit riders, whereas the entertainment area shuttle appears to have been deployed to serve tourists. In summary, it can be concluded that the circulation patterns, service frequencies, and ridership markets vary depending on the deployment location. Based on these case studies, a typology of automated shuttle deployment locations was developed. This is shown in **Table 4.3**, and it reflects important choices among the six different location types for future deployments.

Table 4.1. Timeline of automated shuttle deployments

#	Place/City, State	Deploying Organization	Shuttle Manufacturer	Launch Date	Deployment Location	References
1	National Harbor, Maryland	Local Motors, IBM	Local Motors	16 th June, 2016	Entertainment Area	(119; 120)
2	San Ramon, California	Contra Costa Transportation Authority (CCTA), Bishop Ranch Office Park	EasyMile	6 th March, 2017	Office Park	(107-112)
3	Jacksonville, Florida	Jacksonville Transportation Authority (JTA)	EasyMile	One-day demonstration on 8 th March, 2017; 2 nd phase began on 27 th March, 2018	Stadium	(121; 122)
4	Greenville, South Carolina	Greenville County	Cushman	August 2017	University Campus	(99; 123)
5	Arlington, Texas	City of Arlington	EasyMile	26 th August 2017	Stadium	(105)(108)
6	Las Vegas, Nevada	American Automobile Association (AAA)	Navya	8 th November 2017	Entertainment Area	(116; 117)
7	Weymouth, MA	LStar Ventures	Optimus Ride	2018	Transit Station	(99)
8	Minneapolis, Minnesota	Minnesota Department of Transportation (MnDOT)	EasyMile	26 th to 28 th January, 2018	Downtown	(118)
9	Gainesville, Florida	Gainesville Regional Transit System; Florida Department of Transportation; University of Florida	EasyMile	May 2018	University Campus	(124-126)
10	Ann Arbor, Michigan	University of Michigan, Mcity	Navya	4 th June, 2018	University Campus	(102-104)
11	Dublin, California	Livermore Amador Valley Transit Authority (LAVTA)	EasyMile	22 nd June, 2018	Transit Station	(127)
12	Detroit, Michigan	Bedrock, a property management company	MayMobility	27 th June, 2018	Downtown /Office Park	(128; 129)
13	Lincoln, Nebraska	City of Lincoln	Navya	28 th June, 2018	University Campus	(130; 131)
14	Columbus, Ohio	City of Columbus	MayMobility	10 th December, 2018	Entertainment/ Downtown	(128; 132)

Table 4.1 Continued.

#	Place/City, State	Deploying Organization	Shuttle Manufacturer	Launch Date	Deployment Location	References
15	Denver, Colorado	Denver Regional Transportation District (RTD), City of Denver	EasyMile	29 th January, 2019	Transit Station	(113; 114)
16	Tampa, Florida	University of South Florida's (USF) Center for Urban Transportation Research	Coast Automated	11 th – 15 th February, 2019	University Campus	(133)
17	Sacramento, California	Sacramento State University	Local Motors	21 st February, 2019	University Campus	(134)
18	Park City, Utah	Utah Department of Transportation, Utah Transit Authority	EasyMile	20 th – 23 rd May, 2019	Entertainment Area	(135; 136)
19	Providence, Rhode Island	Rhode Island Department of Transportation	MayMobility	15 th May, 2019	Downtown/ Transit Station	(137; 138)
<i>This list of automated shuttle projects was compiled by the authors to include any pilot project launched in the United States that was open to the public prior to June 1, 2019. This list was compiled using publicly available online resources; it is possible that additional pilot projects were launched during this timeframe, particularly those with limited or no online references.</i>						

Dimension 2: Service Characteristics

Analysis of the case studies helped to identify common service characteristics of the automated shuttles; this includes fares, onboard staff, right-of-way, route length, route shape, speed, stop spacing, and vehicle capacity. All of the pilots were found to run for free during the pilot phase. Additionally, all of the automated shuttle pilot projects had at least one onboard staff member to ensure safety of the riders. The shuttles typically ran in three types of right-of-way, namely on roads with mixed traffic, on off-street trails, and in designated lanes. The route lengths of the pilot projects were typically short, ranging from 0.5 mile to 2 miles. The route shapes were either linear or circular. The speeds of these shuttles typically ranged between 10-15 mph. The stop spacing for the automated shuttles were generally 1/10th of a mile or more, and each route had at least 2 two stops. Vehicle capacity of the pilot projects ranged between from 6 to 12 passengers. **Table 4.4** summarizes these characteristics briefly.

Dimension 3: Stakeholders

Four types of stakeholder organizations were identified for these pilot projects, namely public organizations, partner organizations, shuttle manufacturers, and private operators. First, public organizations usually included the city government where the pilot was launched, the Department of Transportation (DOT), and/or the local transit agencies. Examples included the Minnesota Department of Transportation (in Minneapolis) and the RTD transit agency in Denver. Second, partner organizations often included universities, research centers, office parks, and/or other organizations. This group has significant variation; for example, the pilot in Las Vegas included AAA whereas the Ann Arbor pilot included the University of Michigan's Office of Research. Third, shuttle manufacturers were typically private companies that manufactured, sold or leased their shuttles for these pilot projects. Examples included Navya, EasyMile, MayMobility, Local Motors, Auro (139), Coast Autonomous (140), and Continental USA (141). Last, operators were responsible for the operation and maintenance of the shuttles during the project duration. For example, Denver had TransDev; Minneapolis, San Ramon, and Arlington's automated shuttles had First Transit (142; 143); Las Vegas had Keolis North America, and U-M had U-M Logistics, Transportation, and Parking as their shuttle operator. Typically, these four types of stakeholders worked in partnership in each pilot project. The main stakeholders of automated shuttle pilots are shown in **Figure 4.3**.

Areas for improvement and future research

There are numerous areas for improvement and future research that emerged from this study. First, this project did not explicitly consider data sharing during the automated shuttle deployments; however, to the best of the author's knowledge, none of the automated shuttle deployments examined in this study openly provide data about ridership or operations. This type of data is important to conduct project evaluation and should be made publicly available and/or requested by academic partners for future research.

Second, this research relied primarily on publicly available documents and online resources to gather information; however, future research should interview or survey shuttle riders, operators, and other stakeholder organizations. Their opinions will be valuable for identifying other potential locations and service requirements of automated shuttles. Another related avenue for future research is potential equity issues, which should also be considered in automated shuttle projects. Opinions of minority groups can be assessed in a similar way to prior research for evaluating barriers to equity in smart mobility systems (144).

Some information was available about shuttle technology, but further analysis should be done for a deeper understanding of the technology used in each case and its associated costs. For instance, all of the shuttles had Level 4 automation and had onboard staff. The newest technology could offer fully automated modes (Level 5) and may not involve an onboard staff person. Elimination of onboard staff will likely save substantial operating costs. Thus, cost-benefit analysis should be performed to evaluate different fully autonomous shuttle scenarios, in a similar manner to prior research for automated driving (145).

Last, future research should consider operations on longer routes, at different speeds, and for different headways to better understand the potential of automated shuttle system performance. Similarly, other types of automated shuttles, such as smaller sized on-demand vehicles, should be considered in future research.

Conclusions

This study presents a synthesis of the state of the practice of automated shuttles in the United States in order to identify early trends in automated shuttle deployments. The case study approach was undertaken as the method. Initially, 19 pilot projects were identified, and six case studies were then selected for more detailed study, including automated shuttle projects in Ann Arbor, Arlington, San Ramon, Denver, Las Vegas, and Minneapolis. The synthesis was formulated along three key dimensions.

The first dimension was the type of deployment location. Automated shuttle projects were typically deployed in relatedly dense urban areas, like downtowns, entertainment areas, stadiums, office parks, university campuses, and transit stations. Depending on the type of location, the service frequency and the ridership market varied. For example, service frequency was higher on game days at stadiums. Similarly, tourists were likely a primary market of riders in entertainment areas, while employees were more likely the target market in office parks.

The second dimension was service characteristics. Typically, the shuttles were running at low speeds (10-15 mph) and carried six to 12 passengers for free. The routes were up to two miles in length and had stop spacing as low as one stop per tenth of a mile. Right-of-way was either a designated lane, in mixed traffic or an off-street trail, depending on the deployment.

The third dimension was stakeholders involved in the shuttle project. Government agencies, such as state DOTs, city governments, or local transit agencies, often worked in partnership with other organizations, such as universities or office parks, to deploy automated shuttles. Additionally, each project typically involved a private company that manufactured the shuttle, as well as a private operator to manage operations.

In summary, this research identified three key characteristics of automated shuttle deployments in the United States that can help to shed light on early trends and guide other localities who are considering deployment of automated shuttles in the future.

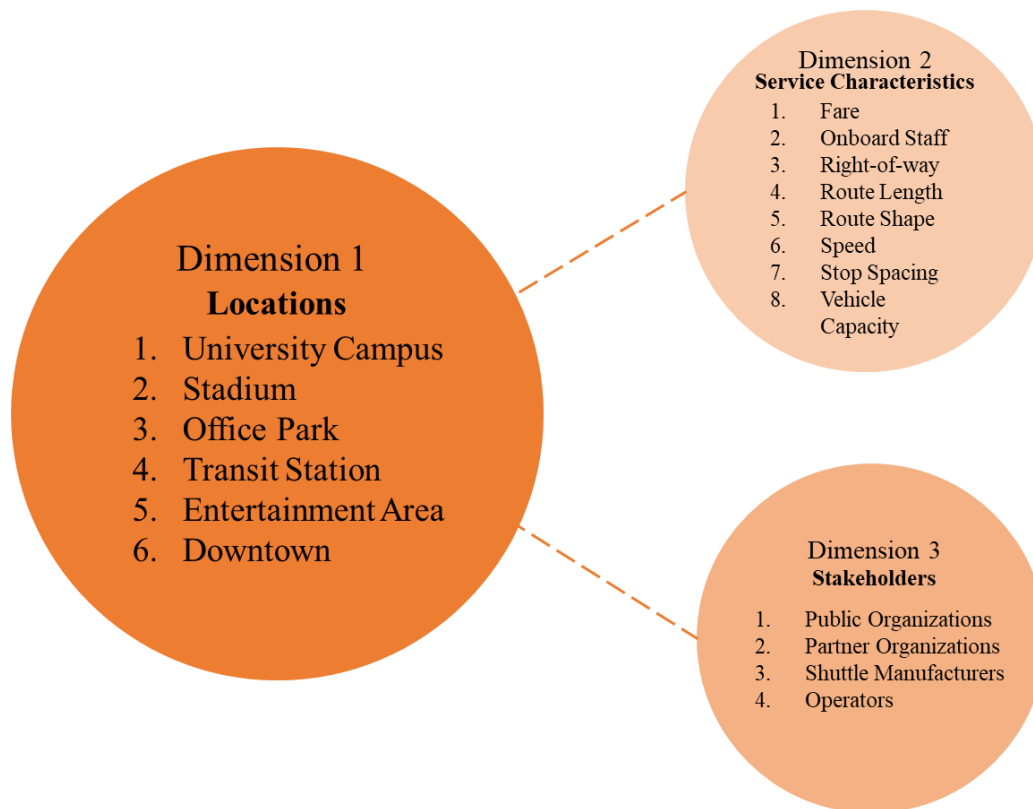


Figure 4.2. Three key dimensions of automated shuttle deployments

Table 4.2. Summary of six case studies

Case	Ann Arbor, MI	Arlington, TX	San Ramon, CA	Denver, CO	Las Vegas, NV	Minneapolis, MN
Locations of Deployment						
Location	University	Stadium	Office	Transit	Entertainment	Downtown
Type	Campus		Park	Station	Area	
Service Characteristics						
Fare	Free	Free	Free	Free	Free	Free
Onboard Staff	Available	Available	Available	Available	Available	Available
Right-of-way	U-M campus roadway shared with other vehicles	Off street trails near Arlington's entertainment district	Parking lots and streets of Bishop Ranch office park	Regular bus route - RTD's newest route 61AV	Downtown mixed traffic road	The road along Nicolett Mall between 3 rd and 4 th Street South
Route Length	1 mile	3 routes in miles: AT&T Stadium route = 0.44; Globe Life Park Events = 0.285; Public demo route = 0.47	Not available	1 mile circulator route	0.6 mile loop	Approximately 380 feet*
Shuttles	2	2	2	1	1	1
Speed (approximate)	12 mph	12 mph	15 mph	12-15 mph	15 mph	12-15 mph
Stops	2	6	Not available	4	3	2
Vehicle Capacity	11	12 or 10+1 wheelchair	12	12	8	12
Stakeholders						
Primary Stakeholders	U-M	City of Arlington	of CCTA; Bishop Ranch Office Park	RTD; City of Denver	AAA	MnDOT
Partner Organizations	Ann Arbor Public Schools and First Responders, NHTSA, Navya	Arlington Convention and Visitors Bureau, EasyMile	GoMentum Station, EasyMile	Panasonic, and L.C. Fulenwider, Inc., EasyMile	City of Las Vegas, The Regional Transportation Commission of Southern Nevada (RTC); Zappos.com, Navya	WSB and Associates, Inc., AECOM, 3M, Colorado Department of Transportation, EasyMile
Operator	U-M Logistics, Transportation and Parking (LTP)	First Transit	First Transit	Transdev	Keolis North America	First Transit

*Estimated by the authors using Google Maps

Table 4.3. Typology based on locations

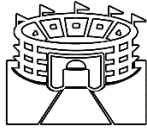
Type of Locations	University Campus 	Stadium 	Office Park 	Transit Station 	Entertainment Area 	Downtown 
Description	Circulation between various campus buildings, libraries, food courts, and gym	Circulation between stadium complex's "outside vehicle prohibited" zone on game days	Circulation between different office buildings and parking lots	Circulation between transit station and nearby destinations and parking areas	Circulation between different tourist spots	Circulation between different buildings in dense downtown areas
Service Frequency	- Seasonal (Semester) -Less frequent (Weekends) -Frequent (Weekdays)	-Seasonal (On game days) -Frequent	-Regular -Frequent (Weekdays) -Less frequent (Weekends)	-Regular -Frequent (Weekdays) -Less frequent (Weekends)	-Regular -Frequent (Tourist season) -Less Frequent (Off season)	-Regular -Frequent (Weekdays) -Less Frequent (Weekends)
Market	-Students -Faculty & staff -Commuters	-Fans -Attendees -Stadium employees	-Office employees -Commuters	-Transit riders -Commuters	-Tourists -Commuters	-Commuters -Some Tourists
Examples	-U-M Ann Arbor, MI -University of South Florida, FL	-AT&T stadium, TX	-Bishop Ranch, CA -Bedrock, MI	-Denver, CO -Providence, RI	-Las Vegas, NV -Columbus, OH	-Minneapolis, MN -Detroit, MI

Table 4.4. Typical service characteristics

Service Characteristics		Description
	Fare	• Usually free • Fare might be introduced when market is established
	Onboard Staff	• No driver • Onboard staff member for safety
	Right-of-way	• Roadway (in mixed traffic) • Off-street trails • Designated lanes (exclusive)
	Route Length	• Typically short (0.5 – 2 miles)
	Route Shape	• Circular • Linear
	Speed	• Typically low speed (10-15 mph)
	Stop Spacing	• 1/10th of a mile or more • Minimum of 2 stops
	Vehicle Capacity	• Low capacity (6-12 passengers)

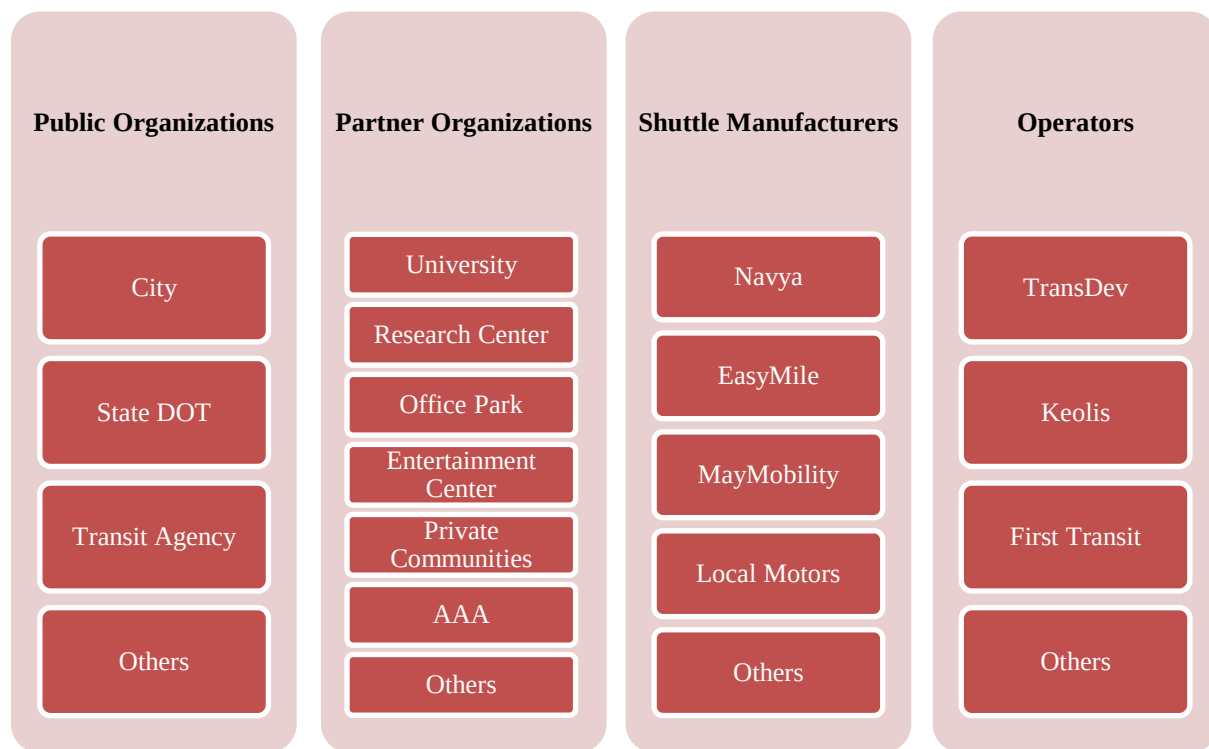


Figure 4.3. Typical stakeholders in automated shuttle projects

CHAPTER 5. INVESTIGATING CRASHES INVOLVING ADVANCED DRIVER ASSISTANCE SYSTEMS (ADAS) EQUIPPED AUTOMATED VEHICLES - THE ROLE OF AUTOMATION

Abstract

While the market penetration of vehicle automation technologies, including advanced driver assistance systems, has increased, many vehicles equipped with such technologies have been involved in traffic accidents. Hence, it is becoming a prime concern for the transportation safety industry to understand the safety of automated vehicles (AVs). This study contributes to AV safety by analyzing crashes of Society of Automotive Engineers Levels two, or lower Advanced Driver Assistance Systems (ADAS) equipped AVs. The objective of this study is to explore the association between crash severity and various factors responsible for AV crashes. To this end, real-world ADAS-equipped AV crash data collected by National Highway Traffic Safety Administration are harnessed. After preprocessing the data, 315 AV crashes from August 2019 to May 2022 are finalized for analysis. An ordered logistic regression model is used to investigate the association of various factors of severity of AV crashes when ADAS system is engaged. Among the crashes, 71% are Property Damage Only, 26% are Injury, and 3% are Fatal or Severe Injury crashes. The results indicate that Tesla involved crashes showed 25% lesser chances of Injury type crash and 7% lesser chances of Fatal or Severe Injury crash compared to other ADAS-equipped AVs. AVs of >4000 pounds is associated with 4% higher chances of Fatal Injury compared to ADAS-equipped AVs of ≤ 3500 pounds. Fatal or severe crash rates (70.00%) are higher for Tesla, but Injury crash rates are (90.48%) higher in non-Tesla AVs. The findings can help AV manufacturers and safety practitioners enhance the safety of AVs.

Introduction

Today, traffic accidents, injuries, and fatalities are the major concern of the transportation industry. An increasing number of severe traffic accidents is one of the adverse impacts of increasing transportation demand. In recent years, traffic fatality has been rising significantly. According to the United States National Highway Transportation Safety Administration (NHTSA), 42,915 people were killed in traffic accidents in 2021 (146), while there were 38,824 fatalities in 2020 and 36,096 fatalities in 2019 (147). This statistic raises concern about the safety of road transportation.

With the increase in emerging vehicle technologies, improvement in transportation safety is highly expected. Many on-road vehicles have recently been equipped with low automation technologies (Society of Automotive Engineers, SAE Level two or lower), including Advanced Driver Assistance Systems (ADAS), to improve their safety performance. Among these vehicles, the market penetration of Tesla, one of the growing automated vehicle manufacturer companies in the US, has considerably increased. Hence, in recent years, the number of Tesla vehicles involved in crashes has risen. Investigating such crashes can provide helpful information about the performance of ADAS technologies. To this end, this study explores the association between crash severity and various factors responsible for crashes in AVs. This study contributes to the knowledge body of automated vehicle safety by conducting a case study on ADAS-equipped AVs. The study uses real-world crash data from NHTSA's website regarding ADAS-equipped AVs. This study explores the association of crash severity with ADAS-equipped AVs, which will help gain a more profound knowledge of ADAS technologies' performance. An ordered logistic regression is developed to investigate the factors associated with the severity of crashes involving ADAS-equipped AVs. The model estimates the chance of various levels of crashes in different AVs (Tesla and Not-Tesla), which might be interesting and valuable for both manufacturers and

safety practitioners. The model can be immediately adopted by automated vehicle safety researchers for similar studies.

Literature Review

As AV technology is emerging, it is becoming a prime concern for the transportation safety industry. SAE has defined six driving automation levels: level zero indicates no driving automation and level five indicates full driving automation in motor vehicles (148). The manufacturers are either technology giants like Google, Uber, and Baidu (149) or traditional vehicle manufacturing companies like Toyota, Nissan, and General Motors (150), who are improving their vehicles constantly to achieve the highest level of automation and deploy fully autonomous vehicles. Currently, SAE level three and level four automation are being tested on public roads (150) and SAE levels one and two are deployed on US roads (151). However, only the California Department of Motor Vehicles has recently launched reports on SAE high-level AV crashes in California. This is one of the most recognized datasets for AV crashes which is utilized by a few studies (150; 152-156). Hence, the existing literature is limited in the safety analysis of lower-level AVs and their crash severities.

The most frequent type of crash happening to AVs is rear-ended collision in mixed-land use areas (154). Crash severity is reported to increase significantly when the AV is responsible and crashes are affected if the AV is in the automated mode, the crashes involve Vulnerable Road Users (VRU) and the roadway environment (150). The analysis of VRU-involved AV crashes revealed that pedestrians and bicyclists are indirectly and directly involved in AV crashes, respectively. It also found that direct VRU-involved AV crashes are more likely to occur when an autonomous mode is engaged than indirectly VRU-involved AV crashes (153). Besides, weather, degree of vehicle damage, crash location, and collision type were also found as contributing factors for AV crash severity (152).

AV crash severity-based studies have utilized various automation factors like Adaptive Cruise Control (ACC), autopilot mode, vehicle characteristics, driver characteristics, road geometry, and environment to realize their impact on crash types and severity. One study on driver's cognitive workload found that the self-reported workload for manual driving is 43.5%, for ACC driving is 38.6%, and for highly automated driving (HAD) is 22.7%. This indicates that drivers of higher level automation are more likely to be disengaged while drivers remain more attentive in vehicles with ACC (157). Another study found that, among different types of maneuvering, stopping at the intersection with having the autonomous mode on has caused more rear-end AV crashes (158). Lane change collision was found to be the second type of pre-crash scenario that mostly happens in AV crashes. The main reasons behind these AV crashes are the gap between AV and human driver's perception-reaction time (PRT), AV's failure to identify the intent of other drivers to change lanes, and miscalculation of time and space in AV path planning (159). Among the road types, arterials were found to have a higher likelihood of injury involved in AV crashes, but collector roads were not found to have a statistically significant association with crash severity in fixed parameter Bayesian models (154). Studies have shown that AVs can better detect pedestrians in daylight compared to poor lighting conditions (160-162). AVs have been compared to conventional vehicles using pre-crash scenarios and AVs were found to be rear-ended by conventional vehicles 1.6 times more than the times the conventional vehicles are rear-ended (159).

The key gap in the literature is that previous studies have not explored the impact of different models (year and make) of the same manufacturer company's automated vehicles on their crash severities. Existing AV crash severity studies are conducted only on California's AV crash data. NHTSA's ADAS-equipped AV crash data is a rich dataset that allows one to study AVs' crash severity, which comes with level 2 automation in the US. This study will help to draw a clear picture of the association of factors in different AVs and the severity of their crashes. The study will also assist researchers in understanding the performance of ADAS in level 2 or lower automation level vehicles. The similarities and differences between all AVs and only Tesla crash severities will be clear from this study.

Methods

In this section, the used methodology of this study is presented. First, the collection of the data and data preprocessing and aggregation are discussed. Second, the cleaned database and the ordered logistic regression to obtain the crash severity likelihood for each category as the modeling approach is presented.

Data

The data is collected from the NHTSA website that recently published ADAS (SAE level two or below) equipped AV crashes that took place in the publicly accessible roads and territories of the US from July 2021 to June 2022. NHTSA issued a Standing General Order that required the manufacturers and operators of ADAS (SAE level two or below) equipped AVs to report crashes that had level two ADAS in use at any time within the 30s of the time of the crash and involved a vulnerable road user (VRU), or caused fatality or major injury, vehicle towing, or deployment of airbags. The crashes are from different locations in the US and not just a particular state (163). Therefore, the results from analyzing this dataset can be generalizable. The data is timely and valid as the reports are from manufacturers and operators.

Initially the dataset had around 900 observations with many duplicates. Duplicates were removed by analyzing the available narratives, and by keeping only the latest report if narratives were not available. Unknown observations for Crash Severity indicator were also deleted. After cleaning and preprocessing the data, 315 AV crashes were finalized for analysis. The dataset contains the month/year, time, and location (city/state) of crashes, crash severity, ADAS engaged information, crash information like severity, pre-crash maneuver of subject vehicle (SV) and crash partner (CP), entity with which the crash occurred, pre-crash speed of the SV, crash contact area of SV and CP, vehicle information like, vehicle make, model, mileage, model year; roadway information like roadway type and speed limit; environmental condition like weather, and lighting, and reporting entities name. Throughout the chapter, subject vehicles are called SV and their crash partners are called CP. The map in **Figure 5.1** shows the states with number of Crashes from this study. It is visible that the highest number of ADAS-equipped AV crashes occurred in Californian (94), followed by Florida (29), Texas (24), and New York (20) which are large, populated states of the US. However, land use type of crash location is not mentioned in the data.

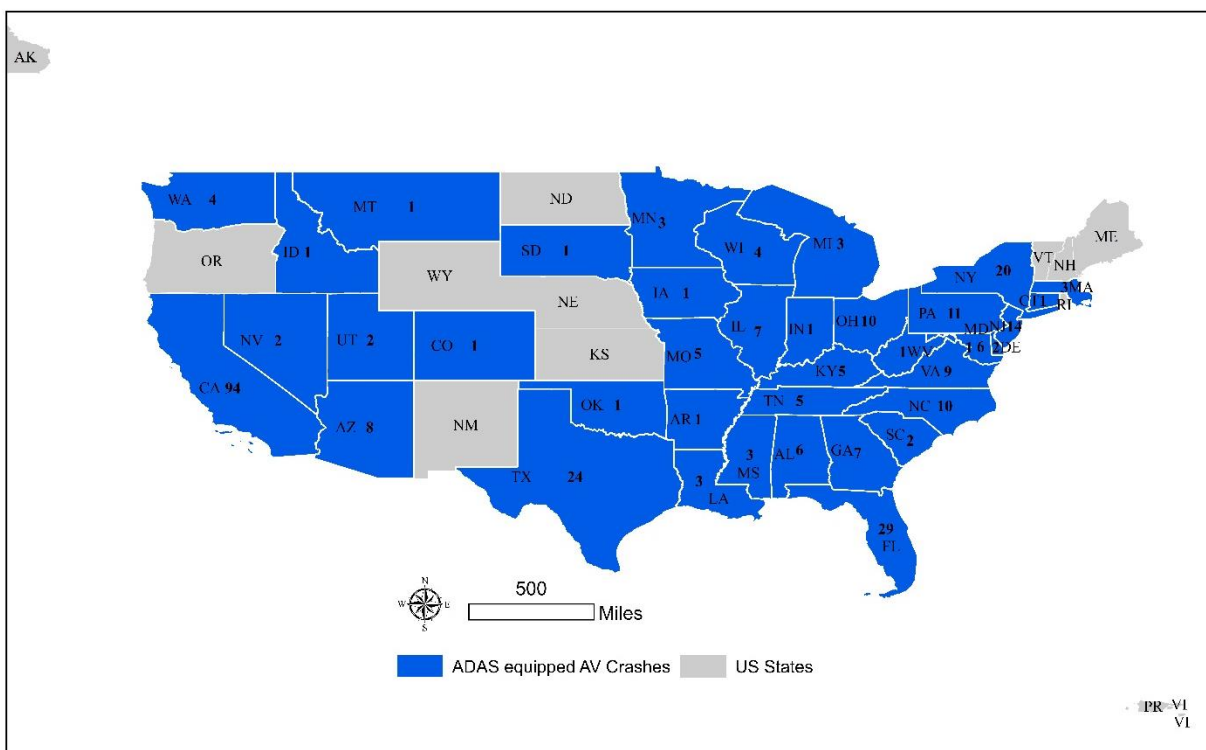


Figure 5.1. ADAS-equipped AV crashes in the US (2021-2022)

Recoding of variables

Variables relevant to crash severity were developed from the existing information in the dataset and by searching the internet. For example, Vehicle Size and Weight for specific make and models are collected from the internet. Most of the variables were categorical and had many levels. For the convenience of analysis, similar categories were combined from the existing dataset to create relevant categories. These variables are explained briefly below:

- **Crash Severity:** There are four categories of crash severity: Serious, Moderate, Minor, Fatality in one variable called “Highest Injury Severity.” Another variable was called “Property Damage” reporting property damage by the crash. Due to fewer data in the Fatality category Fatality and Serious were combined as “Fatal or Severe Injury,” Minor and Moderate were combined as “Injury,” and rest of the observations were considered as “Property Damage Only (PDO), if reported by the “Property Damage” column. Rest were left as Unknown and then removed to have a cleaned data for crash severity.
- **Collision Type:** This variable was made by understanding the narratives, the “CrashWith” variable, and SV and CP contact areas. Collision Type was “Unknown” if these observations in these variables were blank or Unknown. Initially, many types of collisions were identified, such as Angle, Head-on, Sideswipe, Hit animal, Hit fixed objects, Hit pedestrians, Hit trees or pole, Rear-ending CP, being Rear-ended. These were then categorized into six types. Rear-end included both being Rear-ended by other vehicles and Rear-ending other vehicles. Hit animal and pedestrians were combined into one category. Hit fixed object, and Hit trees or poles were combined into one category. Head-on and sideswipe are considered as “Other” category.
- **Pre-crash maneuver of CP:** Pre-crash maneuvers of crash partner vehicle was categorized into six levels. Proceeding straight was one level. Changing lanes, Lane or Road departure, merging, and passing were combined into another level. Left turn and right turns were considered another category. The Other level was formed of all other maneuvers like Backing, crossing, opposing lane, entering traffic, making U-turn, parked, and others. Blanks and Unknowns were combined as the “Unknown”.
- **Vehicle Weight:** At first, weights of the different make and model of cars were search and gathered from the internet. Then three categories were made considering the distribution of the datasets vehicle weights: ≤ 3500 lbs, 3500-4000 lbs, and > 4000 lbs.
- **Vehicle Make:** Out of 315 observations, 223 crashes were by AVs of Tesla company and the rest were AVs from various other companies like Acura, Audi, BMW, Cadillac, Ford, Genesis, Honda, Hyundai, International, Jeep, Kia, Lucid Motors, Mercedes-Benz, Porsche, Subaru, and Toyota. Based on the distribution of AVs involved in crashes, two categories were made: Tesla and Non-Tesla.
- **Pre-crash maneuver of SV:** Pre-crash maneuver of subject vehicles were combined into four categories: Proceeding Straight, Lane Changing/Departure/Merging, Others, and Unknown. Others included Backing, U-turn, Parking, Stopped, Travelling Wrong way, Swerved, and Other.
- **Daytime/Nighttime:** It is divided into two categories: Day and Night. 6:00 am to 7:59 pm is considered day, and 8:00 pm to 5:59 pm is considered night.

- **Peak/Off-peak:** It is also divided into two categories: Peak and Off-peak. Peak hours are 5:00 am-9:00am, 12:00pm-1:00pm, and 4:00pm-8:00pm. The rest is considered Off-peak hours.
- **Roadway Type:** The NHTSA dataset had five road types and many observations were Unknown. Highways were considered one category (Highway/freeways), and intersections, Parking Lot, Rural Road, and Streets were considered the Non-highway category.
- **Subject Vehicle Speeding:** This variable was made from the Posted Speed Limit and Pre-crash speed of the subject vehicles. When Pre-crash speed limit was greater than the posted speed limit it is considered as Yes, otherwise No, and rest are Unknown.
- **Roadway Surface:** Four categories are made: Dry, Snow/Slush/Ice, Wet, and Unknown. “Previously mentioned Other, see narrative” named observations were checked and upon not finding any information, they were categorized as Unknown.
- **Lighting:** Three categories are made for lighting: Daylight, Dawn/Dusk, Dark, and Unknown.
- **Weather condition:** Weather condition categories are kept as they appeared in the NHTSA dataset: Clear, Cloudy, Rain, Snow, Other, and Unknown.
- **Crash With:** Five categories were made: Passenger Car, Fixed Object, Non-Motorist, Other, and Unknown. Hitting fixed objects and hitting pole/tree were considered the Fixed Object category. Non-Motorist category includes Pedestrians and Animals. Other category includes SUV, Van, Bus, Heavy Truck, Pickup Truck, and Motorcycle.
- **Mileage:** Mileage of the vehicles were categorized in intention to capture their usage and age. Four categories were made: <30000 miles, 30000-75000 miles, and >75000 miles, and Unknown.
- **Model Year:** The dataset had AV Models of 2014-2022. Two categories were made: 2014-2018, and 2019-2022.
- **Crash Type:** This variable was made based on the number of entities involved in a crash: single-vehicle and multi-vehicle.
- **Subject Vehicle Damage:** Subject vehicle’s damage condition is developed based on the number of sides of the vehicle that were damaged and whether the vehicle was towed from the crash location or not. If the SV was towed away or towed away and got damaged on more than two sides, or just damaged on more than two sides is considered Major Damage, and the rest is considered Minor or No Damage.
- **Crash Partner Vehicle Damage:** Like the Subject Vehicle Damage, Crash Partner Damage is categorized as Major Damage and Minor or No Damage.

Ordered Logistic Regression Model

The response variable to be applied in the Ordered Logistic Regression (OLS) model needs to be ordered since the nature of the injury severity (dependent variable) in our dataset is ordinal. One of the generalized linear model cases that can account for the ordinal nature of the data is the ordered logistic regression (164). This model is also known as proportional odds logistic

regression. This model provides an exponential relationship between a set of k predictors, $X = (x_1, \dots, x_k)$ and $P = (p_1, \dots, p_n)$ and our response variable $Y = (y_1, \dots, y_n)$. The relationship is as follows:

$$y_i | p_i \sim \text{Bernoulli}(p_i), i = 1, \dots, n \quad (1)$$

$$\text{Logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad (2)$$

The probability of crash severity occurrence for each level i in the model can be estimated, denoted by p_i , by using the maximum likelihood method and solving for $\beta = (\beta_1, \dots, \beta_k)$ (165).

$$p_i = \frac{\exp\{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}\}}{1 + \exp\{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}\}} \quad (3)$$

The crash severity can be estimated after calculating the p_i based on the above formula. This logistic regression can be either binomial regression or multinomial regression. When the response variables are multi-class, multinomial regression can be used. The multiple predictor variables and multinomial response variable in the multinomial regression model have $C-1$ non-overlapping logit models. The c^{th} level of the crash severity occurrence probability, p_{ic} , can be estimated (1).

$$p_{ic} = P(y_i = c | X) = \frac{\exp\{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}\}}{\sum_{c=1}^{C-1} 1 + \exp\{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}\}}, p, c < C \quad (4)$$

The Stata/SE 17.0 is used in this study to analyze the data, and the crash severity is predicted based on the estimated probability, p_{ic} .

Marginal Effect

The results table also highlights the marginal effects of the variables. There are different methods to obtain the marginal effects of the variables. The Average Marginal Effect (AME) method is used in this study. Calculating the marginal effect enables analysts to obtain how much a change in a variable is associated with the expected probability of the response variable.

Results

Frequency Distribution of variables

Table 5.1 shows the frequency distribution of the variables used in the study. The dependent variable is Crash Severity which contains three levels of severity. The full dataset contains 315 observations and among these, 159 observations are from highway/freeways. Among 315 AV crashes, 83.49% are Property Damage Only (PDO) and only 3.17% are fatal or severe injury crashes. Independent variables are broadly categorized under Crash Characteristics, Vehicle Characteristics, Driver Characteristics, and Roadway Characteristics.

Under Crash Characteristics, Collision Type and Subject Vehicle Damage are used. The frequency distribution shows that most crashes were (22.54%) with fixed objects, trees, or poles, followed by Rear-end (19.68%) type collisions. Subject Vehicle Damage indicates that in 50.79% of the cases, the subject vehicle suffered major damage.

Under Vehicle Characteristics, AV, Vehicle Mileage, and Vehicle Weight are used. The frequency distribution shows that in 70.79% of crashes the subject vehicle was Tesla. Distribution of Vehicle Mileage indicates that most crashes involve vehicles whose mileage was ≤ 30000 miles, and only 8% of the subject vehicles had mileage over 75000 miles. This indicates that most of the subject vehicles were comparatively new or less driven. The distribution of Vehicle Weight shows that

45.71% of vehicles weighed >4000 lbs, and 35.87% weighed 3500-4000 lbs. This means that most crashes involved comparatively heavier vehicles.

Under Driving Characteristics, Subject Vehicle Speeding, Precrash Maneuver of SV, and Precrash Maneuver of CV are used. Subject Vehicle Speeding indicates that only 12.06% of the time, SV drivers were driving over the speed limit, and 47.03% of the time, they were not speeding. Of 315 crashes, the vehicles drivers of the subject vehicles were going straight in 54.29% of the cases and only 4% time they were turning. Crash partners were proceeding straight most of the time and most of the time crash partner vehicle's driver maneuvers were unknown. The vehicles were stopped in a traffic lane, slowing or starting in a traffic lane, turning left or right, and changing lanes or merging in 18.13%, 10.94%, 10%, and 1.88% of cases, respectively. It is noteworthy that, in most crashes, both SV and CP were Proceeding Straight (54.29% and 10.48%, respectively) right before the crash (Pre-crash Maneuver).

Under Roadway Characteristics, Roadway Type and Roadway Surface are used. Frequency distribution of Roadway Type indicates that half (50.48%) of crashes happened on highway/freeways and the rest happened in intersections, rural roads, streets, or parking lots. This makes sense because vehicles drive at high speed on highway/freeways and have a higher chance of severe crashes. The frequency distribution of the variables follows a similar pattern on the highways, yet more crash frequency on the highways requires in-depth analysis.

The frequency distribution of some other important variables that are not used in the models is listed in **Table 5.2**. In contrast to Subject Vehicle Damage, Crash Partner Damage values indicate that fewer CPs had major damage (8.89%). Among the crashes, Single-Vehicle crashes were 26.67% and Multi-Vehicle crashes were 45.71%. In most crashes, SVs hit fixed object, trees, or poles and only 16.51% time, they hit another passenger vehicle. The vehicles involved in the crashes were mainly SUVs and Sedans. Also, most of the crashes happened in daylight conditions, clear weather, and off-peak hours.

Ordered Logistic Regression Model Results

Crashes in all road types

Table 5.3 shows the result of the Ordered Logistic Regression (OLR) model. The likelihood ratio test shows that the OLR model is meaningful and better than the null model ($LR\ chi^2 = 85.03$, $Prob > \chi^2 = 0.00$). Pseudo $R^2 = 0.26$ indicates that the model fits the data reasonably well. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were used as model selection criteria to compare models with different variables and select the best one.

The OLR represents the relationship between the contributing factors and severity of AV-involved crashes. The variables with a minimum of one statistically significant category (p-value <0.05) with acceptable signs and magnitude are kept in the model. Marginal effects (ME) are calculated on the model estimates for all categories of crash severity and shown in **Table 5.3** and **Figure 5.2**. The marginal effect will indicate the percentage point change in the probability of a certain level of injury if an explanatory variable is changed by one unit or moved to the target level from the base level (166).

For example, among the vehicle characteristics, Tesla has a coefficient of -2.52 which is statistically significant. This indicates that if the AV is Tesla, it is less likely to be associated with fatal or severe injury. The marginal effects of AV indicate that if an AV is Tesla, it has 31% higher

chances of Property Damage Only crash compared to other ADAS-equipped motor vehicles (Not Tesla). However, Tesla showed 25% lesser chances of Injury type crashes and 7% lesser chances of Fatal or Severe Injury crashes compared to other ADAS-equipped AVs. It infers that ADAS-equipped Tesla vehicles have lesser chances of a severe crash, given a crash. The Vehicle Weight variable indicates that an ADAS-equipped AV of >4000 pounds is associated with 4% higher chances of Fatal Injury, 8% higher chances of Injury, and 11% lesser chances of PDO crashes compared to ADAS-equipped AVs that are ≤ 3500 pounds. This is mention worthy that NHTSA reports severity of the crash but does not specify who survived or suffered the injury between the subject vehicle, crash partner or non-motorists. Although heavier vehicles might suffer less damage, the crash partner vehicle might suffer more injury and damage from a heavier vehicle.

Among the Driver characteristics, Subject Vehicle Speeding indicates that driving above the posted speed limit is associated with a 3% higher chance of fatal or severe injury, 11% higher chance of injury, and 14% lesser chance of PDO crash than compared to ADAS-equipped AVs driving below the speed limit. This is similar to conventional vehicles also (167). On the other hand, marginal effects of the Pre-crash maneuver of CP indicate that compared to CPs that have been driving straight, CPs that were at a stopped position just before the crash was associated with a 5% lesser chance of fatal injury and 12% lesser chance of injury crashes. This also makes sense because driving straight has been the cause of the highest number of fatal injuries in 2019 in the US and it has the risk of rear-end and head-on collisions, which are often fatal (168).

Among Crash characteristics, the Collision Type variable indicates that compared to Angle crashes, a Hit pedestrian or animal type crash is associated with a 7% higher chance of a fatal crash and an 8% higher chance of an Injury crash. This makes sense because when an animal or a pedestrian gets hit by a vehicle, it will have a higher chance of injury or fatality.

Discussion

Factors Associated with AV Crash Severity

Automated vehicles are expected on the roads more when they will reach the entire level of automation. However, it is still essential to know where low-level automated vehicles like ADAS-equipped vehicles stand to achieve road safety. Margins plots in **Figure 5.2** give a graphical representation of the probability of different levels of crash severity for ADAS-equipped AVs for the full dataset.

Association of Vehicle Characteristics and Crash Severity

It is observed among the ADAS-equipped AVs that Tesla vehicles have a lower chance of Injury crash than Fatal crashes and higher chance of PDO crash than other crashes overall. A previous study found 7.5% of fatal and severe crashes can be avoided using adaptive cruise control (ACC) systems (169). It is apparent from the marginal plots that heavier-weight AVs (>4000 lbs.) have a higher chance of association with injury and fatal crashes. It was observed in a previous study that in the event of a collision between two unequal mass-based cars, the fatality risk ratio of the lighter car increases as a power function of mass ratio (170). As NHTSA specifically does not mention the entity who suffered the crash severity, it can be deduced that whenever a heavy AV collides with another lighter vehicle, the heavier vehicle is expected to cause more severe injury to the lighter vehicle.

Table 5.1. Frequency Distribution of Important Variables

Variables		Categories	Full Dataset (N=315)		Highway/Freeways (N=159)		Non-Highway (N=92)	
			Freq.	(%)	Freq.	(%)	Freq.	(%)
Crash Characteristics	Crash Severity	Fatal or Severe Injury	10	3.17	4	2.52	1	1.09
		Injury	42	13.33	14	8.81	20	21.74
		Property Damage Only (PDO)	263	83.49	141	88.68	71	77.17
	Collision Type	Hit pedestrian or animal	13	4.13	9	5.66	1	1.09
		Angle	47	14.92	8	5.03	33	35.87
		Hit fixed object, trees or pole	71	22.54	46	28.93	17	18.48
		Other	20	6.35	12	7.55	7	7.61
		Rear-end	62	19.68	33	20.75	16	17.39
		Unknown	102	32.38	51	32.08	18	19.57
	Subject Vehicle Damage	Major Damage	160	50.79	74	46.54	54	58.70
		Minor or No Damage	78	24.76	48	30.19	26	28.26
		Unknown	77	24.44	37	23.27	12	13.04
Vehicle Characteristics	AV	Non-Tesla	92	29.21	27	16.98	30	32.61
		Tesla	223	70.79	132	83.02	62	67.39
	Vehicle Mileage	<=30000 miles	133	42.22	85	53.46	28	30.43
		30000-75000 miles	87	27.62	46	28.93	33	35.87
		>75000 miles	28	8.89	19	11.95	5	5.43
		Unknown	67	21.27	9	5.66	26	28.26
	Vehicle Weight	<=3500 lbs	58	18.41	16	10.06	18	19.57
		3500-4000 lbs	113	35.87	54	33.96	44	47.83
		>4000 lbs	144	45.71	89	55.97	30	32.61
Driver Characteristics	Subject Vehicle Speeding	No	149	47.30	99	62.26	48	52.17
		Unknown	128	40.63	33	20.75	33	35.87
		Yes	38	12.06	27	16.98	11	11.96

Table 5.1 Continued.

Variables		Categories	Full Dataset (N=315)		Highway/Freeways (N=159)		Non=Highways (N=92)	
			Freq.	(%)	Freq.	(%)	Freq.	(%)
Roadway Characteristics	Pre-crash maneuver of SV	Proceeding Straight	171	54.29	100	62.89	57	61.96
		Other	41	13.02	25	15.72	11	11.96
		Turning	14	4.44	1	0.63	11	11.96
		Unknown	89	28.25	33	20.75	13	14.13
	Pre-crash maneuver of CP	Changing lane, Departure, Merging, Passing	18	5.71	15	9.43	2	2.17
		Other	19	6.03	10	6.29	7	7.61
		Proceeding Straight	33	10.48	15	9.43	15	16.30
		Stopped	21	6.67	9	5.66	8	8.70
		Turning	16	5.08	0	0.00	16	17.39
		Unknown	208	66.03	110	69.18	44	47.83
	Roadway Type	Highway/Freeway	159	50.48	-	-	-	-
		Other	92	29.21	-	-	-	-
		Unknown	64	20.32	-	-	-	-
	Roadway Surface	Dry	137	43.49	82	51.57	54	58.70
		Unknown	135	42.86	41	25.79	32	34.78
		Wet/Snow/Slush/Ice	43	13.65	36	22.64	6	6.52

Table 5.2. Frequency Distribution of Other Important Variables

Variables		Categories	Full Dataset (N=315)		Highway/Freeways (N=159)		Non-Highway (N=92)	
			Freq.	(%)	Freq.	(%)	Freq.	(%)
Crash Characteristics	Crash Partner Damage	Major Damage	28	8.89	17	10.69	10	10.87
		Minor or No Damage	155	49.21	86	54.09	51	55.43
		Unknown	132	41.90	56	35.22	31	33.70
	Crash Type	Single Vehicle	84	26.67	55	34.59	18	19.57
		Multi-Vehicle	144	45.71	56	35.22	60	65.22
		Unknown	87	27.62	48	30.19	14	15.22
	Crash With	Fixed Object	71	22.54	46	28.93	17	18.48
		Non-Motorist	13	4.13	9	5.66	1	1.09
		Other	68	21.59	40	25.16	23	25.00
		Passenger Car	52	16.51	25	15.72	22	23.91
Vehicle Characteristics	Model Year	Unknown	111	35.24	39	24.53	29	31.52
		2014-2018	77	24.44	37	23.27	24	26.09
		2019-2021	238	75.56	122	76.73	68	73.91
	Vehicle Size	Hatchback	32	10.16	18	11.32	8	8.70
		Passenger Van	4	1.27	0	0.00	1	1.09
		Pickup Truck	4	1.27	2	1.26	1	1.09
		SUV	127	40.32	81	50.94	26	28.26
		Sedan	145	46.03	57	35.85	56	60.87
		Subcompact SUV	3	0.95	1	0.63	0	0.00
		Dark Condition	70	22.22	48	30.19	21	22.83
Environmental Characteristics	Lighting	Daylight	104	33.02	68	42.77	36	39.13
		Unknown	141	44.76	43	27.04	35	38.04
	Weather	Clear/Cloudy	141	44.76	84	52.83	55	59.78
		Other/Unknown	133	42.22	2	1.26	32	34.78
		Rain/Snow	41	13.02	35	22.01	5	5.43
Temporal Characteristics	Day/Night	Day	128	40.63	79	49.69	38	41.30
		Night	115	36.51	70	44.03	27	29.35
		Unknown	72	22.86	10	6.29	27	29.35
	Peak/Offpeak	Peak	93	29.52	56	35.22	29	31.52
		Off-peak	150	47.62	93	58.49	36	39.13
		Unknown	72	22.86	10	6.29	27	29.35

Table 5.3. Ordered Logistic Regression Model

Variables		Categories		Coef.	Std.err.	P> z	Marginal Effects				
Dependent Variable	Crash Severity						Property Damage Only	Injury	Fatal or Severe Injury		
Independent Variable	Vehicle Characteristics	Vehicle Make	Not Tesla (Base)								
			Tesla	-2.52	0.57	0.00	0.31	-0.25	-0.07		
		Vehicle Weight	≤3500 lbs (base)								
			3500-4000 lbs	0.75	0.61	0.22	-0.06	0.05	0.02		
	>4000 lbs		1.27	0.47	0.01	-0.11	0.08	0.04			
	Driver Characteristics	Subject Vehicle Speeding	No (Base)								
			Unknown	1.55	0.62	0.01	-0.14	0.11	0.03		
			Yes	1.56	0.65	0.02	-0.14	0.11	0.03		
		Pre-crash maneuver of CP	Proceeding Straight (Base)								
			Changing lane, Departure, Merging, Passing	-0.96	1.00	0.34	0.10	-0.07	-0.03		
			Other	-0.52	0.80	0.52	0.06	-0.04	-0.02		
			Stopped	-1.76	0.87	0.04	0.17	-0.12	-0.05		
			Turning	-0.37	1.26	0.77	0.04	-0.03	-0.02		
			Unknown	-0.88	0.61	0.15	0.10	-0.07	-0.03		
			Crash Characteristics	Collision Type	Angle (Base)						
					Hit fixed object, trees or pole	1.06	0.66	0.11	-0.11	0.08	0.03
					Hit pedestrian or animal	1.75	0.84	0.04	-0.20	0.13	0.07
					Other	-0.58	0.91	0.52	0.04	-0.03	-0.01
					Rear-end	0.55	0.60	0.36	-0.05	0.04	0.01
	Unknown	0.11			0.64	0.86	-0.01	0.01	0.00		
	cut 1	1.88			0.89						
	cut 2	4.11	0.94								
Number of observations = 315, DF = 17											
LR chi² (15) = 85.03, Prob > chi² = 0.00											
LL(null) = -166.58, LL (Model) = -124.06											
Pseudo R² = 0.26, AIC = 282.12, BIC = 345.91											

The level two driver assistance system's role is unclear from here. Three of four Tesla models (S, X, Y) are heavier than 4000 pounds and only Tesla Model 3 is less than 4000 lbs. Even though Tesla cars can be involved in injury or severe crashes, it does not indicate that their drivers or occupants will suffer more injury. A recent study on California DMV's data on autonomous vehicles showed a significant strong positive correlation of 0.73 ($p < 0.1$) between monthly accidents and autonomous miles traveled. More miles driven in autonomous mode means more exposure and hence more crashes, and the crash could be due to other vehicles being at fault or due to failure of AVs automation features (171). The study on ADAS AVs finds that AVs with >75000 miles mileage have higher chances of injury and fatal crash but a lesser chance of PDO crash compared to the mileage of ≤ 3500 miles. Along with indicating more miles driven, mileage can also indicate older cars, and if newer cars are likely to have lesser severe crashes, then this can be an indication of better functioning AV features in newer AVs. However, it needs more data and in-depth analysis to come to this conclusion.

Association of Driver Characteristics and Crash Severity

In conventional vehicles, common driving behaviors that cause fatal crashes are impaired driving, inattentive driving, speeding, drowsy driving, and not wearing seatbelts, according to NHTSA website (172). Similar to conventional vehicles, ADAS-equipped AVs also showed higher chances of injury and severe injury crashes but lesser chances of PDO crashes when the subject vehicle drives above the posted speed limit. Speeding is a human error that leads to many crashes (173). Automation safety features like ACC are recently designed to operate at high and low speeds and in dense and non-critical traffic conditions (174). As 71% of (27 or 38) speeding-involved crashes occurred on Highway/Freeways, according to the dataset, AV manufacturers should investigate more about these high-speed crashes to improve their cruise control safety features. Another pre-crash maneuver of SV compared to proceeding straight showed a lesser chance of Injury or Fatal/severe injury in ADAS-equipped AVs. This can make sense because severe injury or fatalities are caused due to head-on and rear-end crashes and these crashes occur when vehicles proceed straight and collide with each other. During AV subject vehicles were proceeding straight (171 of 315), 25% (43 of 171) times they hit a fixed object, trees, or poles, 22% times they caused Rear-end crashes, and 6% (11 of 171) times they caused either Head-on or sideswipe. In-depth analysis should be conducted on AV drivers when driving straight. Proceeding or driving straight for a long time can be tedious and drivers might get distracted and fall into a crash. Researchers can study ways to improve AV safety features, so drivers do not get bored or distracted. Crash partners of AVs with stopped pre-crash maneuvers compared to those proceeding straight showed lesser chances of Injury or Fatal/severe crashes. In this dataset, 21 crashes had stopped CP, but only 3 were Injury crashes, and the rest were PDO, and of 21 crashes, most were Rear-end crashes. A plausible explanation could be the fact that most of the time, CP had no occupants while it was stopped on the road, or it was a heavier vehicle than the SV. The limitation of the dataset was that out of 315 observations, data on CP was unknown for 208 cases. Robust data on CP could have provided more insights.

Association of Crash Characteristics and Crash Severity

Compared to Angle crashes, if AVs hit a pedestrian or animal, that was associated with higher chances of injury or fatal crashes. This is evident that pedestrian is most vulnerable as they are more exposed than any other entity, and ADAS safety features like Adaptive Cruise Control or Lane Keeping System (LKS) do not disengage for pedestrians or animals. Hence, injury or severe

injury is inevitable if the driver do not take control and take action. In this situation, ADAS-equipped AVs are not different than conventional vehicles if they do not have a pedestrian detection system. Only a few observations in the dataset had narratives, and none of the narratives showed that these AVs had pedestrian detection systems. Tesla vehicles had more severe or fatal crashes on highways/freeways, but it was found in a previous study that severe AV crashes occur on highways (150). Overall, the crash severity results are similar to the results reported by existing literature on AV crashes, except Tesla vehicles showed lesser association with injury or fatality, and heavier weight AV vehicles showed a higher association with Injury and Fatal injury crashes.

AV Crashes at Different Road Types

Important variables are observed across Highway/Freeways and Non-Highways for comparison. **Figure 5.3 (a)** shows AV crashes of different Vehicle Make (Tesla vs. Non-Tesla) across roadway types. It is seen that highway/freeways have more Tesla crashes (132 of 159). There are 30 and 27 Non-Tesla vehicle crashes on Non-Highways and in highway/freeways. This means that highways/freeways have more crashes by Tesla than other AVs, but non-highways have more crashes by other AVs than Tesla. Next, it is essential to see what types of crashes occur on highways and non-highways by these AVs.

Figure 5.3 (b) shows that among these AV injury crashes, 14 are on highway/freeways and 20 are on non-highways. Non-highways have only one fatal or severe crash and highway/freeways have four fatal or severe crashes. The percentage of fatal crashes are not as high as PDO or Injury crashes on neither of highway/freeways nor non-highways. Injury crashes happened more in non-highway areas than in highways/freeways, and fatal or severe injuries happened more in the highway/freeways than in non-highways. This is similar to conventional vehicles. However, **Figure 5.5** shows that crash rates for PDO crashes (80.61%) and Fatal or Severe crashes (70%) are higher for Tesla, but Injury crash rates are (90.48%) higher in non-Tesla AVs. Despite this distribution, the model showed a negative association between crash severity and Tesla vehicles compared to non-tesla vehicles. This could be because the ratio of Fatal or Severe crashes (7 of 223 Tesla crashes) and Total crashes is smaller for Tesla and higher for Non-Tesla (3 of 92 Non-Tesla crashes). Next, knowing what entities are involved in these crashes is essential. **Figure 5.3 (c)** indicates that highway/freeways have a higher percentage of crashes with everything listed under the Crash With variable. But, highway/freeways and non-highway, both these categories have a similar number of crashes with Passenger Cars (48.08% on highway/freeways and 42.31% on Non-Highways).

Figure 5.4 (a) shows the collision types in highway/freeways and non-highways. The highway/freeways have mostly been hit with fixed object/trees/pole type of crashes followed by Rear-end crashes. Non-highways have mostly Angle crashes followed by Hit fixed object/trees/pole and Rear-end crashes. For AVs in this dataset, the number of Head-on or Sideswipe crashes was low. **Figure 5.4 (b)** indicates that in most of the vehicle-involved crashes, the crash partner vehicle was proceeding straight or changing lanes/departure/merging in the highways/freeways. But in the non-highway AV crashes, the crash partners were mostly turning, followed by proceeding straight just before the crash. This also makes sense and matches conventional vehicles' situations.

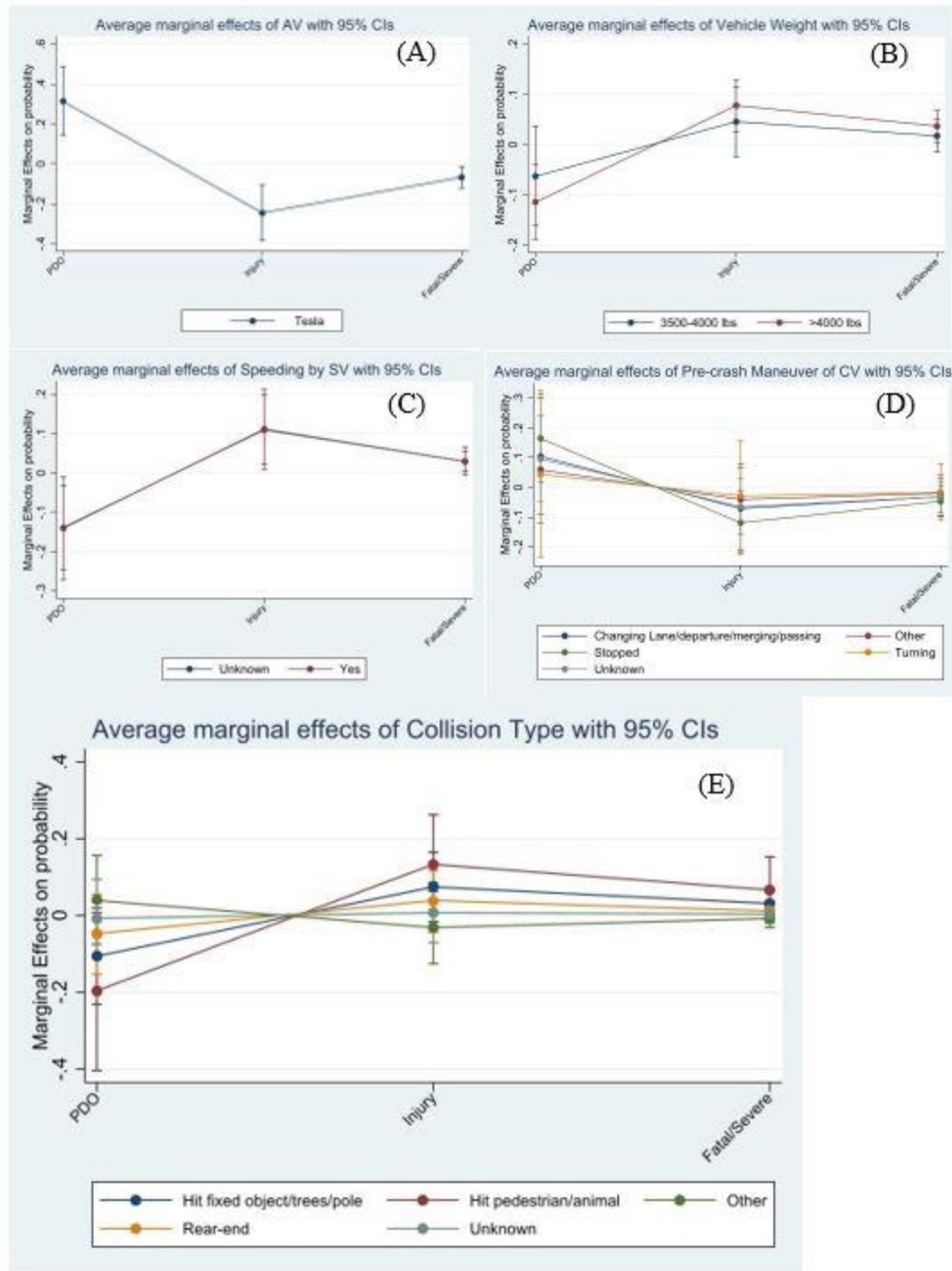


Figure 5.2 Margins plot of the probability of crashes-based on (A) Vehicle Make, (B) Vehicle Weight, (C) Speeding of Subject Vehicle, (D) Pre-crash maneuver of Crash Partner Vehicle, and (E) Collision Type

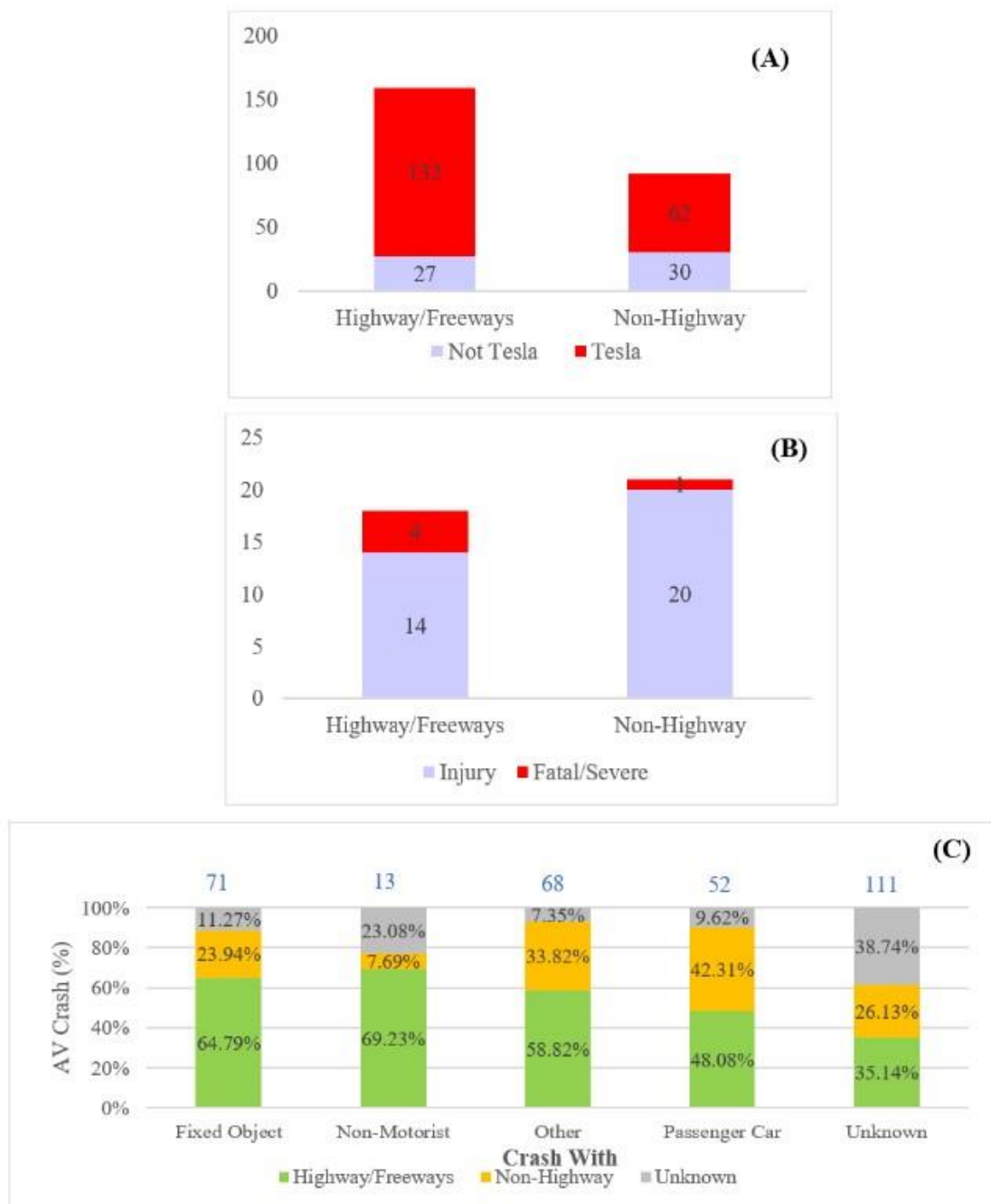


Figure 5.3 (A) Vehicle Make, (B) AV Crash Severity, and (C) Crash With at Different Roadway Types

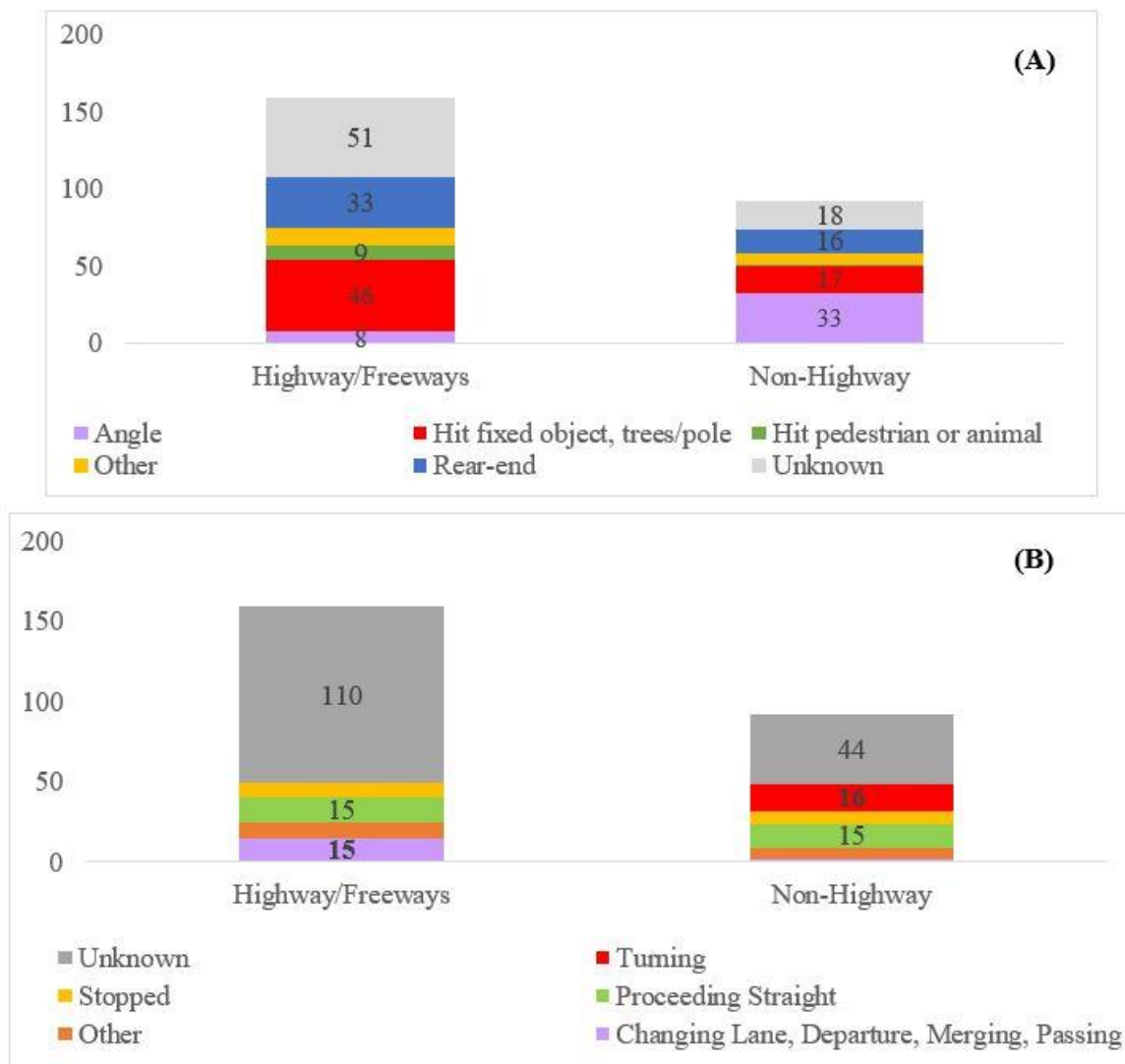


Figure 5.4 (A) Collision Type and (B) Pre-crash maneuver of Crash Partner Vehicle at Different Roadway Types

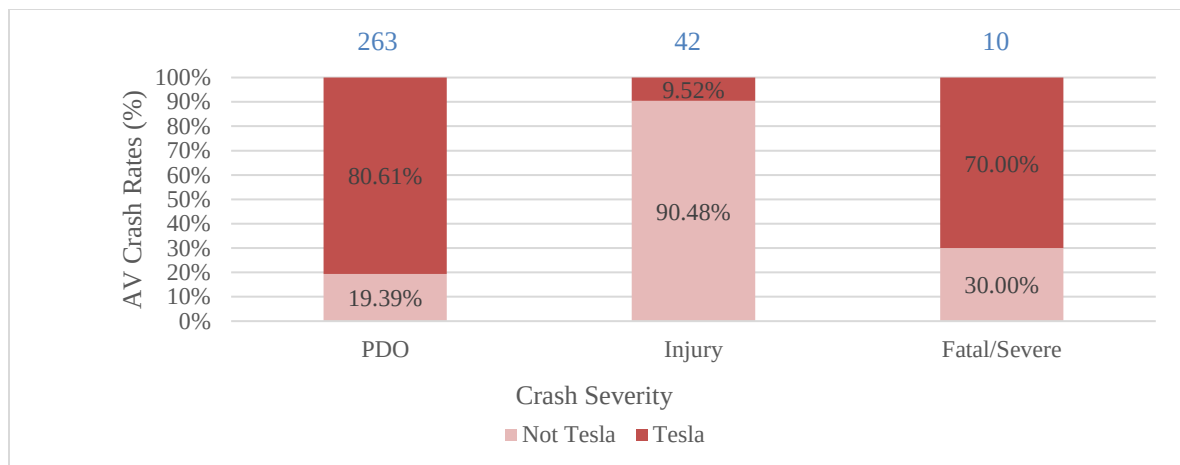


Figure 5.5 Crash Rates by AV Type

Limitations

One limitation of this dataset is that many observations of most variables had “Unknown” or blanks. Many observations and variables had to be discarded to get a complete dataset with all variables. But removing those observations would have reduced the sample size a lot.

Hence, models were developed keeping some of those “Unknown” values in the dataset. Ideally, a large dataset with a complete set of observations would have made the analysis more precise and accurate. The data did not specify which entity suffered the crash and whose crash severity was reported. NHTSA published the injury of involved entities, and the sufferer could be either the subject vehicles’ or crash partner vehicles’ driver or occupants, pedestrian, or animal. Specification of such information would have given the study more clarity. The dataset does not contain the exact crash location (latitude/longitude), demographic information on drivers or occupants, damage severity of the vehicles, and land use information. NHTSA could have provided more information about the crash, like collision type and damage percentage. Although these variables were developed using other information like crash with, how many sides and which sides of the vehicles were damaged, and whether the vehicles were towed or not. The analysis in this could have been made more robust if this information had been present in the dataset. The dataset contains a variable called “Automation System Engaged” and reports the engagement of ADAS.

The dictionary explains that this variable indicates engagement of the highest level of driving automation at any time during the 30 seconds before the crash. However, it was found in a report published on the ADAS dataset by NHTSA that it is not confirmed what system was engaged during the crash. This information had to be retrieved from the narratives of the crash reports in the dataset, but only a few observations could be confirmed about having ADAS engaged. From the narratives, only a few crashes were confirmed to have disengaged. Therefore, assumptions were made that apart from those crashes, the rest of the dataset contained crashes in which the ADAS was engaged. In the case of roadway type, most of the crashes were from highways/freeways, and very few observations were available from intersections, rural roads, streets, or parking lots. A smaller sample size of these non-highway locations made it difficult to develop suitable models on these segments from which conclusions could be drawn.

The results from a Ordered Logistic Regression model were presented as the final model in this paper. But more sophisticated models were also applied to the dataset to get more insight. For example, Gologit, MNL, Mixed-Effects Ordered Logistic, and Bayesian models were also applied. However, due to the data limitation in each level of severity in many levels of the categorical variables the sophisticated models did not converge. Each sophisticated model converged at different specification. Hence, a comparison of different models was not possible. Each sophisticated model converged when it had only 2-3 variables which appeared as weak models compared to the Ordered Logistic model.

Moreover, the Segmented model of the exact specification as the full model did not provide good results. All these issues could have been addressed if the data had been more robust and more significant in sample size. AV crashes were observed the most in Tesla vehicles compared to other ADAS-equipped AVs. Therefore, the Vehicle Make variable was only categorized as Tesla vs. Non-Tesla. However, the study does not intend to emphasize one AV make over the others. Here, the categorization is influenced by the distribution of the samples in the data.

Conclusions

While the market penetration of vehicle automation technologies (SAE Levels one, two, and three) has increased, many vehicles equipped with such technologies have been involved in traffic accidents. Hence, it is becoming a prime concern for the transportation safety industry to understand AV safety clearly. Investigating such crashes can provide helpful information about the performance of ADAS technologies. To this end, this study explores the association between crash severity and various factors responsible for crashes in AVs. The study uses real-world crash data from the NHTSA database, including 315 ADAS-equipped crashes from August 2019 to May 2022 and explores the association of crash severity with these AVs. The model results are generalizable for all AV crashes in the US, and the method can be reapplied in similar studies.

The key findings of the study are as follows:

1. The marginal effects of the Vehicle Make variable indicate that if the vehicle is Tesla, then it has 31% higher chances of Property Damage Only crash compared to other ADAS equipped AVs. However, Tesla showed 25% lesser chances of Injury type crash and 7% lesser chances of Fatal or Severe Injury crash compared to other ADAS equipped AVs.
2. The Vehicle Weight variable indicates that an ADAS-equipped AV of >4000 pounds is associated with 4% higher chances of Fatal Injury, 8% higher chances of Injury, and 11% lesser chances of PDO crashes compared to other ADAS-equipped AVs that are ≤ 3500 pounds.
3. Subject Vehicle Speeding indicates that driving above the posted speed limit is associated with 3% higher chance of fatal or severe injury, 11% higher chance of injury, and 14% lesser chance of PDO crash than compared to ADAS-equipped AVs that drive below the speed limit.
4. The Pre-crash maneuver of CP indicates that compared to CPs that have been driving straight CPs that were at a stopped position just before the crash were associated with a 5% lesser chance of fatal injury and a 12% lesser chance of injury crashes.
5. The Collision Type variable indicates that compared to Angle crashes, Hit pedestrian or animal is associated with a 7% higher chance of a fatal crash and an 8% higher chance of an Injury crash.
6. Crash rates for PDO crashes (80.61%) and Fatal or Severe crashes (70%) are higher for Tesla, but Injury crash rates are (90.48%) higher in Non-Tesla AVs.
7. Highways/Freeways have more crashes by Tesla than other AVs, but non-Highways have more crashes by other AVs than Tesla.

Future research can include more spatial and temporal information on the crashes, driver demography and behavior information, road geometry information, and more on AV technologies for better analysis of crash severity. The findings from the paper might be interesting and insightful for both the manufacturers and users of Tesla, along with other AVs, as they are constantly attempting to achieve a higher automation level. The model can be adopted by automated vehicle safety researchers immediately for similar studies. This chapter can be helpful for traffic safety practitioners, engineers, policymakers, and city planners who work to prepare their cities for autonomous vehicles.

Acknowledgments

The research was partially supported by the Collaborative Sciences Center for Road Safety, a US Department of Transportation National University Transportation Center promoting safety.

**CHAPTER 6. UNDERSTANDING THE SURVIVABILITY OF DRIVERS
FROM MEDIA-REPORTED AUTOMATED VEHICLE-INVOLVED
FATAL CRASHES: A CASE STUDY OF TESLA-INVOLVED FATAL
CRASHES**

Abstract

As on-road automated vehicles (AV) are on the rise, evaluating their safe operation becomes critical. Generally, AVs can be relatively safer than human-driven vehicles because they do not get distracted, violate laws, or make errors. Conversely, AVs can malfunction for various reasons and cause crashes. This study investigates the role of Tesla Autopilot engagement and human behaviors on the survivability of the Tesla driver in Tesla-involved media-reported fatal crashes. A unique dataset is created from a Tesla fatal crash data source (tesladeaths.com). Several new variables are coded using links to the news narratives for each fatal crash. The descriptive analysis provides insights into the survivability of the Tesla driver in 46 single-vehicle and 80 multi-vehicle crashes in the US. When the Autopilot was not claimed in the single-vehicle crashes, Tesla drivers died in 75% (N=30) cases and survived 25% of cases (N=10). For multi-vehicle crashes, 75% (N=3) of the time driver was killed and 25% (N=1) time they survived when Autopilot was claimed. Tesla drivers survived in 20% of cases of fire-involved multi-vehicles crashes. The binary logistic regression model suggests that if a fatal crash happens to Tesla Vehicles, the Tesla Driver is more likely to be killed when the Tesla Vehicle gets fire, the Tesla Driver makes error, or if the Tesla Drivers' age increases. Future research can investigate the change in Tesla driver behaviors in non-fatal crashes. This study will be helpful for AV manufacturers, notably Tesla, who are working to improve the safety features of their vehicles.

Introduction

Human factors, including driving errors and violations, contribute to more than 90% of road crashes (175-177). Due to technological advancements, automated vehicles will significantly change driver-vehicle interaction in the future (178). The automated vehicles include safety features such as one or more driver assistance systems, which have the potential to reduce a significant percentage of crashes while eliminating or reducing human control (179-182). Some of the driver assistance features include adaptive cruise control (183; 184), a forward-collision warning system (185; 186), a dilemma zone mitigation system (183), a blind-spot assistance system (187-189), a lateral vehicle detection system (190; 191), and a low-speed automatic brake system (192). One of the key advantages of automated vehicles with driver assistance technologies is that these vehicles promise to significantly reduce both crash risk and severity by eliminating or reducing a significant portion of one or more driving errors and/or violations (193-195). Tesla has become largely popular partly for their integration of "Autopilot" features in each of its vehicle models. The baseline level of Autopilot that is included with each model includes the ability of the Tesla vehicle to steer, accelerate, and brake autonomously and automatically within its lane. This feature is a bit of a misnomer, as Tesla's Autopilot system requires driver supervision during operation, meaning that the vehicle is not fully autonomous. Vehicle automation is predicted to increase road safety by removing human errors. However, new driving technology can also present safety risks since drivers may need time to learn how to properly use the systems. It is considered that more safety benefits could be achieved if the automated vehicles are deployed on a large scale (175). For instance, many vehicles in the existing transportation system are still human-controlled, meaning that automated vehicles may not fully achieve the safety benefits. Furthermore, most of the automated vehicles in the market are level one or two, which shows that human control still exists in these vehicles and could contribute to crashes. Introducing these automated vehicles onto the roadways (where most of the vehicles are conventional or human-controlled) necessitates an analysis of crashes involving Tesla to identify interesting data trends that may help assess the

safety of Tesla vehicles. While the small number of previous studies related to Automated Vehicles (AV) crashes provide useful insights, there is a need for further research on the topic, which could generate new knowledge about AV safety. The role of the Autopilot system, driver behavior, vehicle, and other roadway and environmental factors in AV crashes has been lightly addressed due to the limited AV crash data (150). This study contributes to the limited research by exploring how the Autopilot performance and behaviors of Tesla drivers relate to survivability (killed versus survived) of Tesla drivers involved in a fatal crash that involves only Tesla vehicles-single-vehicle crashes. Investigating the relationship of Autopilot performance, behaviors of Tesla drivers, and other key roadway and vehicular factors with the chance of Tesla driver survivability is considered to provide a deeper and clearer picture of the role of these factors in single-vehicle crashes, especially when another vehicle or driver is not involved.

The conceptual framework of this study is shown in **Figure 6.1**. This study aims to understand how autopilot engagement of the vehicle, errors, and violations by Tesla drivers (in single-vehicle and multi-vehicle crashes), and other environmental and vehicle attributes relate to the survivability of Tesla drivers. To meet the study objectives, Tesla driver fatality is coded as 0 and 1, where 1 indicates crashes in which the Tesla driver was killed. To quantify the effect of a specific explanatory variable on the Tesla driver fatality, a directional (asymmetric) Somers' Delta (D) test and a binary logistic regression model are used.

Literature Review

There is a limited amount of research that investigates the contributing factors in AV crashes (152; 196). These studies generally have low sample sizes due to the relative novelty of AV technology. Like this study, many researchers have pulled data from narrative sources such as crash reports (150; 152; 197; 198). Of the few studies investigating the impact of human behavior in autonomous vehicle crashes, the consensus is that drivers must remain alert behind the wheel while Autopilot is in use, as automation has not yet reached the capacity for fully autonomous driving and human errors and violations are still plausible (199; 200). Other investigated contributing factors include weather conditions, roadway infrastructure, crash type, and autopilot engagement. It is found that many of these factors are also at play during autonomous vehicle crashes (150; 152; 198; 200). Due to the impacts of roadway infrastructure on crash severity, the importance of timely roadway maintenance and management is critical as the number of AVs on the roadways increases (200). CAV driving mode, i.e., whether Autopilot is engaged, is also found to be a contributing factor in AV crashes (198). There is a need to further investigate AV crashes, especially concerning the roles of human behavior and Autopilot engagement. **Table 6.1** gives a summary of the reviewed literature.

Data

One of the reasons for the limited research related to the safety of AVs is that sufficient field data about AV crashes is not available (150). This study takes advantage of the unique crash data source (201), which includes information and weblinks for media reports on Tesla-involved Fatal crashes. This website was developed privately by people interested in Tesla car technology and Media-reported fatal Tesla crashes. They collect information on media-reported fatal Tesla crashes or incidents from local news archives and government databases that are in the English language.

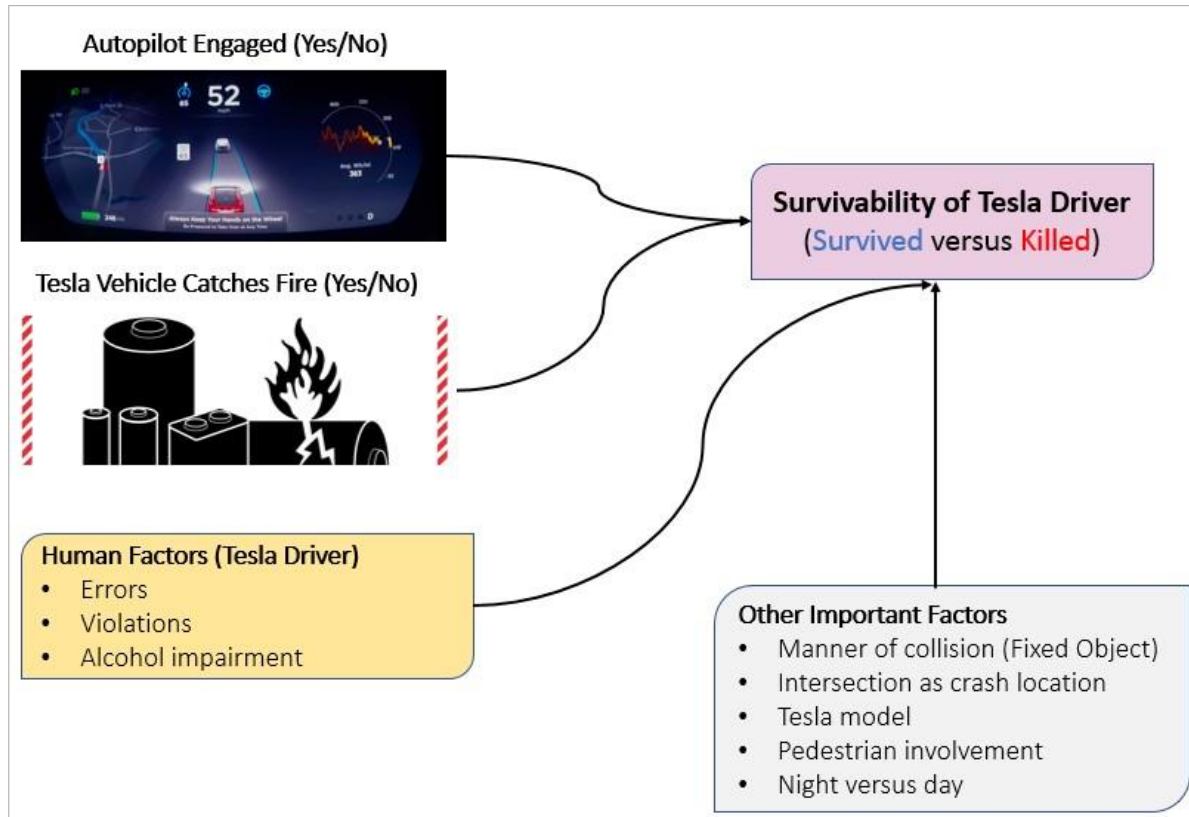


Figure 6.1. Conceptual Framework

The dataset undercounts total deaths involving Tesla because it mainly does not obtain information from non-English speaking geographies. However, this dataset was mentioned in a published paper and a Doctoral Dissertation as a source of Tesla death or fatal crash reports (202; 203). The records for a total of 225 Media-reported fatal Tesla crashes were available on the website (as reported through June 2022). The website provides information about the year, date, location, model of Tesla, whether or not (and how many) drivers and occupants (in Tesla or other vehicles) were killed in each crash, details about the involvement and injury to pedestrians, and information on whether the Autopilot was claimed at the time of the crash. While the website provides useful information, the number of the above-mentioned factors could be limited for a comprehensive analysis. A key resource on the website provides links to the news articles for each crash, which could be used to code new variables embedded in the news articles.

To check the credibility of the website (tesladeaths.com), the available coded variables on the website (year, date, country or state in the U.S., model of Tesla vehicle, and death records) have been verified and compared for all the crashes with the news articles from which the variables were coded. The available information about the coded variables on the website matched with the narratives in the corresponding news articles for each crash, justifying the credibility of the website. A unique and comprehensive database was created by utilizing the website. Additional 29 variables were coded manually via visiting the news article weblink for each crash. Some of the links to the news articles were missing or inaccessible. Therefore, the cleaned dataset had a total of 126 crashes, which include 46 single-vehicle and 80 multi-vehicle crashes. Note that in this chapter, single-vehicle and multi-vehicle Tesla-involved Fatal crashes are studied in depth.

- **Dichotomous variables:** Variables in this category were coded on a Yes/No scale. These new variables include speeding by Tesla driver, speeding by other drivers, Tesla driver impaired (driving under the influence), other driver impaired (driving under the influence), intersection-related crash, Tesla struck other vehicles, other vehicle struck Tesla, gender of Tesla driver, gender of Tesla occupant, gender of other vehicle drivers, Tesla vehicle caught fire, other vehicle caught fire, and time of a crash (AM/PM). While most variable names are self-explanatory, some need explanation. For example, speeding by a Tesla driver and speeding by other drivers are coded as “Yes” if news articles reported that the crash involved drivers exceeding the speed limit of the corresponding road while driving. The variable Tesla driver impaired and other driver impaired are coded as “Yes” if the news articles reported that the drivers were driving under the influence of drugs or alcohol for which the drivers became incapable of driving the cars. There are four models of Tesla: S, X, Y, and 3. All these vehicles have various trim levels, but for convenience, their models are considered by their basic model names. A dummy variable is made for each model where one indicates that the observation is on that car model, and zero (0) indicates that the observation is not about that respective car model.
- **Continuous variables:** These variables include the number of vehicles involved, age of tesla driver, age of tesla occupants, age of other drivers, age of other vehicle occupants, and pedestrian age. Note that age information was not available in the articles for all observations.
- **Categorical variables:** These variables include route type (e.g., local route, state route, highway, and expressway), maneuver by Tesla driver, maneuver by other drivers, type of other vehicles, type of crash (e.g., rear-end, head-on), and day of the week on which crash have occurred.
- **Ordinal variables:** Note that all Tesla crashes reported on the website were fatal. However, the injury severity for drivers and occupants (both in Tesla and other vehicles) has been

coded on a four-level ordinal scale using the information in the articles. These new levels are no injury, minor injury, moderate or serious injury, and fatality for each driver and occupant involved in Tesla crashes.

Methods

Human Crash Contributing Factors

Investigating the relationship between Autopilot performance, driver behaviors, and other factors with the survivability of Tesla drivers can provide a clearer picture of automation. The study classifies behaviors into two broader categories, which include driving errors and violations. Driving errors are mistakes that drivers make while driving. The driving errors are not classified into subcategories (e.g., recognition errors, decision errors, and performance errors) due to the lower sample size. However, if drivers disobey some traffic rules, regulations or law that is considered violations. It is found that errors and violations by Tesla drivers have collectively contributed to 84.78% ($N = 32+7$) of the overall single-vehicle media-reported fatal Tesla crashes ($N = 46$) and 28.75% ($N=5+18$) of overall multi-vehicle media-reported fatal Tesla crashes ($N=80$). Speeding was one of the key violations by Tesla drivers, contributing to 28.57% (2 out of 7) of the overall single-vehicle crashes and 33.33% (6 out of 80) of the overall multi-vehicle crashes. Driver lost control of the vehicle was a common error by Tesla drivers in both single-vehicle and multi-vehicle crashes which is a concerning factor regarding Tesla's automation features. For details about errors and violations by Tesla drivers, please refer to **Table 6.2** and **Table 6.3**.

Correlation Analysis: Somers' D Test

Numerous statistical measures, including gamma, Kendall's Tau b, Stuart's Tau c, Spearman correlation, and Spearman rank correlation coefficient, can be used to understand the association between two ordinal (also applicable to dichotomous) variables (204). While the above tests provide useful insights into the potential association between two variables, these tests are symmetrical where no distinction between the dependent and independent variables can be made (204). On the other hand, in the Somers' D test, one of the variables can be specified as a dependent variable that provides a coefficient indicating the association of a particular independent variable with the dependent variable (204; 205). Given two variables, including Y and X:

- Two pairs (X_i, Y_i) and (X_j, Y_j) are said to be concordant if ranks of both the elements agree
- Two pairs (X_i, Y_i) and (X_j, Y_j) are said to be discordant if the ranks of both elements do not agree

For Somers' D, the two subsamples from the parent data are considered, where X is a dichotomous independent variable which equals 1 and 0 for observations in subsamples A and B, respectively. Y is also a dichotomous variable that equals 1 and 0 if a Tesla driver was killed and survived, respectively. Based on the values of an independent variable, the probability of a Tesla driver being killed in subsample A and subsample B can be computed as (206).

$$p_A = Pr(Y = 1|X = 1) \quad (1)$$

$$p_B = Pr(Y = 1|X = 0) \quad (2)$$

After computing the probabilities of the Tesla driver being killed in subsample A and subsample B based on the values of a particular independent variable, Somers' D coefficient can be computed as below (30; 31).

$$D(Y|X) = pA - pB \quad (3)$$

The values of Somers' D range from -1 to 1, which indicates if all the pairs of the variables disagree and agree, respectively. Smaller values of Somers' D (approaching 0) in either direction (+/-) show a lower association of a variable with the response outcome. To conclude, Somers' D coefficient is the difference between the proportion of concordant and discordant pairs of the response and a particular independent variable where both variables are binary (206).

Binary Logistic Regression

A Binary logistic regression model (207; 208) contains a dichotomous dependent variable whose value is either 0 or 1. The key quantity in a regression type analysis is the mean of the dependent variable given the values of the independent variable.

$$E\left(\frac{Y}{x}\right) = \beta_0 + \beta_1 x \quad (1)$$

This is also called the conditional mean or the expected value of Y given the value of x. Here, values of β denotes the model parameters, E denotes the conditional mean, Y is the response variable and x is the independent variable. The logistic regression model formula is as follows:

$$\Pi(x) = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}} \quad (2)$$

The logit transformation of this logistic function $\Pi(x)$ can be written as follows:

$$g(x) = \ln\left(\frac{\Pi(x)}{1 - \Pi(x)}\right) = \beta_0 + \beta_1 x_1 \quad (3)$$

Now, the odds of group 1 (Y=1) and group 0 (Y=0) can be written as follows:

$$\text{Odds} = \frac{\Pi(x_1)}{1 - \Pi(x_1)} \quad (4)$$

$$\text{Odds} = \frac{\Pi(x_0)}{1 - \Pi(x_0)} \quad (5)$$

The odds ratio is achieved by dividing the first odds by the second odds.

Text Mining Analysis

Text-mining is a tool that helps discover knowledge from unstructured textual data (209). Text mining analysis was performed on the narratives of the media reports of the Tesla-involved fatal crashes to qualitatively analyze the crash reports at hand.

Results and Discussion

Descriptive Statistics

The distribution of injury severity for drivers and pedestrians in both the single-vehicle and multi-vehicle Media-reported fatal Tesla crashes is shown in **Table 6.4** and **Table 6.5**, respectively.

Table 6.1. Summary of the Literature Review

Ref.	Relevance	Sample Size	Findings	Method	Gap/limitation
(152)	Effect of different factors on AV crash severity	94	1. Weather conditions are found to be the most critical factor. 2. Chance of severe injury increases if AV crashes occur at an intersection in extreme weather conditions (fog and snow) and highly diverse land use pattern locations. 3. Type of collision significantly contribute to AV crash severity	Crash reports from California DMV, XGBoost model	Did not collect data on speed/driver characteristics, limited sample
(150)	Mechanism and contributing factors of AV-related crashes	113	1. Chances of more severe injury increase if the subject vehicle is on autonomous driving mode and is mainly responsible for a crash. 2. Highway is where severe injuries are most likely to occur. 3. AV-related crash collisions are associated with the location, whether crashes are associated with yielding to pedestrians/ cyclists, and driving mode.	Crash reports from California DMV, ordinal logistic regression modeling, CART modeling	Limited sample size
(199)	Tesla-specific human factors leading to crashes	1	1. Schema theory and PCM can help explain driver actions. 2. Human error is inescapable, and humans are unreliable for monitoring automated systems, therefore efforts should be focused on level 4 and 5 automation. 3. The underlying cause of a tragic accident could be a design error because of an unclear perception of responsibility included in the design.	Use Schema theory and Perceptual Cycle Model (PCM) to explain driver actions	Case study, not generalized
(200)	Tesla-specific human factors leading to crashes	1	1. Tesla driver could have survived if the attenuator was functional. 2. Roadway management as well as timely maintenance are important.	Idealized triangular crash pulse is used to estimate peak impact acceleration. Public info from NTSB evaluation was used to evaluate severity. Vehicle crash data from EDRs were compared with test data for the crash attenuator.	Case study, not generalized
(197)	Driver behaviors operating AVs	20	1. Drivers who are among the early adopters of partial automation systems (PAS) have a positive attitude toward the PAS systems. 2. Drivers universally engaged in secondary tasks under safe usage conditions. 3. Some self-regulatory behaviors to avoid risky situations.	Data from interviews	Response and sampling bias, low sample size

Table 6.1 Continued.

Ref.	Relevance	Sample Size	Findings	Method	Gap/limitation
(198)	Characteristics and patterns of CAV involved crashes	72	1. CAV crash severity is associated with driving mode, roadside parking, rear-end collision, collision location, and one-way road. 2. CAV driving mode, CAV stopped or not, CAV turning or not, other vehicle turning or not, and normal vehicle overtaking or not affect the type of crash.	Data collected from CA DMV, descriptive statistics analysis	Limited sample size

In 71.74% (N =33) of the single-vehicle crashes, the Tesla driver was killed and in 28.26% (N=13) time they survived. In single-vehicle Tesla-involved media-reported fatal crashes pedestrians were killed in 10% (N = 5) of the crashes. In 27.5% (N=22) of the multi-vehicle crashes, the Tesla driver was killed, whereas pedestrians were killed in 2.5% (N=2) of crashes. Due to the small sample, a new variable was created to represent Tesla driver fatality which is coded as 1 if the Tesla driver was killed and 0 if the driver survived. Cross-tabulation and correlational analysis is then conducted for the Tesla driver survivability in a single-vehicle and multi-vehicle fatal Tesla crash. The survivability of the Tesla driver is used as the dependent variable in the correlational analysis in this study.

Figure 6.2 shows that the highest percentage of these single-vehicle media-reported fatal Tesla crashes have occurred in California (39.13%), followed by Florida (19.57%), New York (6.52%), and Ohio (6.52%). **Figure 6.3** shows that the highest percentage of the multi-vehicle media-reported fatal Tesla crashes have occurred in California (36.25%), Florida (16.25%), Arizona (7.50%), and Georgia (6.25%). This was expected due to the higher affordability of such expensive vehicles in California and Florida.

Cross Tabulation: Tesla Driver Survivability versus Key Factors

To understand the association of Autopilot use, Tesla driver behavior, and other important factors with Tesla driver survivability in single-vehicle and multi-vehicle (Tesla) fatal crashes, **Table 6.6** and **Table 6.7** presents a cross-tabulation that can provide useful insights (175). For instance, statistics reveal that for single-vehicle crashes, Tesla drivers died 50% (N=3) of the time and survived 50% (N=3) of the time when Autopilot was claimed. For multi-vehicle crashes, 75% (N=3) of the time driver was killed and 25% (N=1) time they survived when Autopilot was claimed. When the Autopilot was not claimed in single-vehicle crashes, in 75% (N=30) cases, Tesla drivers died and in 25% cases (N=10), the drivers survived. This indicates that there could be a lower chance of survival for Tesla drivers when autopilot is not engaged. A deeper analysis and reconstruction of the crash and performance of the Autopilot vision system and vehicle control system will be needed to fully understand and remedy the factors contributing to these crashes.

In all the single-vehicle Tesla crashes when Tesla caught fire, the drivers died 100% (N=18) of the cases, according to this dataset. But there have been some crashes where Tesla caught fire and the driver survived. However, those crashes were not present in the dataset. Among the multi-vehicle crashes when Tesla caught fire, in 80% (N=3) cases, Tesla drivers died and in 20% (N=1) cases, the Tesla drivers survived. This is high compared with Fatality Analysis Reporting System (FARS) data on all vehicles.

The FARS datasets from 2013 to 2020 are collected from the NHTSA website to calculate the fire-involved fatal crash rates and driver fatality in those crashes. The results are shown in **Table 6.8**. FARS data revealed that the fire-involved fatal crash rate for all vehicles ranged between 4.15% - 4.79% from 2013 to 2020. Compared to the fatal crash rates of all vehicles, Tesla-involved vehicles' fire-involved fatal crash rates is high (18.25%). This indicates that there may be some underlying factors for which Tesla has a higher percentage of fire-involved fatal crashes.

Table 6.2. Driving Errors and Violations by Tesla Driver in Single-Vehicle Crashes

Crash Types	Single Vehicle (N=46)	
Human Factors (Tesla Driver)	Frequency	Percentage
Errors by Driver		
Driver lost control	31	96.88
Failed to negotiate curve	0	0.00
Swerved into crosswalk	1	3.13
Fell asleep at wheel	0	0.00
Total	32	100.00
Violations		
Driver did not take action as per warning	1	14.29
Failed to stop at Intersection	2	28.57
Hit and run	2	28.57
Tesla Driver Speeding	2	28.57
Ran red light	0	0.00
U-turn	0	0.00
Attempted to pass two lane road	0	0.00
Unprotected left turn	0	0.00
Did not yield	0	0.00
High speed police chase	0	0.00
Slowed to a stop	0	0.00
Total	7	100.00

Table 6.3 Driving Errors and Violations by Tesla Driver in Multi-Vehicle Crashes

Crash Types	Multi-Vehicle(N=80)	
Human Factors (Tesla Driver)	Frequency	Percentage
Errors by Driver		
Driver lost control	3	60.00
Failed to negotiate curve	1	20.00
Swerved into crosswalk	0	0.00
Fell asleep at wheel	1	20.00
Total	5	100.00
Violations		
Driver did not take action as per warning	0	0.00
Failed to stop at Intersection	0	0.00
Hit and run	1	5.56
Tesla Driver Speeding	6	33.33
Ran red light	5	27.78
U-turn	1	5.56
Attempted to pass two lane road	1	5.56
Unprotected left turn	1	5.56
Did not yield	1	5.56
High speed police chase	1	5.56
Slowed to a stop	1	5.56
Total	18	100.00

Table 6.4. Distribution of Injury Severity in Single-Vehicle Crashes

Injury Severity	Single Vehicle Crashes (46)			
	Tesla Vehicles			
	Driver		Pedestrian	
	Freq.	(%)	Freq.	(%)
Not Applicable	3	6.52	39	84.78
No Injury	7	15.22	0	0.00
Suspected Minor Injury	1	2.17	2	4.35
Moderate or Serious Injury	2	4.35	0	0.00
Fatality	33	71.74	5	10.87
Total	46	100	46	100

Table 6.5 Distribution of Injury Severity in Multi-Vehicle Crashes

Injury Severity	Multi Vehicle Crashes (80)			
	Tesla Vehicles			
	Driver		Pedestrian	
	Frequency	Percentage	Frequency	Percentage
Not Applicable	3	3.75	76	95
No Injury	33	41.25	0	0
Suspected Minor Injury	14	17.5	0	0
Moderate or Serious Injury	8	10	2	2.5
Fatality	22	27.5	2	2.5
Total	80	100	80	100

Statewise Distribution of Single Vehicle (Tesla) Crashes in U.S. (N = 46)

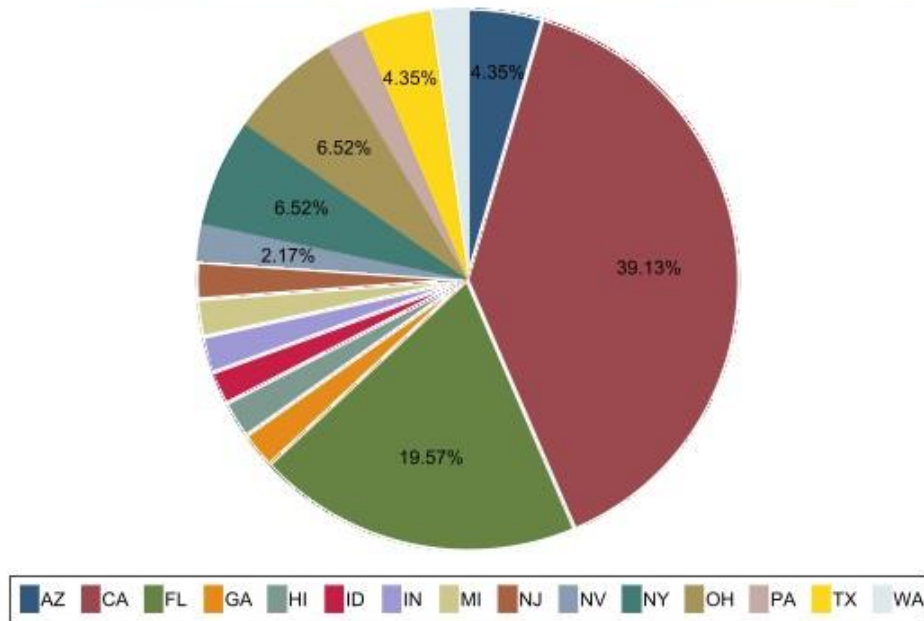


Figure 6.2. State wise Distribution of Single-Vehicle Fatal Crashes involving Tesla Vehicle in the US

Statewise Distribution of Multi-Vehicle (Tesla) Crashes in U.S. (N = 80)

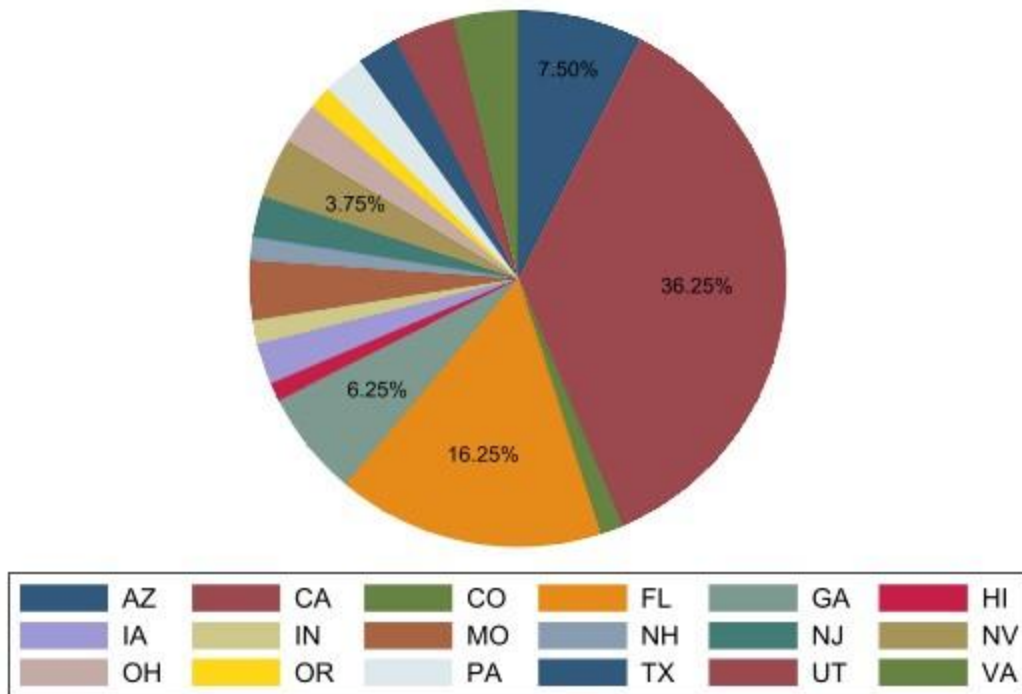


Figure 6.3. State wise Distribution of Multi-Vehicle Fatal Crashes involving Tesla Vehicle in the US

Next, the fatality rate of the drivers involved in the fire-involved fatal crashes of all vehicles is compared to that of the Tesla drivers. It is identified that 95.65% of the Tesla drivers died in these cases, which is higher than all vehicles' drivers' fatality percentage (between 73.04% to 78.41%). This analysis infers that fire involvement in Tesla vehicles is crucial. Tesla manufacturers and researchers should investigate this issue to reduce driver fatality in fire-involved Tesla crashes.

In single-vehicle crashes, Tesla drivers have made an error in 69.56% (N = 32) of the crashes. The driver was killed in 78.13% (N = 25 out of 32) of these crashes, which is greater than the percentage of 57.14% (N = 8 out of 14) of crashes where the Tesla driver died without making any error. In multi-vehicle crashes, Tesla drivers made an error in only 6.67% (N=5) of the crashes. But in the crashes where the Tesla driver did not make any error (N=75), the driver died in 24% (18 of 75) of those crashes. Interestingly, the percentage of Tesla driver deaths is lower in crashes where the Tesla driver has made a violation or was impaired with alcohol than in crashes where the driver was neither impaired nor making violations, according to this dataset in both the single-vehicle and multi-vehicle crashes.

Table 6.6 illustrates that, of the 46 single-vehicle crashes, five are Tesla Model 3, no Model X, one is Model Y, ten are Model S, and 24 observations did not contain any model name. Of the Tesla Model 3 involved crashes (N=5), Tesla driver died 40% (N=2) times. Of the Tesla Model Y involved crashes (N=1), Tesla driver did not die. Of the Tesla Model S involved crashes (N=10), Tesla driver was killed 70% times. Similarly, the survivability of Tesla drivers is higher in crashes when a pedestrian is involved but not when it is nighttime in single-vehicle crashes. Furthermore, Tesla drivers are less likely to survive if a Tesla vehicle collides with a fixed object in single-vehicle crashes. **Table 6.6** and **Table 6.7** show that the percentage of Tesla driver deaths in single-vehicle crashes (25%) and multi-vehicle crashes (13.64%) occurring at intersections are lower than their corresponding percentage (81.58% and 32.76%, respectively) on roadway segments.

Correlation analysis through Somers' D Tests

Besides the mean coefficient of Somers' D, its confidence interval can also be obtained, which shows the range of association of a specific independent variable with the response outcome. The coefficients for both symmetric (similar to gamma and tau tests where no distinction between the dependent and independent variable is made) and asymmetric Somers' D (an asymmetric extension of gamma and tau tests) are presented. The latter one is more relevant for this study. Due to the lower sample size in single-vehicle (N = 46) and multi-vehicle (N=80) crashes, most of the key variables do not show a statistically significant correlation with the Tesla driver fatality. Still, the direction of association of all such insignificant variables is intuitive.

In the single-vehicle crashes provided in **Table 6.9**, the intersection indicator shows a statistically significant and negative association (t-stat = -2.39) with Tesla driver fatality with a mean asymmetric Somers' D coefficient of -0.40. This finding suggests that the chance of a Tesla driver fatality is 40% lower compared to crashes that occur at non-intersection locations. Importantly, the Autopilot use did not show statistical significance with Tesla driver fatality; however, its direction of association was intuitive, and its statistical significance may improve with a larger sample size. Referring to the effect of Autopilot use on Tesla driver fatality in single-vehicle crashes, it is observed that there is a 14% lower chance of a Tesla driver fatality if the Autopilot

was engaged compared to crashes where autopilot was disengaged. Moreover, results indicate that the Tesla driver is less likely to survive a crash with fixed objects.

In the multi-vehicle crashes provided in **Table 6.10**, statistically significant association was found for Driver impairment and Tesla model X. There was 9% less chance of Tesla driver fatality if Tesla driver was impaired compared to if Tesla driver was not impaired. There was 7% less chance of Tesla driver fatality if the Tesla vehicle was model X compared to if it was not model X.

Binary Logistic Regression Model on Tesla Drivers' Fatality

Table 6.11 illustrates the results of the Binary Logistic Regression model to show an association between various factors and Tesla-driver fatality in media-reported Tesla-involved fatal crashes. The model has a pseudo R^2 of 0.43, which indicate a good fit. Lower values of AIC and BIC are used as the model selection criteria to find the best model. The dependent variable is Tesla Driver Killed, and its value is 1 if the Tesla driver dies and 0 if the Tesla driver survives. Independent variables include Autopilot Claimed; Tesla Vehicle Caught Fire; variables included in Human Factors, Tesla Models, Collision Type, and Other Key Factors; and Tesla Driver and Other Vehicles' Drivers' Age. Only statistically significant independent variables with correct signs are kept in the model.

The odds ratios convey that the odds of a Tesla Driver being killed in a fatal crash is 50.21 times more likely when Tesla Vehicle catches fire. This matches with the data from the FARS dataset, which confirms that the Tesla Driver Fatality rate is higher than the average fatality rate of all vehicles in fire-involved crashes in the US. The odds of a Tesla Driver being killed in a fatal crash is 3.29 times more likely when Tesla Driver makes an error. Tesla drivers are supposed to disengage the autopilot and take charge in a risky environment or when the car requires it. Losing control of the car could have happened because the drivers could not take control quickly or the car's features malfunctioned. The odds of a Tesla Driver being killed in a fatal crash is 5.26 times less likely when the crash occurred at intersections. Although intersections are hazard-prone locations, Tesla vehicles may have less fatal crashes there because drivers may be more alert in the intersections and drive slowly or not using the autopilot feature. The odds of a Tesla Driver being killed in a fatal crash is 1.03 times more likely when driver age increase by one unit. This indicates that more elderly drivers are prone to fatal crashes, which is logical as their reflexes and vision are weaker than younger drivers. The odds of a Tesla Driver being killed in a fatal crash is 1.04 times less likely when the Other Vehicle Driver's Age increase by one unit. This is also logical because in a fatal crash if a Tesla driver is younger and the other vehicle's driver is older, the young Tesla driver will have a comparatively more chance of survival.

In summary, it was found from the model that if a fatal crash happens to Tesla Vehicle, the Tesla Driver is more likely to be killed when the Tesla Vehicle gets fire, the Tesla Driver makes error, or if the Tesla Drivers' age increases. Conversely, suppose a fatal crash happens to Tesla Vehicles. In that case, the Tesla Driver is less likely to be killed when the crash occurs at an intersection or when other vehicle's driver's age increases.

Table 6.6. Cross Tabulation of Fatality of Tesla Driver in Single-Vehicle Crashes Versus Human Factors

Key Variables	Category	Tesla Driver Survived (N=13) Freq. (%)	Tesla Driver Killed (N=33) Freq. (%)	Total (N=46) Freq. (%)
Autopilot Claimed	No (0)	10(25.00%)	30(75.00%)	40(100.00%)
	Yes (1)	3(50%)	3(50%)	6(100.00%)
Tesla Vehicle Caught Fire	No (0)	13(46.43%)	15(53.57%)	28(100.00%)
	Yes (1)	0(0.00%)	18(100.00%)	18(100.00%)
Human Factors				
Errors by Tesla Driver	No (0)	6 (42.86%)	8 (57.14%)	14(100.00%)
	Yes (1)	7 (21.88%)	25(78.13%)	32(100.00%)
Violations by Tesla Driver	No (0)	8(20.51%)	31(79.49%)	39(100.00%)
	Yes (1)	5(71.43%)	2(28.57%)	7(100.00%)
Tesla Driver being Impaired with Alcohol	No (0)	11(25.00%)	33(75.00%)	44(100.00%)
	Yes (1)	2(100.00%)	0(0.00%)	2(100.00%)
Model of Tesla Vehicle				
Tesla Model 3	No (0)	10(24.39%)	31(75.61%)	41(100.00%)
	Yes (1)	3(60.00%)	2(40.00%)	5(100.00%)
Tesla Model X	No (0)	13(28.26%)	33(71.74%)	46(100.00%)
	Yes (1)	0(0.00%)	0(0.00%)	0(0.00%)
Tesla Model Y	No (0)	12(26.67%)	33(73.33%)	45(100.00%)
	Yes (1)	1(100.00%)	0(0.00%)	1(100.00%)
Tesla Model S	No (0)	10(27.78%)	26(72.22%)	36(100.00%)
	Yes (1)	3(30.00%)	7(70.00%)	10(100.00%)
Collision Type				
Fixed Object	No (0)	7(35.00%)	13(65.00%)	20(100.00%)
	Yes (1)	6(23.08%)	20(76.92%)	26(100.00%)
Other Key Factors				
Tesla Driver Male	No (0)	6(33.33%)	12(66.67%)	18(100.00%)
	Yes (1)	7(25.00%)	21(75.00%)	28(100.00%)
Night time indicator	No (0)	6(24.00%)	19(76.00%)	25(100.00%)
	Yes (1)	7(33.33%)	14(66.67%)	21(100.00%)
Pedestrian Involved	No (0)	8(20.51%)	31(79.49%)	39(100.00%)
	Yes (1)	5(71.43%)	2(28.57%)	7(100.00%)
Crash occurred at intersection	No (0)	7(18.42%)	31(81.58%)	38(100.00%)
	Yes (1)	6(75.00%)	2(25.00%)	8(100.00%)

Table 6.7 Cross Tabulation of Fatality of Tesla Driver in Multi-Vehicle Crashes Versus Human Factors

Key Variables	Category	Tesla Driver Survived (N=58) Freq. (%)	Tesla Driver Killed (N=22) Freq. (%)	Total (N=80) Freq. (%)
Autopilot Claimed	No (0)	57(75.00%)	19(25.00%)	76(100.00%)
	Yes (1)	1(25.00%)	3(75.00%)	4(100.00%)
Tesla Vehicle Caught Fire	No (0)	57(76.00%)	18(24.00%)	75(100.00%)
	Yes (1)	1(20.00%)	4(80.00%)	5(100.00%)
Human Factors				
Errors by Tesla Driver	No (0)	57(76.00%)	18(24.00%)	75(100.00%)
	Yes (1)	1(20.00%)	4(80.00%)	5(100.00%)
Violations by Tesla Driver	No (0)	44(70.97%)	18(29.03%)	62(100.00%)
	Yes (1)	14(77.78%)	4(22.22%)	18(100.00%)
Tesla Driver being Impaired with Alcohol	No (0)	53(70.67%)	22(29.22%)	75(100.00%)
	Yes (1)	5(100.00%)	0(0.00%)	5(100.00%)
Model of Tesla Vehicle				
Tesla Model 3	No (0)	54(77.14%)	16(22.86%)	70(100.00%)
	Yes (1)	4(40.00%)	6(60.00%)	10(100.00%)
Tesla Model X	No (0)	54(71.05%)	22(28.95%)	76(100.00%)
	Yes (1)	4(100.00%)	0(0.00%)	4(100.00%)
Tesla Model Y	No (0)	55(73.33%)	20(26.67%)	75(100.00%)
	Yes (1)	3(60.00%)	2(40.00%)	5(100.00%)
Tesla Model S	No (0)	50(71.43%)	20(28.57%)	70(100.00%)
	Yes (1)	8(80.00%)	2(20.00%)	10(100.00%)
Collision Type				
Fixed Object	No (0)	57(72.15%)	22(27.85%)	79(100.00%)
	Yes (1)	1(100.00%)	0(0.00%)	1(100.00%)
Other Key Factors				
Tesla Driver Male	No (0)	28(77.78%)	8(22.22%)	36(100.00%)
	Yes (1)	30(68.18%)	14(31.82%)	44(100.00%)
Night time indicator	No (0)	37(68.52%)	17(31.48%)	54(100.00%)
	Yes (1)	21(80.77%)	5(19.23%)	26(100.00%)
Pedestrian Involved	No (0)	55(72.37%)	21(27.63%)	76(100.00%)
	Yes (1)	3(75.00%)	1(25.00%)	4(100.00%)
Crash occurred at intersection	No (0)	39(67.24%)	19(32.76%)	58(100.00%)
	Yes (1)	19(86.36%)	3(13.64%)	22(100.00%)

Table 6.8 Comparison of Fire-involved Fatal Crashes between Drivers of All Vehicles and Tesla in the US

FARS Dataset	Year	Fatal Crashes			Drivers in Fire-involved Fatal Crashes (multiple drivers in each crash)			
		Total Fatal Crashes	Total Fire-Involved Fatal Crash	Fire-involved Fatal Crash Rates	Total Drivers	Fatality	Survivability	Driver Fatality (%) in fire-involved fatal crashes
	2013	30202	1253	4.15%	1409	1074	335	76.22%
	2014	30056	1288	4.29%	1465	1114	351	76.04%
	2015	32538	1397	4.29%	1582	1187	395	75.03%
	2016	34748	1506	4.33%	1686	1282	404	76.04%
	2017	34560	1551	4.49%	1770	1321	449	74.63%
	2018	33919	1528	4.50%	1757	1316	441	74.90%
	2019	33487	1477	4.41%	1695	1238	457	73.04%
	2020	35766	1714	4.79%	1904	1493	411	78.41%
Tesla Dataset	2013-2022	126	23	18.25%	23	22	1	95.65%

Table 6.9. Results of Correlational Analysis: Tesla Driver Survivability versus Key Factors in Single-Vehicle Crashes

Key Variables	Single Vehicle Crashes (N=46)					
	Symmetric Somers' D			Asymmetric Somers' D		
	Coef.	t	P> t	Coef.	Min.	Max.
Autopilot Claimed (1/0)	-0.14	-1.05	0.30	-0.14	-0.39	0.13
Tesla Vehicle Caught Fire (1/0)	0.61	4.80	0.00	0.55	0.34	0.70
Human Factors						
Errors by Tesla Driver (1/0)	0.22	1.31	0.20	0.22	-0.12	0.51
Violations by Tesla Driver (1/0)	-0.34	-2.06	0.05	-0.32	-0.58	-0.01
Tesla Driver being Impaired with Alcohol (1/0)	-0.16	-1.46	0.15	-0.15	-0.35	0.06
Model of Tesla Vehicle						
Tesla Model 3 (1/0)	-0.17	-1.30	0.20	-0.17	-0.41	0.09
Tesla Model X (1/0)	0.00	0.00	0.00	0.00	0.00	0.00
Tesla Model Y (1/0)	-0.08	-1.00	0.32	-0.08	-0.23	0.08
Tesla Model S (1/0)	-0.02	-0.13	0.90	-0.02	-0.29	0.26
Collision Type						
Fixed Object (1/0)	0.15	0.85	0.40	0.14	-0.20	0.45
Other Key Factors						
Tesla Driver Male (1/0)	0.10	0.58	0.56	0.10	-0.24	0.41
Night time indicator (1/0)	-0.11	-0.67	0.51	-0.11	-0.43	0.22
Pedestrian Involved (1/0)	-0.34	-2.06	0.05	-0.32	-0.58	-0.01
Crash occurred at intersection (1/0)	-0.42	-2.39	0.02	-0.40	-0.65	-0.07

Table 6.10 Results of Correlational Analysis: Tesla Driver Survivability versus Key Factors in Multi-Vehicle Crashes

Key Variables	Multi-Vehicle Crashes (N=80)					
	Symmetric Somers' D			Asymmetric Somers' D		
	Coeff.	t	P> t	Coefficient	Min.	Max.
Autopilot Claimed (1/0)	0.12	1.54	0.13	0.12	-0.03	0.27
Tesla Vehicle Caught Fire (1/0)	0.17	1.89	0.06	0.16	-0.01	0.33
Human Factors						
Errors by Tesla Driver (1/0)	0.17	1.89	0.06	0.16	-0.01	0.33
Violations by Tesla Driver (1/0)	-0.06	-0.59	0.56	-0.06	-0.26	0.14
Tesla Driver being Impaired with Alcohol (1/0)	-0.09	-2.28	0.03	-0.09	-0.16	-0.01
Model of Tesla Vehicle						
Tesla Model 3 (1/0)	0.21	1.93	0.06	0.20	-0.01	0.40
Tesla Model X (1/0)	-0.07	-2.03	0.05	-0.07	-0.14	0.00
Tesla Model Y (1/0)	0.04	0.57	0.57	0.04	-0.10	0.18
Tesla Model S (1/0)	-0.05	-0.60	0.55	-0.05	-0.20	0.11
Collision Type						
Fixed Object (1/0)	-0.02	-0.99	0.33	-0.02	-0.05	0.02
Other Key Factors						
Tesla Driver Male (1/0)	0.12	0.95	0.35	0.12	-0.13	0.35
Night time indicator (1/0)	-0.14	-1.19	0.24	-0.13	-0.35	0.09
Pedestrian Involved (1/0)	-0.01	-0.12	0.91	-0.01	-0.11	0.10
Crash occurred at intersection (1/0)	-0.19	-1.91	0.06	-0.19	-0.38	0.01

Table 6.11 Results of the Binary Logistic Regression Model

Dependent Variable	Category	Coef.	Std. err.	P> z	Odds Ratio
Tesla Driver Killed					
Independent Variable					
Intercept		-1.05	0.50	0.04	0.35
Tesla Vehicle Caught Fire	No (Base)				
	Yes	3.92	1.13	0.00	50.21
Errors by Tesla Driver	No (Base)				
	Yes	1.19	0.56	0.03	3.29
Crash occurred at intersection	No (Base)				
	Yes	-1.64	0.75	0.03	0.19
Tesla Driver Age	Continuous	0.03	0.01	0.00	1.03
Other Vehicle Driver Age	Continuous	-0.04	0.01	0.01	0.96
N = 126					
Likelihood Ratio χ^2 (5) = 74.30					
Prob > χ^2 = 0.0000					
Pseudo R^2 = 0.4304					
LL (null) = -86.32					
LL (model) = -49.17					
df = 6					
AIC = 110.33					
BIC = 127.35					

Text Analysis on Crash Reports

Word clouds were made for both single-vehicle and multi-vehicle crashes to identify the presence of any pattern in these crashes or any issue that could be considered necessary for these different types of crashes. **Figure 6.4 (a)** and **Figure 6.4 (b)** show the word cloud for single-vehicle and multi-vehicle Tesla-involved media-reported fatal crashes, respectively. Both **Figure 6.4 (a)** and **Figure 6.4 (b)** have many common words but only **Figure 6.4 (b)**, which is multi-vehicle related, had “Autopilot.” This indicates that AV-based crash reports under report about Autopilot, whereas Autopilot is a vital feature for AVs like Tesla and should have been reported in detail. “Fire” is mentioned more in single-vehicle Tesla crash media reports compared to multi-vehicle ones. “Lane” is mentioned more in multi-vehicle Tesla crash media reports. Only multi-vehicle crash reports mentioned “Truck.”

Comparison of Tesla of Luxury Electric Vehicles

Figure 6.5 shows the yearly sales figures of three Tesla models and some other similar luxury electric cars in the US. It is observed that Tesla has higher sales than any other similar car manufacturer brand, indicating that their exposure is higher. Therefore, the higher fatality rates of Tesla would be because of higher exposure and does not mean that Tesla is any worse or superior to similar cars. To understand the complexity of the situation, FARS data were checked for validity. In the 2019 FARS database, 40 fatal crashes involving Tesla were found. The fatal crash for the BMW i3 was one, and Nissan Leaf was six in 2019 in the US (210). These numbers do not signal less chance of fatality for Tesla crashes compared to other similar cars. But the sales figures indicate that Tesla has more exposure and, therefore, more cars have chances of crashes as they are more in number on the roads. These critical issues should be studied further.

Limitations

The study may add useful knowledge to the existing literature on the safety of AVs like Tesla. However, the main limitation of this study is that the results are based on fatal crashes that involved only Tesla. Hence, the results cannot be generalized to all AVs. Another critical issue is the lower sample size of AV field crashes which is also highlighted in the existing literature. Although the data source reported more than 200 Tesla-involved fatal crashes, only 46 were single-vehicle and 80 were multi-vehicle crashes. They are still not large enough to do robust statistical analysis. The website only reports Tesla involved in Fatal crashes, but there might be Tesla involved in other crashes and a study of autopilot with those crashes would provide more insight. The frequency of driver errors, violations, and crashes at intersections is shallow, and it is difficult to infer a conclusion from their analysis. Another limitation of AV data is also observed in the FARS dataset, where Tesla is coded as “Other Domestic manufacturer-Tesla.” Road geometry and land use information were not available in the data. Better coding of this model is necessary, like other vehicle manufacturers.

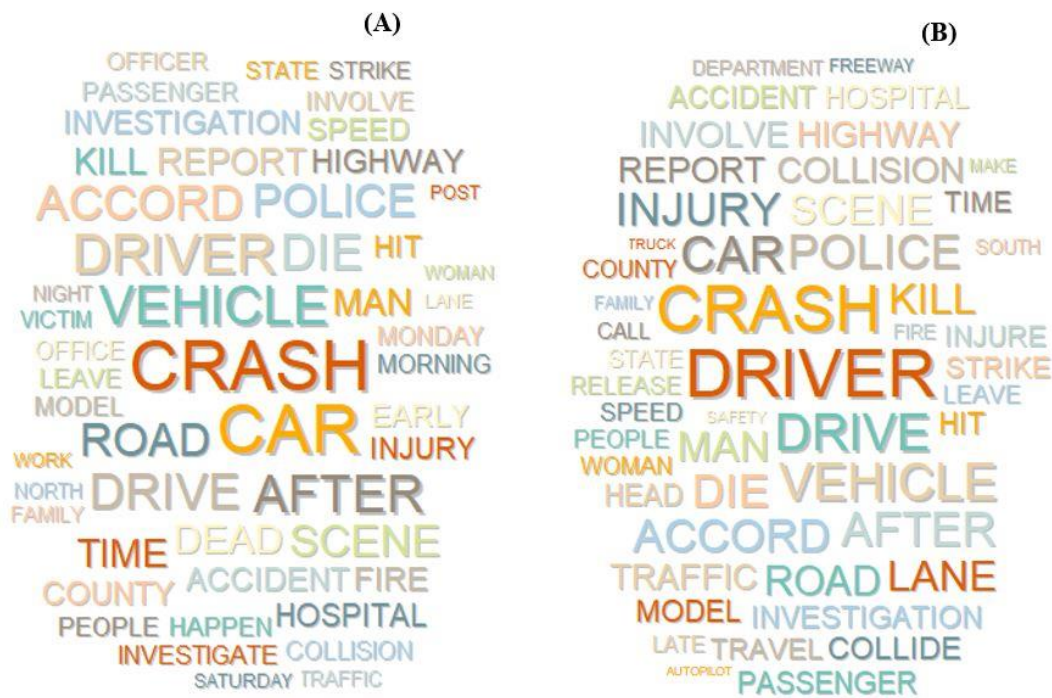


Figure 6.4 Word Cloud for Media Reported Tesla-Involved (a) Single-Vehicle and (b) Multi-Vehicle Fatal Crashes

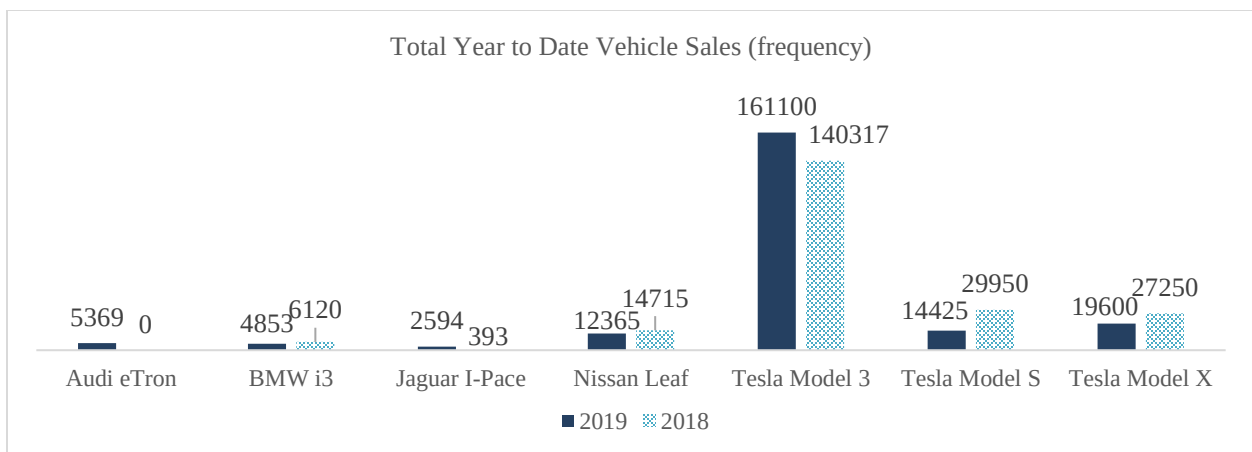


Figure 6.5 Year to Date Sales of Luxury Electric Vehicles

Data Source: (211)

On the other hand, AVs can malfunction for various reasons, causing crashes.

Conclusions

In general, automated vehicles can be relatively safer than human-driven vehicles because artificial intelligence offers a driving experience free of distractions, errors, and traffic violations. This study contributes to the existing body of AV-based literature by evaluating the safety of Tesla cars. A unique database was created from Tesla's fatal crash data (tesladeaths.com). Using links to the news narratives for each fatal crash, several new variables were coded. Several interesting insights are found focusing on 46 single-vehicle crashes (some involving pedestrians) and 80 multi-vehicle Tesla crashes which have occurred in the US. The key findings of this study are as follows:

1. For single vehicle crashes, Tesla drivers died 50% (N=3) of the time and survived 50% (N=3) of the time when Autopilot was claimed. For multi vehicle crashes, 75% (N=3) of the time driver was killed and 25% (N=1) time they survived when Autopilot was claimed.
2. When the Autopilot was not claimed in single vehicle crashes, in 75% (N=30) cases Tesla drivers died and in 25% cases (N=10) the drivers survived.
3. Referring to the Somers' D test to observe the relationship of the Autopilot with Tesla driver survivability, Autopilot engagement does not seem to correlate with Tesla driver survivability.
4. Somers' D coefficient of -0.40 suggests that the chance of a Tesla driver fatality is 40% lower compared to crashes that occur at non-intersection locations for single vehicle crashes.
5. Somers' D test indicate that Tesla driver is less likely to survive in a crash with fixed objects in single vehicle crashes.
6. In the multi-vehicle crashes, statistically significant association was found for Driver impairment and Tesla model X. There was 9% less chance of Tesla driver fatality if Tesla driver was impaired compared to if not impaired and 7% less chance of Tesla driver fatality is Tesla model X compared to if it is not.
7. The Binary logistic regression model suggest that if a fatal crash happens to Tesla Vehicles, the Tesla Driver is more likely to be killed when the Tesla Vehicle gets fire, the Tesla Driver makes error, or if the Tesla Drivers' age increases. Conversely, if a fatal crash happens to Tesla Vehicles, the Tesla Driver is less likely to be killed when the crash occurs at an intersection or when other vehicle's driver's age increases.

In this study, the relationship of Autopilot engagement and Tesla driver behaviors with the survivability of Tesla drivers is investigated. It would be interesting to see how the role of Autopilot and Tesla driver behaviors change in non-fatal crash scenarios. Future studies can further investigate the pedestrian identification system in the Autopilot feature. Furthermore, note that the website (tesladeaths.com) has not included Media-reported non-fatal Tesla crashes and it would be interesting in the future to see the role of Autopilot engagement and Tesla driver behavior in crashes that resulted in severe injuries and minor injuries. In the future, the availability of larger sample sizes may overcome some of the limitations. Although the results are not generalizable, future studies on AV can be benefitted from the method, results, and insights of this study. This study will be helpful for AV manufacturers, notably Tesla, who are working to improve the safety features of their vehicles.

Acknowledgments

This study was partially supported by the Collaborative Sciences Center for Road Safety, CSCRS, a US. DOT-funded National Transportation Center.

CHAPTER 7. CONCLUSION

Connected and Automated Vehicles are still at the early stage of test and deployment. Therefore, their safety performance analysis is crucial for improving the technology and ensuring people's safety. This dissertation intends to examine connectivity and automation technologies to gain comprehensive knowledge of their safety performances. Existing literature showed that connectivity application is new in transit buses and research is scarce in this area. Existing literature is limited in examining automation technology as AV data is available mainly from simulation-based studies and test-bed data are not publicly available. As automation technology is still in the early stage, manufacturers have achieved lower-level automation in personal vehicles so far. This dissertation addresses these gaps and analyzes both technologies at their early stages of deployment, which can be helpful in their advancement and eventually assure multimodal safety. A summary of the findings of this dissertation is provided below. This will be followed by research implications and directions for future research.

Chapter 2's goal is to enhance transit and pedestrian safety by analyzing BSMs and alerts. Data is collected from a CV pilot project that deployed connectivity technology in transit buses and intersections. After integrating this big data from CVs with road geometry and crash information, a unique database is developed for safety analysis. It is identified that the EPCW alerts are negatively associated with bus-involved crashes and driving volatility is positively associated with bus-involved crashes. CV alerts have shown the potential to identify high-risk hazardous locations for transit buses.

Chapter 3's goal is to enhance pedestrian safety by exploring pedestrian-based alerts at different intersections. Pedestrian-based alerts at three different intersections are studied in depth and novel safety surrogate measures are developed to assess the performance of these pedestrian-based alerts. The SSM (MTDI) and Alert Zone have shown early evidence of the performance of EPCW technologies.

Chapter 4's goal is to understand the state of the art of automated shuttle pilot projects in the US. Information on completed and ongoing automated shuttle pilot projects in the US from 2016 to 2019 is reviewed. It is identified that smart cities can consider University Campuses, Downtowns, Office Parks, Stadiums, Transit Stations, and Entertainment Areas for early deployment of automated shuttles.

Chapter 5's goal is to contribute to AV safety by analyzing ADAS-equipped and engaged AV crashes. 315 AV crashes that occurred in the US between July 2021 and May 2022 are utilized for the analysis. The statistical significant associations are found between crash severity and AV features and their drivers. Results suggest that, compared to Non-Tesla AVs, Tesla AVs are associated with a 25% lesser chance of injury crashes and a 7% lesser chance of fatal or severe crashes. AVs heavier than 4000 pounds are associated with 8% higher chance of injury crashes and 4% higher chance of fatal or severe crashes compared to AVs weighing less than 3500 pounds. Highway/freeways contained more fatal or severe crashes and non-highways contained more injury crashes in these years. This study attempts to understand the effectiveness of ADAS technology in SAE level 2 automated vehicles in preventing crash severity.

Chapter 6's goal is to contribute to the knowledge body of AV safety by investigating Tesla autopilot engagement. 46 single-vehicle and 80 multi-vehicle Tesla-involved fatal crashes that

occurred in the US between 2013 and June 2022 are extracted from media reports. It is observed that Tesla drivers died more in fatal crashes when autopilot is not claimed compared to fatal crashes where autopilot is claimed. This showed some potential of the Tesla autopilot feature to provide safety, although no statistically significant associations are found between Tesla driver survivability and autopilot engagement. Text analysis shows a scarcity of mention of the Autopilot feature in media reports which should be brought to attention. Text analysis shows mention of the word Fire more in single-vehicle crashes. Fire-involved fatal crashes are found to be higher in percentage in Tesla vehicles compared to all other vehicles in the US.

Implications

The implications of this dissertation's findings can assist in advancing multimodal safety. Chapter 2 and chapter 3 show that deploying CV and ADAS technologies can help reduce pedestrian and bus-involved crashes. The CV technology can improve transit drivers' awareness of pedestrians at intersections. The CV alerts can be used to find the high-risk hazardous location of buses. The findings from these chapters provide the motivation for further investigation of CV technology. Policies can be developed to encourage more research to improve CV technology to accelerate the deployment of CV technologies. CV researchers can utilize MTDI and Alert Zone to evaluate the impact of CVs. Transit agencies and cities can deploy these technologies to reduce bus-involved crashes at their cities' critical intersections and improve the safety of vulnerable road users. Traffic engineers and road designers can consider installing roadside units for CVs and develop road designs welcoming these emerging technologies. Chapter 4 identified three critical characteristics of automated shuttle deployments in the United States that shows early trends and guides other localities considering the deployment of automated shuttles soon. Chapter 5's key findings are helpful for Tesla and other AV manufacturers and safety researchers who work for safety improvements of AVs. The insights on AV crashes in highways/freeways and non-highways will help city planners, traffic engineers, and designers develop better and safer road designs for AV users. Key findings from Chapter 6 infer that Tesla manufacturers and AV safety researchers should investigate more on Tesla's autopilot feature and fire-involved fatal crashes of Tesla. More in-depth research on fire-involved Tesla crashes will help reduce the predominance of fire-involved crash fatalities among Tesla drivers.

Overall, this dissertation can be helpful for traffic safety practitioners, engineers, policymakers, and city planners who work to prepare their cities for connected and automated vehicles. Implication of the findings and insights can assist cities that are planning to achieve vision zero goals.

Future Research

This dissertation provides timely suggestions for CAV research for the future. In the future, researchers can examine Basic Safety Messages and Alerts at various intersections and bus stops in busy metropolitan cities and rural areas to develop generalizable models for CV technologies. Alerts and BSMs from CVs can be investigated to find an association with crash severity. The EPCW alert should be studied in detail at different intersections and roundabouts over a larger geographic area to further the understanding of bus-involved pedestrian crash studies. Future

efforts should focus on associating CV alerts with traffic signal design and exposure data like pedestrian and vehicle flow at intersections, bus driver characteristics, driver reaction time, stopping sight distance of CVs, and socio-economic characteristics of the area. Transit riders and commuters of the locations identified to have automated shuttle pilot projects can be surveyed to understand user opinions about this shuttle.

Future research can include more spatial and temporal information, driver demography and behavior information, and road geometry information on Automated Vehicle crashes for better insight into various safety features of AVs. Tesla motor vehicle's Autopilot should be studied in depth. It would be interesting in the future to see the role of Autopilot engagement and Tesla driver behavior in crashes that resulted in severe injuries and minor injuries. In the future, the availability of larger AV sample sizes may overcome some limitations. Although all results are not generalizable, future studies on AVs can be benefitted from the methods, results, and insights of this dissertation.

LIST OF REFERENCES

- [1] US Census Bureau. New Census Report Shows Public Transportation Commuters Concentrated in Large Metro Areas of the United States. 2021.
- [2] FHWA. Multiple Sources of Safety Information From V2V and V2I: Redundancy, Decision Making, and Trust - Safety Message Design Report CHAPTER 2. OVERVIEW OF CV APPLICATIONS. 2015.
- [3] Arvin, R., M. Kamrani, and A. J. Khattak. How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data. *Accident Analysis & Prevention*, Vol. 127, 2019, pp. 118-133.
- [4] Liu, J., and A. J. Khattak. Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles. *Transportation research part C: emerging technologies*, Vol. 68, 2016, pp. 83-100.
- [5] Khattak, A. J., and B. Wali. Analysis of volatility in driving regimes extracted from basic safety messages transmitted between connected vehicles. *Transportation research part C: emerging technologies*, Vol. 84, 2017, pp. 48-73.
- [6] Wali, B., A. J. Khattak, H. Bozdogan, and M. Kamrani. How is driving volatility related to intersection safety? A Bayesian heterogeneity-based analysis of instrumented vehicles data. *Transportation Research Part C: Emerging Technologies*, Vol. 92, 2018, pp. 504-524.
- [7] Kamrani, M., R. Arvin, and A. J. Khattak. Extracting useful information from Basic Safety Message Data: an empirical study of driving volatility measures and crash frequency at intersections. *Transportation Research Record*, 2018, p. 0361198118773869.
- [8] Kadam, A. J., V. Patil, K. Kaith, and D. Patil. Developing a Smart Bus for Smart City using IOT Technology. *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, 2018. pp. 1138-1143.
- [9] Tostado, R., A. Roviroso, J. Velázquez, E. Castillo, and O. Cervantes. A first experience with a Smart bus for improving public transportation. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 4, 2017.
- [10] Sungur, C., I. Babaoglu, and A. Sungur. Smart bus station-passenger information system. *2015 2nd International Conference on Information Science and Control Engineering*, IEEE, 2015. pp. 921-925.
- [11] Sridevi, K., A. Jeevitha, K. Kavitha, K. Sathya, and K. Narmadha. Smart bus tracking and management system using IoT. *Asian Journal of Applied Science and Technology (AJAST) Volume*, Vol. 1, 2017.
- [12] USDOT. *Arc-IT Service Packages*. <https://www.arc-it.net/html/servicepackages/servicepackages-areaspsort.html>. Accessed 1/4/2022.
- [13] Machek, E., E. Burkman, T. Crayton, J. Cregger, D. Diggs, S. Fischer, S. Mortensen, S. Peirce, H. Richardson, and A. Thomas. Strategic Transit Automation Research Plan. United States. Federal Transit Administration. Office of Research ..., 2018.
- [14] Valentine, D., A. Polinori, M. Windholtz, W. Gibbs, D. Neumeister, G. Baumgardner, and B. Paselsky. Connected Vehicle (CV) Infrastructure–Urban Bus Operational Safety Platform Project Report. United States. Federal Transit Administration. Office of Research, Demonstration, and Innovation, 2019.
- [15] Song, Z. Transition to a transit city: case of Beijing. *Transportation Research Record*, Vol. 2394, No. 1, 2013, pp. 38-44.

- [16] USDOT-FMCSA. Commercial Motor Vehicle Traffic Safety Facts. 2017.
- [17] Chimba, D., T. Sando, and V. Kwigizile. Effect of bus size and operation to crash occurrences. *Accident Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 2063-2067.
- [18] Goh, K. C. K., G. Currie, M. Sarvi, and D. Logan. Bus accident analysis of routes with/without bus priority. *Accident Analysis & Prevention*, Vol. 65, 2014, pp. 18-27.
- [19] Huting, J., J. Reid, U. Nwoke, E. Bacarella, and K. E. Ky. Identifying factors that increase bus accident risk by using random forests and trip-level data. *Transportation Research Record*, Vol. 2539, No. 1, 2016, pp. 149-158.
- [20] Strathman, J. G., P. Wachana, and S. Callas. Analysis of bus collision and non-collision incidents using transit ITS and other archived operations data. *Journal of safety research*, Vol. 41, No. 2, 2010, pp. 137-144.
- [21] Elvik, R. *State-of-the-art approaches to road accident black spot management and safety analysis of road networks*. Transportøkonomisk institutt Oslo, 2007.
- [22] Montella, A. A comparative analysis of hotspot identification methods. *Accident Analysis & Prevention*, Vol. 42, No. 2, 2010, pp. 571-581.
- [23] Hauer, E., J. Kononov, B. Allery, and M. S. Griffith. Screening the road network for sites with promise. *Transportation Research Record*, Vol. 1784, No. 1, 2002, pp. 27-32.
- [24] Hauer, E., and B. Persaud. How to estimate the safety of rail-highway grade crossings and the safety effects of warning devices. *Transportation Research Record*, Vol. 1114, 1987, pp. 131-140.
- [25] Miaou, S.-P. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, Vol. 26, No. 4, 1994, pp. 471-482.
- [26] Shankar, V., F. Mannering, and W. Barfield. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention*, Vol. 27, No. 3, 1995, pp. 371-389.
- [27] Maher, M. J., and I. Summersgill. A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis & Prevention*, Vol. 28, No. 3, 1996, pp. 281-296.
- [28] Abdel-Aty, M. A., and A. E. Radwan. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, Vol. 32, No. 5, 2000, pp. 633-642.
- [29] USDOT. Enhanced Transit Safety Retrofit Package (E-TRP). U.S. Department of Transportation Intelligent Transportation Systems Joint Program Office, 2018.
- [30] DeWine, M., and T. J. Stickrath. *Welcome to the Crash Statistics System*. Ohio Department of Public Safety. <https://services.dps.ohio.gov/CrashOnline/CrashStatistics/Home>. Accessed 4/26/2019, 2019.
- [31] NOACA. *Northeast Ohio Areawide Coordinating Agency GIS Portal*, Ohio. <https://gis.noaca.org/portal/>. Accessed 7/15/2020.
- [32] Cuyahoga County Open Data. *Land Use 2014 WGS84*. ArcGIS.com. <https://data-cuyahoga.opendata.arcgis.com/datasets/cuyahoga:land-use-2014-wgs84/about>. Accessed 1.20.2022, 2022.
- [33] Transitfeeds. *GCRTA GTFS*. <https://transitfeeds.com/p/greater-cleveland-regional-transit-authority/214?p=5>. Accessed 6.9.2020.
- [34] Aly, H., A. Basalamah, and M. Youssef. Accurate and energy-efficient GPS-less outdoor localization. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, Vol. 3, No. 2, 2017, pp. 1-31.

- [35] u-blox. NEO-7P u-blox 7 Precise Point Positioning GNSS Module. 2014.
- [36] Kim, K., I. Brunner, E. Y. Yamashita, G. Nagamine, and J. Burke. Guilty by Association: A Spatial Analysis of Bus Stop Locations and Pedestrian Accidents. 2010.
- [37] Cui, Y., and S. S. Ge. Autonomous vehicle positioning with GPS in urban canyon environments. *IEEE transactions on robotics and automation*, Vol. 19, No. 1, 2003, pp. 15-25.
- [38] Kim, Y., and H. Bang. Introduction to Kalman filter and its applications. *Introduction and Implementations of the Kalman Filter*, Vol. 1, 2018, pp. 1-16.
- [39] Torres-Reyna, O. Panel Data Analysis Fixed and random Effects using Stata. 2007.
- [40] Greene, W. H. *Econometric analysis*. Pearson Education India, 2003.
- [41] SAS Institute Inc. *The COUNTREG Procedure Panel Data Analysis*. https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/etsug/etsug_countreg_details34.htm#etsug_countreg008815. Accessed 1/4/2022.
- [42] Peeters, A., M. Zude, J. Käthner, M. Ünlü, R. Kanber, A. Hetzroni, R. Gebbers, and A. Ben-Gal. Getis–Ord’s hot-and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. *Computers and Electronics in Agriculture*, Vol. 111, 2015, pp. 140-150.
- [43] ESRI. *Hot Spot Analysis (Getis-Ord Gi*)*. <https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm#>. Accessed 1.10.2022, 2022.
- [44] ---. *How IDW Works*. <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/3d-analyst/how-idw-works.htm>. Accessed 2.26.2022, 2022.
- [45] Johnston, K., J. M. V. Hoef, K. Krivouruchko, and N. Lucas. Using ArcGIS Geostatistical Analyst. ESRI, Redlands, CA, 2001.
- [46] Federal Motor Carrier Safety Administration. *Large Truck and Bus Crash Facts 2019*. USDOT. <https://www.fmcsa.dot.gov/safety/data-and-statistics/large-truck-and-bus-crash-facts-2019>. Accessed 5.26.2022, 2022.
- [47] World Health Organization. Global status report on road safety 2013: supporting a decade of action: summary. World Health Organization, 2013.
- [48] NHTSA. Traffic Safety Facts 2018 Data. USDOT, Washington DC, 2018.
- [49] AASHTO. *A Policy on Geometric Design of Highways and Streets*. 2018.
- [50] Young, K. L., P. M. Salmon, and M. G. Lenné. At the cross-roads: An on-road examination of driving errors at intersections. *Accident Analysis & Prevention*, Vol. 58, 2013, pp. 226-234.
- [51] Chandler, B. E., M. Myers, J. E. Atkinson, T. Bryer, R. Retting, J. Smithline, J. Trim, P. Wojtkiewicz, G. B. Thomas, and S. P. Venglar. Signalized intersections informational guide. United States. Federal Highway Administration. Office of Safety, 2013.
- [52] Van Maren, P. A. Correlation of design and control characteristics with accidents at rural multi-lane highway intersections. 1977.
- [53] de Lavalette, B. C., C. Tijus, S. Poitrenaud, C. Leproux, J. Bergeron, and J.-P. Thouez. Pedestrian crossing decision-making: A situational and behavioral approach. *Safety science*, Vol. 47, No. 9, 2009, pp. 1248-1253.
- [54] Schorr, J. P., and S. H. Hamdar. Safety propensity index for signalized and unsignalized intersections: Exploration and assessment. *Accident Analysis & Prevention*, Vol. 71, 2014, pp. 93-105.
- [55] Litman, T. Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. Victoria Transport Policy Institute, 2021.

- [56] Phan, M. T., I. Thouvenin, and V. Frémont. Enhancing the driver awareness of pedestrian using augmented reality cues. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2016. pp. 1298-1304.
- [57] Fu, X., Q. Nie, J. Liu, A. Khattak, A. Hainen, and S. Nambisan. Constructing spatiotemporal driving volatility profiles for connected and automated vehicles in existing highway networks. *Journal of Intelligent Transportation Systems*, 2021, pp. 1-14.
- [58] Clark, K., M. P. Clamann, and A. J. Khattak. Advancing Crash Investigation with Connected and Automated Vehicle Data. 2021.
- [59] Kamrani, M., R. Arvin, and A. J. Khattak. Extracting useful information from basic safety message data: An empirical study of driving volatility measures and crash frequency at intersections. *Transportation research record*, Vol. 2672, No. 38, 2018, pp. 290-301.
- [60] Karbasi, A., and S. O'Hern. Investigating the impact of connected and automated vehicles on signalized and unsignalized intersections safety in mixed traffic. *Future transportation*, Vol. 2, No. 1, 2022, pp. 24-40.
- [61] Li, T., and K. M. Kockelman. Valuing the safety benefits of connected and automated vehicle technologies. *Transportation Research Board 95th Annual Meeting*, No. 1, 2016.
- [62] Wang, C., Y. Xie, H. Huang, and P. Liu. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accident Analysis & Prevention*, Vol. 157, 2021, p. 106157.
- [63] Wu, J., H. Xu, Y. Zheng, and Z. Tian. A novel method of vehicle-pedestrian near-crash identification with roadside LiDAR data. *Accident Analysis & Prevention*, Vol. 121, 2018, pp. 238-249.
- [64] Tak, S., S. Kim, D. Lee, and H. Yeo. A comparison analysis of surrogate safety measures with car-following perspectives for advanced driver assistance system. *Journal of Advanced Transportation*, Vol. 2018, 2018.
- [65] Monreal, C. O., M. Pichler, G. Krizek, and S. Naumann. Shadow as route quality parameter in a pedestrian-tailored mobile application. *IEEE Intelligent Transportation Systems Magazine*, Vol. 8, No. 4, 2016, pp. 15-27.
- [66] Tsuji, T., H. Hattori, M. Watanabe, and N. Nagaoka. Development of night-vision system. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 3, No. 3, 2002, pp. 203-209.
- [67] Keller, C. G., T. Dang, H. Fritz, A. Joos, C. Rabe, and D. M. Gavrila. Active pedestrian safety by automatic braking and evasive steering. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 4, 2011, pp. 1292-1304.
- [68] Hadi, M., M. A. Islam, S. Afreen, and T. Wang. Evaluation of an advanced driver-assistance system to reduce pedestrian and rear-end crashes of transit vehicles. *Transportation research record*, Vol. 2675, No. 11, 2021, pp. 1301-1309.
- [69] Matsuo, K., N. Chigai, M. I. Chattha, and N. Sugiki. Vulnerable road user safety evaluation using probe vehicle data with collision warning information. *Accident Analysis & Prevention*, Vol. 165, 2022, p. 106528.
- [70] Banerjee, S., M. Jeihani, N. K. Khadem, and M. Kabir. Influence of Pedestrian Collision Warning Systems on Driver Behavior: A Driving Simulator Study. *arXiv preprint arXiv:2112.09074*, 2021.

- [71] Jung, H., M.-K. Choi, K. Soon, and W. Y. Jung. End-to-end pedestrian collision warning system based on a convolutional neural network with semantic segmentation. *2018 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, 2018. pp. 1-3.
- [72] Zhou, Q., and X. Wang. Vehicle-Pedestrian Conflict Scenarios Easily Missed by Pedestrian Collision Warning System. 2019.
- [73] Hussein, A., F. Garcia, J. M. Armingol, and C. Olaverri-Monreal. P2V and V2P communication for pedestrian warning on the basis of autonomous vehicles. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2016. pp. 2034-2039.
- [74] Hasan, R., M. A. Hoque, Y. Karim, R. Griffin, D. Schwebel, and R. Hasan. Smartphone-based Distracted Pedestrian Localization using Bluetooth Low Energy Beacons. *2020 SoutheastCon*, IEEE, 2020. pp. 1-2.
- [75] Agresti, A. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, 2007.
- [76] Howell, D. C. Chi-Square Test: Analysis of Contingency Tables. *International Encyclopedia of Statistical Science*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. pp. 250-252.
- [77] NOACA. *Northeast Ohio Areawide Coordinating agency GIS Portal*. <https://gis.noaca.org/portal/>. Accessed 7/15/2020.
- [78] Arroyo, V. A., S. E. Bennett, W. A. V. Clark, J. M. Crites, M. Dougherty, J. S. Halikowski, M. W. Hancock, S. Hanson, S. Heminger, C. T. Hendrickson, J. D. Holt, G. P. Lagrange, M. P. Lewis, J. McDonald, D. A. Osterberg, S. Palmer, S. Rosenbloom, H. G. Schwartz Jr, K. C. Sinha, D. Sperling, G. C. Thomas, and P. A. Washington. TCRP Report 165: Transit Capacity and Quality of Service Manual. Washington DC, 2013.
- [79] Valdes, V., G. W. Torn, S. Mortensen, and D. Diggs. Strategic Transit Automation Research Plan. Federal Transit Administration (FTA), 2018.
- [80] USDOT. *Preparing for the Future of Transportation: Automated Vehicle 3.0*. USDOT. <https://www.transportation.gov/av/3>. Accessed Nov. 20, 2018.
- [81] Azad, M., N. Hoseinzadeh, C. Brakewood, C. R. Cherry, and L. D. Han. Fully Autonomous Buses: A Literature Review and Future Research Directions. *Journal of Advanced Transportation*, Vol. 2019, 2019.
- [82] Alessandrini, A., R. Alfonsi, P. Delle Site, and D. Stam. Users' preferences towards automated road public transport: results from European surveys. *Transportation research procedia*, Vol. 3, 2014, pp. 139-144.
- [83] Pessaro, B. Evaluation of Automated Vehicle Technology for Transit. 2015.
- [84] Weele, M. V. *WEpods project*. DAVI on the road. <http://davi.connekt.nl/wepods-project/>. Accessed May 21, 2019, 2019.
- [85] Navya. *Carpostal and the city of Sion extend the Autonom shuttle experimentation*. Navya. <https://navya.tech/en/carpostal-and-the-city-of-sion-extend-the-navya-shuttle-experimentation/>. Accessed May 21, 2019, 2019.
- [86] Nordhoff, S., B. V. Arem, N. Merat, R. Madigan, L. Ruhrort, A. Knie, and R. Happee. User acceptance of driverless shuttles running in an open and mixed traffic environment. *Proceedings of the 12th ITS European Congress, Strasbourg, France*, 2017. pp. 19-22.
- [87] Salonen, A. O. Passenger's subjective traffic safety, in-vehicle security and emergency management in the driverless shuttle bus in Finland. *Transport policy*, Vol. 61, 2018, pp. 106-110.

- [88] Portouli, E., G. Karaseitanidis, P. Lytrivis, A. Amditis, O. Raptis, and C. Karaberi. Public attitudes towards autonomous mini buses operating in real conditions in a Hellenic city. *2017 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2017. pp. 571-576.
- [89] Yap, M. D., G. Correia, and B. Van Are. Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips. *Transportation research part a: policy and practice* Vol. 94, 2016, pp. 1-16.
- [90] Piao, J., M. McDonald, N. Hounsell, M. Graindorge, T. Graindorge, and N. Malhene. Public views towards implementation of automated vehicles in urban areas. *Transportation research procedia*, Vol. 14, 2016, pp. 2168-2177.
- [91] Sinner, M., S. Brawand, and U. Weidmann. Networking planning with autonomous buses. *97th Annual Meeting of the Transportation Research Board (TRB 2018)*, TRB Annual Meeting, 2018.
- [92] Ceder, A. A., T. Liu, A. Baravian, E. Purguette, and V. Bouthiere. Shared and Individual Autonomous Public-Transport Vehicles to Serve Air Passengers: A Comparison. 2018.
- [93] Winter, K., O. Cats, G. Correia, and B. van Are. Performance analysis and fleet requirements of automated demand-responsive transport systems as an urban public transport service. *International journal of transportation science and technology* Vol. 7, No. 2, 2018, pp. 151-167.
- [94] Bösch, P. M., F. Becker, H. Becker, and K. W. Axhausen. Cost-based analysis of autonomous mobility services. *Transport policy*, Vol. 64, 2018, pp. 76-91.
- [95] Dong, X., M. DiScenna, and E. Guerra. Transit user perceptions of driverless buses. *Transportation*, Vol. 46, No. 1, 2019, pp. 35-50.
- [96] Anania, E., S. Rice, S. Winter, M. Milner, N. Walters, and M. Pierce. Why People Are Not Willing to Let Their Children Ride in Driverless School Buses: A Gender and Nationality Comparison. *Social Sciences*, Vol. 7, No. 3, 2018, p. 34.
- [97] Lutin, J. M. Not If, but When: Autonomous Driving and the Future of Transit. *Journal of Public Transportation*, Vol. 21, No. 1, 2018, p. 10.
- [98] Lewis, P., and A. Grossman. Beyond Speculation 2.0: Automated Vehicles and Public Policy. 2019.
- [99] Cregger, J., M. Dawes, S. Fischer, C. Lowenthal, E. Machek, and D. Perlman. Low-Speed Automated Shuttles: State of the Practice Final Report. United States. Joint Program Office for Intelligent Transportation Systems, 2018.
- [100] Yin, R. K. *Case Study Research: Design and Methods*. California: Sage Publications, Newbury Park, 1989.
- [101] Brooks, D. Driverless shuttle at Fort Bragg is ‘Army’s future’. *The Fayetteville Observer*, 2017.
- [102] Mcity. Mcity Driverless Shuttle: A Case Study. University of Michigan, Ann Arbor, MI, 2018.
- [103] Moore, N. C. *Mcity Driverless Shuttle launches on U-M’s North Campus*. Michigan News. <https://news.umich.edu/mcity-driverless-shuttle-launches-on-u-ms-north-campus/>. Accessed Oct. 21, 2018, 2018.

- [104] Carney, S. *McCity Driverless Shuttle launches on U-M's North Campus*. The Michigan Engineer News Center, The University of Michigan Ann Arbor, MI <https://news.engin.umich.edu/2018/06/mcity-driverless-shuttle-launches-on-u-ms-north-campus/>. Accessed Feb. 4, 2019, 2019.
- [105] The city of Arlington. *Milo Pilot Program Closeout Report*. Arlington, Texas, n.d.
- [106] Schrock, S. *Schedule your free ride on Milo, Arlington's Autonomous Shuttle*. City of Arlington, Texas, Arlington, Texas. https://www.arlingtontx.gov/news/my_arlington_tx/news_archive/2017_archived_news/august_2017/schedule_your_free_ride_on_milo. Accessed Oct. 5, 2018, 2018.
- [107] Berger, N. Autonomous vehicles to hit the road in Bishop Ranch. *The Press*, thepress.net, 2018.
- [108] Green, A. Self-driving corporate shuttles hit the road in the East Bay. *San Francisco Business Times*, 2018.
- [109] Willis, L. Testing of first autonomous shuttle on public roads in California begins. Contra Costa Transportation Authority, Contra Costa, California, 2018.
- [110] Said, C. *San Ramon office park testing 2 self-driving shuttles*. SFGATE. <https://www.sfgate.com/news/article/San-Ramon-office-park-testing-2-self-driving-10981577.php>. Accessed May 17, 2019, 2019.
- [111] Noebel, A. California's First Driverless Bus Hits the Road in San Ramon. Bishop Ranch, San Ramon, California, 2018.
- [112] Descant, S. *Driverless Shuttle Pilot Enters Final-Phase Testing in California*. <https://www.govtech.com/fs/automation/Driverless-Shuttle-Pilot-Enters-Final-Phase-Testing-in-California.html>. Accessed Feb. 10, 2019, 2019.
- [113] Easymile. *RTD Launches the First Autonomous Shuttle in Denver*, Denver, Colorado. <https://easymile.com/rtd-launches-the-first-on-road-deployment-of-an-autonomous-shuttle-ez10-in-denver/>. Accessed May 7, 2019, 2019.
- [114] Murray, J. *Self-driving shuttle offers Denver a glimpse of its future — but will riders jump on board?* <https://www.denverpost.com/2019/01/31/denver-airport-self-driving-shuttle-easymile/>. Accessed May 7, 2019, 2019.
- [115] Genova, D. *Route 61AV: What RTD has learned*. RTD. <https://www.rtd-denver.com/news-stop/blogs/route-61av-what-rtd-has-learned>. Accessed 12/18/2019, 2019.
- [116] AAA. *AAA Free Self-Driving Shuttle Pilot Program*. <http://www.aaahoponlasvegas.com/>. Accessed Feb. 19, 2019, 2019.
- [117] Akers, M. How does downtown's autonomous bus work? *Las Vegas Sun*, Las Vegas, 2018.
- [118] WSB, and AECOM. *MnDOT Autonomous Bus Pilot Project Testing and Demonstration Summary* Minnesota, 2018.
- [119] IBM. *Local Motors Debuts "Olli", the First Self-driving Vehicle to Tap the Power of IBM Watson*. IBM. <https://www-03.ibm.com/press/us/en/pressrelease/49957.wss>. Accessed June 25, 2019, 2019.
- [120] Local Motors. *National Harbor*. Local Motors. <https://localmotors.com/national-harbor/>. Accessed June 25, 2019, 2019.
- [121] JTA. *JTA Holds Autonomous Vehicle Demonstration*. Jacksonville Transportation Authority. <https://www.jtafla.com/media-center/press-releases/jta-holds-autonomous-vehicle-demonstration/>. Accessed May 20, 2019, 2019.

- [122] Transdev. *Autonomous Mobility*. <https://www.transdev.com/en/our-innovations/shared-autonomous-mobility/>. Accessed May 10, 2019, 2019.
- [123] Ban, C. *Autonomous vehicles hitting Greenville County, S.C. roads*. National Association of Counties. <https://www.naco.org/articles/autonomous-vehicles-hitting-greenville-county-sc-roads>. Accessed June 26, 2019, 2019.
- [124] Transdev. *Transdev Launches Third AV Operation in Florida*. Transdev. https://www.transdevna.com/news/2018/05/21/transdev-launches-third-av-operation-in-florida/?fbclid=IwAR2wwqxLXBEJeGRy-oSURRWPrbTp85Hwp_7MoohGQOTL68ktaYF9tAdHixg. Accessed May 21, 2019, 2019.
- [125] Caplan, A. *Self-driving shuttle hits the streets*. Gainesville.com. <https://www.gainesville.com/news/20180503/self-driving-shuttle-hits-streets>. Accessed May 21, 2019, 2019.
- [126] Mcloud, D. *Driverless shuttles planned for Gainesville, Fla.* Equipment World.com. <https://www.equipmentworld.com/driverless-shuttles-planned-for-gainesville-fla/>. Accessed May 21, 2019, 2019.
- [127] Degan, R. J. *Local BART station celebrates first autonomous vehicle*. Pleasanton Weekly. <https://pleasantonweekly.com/news/2018/06/26/dublin-bart-receives-first-autonomous-vehicle>. Accessed June 28, 2019, 2019.
- [128] Wiggers, K. *May Mobility raises \$22 million to deploy autonomous shuttles in cities across the U.S.* <https://venturebeat.com/2019/02/12/may-mobility-raises-22-million-to-deploy-autonomous-shuttles-in-cities-across-the-u-s/>. Accessed April 19, 2019, 2019.
- [129] Noble, B. *Self-driving shuttles begin running in downtown Detroit*. *The Detroit News*, 2018.
- [130] Insurance Journal. *Lincoln, Nebraska, to Test All-Electric, Driverless Shuttle Technology*. Insurance Journal. <https://www.insurancejournal.com/news/midwest/2018/06/27/493424.htm>. Accessed Feb. 3, 2019, 2019.
- [131] City of Lincoln Nebraska. *Autonomous Shuttle Project*. <https://lincoln.ne.gov/city/ltu/shuttle/>. Accessed Oct. 10, 2018, 2018.
- [132] Musulin, K. *Columbus, OH launches first self-driving shuttle service Smart Circuit*. Smartcitiesdive. <https://www.smartcitiesdive.com/news/self-driving-shuttles-columbus-ohio-may-mobility/532664/>. Accessed May 6, 2019, 2019.
- [133] Nghiem, A. *USF Hosts Its First Ever Autonomous Vehicle Demonstration*. Globe Newswire. <https://www.globenewswire.com/news-release/2019/02/13/1724305/0/en/USF-Hosts-Its-First-Ever-Autonomous-Vehicle-Demonstration.html>. Accessed May 8, 2019, 2019.
- [134] Moleski, V. *Meet Olli: Sacramento State unveils new 3D-printed, electric-powered autonomous shuttles*. The Sacramento Bee. <https://www.sacbee.com/news/local/transportation/article226544445.html>. Accessed May 8, 2019, 2019.
- [135] UDOT. *2019 Tour Schedule*. UDOT
- UTA, Utah. <http://www.avshuttleutah.com/locations.aspx>. Accessed May 8, 2019, 2019.
- [136] Giardinelli, C. *Driverless shuttle debuts at Station Park*. KSL. <https://www.ksl.com/article/46576021/driverless-shuttle-debuts-at-station-park?fbclid=IwAR2iESpQWII24N-ZNphsFjSgtt9frBxAZQTIEDvOdMNQQNZbj17Dpz1hwcE>. Accessed June 24, 2019, 2019.

- [137] Bigelow, P. *May Mobility to deploy self-driving shuttles in Rhode Island*. Automotive News. <https://www.autonews.com/mobility-report/may-mobility-deploy-self-driving-shuttles-rhode-island>. Accessed May 20, 2019, 2019.
- [138] O'Brien, M. *Self-Driving Shuttle Pulled Over in Rhode Island*. Transport Topics News. <https://www.ttnews.com/articles/self-driving-shuttle-pulled-over-rhode-island>. Accessed May 21, 2019, 2019.
- [139] CB Insights. *5 Companies Working On Driverless Shuttles And Buses*. CBINSIGHTS. <https://www.cbinsights.com/research/driverless-shuttle-companies/>. Accessed May 20, 2019, 2019.
- [140] Coast Autonomous. *Coast Autonomous*. <https://coastautonomous.com/press>. Accessed May 20, 2019, 2019.
- [141] Continental. *Driverless Mobility*. Continental USA. <https://www.continental-automotive.com/Landing-Pages/CAD/CUE/Driverless-Mobility/Driverless-Mobility>. Accessed May 20, 2019, 2019.
- [142] First Transit. *Shared Autonomous Vehicle (SAV) Mobility Solutions*. First Transit. [https://www.firsttransit.com/services/shared-autonomous-vehicle-\(sav\)](https://www.firsttransit.com/services/shared-autonomous-vehicle-(sav)). Accessed June 10, 2019, 2019.
- [143] Haque, A. M., and A. J. Khattak. Evaluating Pedestrian Safety in Bus-Pedestrian Interactions with the Application of Enhanced Pedestrian Collision Warning Systems: Analysis of Large-Scale Basic Safety Message Data. Presented at 2022 TRB Annual Meeting, Washington DC, 2021.
- [144] Golub, A., V. Satterfield, M. Serritella, J. Singh, and S. Phillips. Assessing the barriers to equity in smart mobility systems: A case study of Portland, Oregon. *Case Studies on Transport Policy*, 2019.
- [145] Madadi, B., R. van Nes, M. Snelder, and B. van Arem. Assessing the travel impacts of subnetworks for automated driving: An exploratory study. *Case Studies on Transport Policy*, Vol. 7, No. 1, 2019, pp. 48-56.
- [146] NHTSA. Early Estimates of Motor Vehicle Traffic Fatalities And Fatality Rate by Sub-Categories in 2021. 2022.
- [147] ---. Overview of Motor Vehicle Crashes in 2020. Washington DC, 2022.
- [148] International, S. *SAE Levels of Driving Automation™ Refined for Clarity and International Audience*. SAE. <https://www.sae.org/blog/sae-j3016-update>. Accessed 7.29.2022.
- [149] Clements, L. M., and K. M. Kockelman. Economic effects of automated vehicles. *Transportation research record*, Vol. 2606, No. 1, 2017, pp. 106-114.
- [150] Wang, S., and Z. Li. Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PloS one*, Vol. 14, No. 3, 2019, p. e0214550.
- [151] FHWA, U. Impacts of Automated Vehicles on Highway Infrastructure. 2020.
- [152] Chen, H., H. Chen, Z. Liu, X. Sun, and R. Zhou. Analysis of factors affecting the severity of automated vehicle crashes using XGBoost model combining POI data. *Journal of Advanced Transportation*, Vol. 2020, 2020.
- [153] Kutela, B., S. Das, and B. Dadashova. Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors. *Accident Analysis & Prevention*, Vol. 165, 2022, p. 106473.

- [154] Boggs, A. M., B. Wali, and A. J. Khattak. Exploratory analysis of automated vehicle crashes in California: A text analytics & hierarchical Bayesian heterogeneity-based approach. *Accident Analysis & Prevention*, Vol. 135, 2020, p. 105354.
- [155] Sinha, A., V. Vu, S. Chand, K. Wijayarathna, and V. Dixit. A crash injury model involving autonomous vehicle: Investigating of crash and disengagement reports. *Sustainability*, Vol. 13, No. 14, 2021, p. 7938.
- [156] Boggs, A. M., R. Arvin, and A. J. Khattak. Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis & Prevention*, Vol. 136, 2020, p. 105406.
- [157] De Winter, J. C., R. Happee, M. H. Martens, and N. A. Stanton. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, Vol. 27, 2014, pp. 196-217.
- [158] Ashraf, M. T., K. Dey, S. Mishra, and M. T. Rahman. Extracting rules from autonomous-vehicle-involved crashes by applying decision tree and association rule methods. *Transportation research record*, Vol. 2675, No. 11, 2021, pp. 522-533.
- [159] Liu, Q., X. Wang, X. Wu, Y. Glaser, and L. He. Crash comparison of autonomous and conventional vehicles using pre-crash scenario typology. *Accident Analysis & Prevention*, Vol. 159, 2021, p. 106281.
- [160] Tarmizi, I. A., and A. Abd Aziz. Vehicle detection using convolutional neural network for autonomous vehicles. *2018 international conference on intelligent and advanced system (ICIAS)*, IEEE, 2018. pp. 1-5.
- [161] Wiseman, S., S. Adler-Golden, E. Ientilucci, and T. Perkins. Enhanced Target Detection Under Poorly Illuminated Conditions. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 2021. pp. 1425-1428.
- [162] Wang, H., Y. Yu, Y. Cai, X. Chen, L. Chen, and Q. Liu. A comparative study of state-of-the-art deep learning algorithms for vehicle detection. *IEEE Intelligent Transportation Systems Magazine*, Vol. 11, No. 2, 2019, pp. 82-95.
- [163] NHTSA. Summary Report: Standing General Order on Crash Reporting for Level 2 Advanced Driver Assistance Systems. 2022.
- [164] Jeong, H., Y. Jang, P. J. Bowman, and N. Masoud. Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. *Accident Analysis & Prevention*, Vol. 120, 2018, pp. 250-261.
- [165] Dobson, A. J., and A. G. Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.
- [166] Liu, J., A. Hainen, X. Li, Q. Nie, and S. Nambisan. Pedestrian injury severity in motor vehicle crashes: an integrated spatio-temporal modeling approach. *Accident Analysis & Prevention*, Vol. 132, 2019, p. 105272.
- [167] Liu, C., and R. Subramanian. Factors related to fatal single-vehicle run-off-road crashes. 2009.
- [168] Associates, P. a. *What Are the Most Common Types of Car Accidents? Nationwide Statistic Overview*. <https://www.prnewswire.com/>. <https://www.prnewswire.com/news-releases/what-are-the-most-common-types-of-car-accidents-nationwide-statistic-overview-301391422.html#:~:text=The%20most%20common%20type%20of%20maneuver%20that%20led%20to%20a,time%20are%20driving%20straight%20ahead>. Accessed 7.29.2022.

- [169] Chira-Chavala, T., and S. Yoo. Potential safety benefits of intelligent cruise control systems. *Accident Analysis & Prevention*, Vol. 26, No. 2, 1994, pp. 135-146.
- [170] Evans, L. Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian mechanics. *Accident Analysis & Prevention*, Vol. 26, No. 5, 1994, pp. 609-616.
- [171] Dixit, V. V., S. Chand, and D. J. Nair. Autonomous vehicles: disengagements, accidents and reaction times. *PLoS one*, Vol. 11, No. 12, 2016, p. e0168054.
- [172] NHTSA. *Risky Driving*. USDOT. <https://www.nhtsa.gov/risky-driving>.
- [173] Mueller, A. S., J. B. Cicchino, and D. S. Zuby. What humanlike errors do autonomous vehicles need to avoid to maximize safety? *Journal of safety research*, Vol. 75, 2020, pp. 310-318.
- [174] Varotto, S. F., C. Mons, J. H. Hogema, M. Christoph, N. van Nes, and M. H. Martens. Do adaptive cruise control and lane keeping systems make the longitudinal vehicle control safer? Insights into speeding and time gaps shorter than one second from a naturalistic driving study with SAE Level 2 automation. *Transportation Research Part C: Emerging Technologies*, Vol. 141, 2022, p. 103756.
- [175] Khattak, A. J., N. Ahmad, B. Wali, and E. Dumbaugh. A taxonomy of driving errors and violations: Evidence from the naturalistic driving study. *Accident Analysis & Prevention*, Vol. 151, 2021, p. 105873.
- [176] Treat, J. R., N. S. Tumbas, S. T. McDonald, D. Shinar, R. D. Hume, R. Mayer, R. Stansifer, and N. J. Castellan. Tri-level study of the causes of traffic accidents: final report. Executive summary. Indiana University, Bloomington, Institute for Research in Public Safety, 1979.
- [177] Singh, S. Critical reasons for crashes investigated in the national motor vehicle crash causation survey. 2015.
- [178] Administration, N. H. T. S. Preliminary statement of policy concerning automated vehicles. *Washington, DC*, Vol. 1, 2013, p. 14.
- [179] Tideman, M., M. C. van der Voort, B. van Arem, and F. Tillemma. A review of lateral driver support systems. *2007 IEEE Intelligent Transportation Systems Conference*, IEEE, 2007. pp. 992-999.
- [180] McKeever, B. B. Estimating the potential safety benefits of intelligent transportation systems. 1998.
- [181] Ervin, R., J. Sayer, D. LeBlanc, S. Bogard, M. Mefford, M. Hagan, Z. Bareket, and C. Winkler. Automotive collision avoidance system field operational test report: methodology and results. 2005.
- [182] Lee, D., S. Oh, S. Heo, and M. Hahn. Drowsy driving detection based on the driver's head movement using infrared sensors. *2008 Second International Symposium on Universal Communication*, IEEE, 2008. pp. 231-236.
- [183] Rakha, H. *Feasibility of using in-vehicle video data to explore how to modify driver behavior that causes nonrecurring congestion*. Transportation Research Board, 2011.
- [184] Kummetha, V. C., A. Kondyli, and S. D. Schrock. Analysis of the effects of adaptive cruise control on driver behavior and awareness using a driving simulator. *Journal of Transportation Safety & Security*, Vol. 12, No. 5, 2020, pp. 587-610.
- [185] Gaspar, J. G., C. W. Schwarz, T. L. Brown, and J. Kang. Gaze position modulates the effectiveness of forward collision warnings for drowsy drivers. *Accident Analysis & Prevention*, Vol. 126, 2019, pp. 25-30.

- [186] Regan, M. A., E. Mitsopoulos, N. Haworth, and K. Young. Acceptability of in-vehicle intelligent transport systems to Victorian car drivers. *Monash University Accident Research Centre*, 2002.
- [187] Cicchino, J. B. Effects of blind spot monitoring systems on police-reported lane-change crashes. *Traffic injury prevention*, Vol. 19, No. 6, 2018, pp. 615-622.
- [188] Farmer, C. M. Crash avoidance potential of five vehicle technologies. *Insurance Institute for Highway Safety*, 2008.
- [189] Jermakian, J. S. Crash avoidance potential of four passenger vehicle technologies. *Accident Analysis & Prevention*, Vol. 43, No. 3, 2011, pp. 732-740.
- [190] Blanco, M., J. S. Hickman, R. L. Olson, J. L. Bocanegra, M. Greening, P. Madison, G. T. Holbrook, R. J. Hanowski, A. Nakata, and D. S. Bowman. Investigating critical incidents, driver restart period, sleep quantity, and crash countermeasures in commercial vehicle operations using naturalistic data collection. United States. Department of Transportation. Federal Motor Carrier Safety ..., 2016.
- [191] Lee, S. E., E. C. Olsen, and W. W. Wierwille. A comprehensive examination of naturalistic lane-changes. United States. National Highway Traffic Safety Administration, 2004.
- [192] Watanabe, Y., and S. Ito. Influence of vehicle properties and human attributes on neck injuries in rear-end collisions. *Proceedings of the 20th International Technical Conference on the Enhanced Safety of Vehicles (CD-ROM)*. Washington, DC: National Highway Traffic Safety Administration, 2007.
- [193] Litman, T. *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, BC, Canada, 2017.
- [194] Hayes, B. Leave the driving to it. *American Scientist*, Vol. 99, No. 5, 2011, p. 362.
- [195] Fagnant, D. J., and K. Kockelman. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, Vol. 77, 2015, pp. 167-181.
- [196] Favarò, F. M., N. Nader, S. O. Eurich, M. Tripp, and N. Varadaraju. Examining accident reports involving autonomous vehicles in California. *PloS one*, Vol. 12, No. 9, 2017, p. e0184952.
- [197] Lin, R., L. Ma, and W. Zhang. An interview study exploring Tesla drivers' behavioural adaptation. *Applied ergonomics*, Vol. 72, 2018, pp. 37-47.
- [198] Xu, C., Z. Ding, C. Wang, and Z. Li. Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. *Journal of safety research*, Vol. 71, 2019, pp. 41-47.
- [199] Banks, V. A., K. L. Plant, and N. A. Stanton. Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016. *Safety science*, Vol. 108, 2018, pp. 278-285.
- [200] Barth, T., M. P. McKay, and R. Smith. Survivability of a Tesla collision into a non-operational crash attenuator in Mountain View, CA. *Traffic injury prevention*, Vol. 21, No. sup1, 2020, pp. S154-S156.
- [201] tesladeaths.com. *Tesla Deaths*. <https://www.tesladeaths.com/#totals>. Accessed 5.10.2021.
- [202] Horel, J.-B., C. Laugier, L. Marsso, R. Mateescu, L. Muller, A. Paigwar, A. Renzaglia, and W. Serwe. Using formal conformance testing to generate scenarios for autonomous vehicles. *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE, 2022. pp. 532-537.

- [203] Ma, W. Next Generation Mutation Testing: Continuous, Predictive, and ML-enabled. University of Luxembourg, Luxembourg, Luxembourg, 2022.
- [204] Göktaş, A., and Ö. İşçi. A comparison of the most commonly used measures of association for doubly ordered square contingency tables via simulation. *Advances in Methodology and Statistics*, Vol. 8, No. 1, 2011, pp. 17–37-17–37.
- [205] Newson, R. Parameters behind “nonparametric” statistics: Kendall's tau, Somers' D and median differences. *The Stata Journal*, Vol. 2, No. 1, 2002, pp. 45-64.
- [206] Newson, R. B. Interpretation of Somers' D under four simple models. 2014.
- [207] Al-Ghamdi, A. S. Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention*, Vol. 34, No. 6, 2002, pp. 729-741.
- [208] Kleinbaum, D. G., K. Dietz, M. Gail, M. Klein, and M. Klein. *Logistic regression*. Springer, 2002.
- [209] Feldman, R., and I. Dagan. Knowledge Discovery in Textual Databases (KDT). *KDD*, No. 95, 1995. pp. 112-117.
- [210] NHTSA. *Fatality Analysis Reporting System (FARS)*. <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>. Accessed 6.1.2022.
- [211] GOODCARBADCAR. *2019 US Auto Sales by Model Analysis*. <https://www.goodcarbadcar.net/2019-us-vehicle-sales-figures-by-model/>. Accessed 30.7.2022.

VITA

Antora Mohsena Haque was born in Dhaka, Bangladesh. After completing Higher Secondary School Certificate (HSC) exams, Antora got admission to the Bangladesh University of Engineering and Technology (BUET) through a competitive admission exam in 2008. During her undergraduate studies in the Urban and Regional Planning program, she became interested in transportation planning studies. Soon after graduating from BUET in 2013, she worked as a Junior Planner in her department's research projects. After working for a year, she decided to pursue higher education. She came to the US in 2014 and started her Master's program in Community and Regional Planning at the University of Texas at Austin. She used to commute by bus in Austin, Texas, every day to go to her campus. Since then, she decided to work for the improvement and safety of transit in her career. She finished her Master's degree in 2016 after completing a thesis titled "The Impact of Light Rail Transit on Residential Rental Market: Case Study of Dallas Area Rapid Transit (DART)." Soon after completing her Master's degree, she returned to Bangladesh to gain industry experience. She joined a renowned architecture firm as an Urban Planner in Dhaka and worked on various urban development-based projects. Antora again came to the US in 2018 to pursue a doctorate in Civil Engineering with a concentration in Transportation Engineering with a Master's in Statistics from the University of Tennessee, Knoxville (UTK). Her research interests include transportation safety, connected and automated vehicle technology, and intelligent transportation systems.

Antora served as a Secretary and Vice President at UTK Institute of Transportation Engineers (ITE)/Women in Transportation Seminar (WTS), Student Chapter for the academic years 2019-2020 and 2020-2021. She received the 2022-2023 - Tennessee Section Institute of Transportation Engineer's (TSITE) Billy M. Hart Memorial Scholarship. She was selected as one of the participants in Carnegie Mellon University's Civil and Environmental Engineering Rising Stars workshop in 2022.