

**EXPLORING POTENTIAL IMPACTS OF CONNECTED AND AUTOMATED
VEHICLE TECHNOLOGIES**

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

Iman Mahdinia

December 2021

Copyright © 2021 by Iman Mahdinia
All rights reserved.

*This dissertation is dedicated to my parents and my beloved family
who constantly encouraged me to pursue my dreams and finish my dissertation*

ACKNOWLEDGEMENTS

First and most of all, I am thankful to God for His blessings during my endeavors. Without His continuous guidance, completing my dissertation would have been impossible and my dream of earning a doctoral degree would have remained a mere dream.

I would like to thank everyone who has directly or indirectly supported, encouraged, and guided me in my quest for knowledge. I wish to express my profound appreciation to Dr. Asad J. Khattak. His timely and powerful mentorship in these four years guided and supported me in the completion of my dissertation and research study. It was his leadership and mentorship that enabled me to undertake and contribute to research programs at the University of Tennessee. It has been my honor working under his supervision. A debt of gratitude is extended to my dissertation committee, Dr. Russell Zaretski, Dr. Anahita Khojandi, and Dr. Candace Brakewood. I am honored to have their tremendous guidance, help, and support in my dissertation. Also, special thanks to Dr. Zaretski for supervising my M.S. statistics degree.

To my co-authors, Dr. Asad Khattak, Dr. Ramin Arvin, Dr. Amir Ghiasi, Amin Mohamadnazar, and Sevin Mohammadi, thank you for your valuable knowledge and insights into my research. I also want to thank all my lab mates for their continuous support and memorable experiences. Special thanks to the Department of Civil and Environmental Engineering faculty and staff. My work has been greatly supported through the National Science Foundation, Collaborative Sciences Center for Road Safety, US Department of Transportation, and Tennessee Department of Transportation.

My acknowledgment would be incomplete without thanking my biggest pillar of unconditional support and guidance, my mother, my father, and my family for their valuable supports and precious love for supporting me throughout these years. My Family's involvement, concern, forbearing, and understanding have been the prime support for me throughout these years and without their help, I could not reach this stage in my life.

ABSTRACT

Connected and automated vehicle technologies have the potential to significantly improve transportation system performance. In particular, advanced driver-assistance systems (ADAS), such as adaptive cruise control (ACC), cooperative adaptive cruise control (CACC), pedestrian crash prevention (PCP) system, and advanced automated collision notification (ACN) system may lead to substantial improvements in performance by decreasing driver inputs and taking over control of the vehicle. The main questions that might arise are 1) how we can quantify the potential environmental and safety impacts of these technologies and 2) what the potential for these technologies is to address some of the important transportation-related problems. Due to the limitation in the empirical CAV data, knowledge about their impacts is very limited. This dissertation attempts to fill a portion of the existing gap. The main focus of this study is on CAVs with a low level of automation. The key hypothesis is that the technologies utilized in the CAVs with a low level of automation (i.e., SAE automation levels 1 and 2) available in the market can benefit the environment, traffic safety, and the safety of vulnerable road users. The main contributions of this dissertation are 1) addressing research question one by analyzing new testbed microscopic level data generated by CAVs and 2) addressing the second research question by identifying some important transportation-related problems that potentially can be addressed by CAV technologies. Overall, this dissertation estimates CAV technologies' impacts based on the Haddon Matrix concept in terms of pre-crash, during-crash, and post-crash perspectives. Methodologically speaking, advanced machine learning, data science, and frequentist techniques including truncated regression model, two-stages residual inclusion treatment to address endogeneity, geographically and temporally weighted regression model to address heterogeneity, explainable machine learning methods, XGBoost, and SHAP technique to interpret the result of XGBoost model are utilized to conduct the analysis. The results presented in this dissertation are derived from analyzing CAV testbed data and real-world crash data. Finally, the implications of the findings and future research areas are discussed in each chapter to a comprehensive understanding of the potential impacts of CAVs.

Keywords: Connected and Automated Vehicles, Advanced Driver Assistance Systems, Safety Impact, Environmental Impact.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: SAFETY IMPACTS OF CONNECTED AND AUTOMATED VEHICLES: A SYSTEMATIC LITERATURE REVIEW USING ADVANCED DATA MINING AND TEXT ANALYTICS TOOLS	7
Abstract	8
Introduction.....	8
Study Framework.....	9
Text Analysis of Safety Oriented CAV Big Data Research	12
Conclusions.....	15
CHAPTER 3: SAFETY, ENERGY, AND EMISSIONS IMPACTS OF ADAPTIVE CRUISE CONTROL AND COOPERATIVE ADAPTIVE CRUISE CONTROL	18
Abstract	19
Introduction.....	19
Literature Review.....	21
Methods.....	22
Platooning Control	22
Safety Measures	22
Energy and Environmental Measures	23
Data	24
Results and Discussion	28
Safety Measures	28
Energy and Environmental Measures	33
Limitations	36
Conclusions.....	36
CHAPTER 4: INTEGRATION OF AUTOMATED VEHICLES IN MIXED TRAFFIC: EVALUATING CHANGES IN PERFORMANCE OF FOLLOWING HUMAN-DRIVEN VEHICLES	38
Abstract	39
Introduction.....	40
Literature Review.....	40
Methodology	42
Safety Indicator	44

Environmental Measure	47
Data	48
Results	50
Safety Evaluation	51
Environmental Indicators Evaluation	54
Discussion and Limitations	54
Conclusions	58
CHAPTER 5: HOW EFFECTIVE ARE PEDESTRIAN CRASH PREVENTION SYSTEMS IN IMPROVING PEDESTRIAN SAFETY?.....	60
Abstract	61
Introduction	61
Literature Review	62
Methodology	62
Data	63
Results	64
Heckman Sample Selection Model	70
Limitations	73
Conclusions	73
CHAPTER 6: UNDERSTANDING THE ROLE OF FASTER EMERGENCY MEDICAL SERVICE RESPONSE IN SURVIVAL TIME OF PEDESTRIANS IN FATAL CRASHES....	76
Abstract	77
Introduction	77
Literature Review	78
Methods	80
First Stage: 2SRI Endogenous Treatment (global model)	80
Second Stage: Truncated Regression (global model)	82
Geographically and Temporally Weighted Truncated Regression (local model)	82
Model Selection Approach	84
Data	84
Pre-crash Variables	86
Post-crash Variables	86
Data Transformation	86
Results	89

First Stage: 2SRI Endogenous Treatment Results	89
Second Stage: Truncated Regression Results (global model)	89
Second Stage: Geographically and Temporally Weighted Truncated Regression Results (local model)	92
Discussion and Limitations	97
Conclusions	98
CHAPTER 7: CAN FASTER CRASH NOTIFICATION TIME INCREASE SURVIVAL TIME OF BICYCLISTS IN FATAL CRASHES: APPLICATION OF EXPLAINABLE MACHINE LEARNING	100
Abstract	101
Introduction	101
Literature Review	103
Methods	103
Haddon Matrix	104
Machine Learning Model: XGBoost	104
SHAP Value	106
Data	106
Variables	106
Data Transformation	109
Results and Discussion	109
Conclusions	113
CHAPTER 8: CONCLUSIONS	114
REFERENCES	117
VITA	135

LIST OF TABLES

Table 2-1 Results of Topics Extraction through Factor Analysis.....	14
Table 3-1 VSP modes and average emissions rate (132).....	25
Table 3-2 Descriptive statistics of speed*	29
Table 3-3 A: Average aggregated driving volatility (m/s) - B: Fuel consumption (in ml).....	32
Table 3-4 Vehicle's total emissions (in grams) based on VSP model.....	35
Table 4-1 VSP Modes and Average Emissions Rate (132)	49
Table 4-2 Descriptive Statistics of Speed (m/s) and Spacing (m), Average of All Drivers	52
Table 4-3 Driving Volatility Values, Average of All Drivers	53
Table 4-4 Descriptive Statistics of TTC (sec), Average of All Drivers.....	55
Table 4-5 Vehicle's Fuel Consumption (in ml), Based on Kamal et al. Model (131).....	56
Table 4-6 Vehicle's Total Emissions (in grams) Based on the VSP Model	56
Table 5-1 Descriptive Frequencies of Key Variables	66
Table 5-2 Descriptive Statistics of Variables.....	67
Table 5-3 Estimation of the Risk of Pedestrian Fatality by Impact Speed	69
Table 5-4 Examples of the Performance of PCP Systems.....	69
Table 5-5 Frequency of Tests Resulted in Collisions	71
Table 5-6 Heckman Sample Selection Regression	72
Table 5-7 Marginal Effects of the Probability of being Selected (Probit Model)	74
Table 6-1 Variable statistics-FARS dataset	87
Table 6-2 Descriptive statistics of transformed variables.....	88
Table 6-3 Linear regression estimates for log-EMS response time (Stage 1)	90
Table 6-4 Truncated with 2SRI treatment regression model	91
Table 6-5 Geographically and temporally weighted truncated with 2SRI treatment regression model.....	93
Table 7-1 Classification of the variables based on the Haddon Matrix.....	107
Table 7-2 Variable Statistics.....	108
Table 7-3 Descriptive Statistics of Transformed Variables.....	110
Table 7-4 Difference in the mean of crash notification time in late-night hours, hit and run crashes, and urban or rural areas.....	112
Table 8-1 Classifying CAV technologies' benefits from Haddon's Matrix concept	116

LIST OF FIGURES

Figure 1-1 Framework of the dissertation.....	3
Figure 1-2 Outline of the dissertation	3
Figure 2-1 Study Framework	10
Figure 2-2 Word Cloud and Phrase Cloud of High-repetitive Words and Phrases in studies	13
Figure 2-3 Proximity Plot of CACC vs ACC	16
Figure 3-1 Definition and visualization of scenarios 1 and 2	25
Figure 3-2 Time-Speed diagram of a CACC-mode Test for the lead vehicle	27
Figure 3-3 (a) Time-Speed diagrams of selected tests in scenario 1; (b) Time-Speed diagrams of selected tests in scenario 2; (c) Time-Acceleration diagrams of selected tests in scenario 1; (d) Time-Acceleration diagrams of selected tests in scenario 2	30
Figure 3-4 Time series of the temporal driving volatility	31
Figure 3-5 Descriptive statistics of TTC (sec) for all tests in scenario 1.....	34
Figure 3-6 Time-Speed diagram of the tests during deceleration.....	34
Figure 4-1 Framework of study	43
Figure 4-2 Temporal driving volatility calculation.....	46
Figure 4-3 Vehicles Platooning Scenarios.....	49
Figure 4-4 A Sample of the leading vehicle (AV) Speed Profile (172).....	49
Figure 4-5 Speed-profile 2: Time-Speed and Time-Volatility Diagrams in Scenario 1 (Orange Lines) and Scenario 2 (Blue Lines) Including All Nine Drivers	52
Figure 4-6 Speed-profile 3: Time-Speed and Time-Volatility Diagrams in Scenario 1 (Orange Lines) and Scenario 2 (Blue Lines) Including All Nine Drivers	53
Figure 4-7 A Sample of the Average Relative Speed-Time Diagram for Scenario 1 and 2.....	55
Figure 4-8 Comparison Between Scenario 2 (Human-driven vehicle and AV) and Scenario 1 (Two Human-driven vehicles)	57
Figure 5-1 Test Scenarios in the Insurance Institute for Highway Safety Data (209).....	65
Figure 5-2 IIHS Pedestrian Crash Prevention System Testing, IIHS Images (209).....	65
Figure 5-3 Histogram of the Dependent Variable (Speed at Impact km/h). The Right Histogram Is an Expansion of the Distribution on the Left after Removing Zeros.....	68
Figure 5-4 “Plot for S-shaped Curves for Pedestrian Fatality Risk by Impact Speed” (208).....	68
Figure 5-5 Percentage of Successful Collision Avoiding by the PCP Systems.....	71
Figure 5-6 Density plot of speed at impact (left side) and Box plots of time-to-collision (right side) for different car model years from 2018 to 2021	71
Figure 6-1 Flow diagram of the models.....	81
Figure 6-2 Number of observations in each U.S. state	85
Figure 6-3 Distribution of TTD and EMS response time before and after the transformation	88
Figure 6-4 Scatterplot of pedestrian time-to-death vs EMS response time	88
Figure 6-5 Spatiotemporal variation of the coefficient of EMS response time estimated by the GTWTR model.	94
Figure 6-6 Spatial variation of the coefficient of pedestrian behaviors averaged across all years estimated by the GTWTR model.	95
Figure 6-7 Spatial variation of some coefficients averaged across all years estimated by GTWTR model.....	96

Figure 7-1 Study Framework	105
Figure 7-2 Shap Summary Plot of XGBoost Model for Bicyclist Time-to-death	110
Figure 7-3 SHAP value Plot of Natural Logarithm of Crash Notification Time.....	112

CHAPTER 1: INTRODUCTION

With the emergence of connected and automated vehicles (CAVs) equipped with wireless communication technologies, it is important to consider the large amount of information generated by electronic sensors in the combined ecosystems of CAVs. Generated large-scale empirical data has significant potential to facilitate a deeper understanding of what can be achieved from deploying CAVs on roadways and what transportation-related problems can be solved. Connected and automated vehicle technologies have the potential to significantly improve transportation system performance. In particular, advanced driver-assistance systems (ADAS), such as adaptive cruise control (ACC), cooperative adaptive cruise control (CACC), pedestrian crash prevention (PCP) system, and advanced automated collision notification (ACN) system may lead to substantial improvements in performance by decreasing driver inputs and taking over control of the vehicle. The main questions that might arise are 1) how we can quantify the potential environmental and safety impacts of these technologies and 2) what the potential for these technologies is to address some of the important transportation-related problems. Due to the limitation in the empirical CAV data, knowledge about their impacts is very limited.

As such, this research attempts to fill a portion of the existing gap. The main focus of this dissertation is on CAVs with a low level of automation. The key hypothesis of the study is that the technologies utilized in the CAVs with a low level of automation (i.e., SAE automation levels 1 and 2) available in the market can benefit the environment, traffic safety, and the safety of vulnerable road users. Overall, this dissertation estimates CAV technologies' impacts based on the Haddon Matrix concept in terms of pre-crash, during-crash, and post-crash perspectives. The overall framework of the study is provided in Figure 1-1. In summary, the main contributions of this dissertation are 1) addressing research question one by analyzing new test-bed microscopic level data generated by CAVs and 2) addressing the second research question by identifying some important transportation-related problems that potentially can be addressed by CAV technologies. In particular, each chapter presents the results of peer-reviewed papers either quantifying the potential impacts and benefits of an advanced driving assistant system (chapters 2, 3, and 4) or exploring the potential for CAVs to address some important problems mainly related to the vulnerable road users (chapters 5 and 6). Chapter 1 presents a systematic literature review of CAV safety impacts. Figure 1-2 illustrates the outline of the study. In this study, driving volatility and time-to-collision are utilized to evaluate traffic safety. Driving volatility is a surrogate measure of safety that quantifies the deviation of driving behavior from the norm. In other words, driving volatility captures alterations in the longitudinal control of a vehicle. Time-to-collision is another surrogate safety measure used in this dissertation to evaluate rear-end crash risk. Likewise, to investigate the environmental impacts of CAVs, two measures of fuel consumption and emissions are utilized. Methodologically speaking, advanced machine learning, data science, and frequentist techniques including truncated regression model, two-stages residual inclusion treatment to address endogeneity, geographically and temporally weighted regression model to address heterogeneity, explainable machine learning methods, XGBoost, and SHAP technique to interpret the result of XGBoost model are utilized to conduct the analysis. Due to the limitation in the data, making simplifying assumptions in this study is inevitable.

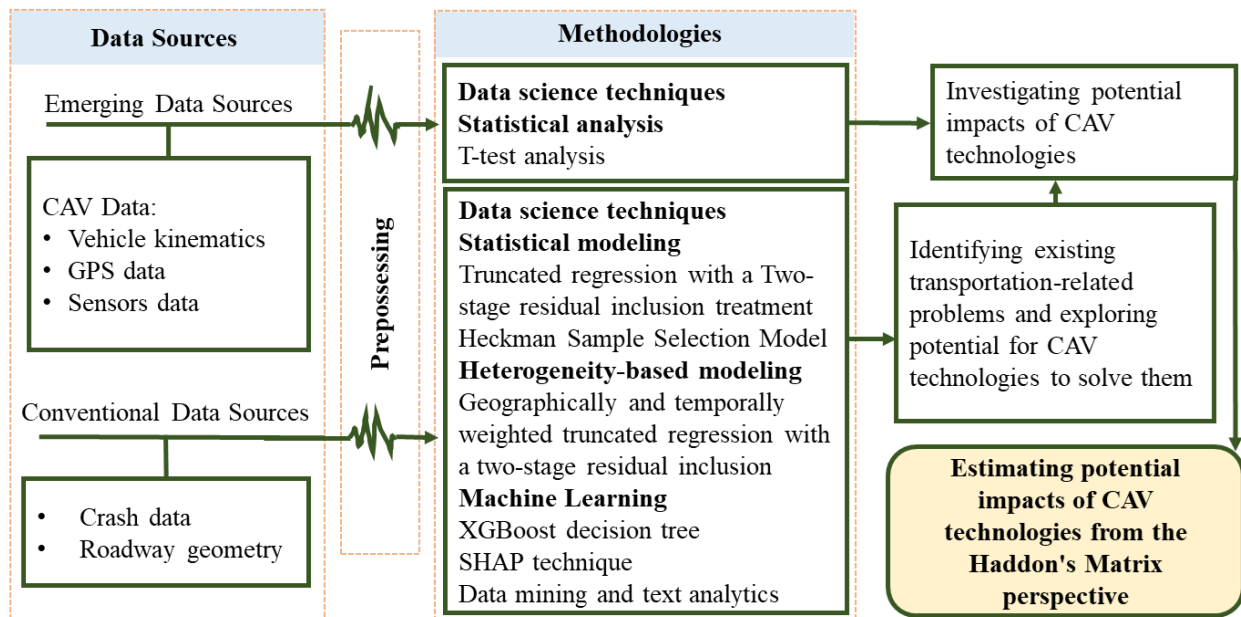


Figure 1-1 Framework of the dissertation

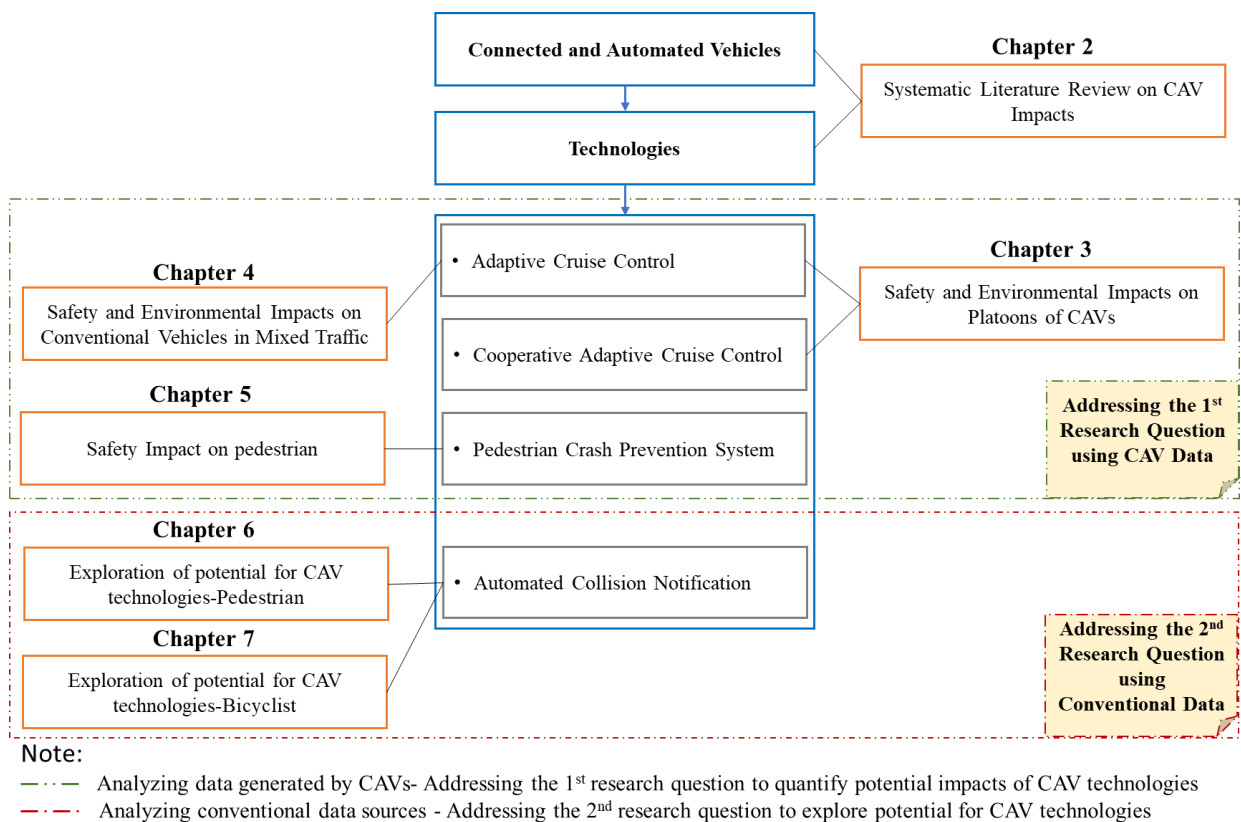


Figure 1-2 Outline of the dissertation

In summary:

Chapter 2 presents a systematic literature review using advanced data mining and text analytics tools to systematically synthesize past studies related to CAVs for safety monitoring and improvement to extract key concepts, patterns, and trends in CAV safety-related studies.

Chapter 3 aims to address the first research question. This chapter quantifies the potential safety, energy efficiency, and emission impacts of automated and cooperative systems in a platoon of five vehicles containing conventional, adaptive cruise control (ACC), and cooperative adaptive cruise control (CACC) vehicles by analyzing experimental data based on real-world conditions in tests conducted by the Federal Highway Administration, U.S. Department of Transportation. ACC and CACC controllers can improve the performance of vehicles by decreasing the driver inputs and taking control of the vehicle. Driving volatility and time-to-collision are utilized as surrogate safety measures to evaluate the safety and fuel consumption and emissions are utilized to evaluate environmental impacts.

Chapter 4 aims to address the first research question. This chapter focuses on the impacts of CAVs on the behavior of conventional vehicles in mixed traffic. This study creates new knowledge by quantifying the behavioral changes caused when conventional human-driven vehicles follow automated vehicles with SAE automation level one equipped with adaptive cruise control and investigating the impact of this behavior change on safety and the environment by analyzing the data obtained from a field experiment by Texas A&M University. Driving volatility and time-to-collision are utilized as surrogate safety measures to evaluate the safety and fuel consumption and emissions are utilized to evaluate environmental impacts.

Chapter 5 aims to address the first research questions. The objective of this chapter is to investigate the effectiveness and reliability of the pedestrian crash prevention (PCP) systems utilized in on-road vehicles in improving pedestrian safety. Therefore, this study analyzes the Insurance Institute for Highway Safety (IIHS) dataset where several vehicles available in the market and their technologies have been evaluated in terms of safety. In this chapter a Heckman sample selection type model is developed to identify key factors affecting the performance of the PCP systems.

Chapter 6 aims to address the second research question by exploring the potential for CAV technologies to address an important transportation-related problem. This chapter examines pedestrian survivability in fatal injury crashes. Fatalities and severe injuries among vulnerable road users, particularly pedestrians, are rising. Pedestrian fatalities, as a transportation-related problem, can potentially be reduced by CAV technologies, particularly the pedestrian crash prevention system and the advanced automated collision notification system (ACN) system utilized in CAVs. This study analyzes pedestrian time-to-death (survival time) in fatal injury crashes by utilizing the Fatality Analysis Reporting System dataset from 2015 to 2018 and developing a truncated regression with a Two-stage Residual Inclusion treatment. This study extracts valuable information that helps to reduce fatal pedestrian crashes particularly the role of faster emergency medical service response time.

Chapter 7 aims to address the second research question by exploring the potential for CAV technologies to address an important transportation-related problem. This study seeks to investigate the effect of faster crash notification time on the survivability of bicyclists in fatal injury crashes by developing an XGBoost machine learning model and interpreting the result using the SHAP technique. This study analyzes bicyclists' time-to-death (survival time) in fatal injury crashes by analyzing the Fatality Analysis Reporting System dataset from 2015 to 2019. Finally, this study shows that a delay in crash notification time is a problem that adversely reduces crash victim survival time. This problem can be alleviated by the advanced automated collision notification system (ACNS) utilized in CAVs.

Chapter 8 presents the overall conclusion of this dissertation. The overall intellectual merit and contribution of this dissertation is a robust analysis of the impacts of vehicles' automation and cooperation technologies using emerging large-scale data obtained from CAVs' field tests. While the benefits of these technologies are expected, this study quantifies and more realistically assesses the expected benefits of the technologies. Furthermore, this study explores the potential for CAV technologies to address some of the important transportation problems related to vulnerable road users.

It is expected that under this dissertation, at least five papers in top transportation conferences and journals be presented and published:

Mahdinia, I., Arvin, R., Khattak, A. J., & Ghiasi, A. (2020). Safety, energy, and emissions impacts of adaptive cruise control and cooperative adaptive cruise control.

- Published in the Journal of Transportation Research Record.
- Peer-reviewed conference paper: Presented at 99th Transportation Research Board Annual Meeting 2020, Washington DC
- Personal contribution ~75%; Study design; Data processing; Statistical modeling; Preparation of manuscript.

Mahdinia, I., Mohammadnazar, A., Arvin, R., & Khattak, A. J. (2021). Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles. Accident Analysis & Prevention, 152, 106006.

- Published in the journal of Accident Analysis & Prevention
- Peer-reviewed conference paper: Presented at 99th Transportation Research Board Annual Meeting 2020, Washington DC
- Personal contribution ~80%; Study design; Data processing; Statistical modeling; Preparation of manuscript.

Mahdinia, I. and Khattak, A. J. (2021). How Effective are Pedestrian Crash Prevention Systems in Improving Pedestrian Safety? Harnessing Large-Scale Experimental Data

- Journal article: Under review in the journal of Accident Analysis and Prevention.
- Peer-reviewed conference paper: Accepted for presentation at 101st Transportation Research Board Annual Meeting 2022, Washington DC

- Personal contribution ~95%; Study design; Data processing; Statistical modeling; Preparation of manuscript.

Mahdinia, I., Mohammadnazar, A., & Khattak, A. J. (2021). Understanding the Role of Faster Emergency Medical Service Response in Survival Time of Pedestrians in Fatal Crashes.

- Journal article: Under review in the journal of Accident Analysis & Prevention.
- Personal contribution ~80%; Study design; Data processing; Statistical modeling; Preparation of manuscript.

Mahdinia, I., Mohammadnazar, A., & Khattak, A. J. (2021). Can Faster Crash Notification Time Increase Survival Time of Bicyclists in Fatal Crashes: Application of Explainable Machine Learning.

- Journal article: Under review in Journal of Safety Research.
- Peer-reviewed conference paper: Presented at 100th Transportation Research Board Annual Meeting 2021, Washington DC.
- Personal contribution ~85%; Study design; Data processing; Statistical modeling; Preparation of manuscript.

CHAPTER 2: SAFETY IMPACTS OF CONNECTED AND AUTOMATED VEHICLES: A SYSTEMATIC LITERATURE REVIEW USING ADVANCED DATA MINING AND TEXT ANALYTICS TOOLS

Abstract

With the emergence of connected and automated vehicles (CAVs) equipped with wireless communication technologies, it is important to consider the large amount of information that is generated by electronic sensors in the combined ecosystems of CAVs. While generated large-scale empirical or simulation-based data has the significant potential to facilitate a deeper understanding of transportation-related problems, understanding what can be achieved from processing and analyzing the data generated by CAVs is still an open challenge. This study focuses on a systematic review of studies that utilized CAV-generated data to answer the fundamental question of how CAVs will improve the safety of road users. Natural language processing techniques, in addition to statistical procedures, can uncover the hidden thematic structure in a collection of textual information. Using advanced data mining and text analytics tools, this study systematically synthesizes past studies related to CAVs for safety monitoring and improvement to extract key concepts, patterns, and trends in CAV-related studies. The identified research challenges are open questions in the field of CAVs.

Keywords: Connected and automated vehicle, CAV, CAVs data, Safety, Text mining

Introduction

Rapid technological advancements in recent years have established the elemental foundation of Cooperative Intelligent Transportation Systems (C-ITS), resulting in the emergence of connected and automated vehicles (CAVs). Digital connectivity between vehicles and transportation infrastructure enables large and varied data sets, or "CAV Data". CAV Data allows transportation engineers and traffic managers to monitor, evaluate, and improve traffic safety. While generated large-scale empirical data has significant potential to facilitate a deeper understanding of transportation-related problems, the fundamental question to be answered is how CAVs will improve road users' safety. Although many studies in the past have utilized big data generated by CAVs to analyze traffic safety, there are uncertainties in the findings and additional uncertainties arise from the transition from conventional vehicles to CAVs. In this regard, this study explores the literature to provide insights into the potential impacts of CAVs on traffic safety. The study also explores the literature for ways in which CAV Data can be harnessed for the safe operation of CAVs and relevant challenges. It merits consideration of articulating prior studies' findings pattern. Within this domain, the study systematically compares CAV Data initiatives nationally and internationally, using advanced data mining and thematic text analytics tools, and provides insights regarding the analysis of CAV Data and evaluation of CAVs in terms of the potential safety impacts.

This study provides content/text analysis of a wide array of diverse papers that contain insight on the key topics of CAVs safety impacts analyzed by data generated by CAVs' deployment, test-bed experiments, and simulations. The identified research challenges are open questions in the field of CAVs. The outcomes of this study highlight the main topics of the literature, but also determine the research-worthy gaps in the field. Furthermore, the results can ultimately help provide a conceptual foundation for the development of new approaches for guiding and tracking the implications of CAV Data to evaluate the impacts of CAVs on traffic safety. The study is original in its application of text analysis on 80 studies in order to give help to a wide-ranging review of CAV safety impacts and infrastructure design requirements. The remainder of this paper encompasses a description of the study framework, the results from text

analysis regarding the safety aspect of CAVs, and the literature review related to the findings are presented and discussed. Finally, the last section comprises the conclusions.

Study Framework

Figure 2-1 illustrates the study framework. The main objective of this research is to systematically review studies, papers, and reports related to the analysis of CAV Data and the safety impacts, and extract key topics, patterns, and trends using advanced data mining and text analysis tools. Data collected from CAVs are one of the most important data sources available today, providing researchers with useful information. This research is mainly focused on safety, although studies that have used CAV Data contain a broad range of concepts such as mobility, security, and accessibility. The study conducted a comprehensive search using different search engines such as Google Scholar, Science Direct, Web of Science, and journal databases to select the intended studies in journal papers, conference papers, and project reports. Example keywords include “connected vehicles”, “automated or autonomous vehicles”, “CAV”, “safety”, and “impact”. Researchers also surveyed the references of each literature in order to find and add relevant sources to the main database.

After gathering a vast database of literature, the study evaluated the studies to determine their relevance and key results/conclusions were extracted from the studies. In addition, this study conducted rigorous and advanced text analyses of research activities on high-quality literature on CAV safety-oriented studies. Advanced text analytics facilitate the systematic analysis of otherwise “random textual data” through the application of statistical pattern learning techniques. This study categorizes and stems a corpus of individual research sources to spot unique concepts and entities to produce granular taxonomies. Specifically, text categorization and stemming techniques help extract the most important key topics and phrases through factor analysis (1). Machine learning can be applied to natural language processing techniques (in addition to statistical procedures) to uncover the hidden thematic structure in a collection of textual information, in this case, individual papers or reports (2). To perform content analysis, an “inclusion dictionary” is developed. Then, frequency analysis is applied to identify shared and the most frequently used keyword and phrases in the literature. After performing an initial analysis, an exclusion list is made to remove the words that either carry little semantic value, such as propositions, conjunctions, etc., or those frequently used words with little discriminative value. The “inclusion dictionary” merges different forms of a word (e.g. “vehicles” and “vehicle”) to consider them as a single word. Various visualization tools such as “word clouds”, “proximity plots”, and “concept maps” display the key concepts of the literature based on co-occurrences and the statistical analysis. Word cloud plots are used to demonstrate the frequency statistics of the word lists. In the word cloud, frequencies are converted to words with different sizes, the more frequently the word appeared in the studies, the larger the word would be in the plot. A proximity plot is a robust visualization tool that extracts information from a large amount of data stored in distance matrices (3). A concept map is a graphic representation of the proximity values calculated on all included words using multidimensional scaling. This study utilized Qualitative Data Analysis (QDA Miner) software for data processing and visualization. Statistical pattern learning analysis was performed with WordStat software on the collected research papers and technical reports.

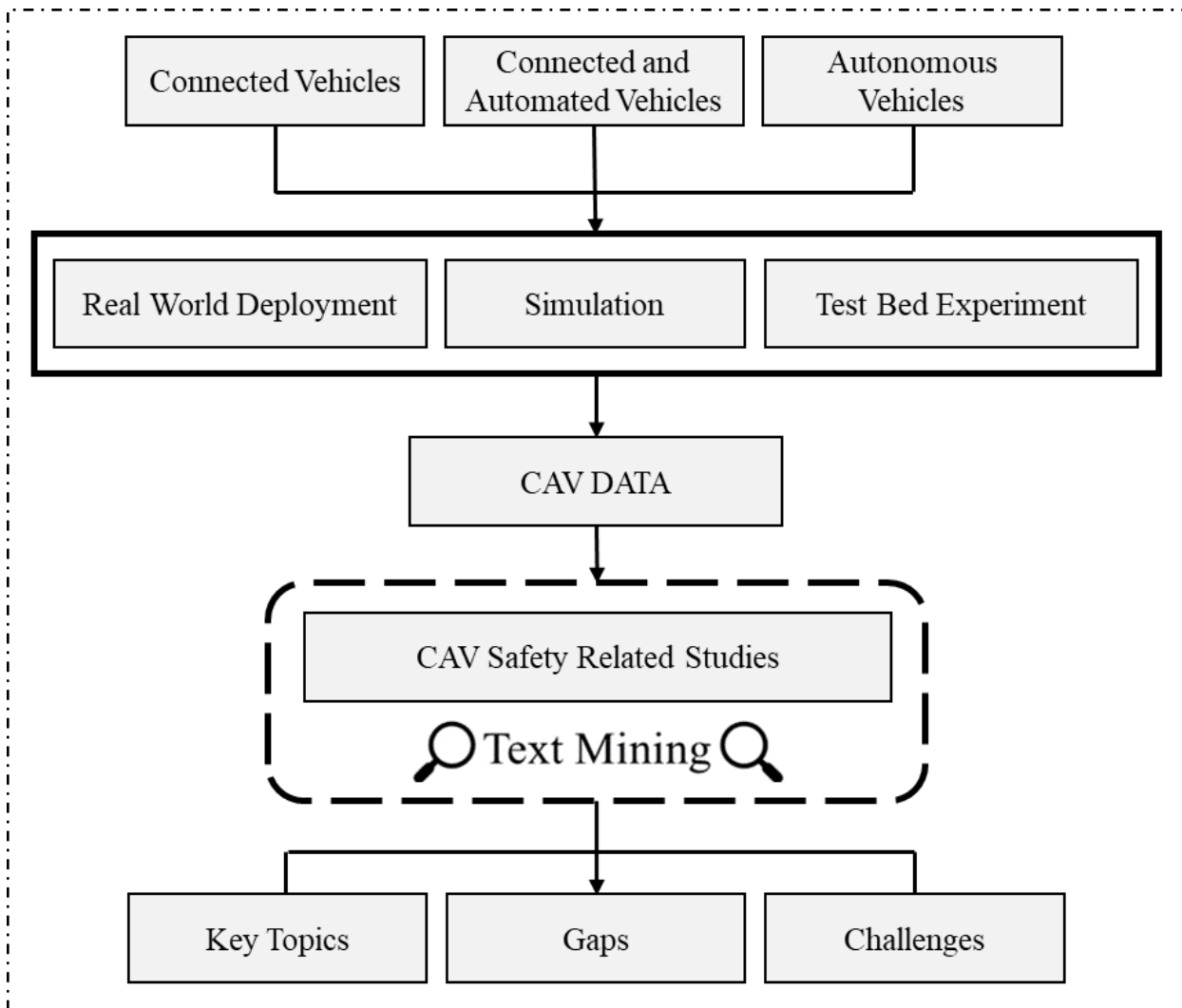


Figure 2-1 Study Framework

Reviewing Safety Oriented CAV Research

After reviewing worldwide peer-reviewed papers related to CAV safety assessment and improvement, it can be inferred that self-driving cars have positive impacts on mobility (4), such as increasing mixed-traffic performance and capacity (5; 6), increasing the efficiency and throughput of intersections (7), and lowering fuel consumption (8). However, CAV deployment offers both positive and negative impacts on safety. CAVs have the potential to reduce conflicts (9; 10) by improving the safety of lane change maneuvers (11-15) through advanced technologies, such as cooperative adaptive cruise control (CACC) (16), driving assistance systems (17; 18), dedicated short-range communication (DSRC) (19), Internet of Cars (20), blind-spot monitoring (21), lane departure warning, and forward-collision warning (22). It is believed that queue warning systems (23), collision notification systems (24; 25), and cooperative adaptive cruise control (26-30) improve the safety conditions of the network by reducing the number of rear-end conflicts. In addition, studies show that “inter-vehicle warning information system” (31) and “forward collision warning and autonomous braking systems” (32; 33) reduce the number of rear-end conflicts (34; 35) - up to approximately 84.3 percent for inter-vehicle warning systems with a 100% of market penetration.

Connected vehicle data have the potential to advance real-time road safety management (36). For the sake of connected vehicle technologies and low-level automation systems, the number of accidents in major categories can be reduced (37) by perceiving and avoiding unsafe events (38; 39). One study shows that intelligent cruise control systems (ICCS) can reduce accidents by up to 7.5 percent (40). Likewise, driving assistance technologies can reduce the number of crashes at least 32.99 and 40.88 percent for light and heavy vehicles, respectively, while a 35 percent reduction in near-crash events can be achieved in adverse weather conditions (41). Thus, the increasing automated vehicle market penetration can decrease the risk of secondary crashes by about 10 percent, decrease crash-related delays, and increase reliability (42-46). Moreover, CAV platooning can improve passenger comfort (47; 48), as well as longitudinal safety (49-52). Also, CAV platooning can positively impact safety for heavy-duty vehicles (53).

In contrast, it has long been debated that CAVs will increase vehicle miles traveled (54), and have negative impacts on safety that include technology failures (55; 56), privacy violations (57; 58), and cyber-attacks (59; 60). Examples of CAV safety issues are targeting the machine vision, Global Positioning Systems (GPS), in-vehicle devices, radar, lidar, maps, acoustic sensors, infrastructure equipment (61). This has prompted many researchers to propose platforms of safety impact analysis (62; 63) benefiting vehicle dynamics, sensing errors, and crash severity quantifications (64). However, safety forecasts for highly and fully automated vehicles are mostly based on assumptions that still need to be refined and validated in a more detailed fashion (65; 66). Automated technologies can reduce the drivers’ workload, provide safer control (67), and decrease driver error (68-70). However, automated vehicles (AVs) can cause passive fatigue and decrease the alertness of drivers, which in turn slows manual takeover (71-75). In addition, drivers’ reaction time in disengagement situations increases with increasing vehicle mile traveled in CAVs, which undermines the safe transition mode (76; 77). Nevertheless, different auditory systems employed in AVs can reduce the “reflecting information processing time” as a part of reaction time; however, the auditory systems do not change the “reflecting allocation of attention” part of reaction time (78).

Higher penetration rates of CAVs require a major revision in safety standards (79). With the emergence of high-level automated vehicles, the responsibility of driving errors moves from

drivers to vehicles and vehicle manufacturers. However, current standards and regulations do not address this issue, including ISO 26262, an international standard for functional safety of electrical and/or electronic systems in the production of automobiles defined by the International Organization for Standardization in 2011 (80). The impacts of CAVs in terms of interactions with other components of the transportation system also remain uncertain (57) while trust in AV technology is relatively low in interactions with pedestrians (81; 82).

Studies have reviewed key aspects of the emergence of CAVs, their impacts, and their requirements (83); in this study, the focus is placed on deriving high-quality structured information from safety-oriented Big Data studies. This helps researchers in systematically structuring and creating a synopsis of the current body of Big Data knowledge. Researchers extract meaningful information and produce new insights by spotting Big Data innovation indicators.

Text Analysis of Safety Oriented CAV Big Data Research

The study aims to systematically review relative studies of CAV Big Data-oriented in order to give help to a wide-ranging review of CAV safety impacts. To this end, advanced text analysis is conducted on the 80 selected studies. Figure 2-2 demonstrates the word and phrase cloud plot based on the frequency statistics of the word list and phrase list regarding the transportation safety-related dictionary. In a word cloud, frequencies are converted to words with different sizes, the more frequently the word appeared in the studies, the larger the word would be in the plot. For example, the word cloud emphasizes the words “control”, “crash”, “technology”, “brake, which are the top three words listed. Likewise, the phrases “lane change”, “real world”, “adaptive cruise control”, “real end”, and “emergency braking” are the top three phrases listed. The frequency of these phrases indicates that the focus of the evaluated studies was lane changing and rear-end collision risk analysis.

While the overall topic was safety-oriented CAV Big Data, the text mining technique was used to select the sub-topics. Table 2-1 shows the results of topic extraction using the non-negative matrix factorization method. Generally, higher coherence indicates higher variability explained by the topic. Higher coherence reflects if a topic represents a single theme or similar concepts by measuring the co-concurrency of the topic descriptors in the whole documents (84). The increased amount of coverage suggests the topic is more important. Therefore, coherence is a good measure to determine which topics are most important (85). The topic of “Technology Cost And Benefit” appears in 91 percent of the studies, and it has the highest coherence of 0.561. The second highest coherence belongs to “Autonomous Emergency Braking” with a coherence of 0.472, which appeared in 99 percent of the studies. Other topics of interest based on decreasing coherence include: “Vehicle Platoons”, “Cooperative Adaptive Cruise Control”, “Driving Task”, “Security”, “Fuel Consumption and Emissions”, “Market Penetration Rate”, “Dynamic Intersection”, and “Risk Assessment”.



Figure 2-2 Word Cloud and Phrase Cloud of High-repetitive Words and Phrases in studies

Table 2-1 Results of Topics Extraction through Factor Analysis

Topic	Keywords	Coherence	Frequency	% Cases
Autonomous Emergency Braking	AEB; FCW; Crash Involvement Rates; Vehicles With FCW; Front Crash Prevention; Acceleration; Deceleration; Braking; Emergency Braking; Time Headway; Longitudinal Control; Control Algorithm; Desired Acceleration;	0.472	2625	98.75%
Cooperative Adaptive Cruise Control	CACC; Collision Risk; CACC System; Collision Avoidance; CACC Platoon; Collision Warning; Crash Risks	0.441	1181	87.50%
Cost and Benefit	Cost; Benefit; Economic; Savings; Private; Technology; Crash Avoidance Technologies; Duty Vehicles; Cost Savings; Societal Benefits; Annual Net Benefit; Technology Purchasing; Safety Benefits; Wide Deployment; Technology Purchasing Costs; Economic Costs;	0.561	2247	91.25%
Risk Assessment	Risk; Prediction; Network; Safe; Risk Assessment; Collision Risk; Safe Distance	0.360	919	90.00%
Fuel Consumption	Fuel; Conventional; Volatility; Fuel Consumption; Human Driver; Driving Volatility; Human Drivers; Conventional Vehicles; Volatility Measures; Emissions; Human Factors; Miles Driven	0.399	562	78.75%
Market Penetration Rate	Penetration; MPR; Simulation; Stream; Market Penetration Rate; Simulation Experiments; Traffic Stream	0.397	878	77.50%
Dynamic Intersection	Intersection; Protocol; Stop; Traffic Throughput; Traffic Light; Cooperative Dynamic Intersection Protocol; Stop or Yield Control Method	0.387	495	67.50%
Driving Task	Fatigue; Passive; Task; Automation; Stress; Loss; Driving Task; Levels Of Automation; Full Automation; Task Engagement; Active And Passive Fatigue; Dynamic Driving Task; Emergency Event	0.440	694	61.25%
Security	Government; Cybersecurity; Privacy; Risks; Strategy; Liability; Security; Safety and Security; AV Testing	0.433	1019	60.00%
Connected Vehicle Platoons	Platoons; Lanes; CV Platoons; Lanes CV Platoons; CV Data; Light Vehicles; Standard Deviation of Speed; Time Exposed Time; Time Integrated Time	0.442	525	48.75%

Proximity plots are an accurate way to visualize keywords that co-occur with one or several target keywords. Figure 2-3 is the proximity plot of keywords considering Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) as the reference target keywords. This figure presents the major keywords on ACC and CACC and their proximity scores. For example, the keywords of “platoons”, “platform”, “Cooperation”, “Control”, “Risk”, and “traffic flow” have the highest proximity with CACC. Similarly, the keywords of “control”, “traffic flow”, “parameters”, “cooperation”, “comfortability”, “human”, and “collision” have the highest proximity with ACC. In addition, a comparison can be drawn between the proximity score of the common keyword in CACC and ACC. From the figure, it can be concluded that in studies in our database conducted on CACC and ACC, Fuel consumption and emissions were not a matter of concern. Likewise, when it comes to vehicle platooning, more studies have been conducted on the CACC compared to ACC technology.

Conclusions

Rapid technological advancements in recent years have manifested in the emergence of CAVs. The next frontier of transportation development is to equip motor vehicles and transportation systems with wireless communication technologies in a bid to establish cooperative, well-informed, and proactive transportation systems. In this context, it is important to consider the increasing amounts of information generated by electronic sensors in combined ecosystems of CAVs. While generated large-scale empirical data has significant potential to facilitate a deeper understanding of transportation-related problems particularly safety problems, integrating and processing large datasets in a meaningful manner is still an open challenge.

CAVs perform a set of actions that are implemented in diverse spatial, economic, and social contexts, and are expected to impact safety outcomes. The fundamental question is how CAVs will affect road users’ safety. An answer can be provided by Big Data as it can be used to integrate CAVs into the existing transportation system. Thereby, this study focuses on data generated from CAVs. The study utilized different search engines to locate several important documents which discuss CAV Big Data and CAV safety. Apart from carefully reviewing the literature (80 peer-reviewed papers and reports), rigorous and advanced machine learning text analytics tools are utilized to extract key concepts, spot patterns, and trends in studies related to CAV technologies and CAV safety monitoring and improvement. The study derives high-quality information from innovative research activities undertaken by various research entities through Big Data initiatives.

Big Data is harnessed in real-time for driver assistance (alerts and warnings) and enables higher levels of connectivity and automation where vehicles can take over control of speed and steering. The specific technologies that are being discussed in the literature include collision avoidance and vehicle control. Based on the text mining results, promising technologies that are utilized in real-time include “cooperative adaptive cruise control”, “adaptive cruise control”, “collision avoidance and warning”, “collision control”, “emergency autonomous braking”, “communication systems and data”, “autonomous lane changing”, and “driving task, engagement/disengagement”.

The text-mining results provide a conceptual foundation for developing new approaches for guiding and tracking the safety implications of Big Data and related innovations for CAVs. Although CAVs have substantial potential to improve traffic safety, the results reveal new

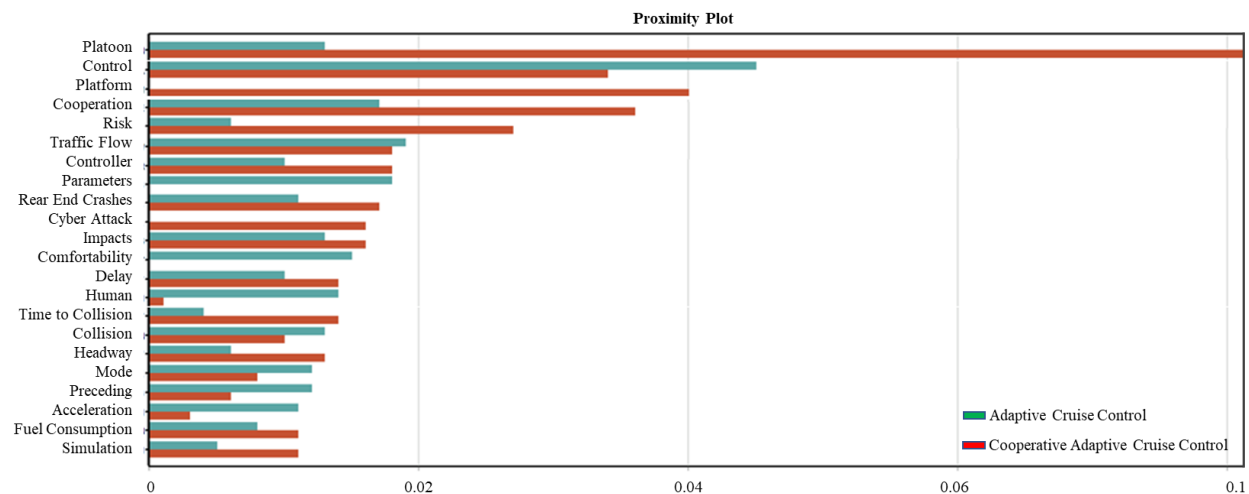


Figure 2-3 Proximity Plot of CACC vs ACC

challenges that CAV technology should address, such as cyber-attacks, privacy violation, invalidated assumptions, uncertain interaction behavior (e.g., between automated vehicles and pedestrians or bicyclists), technology failure, driver inattention (that can increase crash risk due to disengagement), high technology cost, and regulation inconsistency across states/countries. However, the Big Data gathered from CAVs promotes our understanding of transportation systems and traffic networks in detail. For instance, the archived Big Data can be used for accident reconstruction and this can provide a wide understanding of traffic safety. Similarly, the integration of real-time data from drivers, vehicles, and roadways leads us to build a more complete picture of the operational domain and provide warnings or take over control (at low levels of automation).

A wide range of stakeholders from government agencies, private sector companies, and academia are involved in CAV research, development, deployment, and overall CAV safety. Thus, the certification of the safety of the sensors, machine intelligence, and algorithms that create and process big data are of the utmost importance, especially for higher levels of vehicle automation, should be certified. In this regard, more research and development supported by the stakeholders will be critical to share safety practices and harness the true potential of CAV data. The topics that need more research are identified in this study: analysis of archived and real-time CAV data, an increased understanding of safety data to reduce transportation injuries and deaths, and changes to infrastructure geometry needed for successfully facilitating CAV operation.

Limitations regarding the results of this study are that they reflect the inputs; this means the results are affected by the selected literature and searched keywords. Noting that the literature review in this paper is comprehensive, but not exhaustive. Thereby, it is suggested to investigate more related studies in the future.

CHAPTER 3: SAFETY, ENERGY, AND EMISSIONS IMPACTS OF ADAPTIVE CRUISE CONTROL AND COOPERATIVE ADAPTIVE CRUISE CONTROL

This chapter is presented at the 99th Transportation Research Board Conference and is published in the Journal of Transportation Research Record.

Mahdinia, I., Arvin, R., Khattak, A. J., & Ghiasi, A. (2020). Safety, energy, and emissions impacts of adaptive cruise control and cooperative adaptive cruise control. *Transportation Research Record*, 2674(6), 253-267.

Abstract

Connected and automated vehicle technologies have the potential to significantly improve transportation system performance. In particular, advanced driver-assistance systems, such as Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC), may lead to substantial improvements in performance by decreasing driver inputs and taking over control of the vehicle. However, the impacts of these technologies on the vehicle- and system-level energy consumption, emissions, and safety have not been quantified in field tests. The goal of this paper is to study the impacts of automated and cooperative systems in mixed traffic containing conventional, ACC, and CACC vehicles. To reach this goal, experimental data based on real-world conditions are collected (in tests conducted by FHWA, USDOT) with presence of ACC, CACC, and conventional vehicles in a vehicle platoon scenario and a cooperative merging scenario. Specifically, a platoon of five-vehicles with different vehicle type combinations is analyzed to generate new knowledge about potential safety, energy efficiency, and emission improvement from vehicle automation and cooperation. Results show that adopting the CACC system in a five-vehicle platoon substantially reduces the driving volatility and reduces the risk of rear-end collision which consequently improves safety. Furthermore, it decreases fuel consumption and emissions compared with the ACC system and manually-driven vehicles. Results of the merging scenario show that while the cooperative merging system slightly reduces the driving volatility, the fuel consumption and emissions can increase due to sharper accelerations of CACC vehicles compared with manually-driven vehicles.

Keywords: Adaptive Cruise Control, Cooperative Adaptive Cruise Control, Vehicle Platooning, Safety, Big Data, Fuel Consumption, Emissions

Introduction

As connected and autonomous vehicles (CAVs) continue to develop quickly, it is becoming very momentous to consider the impacts of CAVs on traffic networks. In the CAV era, vehicles can communicate and convey their information, e.g., position and performance, in real-time. CAV

technologies have the potential to expressly improve transportation system safety and provide environmental benefits. For instance, advanced driver-assistance systems (ADAS), such as Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC), may lead to substantial improvements. Potentially, ACC and CACC controllers can improve the performance of the system by decreasing the driver inputs and taking control of the vehicle. However, the impacts of these technologies are uncertain and the amount of improvements in safety, energy consumption, and emission reduction is not well-known. Evaluating the performance of a transportation system through monitoring and measuring such impacts has been of vital concern for researchers (86; 87).

CAVs generate useful information (88) which allows for the analysis of some traffic safety factors (89). On-road testing and early deployments are critical to enhance the performance of vehicle automation. Safe deployment of ADAS is crucial and predominant to U.S. DOT's line of action. Real-world AV testing helps developers recognize and fix system limitations and drawbacks, not just on individual vehicles but across fleets (90; 91). The Cooperative Automated Research Mobility Application (CARMA) (92) is a U.S. federal field test of CAVs which "enables Automated Driving Systems (ADS) to navigate more safely and efficiently with other vehicles and roadway infrastructure through communication and cooperation" (92). CARMA proposes an opportunity to collect CAV's instantaneous kinematics that can be used for different studies such as CAVs performance evaluation and driver behavior analysis.

Driver behavior can influence various factors related to safety, energy, and emission (93; 94). Although there are several concerns regarding the cyber-security of CAVs (95; 96), the ACC and CACC have the potential to influence traffic safety, energy, and emission-related issues by reducing human driving tasks (97; 98). At low levels of automation, the ACC controller manages the brake and throttle to control and adjust the vehicle speed based on the lead vehicle speed. ACC capabilities can be further enhanced with the addition of Vehicle-to-Vehicle (V2V) communication, leading to the CACC (99). CACC represents moderate levels of autonomy and cooperation and provides a longitudinal control of speed and distance to the leading vehicle at automation level 1. The main advantage of communication between the vehicles is that the following vehicle receives the information not only from the preceding vehicle, but also from other upstream vehicles which brings about an anticipative reaction/control.

Despite the low market penetration of CAVs, interest in the concept of vehicle connectivity and particularly CACC has been growing in recent years. Although many studies have heretofore investigated the impacts of ADAS, the amount of improvements or deteriorations of using ADAS is an open question to answer. This study aims to generate new knowledge about potential safety and environmental improvements from vehicle automation, which will be useful for researchers, developers, and implementers. To this end, this study analyzes a test-bed dataset to shed light on the potential effects of using cooperative driving automation systems on traffic safety, fuel efficiency, and emission and provides a comparison between ACC and CACC impacts. In this paper, two measures are selected for safety evaluation including temporal driving volatility and Time-to-Collision (TTC). Volatility measures are used to quantify driving variations based on trajectories of vehicles. Volatility can be used to quantify traffic safety (100). Additionally, to

quantify safety, especially rear-end collision risk, TTC can be used as a surrogate measure (26). This measure uses trajectories of following and leading vehicles to quantify risk. It is widely used in the literature to quantify longitudinal crash risk in vehicle platoons (49; 101). Likewise, fuel consumption and emission are used to evaluate the energy and environmental impacts that will be described in detail in the methodology section.

Literature Review

It is expected that CAVs will enhance mixed-traffic performance and increase roadway capacity (5), reduce emission (102), enhance fuel economy (8), enhance public health (103) and decrease traffic conflicts (9) by improving safety through the use of advanced technologies, such as cooperative adaptive cruise control (16). It is believed that CACC improves the safety conditions of the network by reducing the number of rear-end conflicts (26; 29). The advantages of the CACC system can improve the string stability of vehicle platoons (104) and decrease the headway between vehicles (105). The ACC system's impacts on traffic safety (106; 107) and fuel efficiency (108) have been studied vastly (109). However, the amount of total improvement has remained uncertain in studies focusing on various CACC impacts such as improving the string stability (105), improving traffic flow (110), decreasing the risk of rear-end collision (26; 111), reducing fuel consumption (112), reducing the average following gap in a vehicle platoon (113) and several simulations and modeling studies in which the assumptions limit the outcome of the studies.

Emerging big data provides useful insights in traffic safety analysis (114; 115). CAVs generate useful information that allows researchers to analyze some traffic safety factors including driving behavior. From the safety perspective, driving behavior is considered as a leading cause of crashes (116). Studies have shown that driving volatility is an appropriate measure to capture driving behavior and safety performance of the network (117). Driving volatility is defined as "the extent of variations in driving, especially hard accelerations/braking, and frequent switching between different driving regimes" (118; 119). It should be noted that driving regimes refer to abrupt changes in instantaneous driving behavior in a driving cycle. In other words, driving volatility aims to quantify instantaneous variations from stable driving. Previous studies have used diverse measures to quantify driving volatilities (118-121). Several studies used vehicle kinematics to measure volatile behavior or aggressive behavior as a measure of safety (117; 119; 122; 123). Vehicle speed is used to quantify driving volatility (117) as a term of driving behavior. Driving volatility can be studied at different levels including vehicle-level or vehicle fleet-level. In order to understand the phenomena at the micro-level, this study utilized vehicle-level analysis. The data helps perform precise analysis and facilitates the comparison between scenarios in terms of vehicle safety and mobility performance.

As presented in this section, many potential improvements from vehicle automation and especially the CACC system are discussed in the literature. Although many studies evaluate the impacts of CAVs on traffic performance largely through several simulations and modeling techniques, the amount of the improvements based on field experimental data is scarcely provided and there is a need for more empirical evidence regarding the impacts of connectivity and automation for wide-scale deployment. The important contribution of this study is a robust analysis of the impacts of automation and cooperation using large-scale, federally collected big data obtained from the field test (92). This study quantifies the observed benefits of ACC and CACC

from the real-life testing through using key measures of performance to quantify improvements and gain insights into safety and energy/emissions benefits in vehicle platoons and during merging. While the benefits are expected, this study quantifies the benefits and points to a more realistic assessment of benefits.

Methods

This study aims to generate new knowledge about potential safety improvements and environmental benefits from investigating the impacts of ACC and CACC systems. This section first presents the ACC and CACC models and then introduces the measures used to evaluate the safety and environmental impacts.

Platooning Control

In this study, instead of applying typical ACC/CACC models, a platooning control algorithm is implemented, in which the target vehicle listens to all preceding vehicles but follows only one leader. This platooning algorithm originates from a CACC model developed by Milanés et al., (111) and Liu et al., (124), and extended to platooning control by Bujanovic (125). This algorithm provides the means for leader selection when two or more vehicles begin to communicate with each other. In essence, this algorithm finds the dynamic platoon leader for any subject vehicle. By default, the leader will be the first vehicle in the platoon. Then, the algorithm monitors the time headway between all preceding vehicles. If the time headway between any pair of the preceding consecutive vehicles gets less than a predefined lower bound threshold, denoted by h^{lb} , then the leader is changed to the first of those two vehicles. On the other hand, if the time headway between any pair of consecutive preceding vehicles gets greater than an upper bound threshold, denoted by h^{ub} , then the leader is changed to the second of those vehicles. For any subject vehicle n , given the platoon leader, the speed command of the subject vehicle, denoted by $\hat{x}_n$ is determined by the following equation:

$$\hat{x}_n = \hat{x}_l + k_1(x_l - x_n - (n - l)h^d\dot{x}_n) + k_2(\dot{x}_l - \dot{x}_n - (n - l)h^d\ddot{x}_n) \quad (1)$$

where $\hat{x}_l$ is the speed command of the leading vehicle, x_n and $\dot{x}_n$ are the current location and speed of the subject vehicle, respectively, x_l and $\dot{x}_l$ are the current location and speed of the leading vehicle, respectively, h^d is the desired time headway between a pair of consecutive vehicles, and k_1 and k_2 are coefficients. The parameters used in the field experiments are $h^{lb} = 0.4$ sec, $h^{ub} = 4.0$ sec, $h^d = 0.84$ sec, $k_1 = 0.05$, and $k_2 = 0.08$.

Safety Measures

In the literature, several surrogate safety measures are provided and discussed to quantify safety performance both at the macro or micro level. In terms of micro-level analysis, time-to-collision, TTC, is used to quantify longitudinal crash risk (49; 101) and driving volatility is used to quantify traffic safety (126). In this study, we utilized the concept of driving volatility and TTC to quantify the safety performance of the vehicles under different scenarios.

Driving Volatility

Volatility measures are used to quantify driving variation. Volatility measures try to capture variations in longitudinal control of the vehicle. To this end, these measures can be applied to

speed, acceleration/deceleration, and vehicular jerk. It has been shown that the speed volatility, in terms of Mean Absolute Deviation, contributes substantially to crash risk (117). We can write:

$$D_{mean} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (2)$$

where n is the number of observations, x_i is representing the observation i , and $\bar{x}$ is the mean.

In this study, the concept of temporal driving volatility is utilized (127) which reflects the variations in instantaneous driving decisions at the micro-level and at the end, provides time-series data. The advantage of this approach is capturing the time dependency between the observations which can help to detect and identify the times that the driver is exhibiting a volatile behavior (117). Similar to the concept of moving average, a 3-second time window is considered to calculate the temporal-volatility-measure and the measure is assigned to the subject time. The aggregation of the area under the curve of temporal driving volatility relating to a period is the total driving volatility of that period.

Time-to-Collision Measures

Time-to-Collision (TTC) is a surrogate safety measure (26) which is generally defined as “the duration of time before two objects collide with initial certain conditions” (128). This measure has been used vastly to assess the risk of the rear-end collision (26; 49; 128). If two following vehicles continue at their present speed, TTC can be defined as follows:

$$TTC = \frac{X_L(t) - X_F(t) - L}{V_F(t) - V_L(t)}, \quad \forall V_F(t) > V_L(t) \quad (3)$$

where $X_F(t)$ and $X_L(t)$ are the positions of the following and leading vehicles respectively with respect to time, $V_F(t)$ and $V_L(t)$ are the speed of following and leading vehicle with respect to time and L is the length of the leading vehicles.

Energy and Environmental Measures

Fuel Consumption

The fuel consumption in a vehicle equipped with an internal combustion engine is a function of many factors such as speed, acceleration, fuel type, temperature, atmospheric pressure, and gear and engine efficiency. Therefore, the estimation of fuel consumption requires a very complex equation. However, a simplified model of fuel consumption can approximate engine performance and provides a good estimation of real fuel consumption (129). In this study, regarding the similarity of the vehicles and the environment condition in the field test, it is assumed that the only parameters contributing to the fuel consumption are the vehicle's speed and acceleration. This assumption is reasonable because this study aims to compare the fuel consumption of similar vehicles in different modes (i.e. conventional, ACC, and CACC) relatively. To this end, the model proposed by Kamal et al. (130; 131) is used to calculate fuel consumption. Their proposed equation takes advantage of the relation between speed, acceleration and fuel consumption presented as follows:

$$F_v = (b_0 + b_1 v_t + b_2 v_t^2 + b_3 v_t^3 + a_t(c_0 + c_1 v_t + c_2 v_t^2)) \quad (4)$$

where F_v estimates vehicles fuel consumption (milliliter per second), v_t is the vehicle speed at time t (m/s), a_t is acceleration of the vehicle at time t (m/s²), $b = 0.1569$, $b_1 = 2.450 \times 10^{-2}$, $b_2 = -7.415 \times 10^{-4}$, $b_3 = 5.975 \times 10^{-5}$, $c_0 = 0.07224$, $c_1 = 9.681 \times 10^{-2}$, and $c_2 = 1.075 \times 10^{-3}$. It is assumed that the vehicle consumes no fuel during deceleration where $a_t < 0$ (131). Due to the similarity of the vehicles used in the CARMA empirical dataset, the following parameters are assumed to be the default values for all vehicles. The equivalent mass for the vehicles is $M_v = 1200$ kg, the drag coefficient is $C_D = 0.32$, the air density is $P_a = 1.184$ kg/m³, the frontal area is $A_f = 2.5$ m², and the rolling resistance coefficient is $\mu = 0.015$ (130; 131).

Emissions

The vehicle-specific power (VSP) microscopic model is used to estimate emissions regarding vehicle second-by-second speed, acceleration and terrain gradient (132). The VSP is calculated using equation 5. Also, total emission by each vehicle can be estimated as:

$$VSP = v[1.1a + 9.81\sin(\text{Arctan}(\text{grade})) + 0.132] + 0.000302v^3 \quad (5)$$

where VSP is vehicle specific power [kW/metric ton]; v is vehicle speed [m/s]; a is acceleration (+) or deceleration (-) [m/s²]; and grade is terrain gradient [$\pm\%$].

$$\text{Total Emissions} = CO_2^{VSP-Mode} * t + CO^{VSP-Mode} * t + NO_x^{VSP-Mode} * t + HC^{VSP-Mode} * t \quad (6)$$

where $VSP-Mode$ can be derived from Table 3-1; t is the duration vehicle drives in the VSP-Mode; CO_2 is carbon dioxide (g/s); CO is carbon monoxide (g/s); NO_x is oxides of nitrogen (g/s); HC is hydrocarbons (g/s).

Data

This study takes advantage of analyzing a unique dataset collected during USDOT's CARMA program (92). The data represent a proof-of-concept vehicle platooning based on the ACC and CACC applications. CARMA platform version 1 enables the implementation of the proof-of-concept CACC-based platooning in passenger vehicles equipped with production adaptive cruise control, and vehicle-to-vehicle communications using Dedicated Short-Range Communications (DSRC) to enable the automatic synchronization of longitudinal movements of a string of vehicles. The data characterize the state-of-the-art capability of the CACC application based on engineering tests that were performed on closed tracks by professional drivers and using prescribed test procedures. Likewise, CARMA platform version 2 includes the integrated highway prototype that consists of vehicle platooning, speed harmonization, and automated lane change and merge (92; 133). The dataset consists of two scenarios including (i) vehicle platooning scenario and cooperative merge scenario as shown in Figure 3-1. The first scenario contains the variables of speed, acceleration/deceleration and the position of the vehicles in a fleet of five Cadillac SRX vehicles including a lead vehicle and four following vehicles. The data is collected in a 4.5-mile two-lane test track at the Aberdeen Test Center (ATC) in Maryland and is collected with a frequency of 20 Hertz. The data encompass sudden speed change scenarios along with braking and accelerating with several tests of CACC and ACC modes. The two types of platoon modes were selected which are described as follows: (133)

Table 3-1 VSP modes and average emissions rate (132)

VSP-Mode	Definition (kw/t)	CO ₂ (g/s)	CO (mg/s)	NO _x (mg/s)	HC (mg/s)
1	VSP < -2	1.671	7.807	0.901	0.450
2	-2 ≤ VSP < 0	1.458	3.908	0.628	0.257
3	0 ≤ VSP < 1	1.135	3.347	0.346	0.406
4	1 ≤ VSP < 4	2.233	8.335	1.173	0.432
5	4 ≤ VSP < 7	2.920	10.959	1.706	0.530
6	7 ≤ VSP < 10	3.525	17.013	2.368	0.705
7	10 ≤ VSP < 13	4.107	20.026	3.103	0.822
8	13 ≤ VSP < 16	4.635	29.222	4.234	0.976
9	16 ≤ VSP < 19	5.161	35.531	5.069	1.112
10	19 ≤ VSP < 23	5.633	55.068	5.865	1.443
11	23 ≤ VSP < 18	6.535	113.824	7.623	2.061
12	28 ≤ VSP < 33	7.585	207.586	12.149	3.373
13	33 ≤ VSP < 39	9.024	441.775	15.456	4.857
14	39 ≤ VSP	10.088	882.300	17.863	10.948

Notes: The average emission rates are for Tier 1 light-duty gas vehicles with engine displacement <3.5L and odometer< 50,000 miles (132).

- **ACC (Hybrid):** CACC in LV and ACC in FVs.
- **CACC:** LV and FVs are on CACC mode.

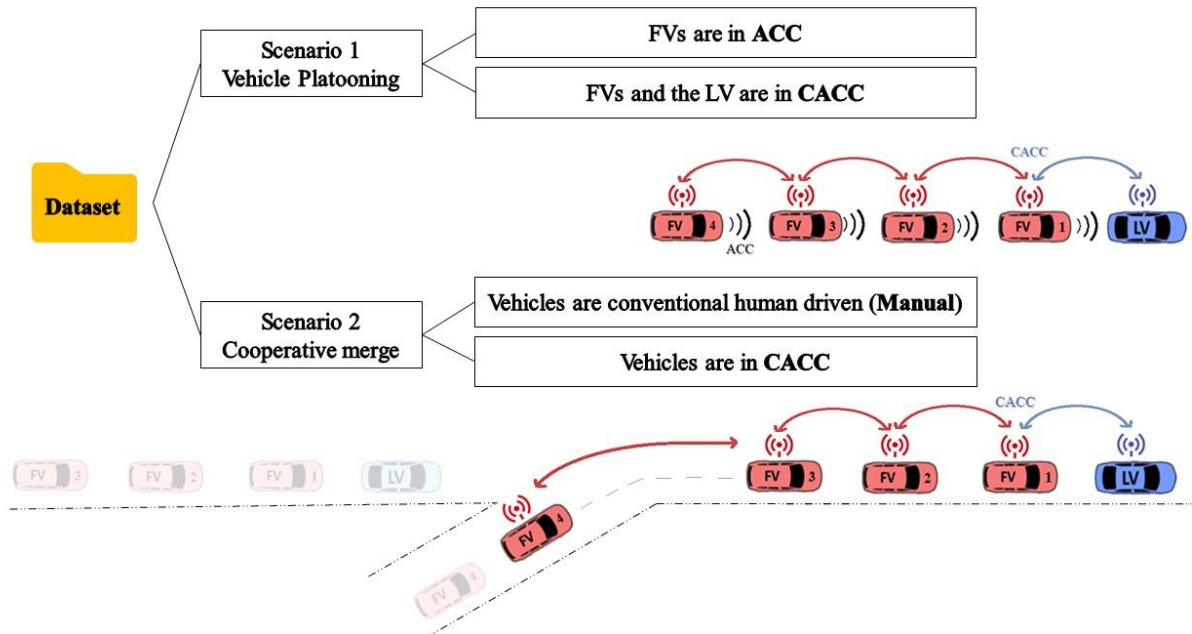


Figure 3-1 Definition and visualization of scenarios 1 and 2

The dataset consists of 72 tests with different time durations. Among them, 11 tests are selected for this study based on the availability of the required data and similarity of the speed profile of the lead vehicles between ACC and CACC modes. After processing the data for each test, a period of 200 seconds is extracted, in which the speed varies between 60 and 45 mph along with a sudden speed change with a deceleration and acceleration period (see Figure 3-2). As the speed data has substantial noise, the moving average method (134) is used to smooth the data with an order of 30 observations. Figure 3-2 illustrates the time-speed graph for the lead vehicle before and after smoothing. Note that due to the noise in the speed data, it was necessary to derive the acceleration from the smoothed speed records. This eliminated the difference between the given acceleration available in the data and the observed acceleration regarding the smoothed speed data.

As mentioned, the speeds of lead vehicles in both CACC and ACC scenarios are pre-determined by a prescribed speed profile. Therefore, the difference between these scenarios is the following vehicles. The algorithm used by the CACC platooning was developed by Bujanovic (125; 135). When vehicles begin to follow and communicate with each other, the algorithm selects the leader. Then, the following vehicles select “a motion profile in close proximity to the lead vehicle or the rearmost vehicle in the lead vehicle’s position” (105). In the platoon, each vehicle receives the command speed of the preceding vehicle and continually receives the position of every downstream vehicle. Once the platoon is formed, the controlling system reports the target gap to the lead vehicle and the following vehicles. In the first scenario, the target time gap for the ACC mode was set at 1.1 seconds and 1.2 seconds for the CACC mode. Also, the lead vehicle for the ACC mode had lower acceleration and deceleration limits (0.25 m/s^2), than the lead vehicle for the CACC mode (1.0 m/s^2 and 0.3 m/s^2 , respectively). This difference may have a small impact on the assessment which will be discussed further. The limits for ACC mode were lower because it could not maintain a stable platoon at higher limits (133).

The merging scenario, referred to as Scenario 2, has four vehicles traveling in a platoon and one vehicle merges into the lane (from an on-ramp) to join the platoon. This scenario has two modes. In the first mode, vehicles are driven manually, and the merging vehicle is released with a conventional metering light. The lead vehicle follows a predetermined trajectory and the following vehicles try to manually maintain a specific gap. In the second case, the cooperative platooning maneuver was used to form a platoon and the merging vehicle was released automatically (CACC) through V2V and V2I communications to enable the merging vehicle to signal other vehicles of its purpose to merge into traffic flow and join the platoon. The lead vehicle follows a predetermined speed profile and the following vehicles try to automatically regulate their speed (105). A test for automated merging mode and a test for manual driven mode are selected and variables of speed, acceleration, deceleration, time to merge and following distance are captured with a resolution of 0.05 seconds (20 Hertz) in this scenario. For more information about the used algorithm and the field experiment please see Bujanovic et al., (135) and Ma et al., (105).

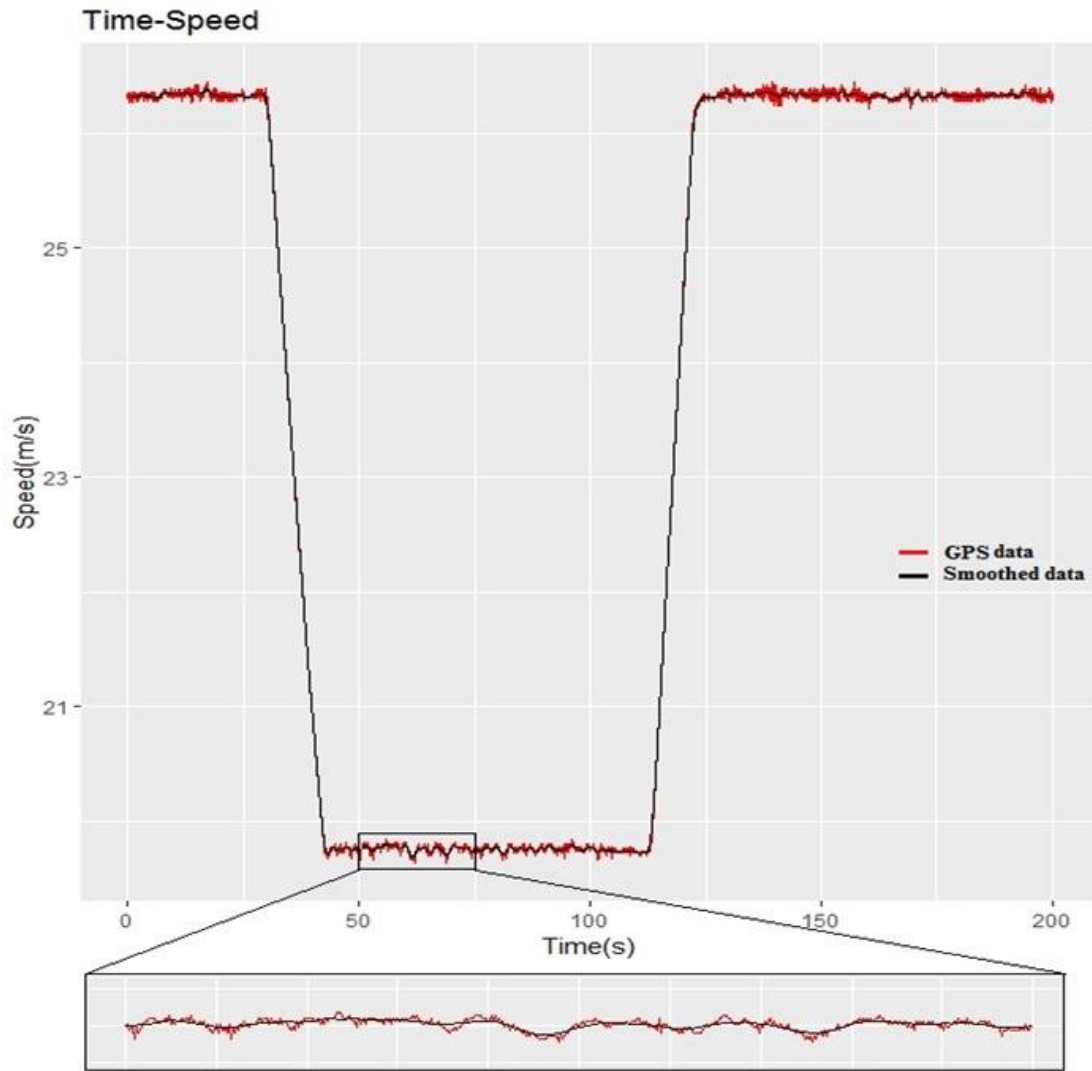


Figure 3-2 Time-Speed diagram of a CACC-mode Test for the lead vehicle

Results and Discussion

The descriptive statistics of the vehicle speeds for the tests are presented in Table 3-2. The descriptive statistics are reasonable and did not show many outliers or major errors and due to an error-check, the data seem valid and reliable. It is worth noting that the standard deviation of the following vehicle (FV) speed in CACC mode is slightly lower than ACC mode in scenario 1. This indicates a higher variation in vehicles' speed in ACC mode and more stability in the CACC mode. Figure 3-3-a and 3-3-b present the time-speed diagrams of sample tests in scenarios 1 and 2 respectively. Similarly, Figure 3-3-c and 3-3-d present the time-acceleration diagrams. It can be seen in the time-speed diagram of the vehicles in the platoon (Figure 3-3-a) that the convergence time of the FVs to adjust their speed to the lead vehicle (LV) in CACC mode is relatively lower than ACC mode. The acceleration rate of the vehicles in the CACC scenario is higher than the ACC scenario, this can be seen from the sharper slope of the Time-Speed diagram of the LVs in the CACC scenario between time 100 to 150 as shown in Figure 3-3-a. Thus, only the first 100 seconds of each test in scenario 1 is considered for further comparison and assessments. Furthermore, it is found in the time-speed graphs that after a change in the LV speeds, the FVs in the CACC mode adjust their speeds more rapidly with the LV compared with the ACC mode.

Safety Measures

Two measures including temporal driving volatility and TTC are selected to evaluate traffic safety. For scenario 1, results are presented for the first 100 seconds of each test where the vehicle's deceleration limits are very similar. The time between 100 to 200 seconds is excluded due to the different limits of acceleration in CACC (1 m/s^2) and ACC (0.25 m/s^2). However, this can be considered an advantage of the CACC algorithm to deal with a higher limit of acceleration.

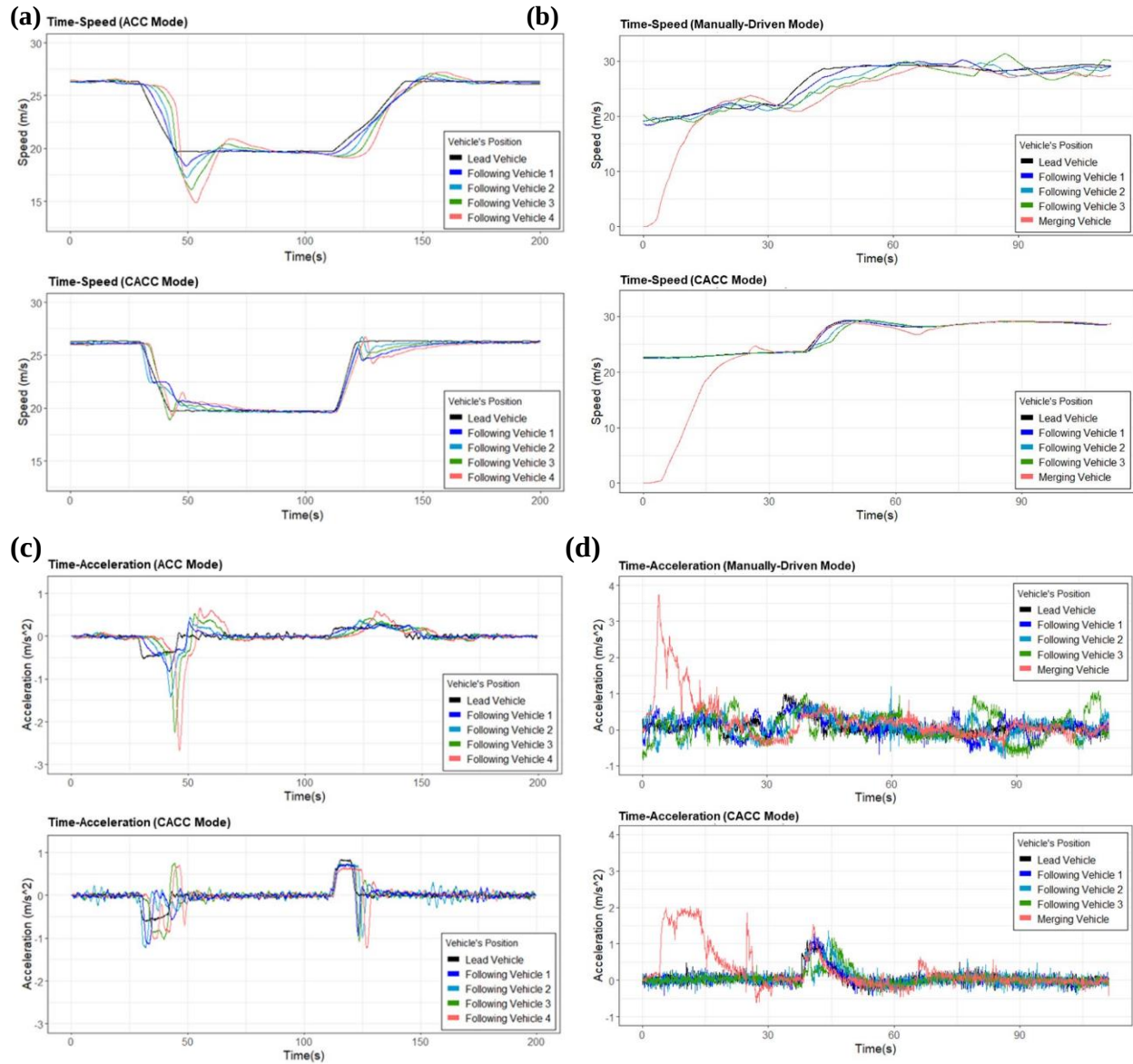
Driving Volatility Results

Using the method discussed earlier, the temporal driving volatility using the mean absolute deviation of vehicle speeds is calculated for both scenarios. The time series of temporal driving volatility are shown in Figure 3-4 for the scenarios. As illustrated in the figure, the CACC mode reduces the volatility noticeably in scenario 1. Note that the volatility after 100 seconds is higher in the CACC mode compared with ACC. This is due to the higher limit of the acceleration of CACC. Regarding the second scenario, the manually-driven (MD) mode shows more values in volatility for the FVs. This is due to smoother following in the CACC mode rather than (MD) mode which leads to less driving volatility. The average of all tests for aggregated values of the temporal driving volatility for each vehicle is presented in Table 3-3. The “% Difference” column is added to the table to show the relative percentage of change in the volatility in scenarios 1 and 2 compared to the ACC and MD modes respectively. The results indicate that on average, CACC compared with ACC reduces the FVs' volatility with a mean of 19.8%, ranging from 13.6% to 29%. It can be concluded that as the number of vehicles increases in the platoon, CACC reduces the volatility with a higher intensity. This can be caused by the connectivity between vehicles in the CACC mode as it allows for smaller reaction time and more coordination in speed compared with the ACC mode. Similarly, in scenario 2 in the CACC mode, the average reduction of volatility was found to be 60.1%, with values ranging from 54.9% to 70%. Likewise, the cooperative merging system reduces the volatility by 6.2 percent.

Table 3-2 Descriptive statistics of speed*

Vehicle No.	Mode	Duration (s)	No. of Observations	Mean	SD	Min	Max
Scenario 1 (10 Tests)							
Lead Vehicle	ACC	200	4000	23.271	3.031	19.588	26.440
	CACC	200	4000	23.652	3.112	19.620	26.432
Following Vehicle 1	ACC	200	4000	23.125	3.124	18.094	26.566
	CACC	200	4000	23.511	3.001	19.508	26.920
Following Vehicle 2	ACC	200	4000	23.076	3.252	17.008	26.871
	CACC	200	4000	23.527	2.946	19.240	26.488
Following Vehicle 3	ACC	200	4000	23.139	3.397	15.857	27.143
	CACC	200	4000	23.535	2.940	18.652	26.964
Following Vehicle 4	ACC	200	4000	23.131	3.572	14.596	27.340
	CACC	200	4000	23.570	2.938	18.128	27.096
Scenario 2 (2 Tests)							
Lead Vehicle	Manual	112	2245	26.288	3.684	19.140	29.396
	CACC	111	2225	26.624	2.706	22.463	29.283
Following Vehicle 1	Manual	112	2244	26.081	3.774	18.365	30.194
	CACC	111	2222	26.585	2.704	22.520	29.231
Following Vehicle 2	Manual	112	2243	25.469	3.699	18.765	31.352
	CACC	111	2225	26.529	2.699	22.640	29.366
Following Vehicle 3	Manual	112	2244	25.683	3.726	19.094	29.704
	CACC	111	2224	26.554	2.731	22.518	29.301
Merging Vehicle	Manual	112	2245	23.914	6.390	0.000	29.158
	CACC	111	2225	24.298	7.648	0.000	29.203

*Note that values of speed are the average of all tests in each scenario



Note: ACC = adaptive cruise control; CACC = cooperative adaptive cruise control

Figure 3-3 (a) Time-Speed diagrams of selected tests in scenario 1; (b) Time-Speed diagrams of selected tests in scenario 2; (c) Time-Acceleration diagrams of selected tests in scenario 1; (d) Time-Acceleration diagrams of selected tests in scenario 2.

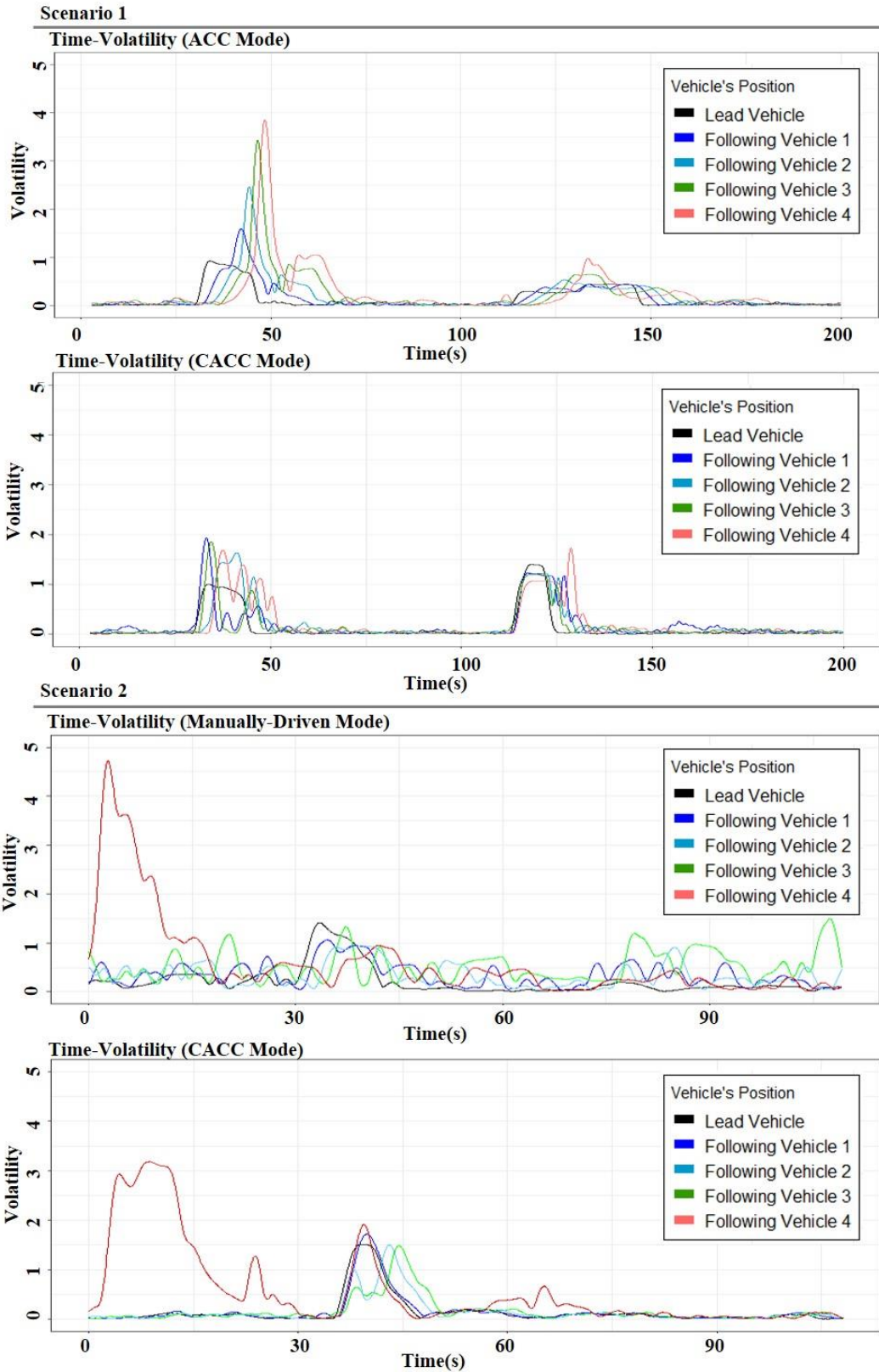


Figure 3-4 Time series of the temporal driving volatility

Table 3-3 A: Average aggregated driving volatility (m/s) - B: Fuel consumption (in ml)

A: Driving Volatility					
Vehicle	Mode	Volatility (m/s)	Mode	Volatility (m/s)	% Difference
Scenario 1					
Lead Vehicle	ACC	268.125	CACC	265.562	-0.96%
Following Vehicle 1	ACC	356.965	CACC	302.214	-15.34%
Following Vehicle 2	ACC	451.511	CACC	390.216	-13.58%
Following Vehicle 3	ACC	566.917	CACC	446.026	-21.32%
Following Vehicle 4	ACC	686.264	CACC	487.402	-28.98%
Scenario 2					
Lead Vehicle	Manually Driven	469.811	CACC	330.45	-29.66%
Following Vehicle 1	Manually Driven	766.928	CACC	341.542	-55.47%
Following Vehicle 2	Manually Driven	738.773	CACC	332.959	-54.93%
Following Vehicle 3	Manually Driven	1144.24	CACC	343.762	-69.96%
Merging Vehicle	Manually Driven	1321.03	CACC	1239.5	-6.17%
B: Fuel Consumption					
Vehicle	Mode	Fuel (ml)	Mode	Fuel (ml)	% Difference
Scenario 1					
Lead Vehicle	ACC	50.072	CACC	50.226	0.31%
Following Vehicle 1	ACC	49.389	CACC	49.144	-0.50%
Following Vehicle 2	ACC	46.945	CACC	44.986	-4.17%
Following Vehicle 3	ACC	46.037	CACC	42.968	-6.66%
Following Vehicle 4	ACC	46.761	CACC	45.034	-3.69%
Scenario 2					
Lead Vehicle	Manually Driven	141.721	CACC	123.062	-13.17%
Following Vehicle 1	Manually Driven	154.676	CACC	121.728	-21.30%
Following Vehicle 2	Manually Driven	151.124	CACC	123.423	-18.33%
Following Vehicle 3	Manually Driven	152.21	CACC	121.638	-20.09%
Merging Vehicle	Manually Driven	151.319	CACC	152.135	0.54%

Note: ACC stands for adaptive cruise control, CACC stands for cooperative adaptive cruise control, and Manually Driven refers to the conventional vehicles.

Time-to-Collision Results

Using Equation 3, TTC was calculated for all tests. The 100 lowest values of TTC are selected for each vehicle as potentially unsafe car following conditions. Figure 3-5 illustrates the descriptive statistics of the average value TTC for each vehicle's position in the fleet. The results indicate that vehicles equipped with ACC have lower minimum and mean values of TTC compared with CACC. On average, there is a 100% increase in the minimum values of TTC in CACC mode. For instance, for the last vehicle in the platoon (FV4), the minimum value of TTC in ACC mode is 4.2 seconds, while the minimum value of TTC in the CACC mode is 8.3 seconds. This reveals that the risk of a rear-end crash is lower in a vehicle platoon using CACC mode. It is worth noting that the V2V communication and speed harmonization between vehicles in CACC mode reduce the difference between the speed of the two following vehicles and as a result increase the TTC values.

Regarding the second scenario, while no critical situation is defined in the lead vehicle's trajectory to provide a sudden speed reduction or increase, the TTC values are very high while vehicles following each other at similar speeds. Thereby, the values of TTC is not worth mentioning for the second scenario.

Energy and Environmental Measures

Fuel Consumption

Using Equation 4, fuel consumption is calculated for the first 100 seconds of two selective tests from ACC and CACC modes. The fuel consumption formula is sensitive to the distance that the vehicle traveled with constant speed or during acceleration. Therefore, among the tests, an ACC and CACC test are selected that travel roughly the same distance. In other words, they are selected based on the minimum distance between the lines in the time-speed diagram while decelerating as shown in Figure 3-6. The results are presented in Table 3-3.

As presented in Table 3-3, the CACC system in scenario 1 on average, reduces the overall fuel consumption of the FVs with 3.8% ranging from 0.5% to 6.7% compared with ACC. Likewise, in scenario 2, the results reveal that CACC reduces fuel consumption compared with manually driven vehicles. The fuel consumption reduction ranges between 18.3% and 21.3% with a mean value of 19.9%. Regarding the merging vehicle, due to the ability of the CACC algorithm to deal with a higher acceleration limit, the amount of fuel consumption slightly increases by 0.54%. But it can be improved if the CACC control algorithm for accelerations is adjusted in a way that allows for energy benefits to be achieved during merging.

Emissions

The average emissions of each vehicle in each scenario are presented in Table 3-4. Regarding scenario 1, the result reveals that the CACC mode reduces the total emissions of the FVs slightly. The total emissions reduction ranges from 3.1% to 4.9% and on average 3.7%. Regarding scenario 2, the FVs in the CACC mode produce fewer emissions ranging from 2.33% to 4.8% with a mean of 3.3%. However, the merging vehicle in the CACC mode emits slightly more gases (4.1%) due to the sharper acceleration compared with the manually-driven mode. Nonetheless, the difference in emissions is not very substantial.

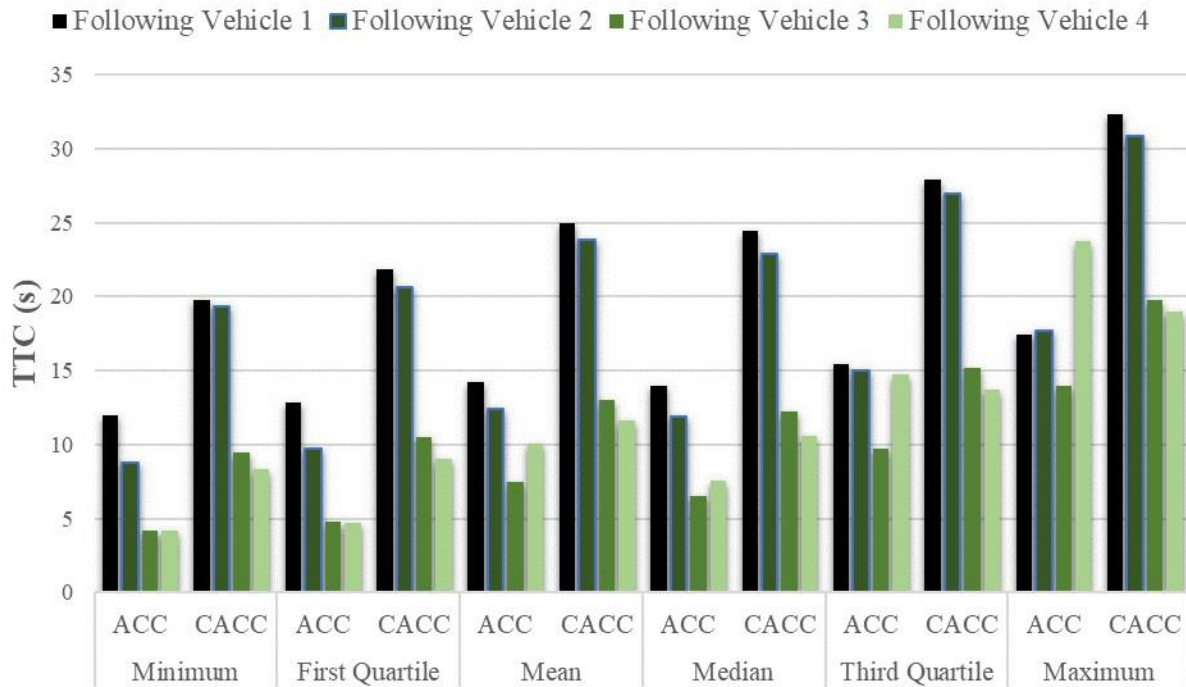


Figure 3-5 Descriptive statistics of TTC (sec) for all tests in scenario 1

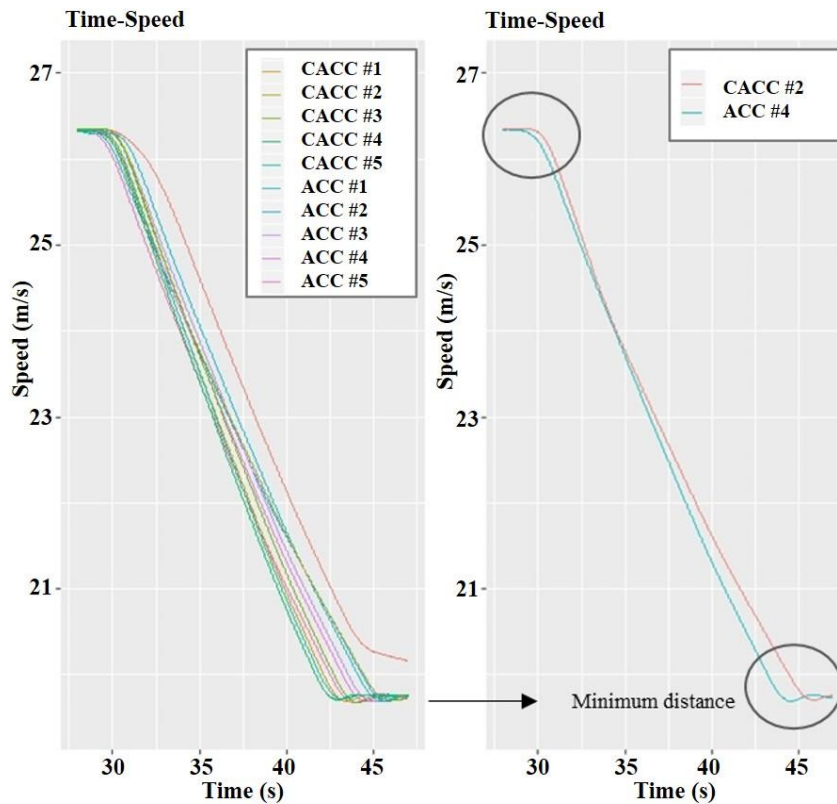


Figure 3-6 Time-Speed diagram of the tests during deceleration

Table 3-4 Vehicle's total emissions (in grams) based on VSP model

	CO ₂	CO	NO _x	HC	Emission	CO ₂	CO	NO _x	HC	Emission	%Diff erence
Scenario 1 ACC						CACC					
LV	290.3	1.221	0.177	0.057	291.8	289.3	1.220	0.177	0.057	290.8	-0.35%
FV1	298.1	1.283	0.186	0.059	299.6	287.6	1.199	0.177	0.056	289.1	-3.53%
FV2	305.4	1.333	0.194	0.060	307	295.9	1.398	0.191	0.060	297.6	-3.06%
FV3	309.7	1.378	0.201	0.061	311.4	294.4	1.370	0.188	0.059	295.9	-4.94%
FV4	315.3	1.462	0.211	0.063	317.1	304.5	1.511	0.204	0.062	306.2	-3.41%
Scenario 2 Manually-driven Mode						CACC					
LV	458.3	3.457	0.389	0.104	462.3	441	3.732	0.367	0.102	445.2	-3.69%
FV1	453.9	3.412	0.394	0.103	457.8	431.2	4.014	0.364	0.100	435.7	-4.83%
FV2	447.6	3.296	0.383	0.101	451.4	436.7	3.657	0.360	0.102	440.9	-2.33%
FV3	448.9	6.659	0.445	0.135	456.2	439.2	3.515	0.356	0.100	443.2	-2.85%
MV	454.1	3.636	0.401	0.106	458.2	470.4	5.878	0.442	0.130	476.9	4.08%

Note: LV denoting the lead vehicle; FV denoting the following vehicle; MV denoting the merging vehicle; CO₂ denoting carbon dioxide; CO denoting carbon monoxide; NO_x denoting oxides of nitrogen; HC denoting hydrocarbons; Emission is the summation of CO₂, CO, NO_x, and HC.

Limitations

One of the key limitations of this study is that the available data includes LV deceleration rates that are slightly different (0.05 m/s^2). It is worth noting that during the ACC tests in the dataset, the speed measurement of one of the vehicles in the fleet exhibited a few seconds of lag. As a result, another sensor (“velocity_fwd_PINPOINT”) was used in place of “speed_CACC” for this vehicle. Given that a different type of sensor was used to capture the speed, the vehicle’s speed exhibits much more noise compared to other vehicles’ speed records. After smoothing the noise, the vehicle’s speed data were not precisely consistent with other variables (e.g. vehicle location) in some tests. To solve this issue and maintain the accuracy of the findings, the values for this vehicle in some tests are not considered in the calculations. Another limitation of this study is that results are derived from simplified scenarios. It is essential to consider more complicated scenarios with higher fluctuations of vehicular speed and interactions with other traffic in the mixed environment to conduct more generalizable conclusions. Also, this study used a limited number of iterations of the tests and there might be some randomness and error in the data.

Conclusions

CAV technologies have the potential to substantially improve transportation system energy efficiency and safety. In particular, ACC and CACC systems may lead to substantial improvements in sustainable mobility. Although many studies evaluate the impacts of vehicle automation on traffic performance, largely through several simulations and modeling techniques, the amount of improvement based on real-world data and field experiments are still scarce. Thus, there is a need for more empirical evidence regarding the impacts of automation for wide-scale deployment.

In this study, the impacts of ACC and CACC on traffic safety and the environment are explored through analysis of field test data based on real-world conditions collected by ACC, CACC, and manually driven vehicles. A temporal driving volatility measure and TTC are calculated to evaluate the safety of vehicles in different modes of platooning. Likewise, fuel consumption and emissions are calculated to quantify the impacts on energy efficiency and the environment.

The results show that the moderate levels of automation through vehicles equipped with CACC substantially reduce driving volatility as a safety measure in a five-vehicle-platoon from 13.6% to 29% compared with the ACC equipped vehicles. Similarly, in scenario 2, the cooperative merging system reduces the volatility of the merging vehicle by 6.2% compared to the merging vehicle in manually-driven. As one of the main features of the CACC system and V2V communication, it is expected that the automatic synchronization of longitudinal movements of a string of vehicles will reduce the driving volatility. Also, the CACC system reduces the risk of rear-end collision by increasing the minimum values of TTC near 100% for the vehicles in the fleet.

One of the reasons that the CACC system can improve vehicle- and system-level fuel efficiency and reduce emissions is to control the magnitudes of the vehicle's acceleration and the variation of acceleration over time. The results show that the CACC technology on average reduces the overall fuel consumption and emissions in a five-car-fleet with 3.7% ranging from 0.5% to 6.7% compared with ACC. This result is consistent with previous studies (102; 112). Also, the

results of scenario 2 reveal that the amount of fuel consumption for the merging vehicle slightly increases by 0.54% compared to manually-driven mode. Regarding the environmental impact, the reduction in total emissions ranges from 3.1% to 4.9% and on average 3.7% in the first scenario. However, in the second scenario, the merging vehicle in the CACC mode emits slightly more gas (4.1%). In addition, fuel consumption and emissions can increase due to the ability of the CACC algorithm to deal with higher acceleration limits to form a stable platoon. This increase can be improved through an adjustment to the CACC control algorithm for accelerations that allows for the energy benefits during acceleration. In general, the results of fuel efficiency discussed in this study are consistent with the outcomes of many simulation-based studies (8; 136; 137). However, the results regarding the driving volatility are innovative in this study. Also, the findings are largely consistent with expectations.

The contribution of this study is reflected in the analysis of vehicle automation, connectivity, and cooperation. The analysis is based on a large-scale federally collected (big) database obtained through field testing. The observed benefits of ACC and CACC from real-life testing are quantified by using key measures of performance. Insights are gained into safety and energy/emissions benefits in vehicle platoons and during merging at on-ramps. This study quantifies the expected benefits of ACC and CACC to estimate the realistic impacts of automation and connectivity. Moreover, the method used in this study is repeatable; and the results reduce the uncertainty of vehicle technologies in terms of their impacts. To further improve the confidence of these results, it will be necessary to collect even larger samples of data. Additionally, the results are obtained from analyzing a limited set of scenarios and field tests in specific circumstances. Therefore, the results are limited to the discussed scenarios and might not be generalizable to other conditions. To have more generalizable results, more data can be collected expanding the scenarios examined in different mixed traffic environments and conditions, e.g., adverse weather, different roadway classifications, and edge cases.

CHAPTER 4: INTEGRATION OF AUTOMATED VEHICLES IN MIXED TRAFFIC: EVALUATING CHANGES IN PERFORMANCE OF FOLLOWING HUMAN-DRIVEN VEHICLES

This chapter is presented at the 99th Transportation Research Board Conference and is published in the Journal of Accident Analysis and Prevention.

Mahdinia, I., Mohammadnazar, A., Arvin, R., & Khattak, A. J. (2021). Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles. *Accident Analysis & Prevention*, 152, 106006.

Abstract

The introduction of Automated Vehicles (AVs) into the transportation network is expected to improve system performance, but the impacts of AVs in mixed traffic streams have not been clearly studied. As AV's market penetration increases, the interactions between conventional vehicles and AVs are inevitable but by no means clear. This study aims to create new knowledge by quantifying the behavioral changes caused when conventional human-driven vehicles follow AVs and investigating the impact of these changes (if any) on safety and the environment. This study analyzes data obtained from a field experiment by Texas A&M University to evaluate the effects of AVs on the behavior of a following human-driver. The dataset is comprised of nine drivers that attempted to follow 5 speed-profiles, with two scenarios per profile. In scenario one, a human-driven vehicle follows an AV that implements a human driver speed profile (base). In scenario two, the human-driven vehicle follows an AV that executes an AV speed profile. In order to evaluate safety, these scenarios are compared using time-to-collision (TTC) and several other driving volatility measures. Likewise, fuel consumption and emissions are used to investigate environmental impacts. Overall, the results show that AVs in mixed traffic streams can induce behavioral changes in conventional vehicle drivers, with some beneficial effects on safety and the environment. On average, a driver that follows an AV exhibits lower driving volatility in terms of speed and acceleration, which represents more stable traffic flow behavior and lower crash risk. The analysis showed a remarkable improvement in TTC as a result of the notably better speed adjustments of the following vehicle (i.e., lower differences in speeds between the lead and following vehicles) in the second scenario. Furthermore, human-driven vehicles were found to consume less fuel and produce fewer emissions on average when following an AV.

Keywords: Automated Vehicle, Safety, Volatility, Time-to-collision, Driving Behavior

Introduction

Increasing participation in social activities increases transportation demand and causes several adverse impacts such as traffic crashes, injuries, fatalities, and air pollution (87; 138). According to the United States National Highway Transportation Safety Administration, 36,560 people were killed in traffic crashes in 2017 (139). Today, Connected and Automated Vehicles (CAVs) are being introduced in the market and are believed to have the potential to mitigate some of the negative impacts of transportation (57). Safe deployments and real-world testing of CAVs are crucial and predominant in U.S. DOT's line of action. Real-world AV testing helps developers recognize benefits and drawbacks, and fix system limitations not just of individual vehicles, but across fleets (90; 91). Because CAVs decrease driver inputs and can maintain periodic control of the vehicle at some levels of automation (140), they can lead to substantial improvements in safety (141) by reducing driver workloads and human error (68; 70), and provide environmental benefits (142; 143) by reducing emissions (144-146) and fuel consumption (8; 147).

Many estimated benefits of CAVs will most likely not be obtained until after full market penetration, but 100% market penetration will not be reached in the immediate future (148). During the transition period, traffic flow will include a mix of CAVs and conventional vehicles. Therefore, interactions between conventional vehicles and CAVs will be inevitable. These interactions will create complicated environments that impact driving behavior patterns in traffic streams and fleets (149). Consequently, it is imperative that we investigate CAVs' influence on the behavior and performance of conventional human-driven vehicles. As real-world CAV data lack detail for in-depth analysis, there are uncertainties about these interactions. Therefore, the fundamental questions to be answered are how will CAVs affect the behavior and performance of conventional vehicles, and will the interactions improve safety? Furthermore, what is the extent of these potential improvements?

To address these questions and develop more knowledge regarding CAV impacts, this study aims to explore the impacts of Automated Vehicles (AVs) on driver behavior in conventional human-driven vehicles and estimate the effects on safety and the environment. The advent of CAVs and location-based services gives useful information about vehicles' kinematics in real-world driving conditions ((150-153). To this end, this study exploits a real-life field-test dataset and adopts a realistic assessment of the impacts of AVs on non-AVs (conventional vehicles) at a microscopic level through quantification of safety and environmental factors. In this study, various measures are used for safety evaluations including temporal driving volatility and Time-to-Collision (TTC) measures. Likewise, fuel consumption and emissions are used to investigate environmental impacts.

The findings of this study can aid predictions of AV impacts on mixed traffic users and analysis of safety and efficiency in mixed traffic environments. Furthermore, the findings of this study can help practitioners and engineers adjust and enhance AV performance based on interactions with non-AVs. The rest of this paper is organized as follows. The next section presents the literature review, followed by the study's framework and methodology. Subsequently, the description of the data is presented. The paper concludes with the results and discussion sections followed by the conclusions.

Literature Review

Driving behavior refers to a broad concept that is defined by a considerable number of variables and measures. From a safety perspective, driving behavior is the leading cause of crashes (116),

majorly contributing to traffic safety (154). Several studies have explored the relationship between driver behavior and road safety by classifying drivers based on their tendency to have risky driving profiles (155-161). Driving behavior can also be evaluated from an environmental view. Faria et al., for instance, assessed the effects of aggressive driving styles on fuel consumption and vehicle emissions (162). After studying different driving styles, Rios-Torres et al. concluded that fuel consumption increases remarkably with volatile driving in conventional vehicles compared to calm driving in conventional vehicles (163).

Various measures were used to evaluate driving behavior in terms of safety. In (117), for example, driving volatility characterized and quantified the volatile behavior of drivers as a measure of safety. In recent literature, driving volatility, taken from the economic field, is a safety surrogate measure that aims to assess crash risk (100; 117; 164; 165). Volatility can be represented by vehicular speed, acceleration, and jerk (164; 166). It was revealed that an increase in speed and acceleration volatility significantly increases crash frequency (117). The associations of speed and acceleration volatility with rear-end, sideswipe, angle, and head-on crashes were found to be positive and significant (117; 164). Moreover, time-to-collision (TTC) was used as a surrogate measure to quantify safety, especially rear-end collision risk (26). This measure uses trajectories of consecutive vehicles to quantify risk. It is broadly used in the literature to quantify longitudinal crash risk in car platoons (49; 101).

While numerous studies have evaluated driver behavior in conventional vehicles, the behavior of drivers in low levels of automation CAVs is expected to be affected by the increasing market penetration of CAV technologies. Some of the studied effects on driving behavior include alertness (71; 72), reaction time (76), various human factors (167), and lane changing behavior in connected environments (168-170). Recently, particular attention has been devoted to the impact of CAVs equipped with higher levels of automation. Studies show that automated technologies can increase safe vehicle control (67) and reduce driver workloads, and therefore reduce human error (68; 70). However, it has also been found that AVs can cause passive fatigue and decrease driver alertness, which in turn slows manual takeover (71-75).

While various studies have investigated the impacts of CAVs on traffic networks, a change in driving behavior in conventional vehicles when interacting with CAVs is expected. As interactions between conventional vehicles and CAVs are inevitable (149; 171), it has been found that human driver behavior changes when following an AV (172; 173). Studies also show that an AV in a platoon of human-driven vehicles can dampen stop-and-go waves (174) and stabilize flow (175). Simulations show that with an increase in the CAV penetration rate, the uncertainty inherent in human driver behavior is reduced, traffic flow stability improves (173), and the portion of smooth driving expands (176).

From a safety perspective, increased CAV penetration can considerably improve traffic safety (176). Papadoulis et al. investigated the impact of CAVs on traffic conflicts in motorways at different market penetration rates using a traffic microsimulation framework. They utilized two surrogate safety measures, Time-to-Collision (TTC) and Post Encroachment Time (PET), to evaluate traffic conflicts. The study found that CAVs can improve safety by mitigating traffic conflicts even at low market penetration rates. Also, they realized that human-driven vehicles appear to modify their driving behavior due to less speed alterations in vehicle platoons involving CAVs (9). No limitation in the platoon size was considered in their study, but there would possibly be some limitation in the size of fleets in the real world. Likewise, Viridi et al. conducted a

microsimulation model to evaluate CAVs' impact on traffic conflicts at a network including intersections or roundabouts with different levels of control. They found that as CAV penetration in mixed traffic increases, traffic conflicts are mitigated. Moreover, they realized that the average headway increases with increasing CAV penetration (177). However, it is expected that at higher traffic flow and larger CAV penetration rates, the average headway will be reduced (178). Ye and Yamamoto investigated the impact of CAVs on traffic safety at different penetration rates using a simulation approach. Their findings show an improvement in safety in mixed traffic at larger CAV penetration rates by growing the small time-to-collision values and lessening the speed difference between vehicles. Their findings indicate that an increase in CAV penetration can ease stop-and-go traffic, smooth traffic flow, and consequently, improve traffic safety (176). Sinha et al. studied the impact of CAVs on crash severity and frequency using a microsimulation modeling exercise. They utilized two surrogate safety measures (TTC and PET) to inspect traffic safety. Their results show that with each incremental increase in CAV penetration, the severity of crashes between conventional vehicles and CAVs is reduced due to the small differences in vehicle speeds. However, the crash rate is not invariably diminished due to the lower magnitude of TTC and PET values at higher penetration rates (149). The reduction in TTC with an increase in CAV penetration rate contradicts other studies (9; 176). The difference in their findings could be due to the different CAV operational characteristics or different simulation assumptions that potentially have a direct impact on the evaluation of safety effects on mixed traffic flow (176). The limitations of the studies based on simulations are in their emulation of both CAV's and human drivers' behavior. However, in the real world, the diversity of human driving behavior in a mixed traffic environment might result in expecting different impacts of CAVs.

There seem to be several gaps in the literature. First, a noticeable limitation is that a large majority of studies have not investigated the impact of AVs on driver behavior in conventional vehicles. Second, the studies that explore driving behavior changes in conventional vehicles focus mostly on impacts on traffic flow dynamics rather than capturing safety and environmental impacts. Third, the findings of previous studies are mostly based on simulations and lack experimental tests. This paper addresses these gaps by investigating the safety and environmental impacts of driving behavior changes in human-driven vehicles when interacting with AVs. To this end, this study uses Texas A&M University data obtained from a field experiment to evaluate the effects of AVs on the behavior of a following human-driven vehicle. While the effects are expected, this study quantifies them and points to a more realistic assessment of expected benefits.

Methodology

This paper aims to create new knowledge about the AVs' impact on driving behavior of the following vehicle in a fleet and evaluate the resulting impacts at a microscopic level in terms of safety and environmental factors. Given the objective of this study, several measures are selected and will be described in this section. To that end, driving volatility and time-to-collision are selected to evaluate traffic safety, fuel consumption and emissions are used as environmental indicators. The framework of this study is illustrated in Figure 4-1.

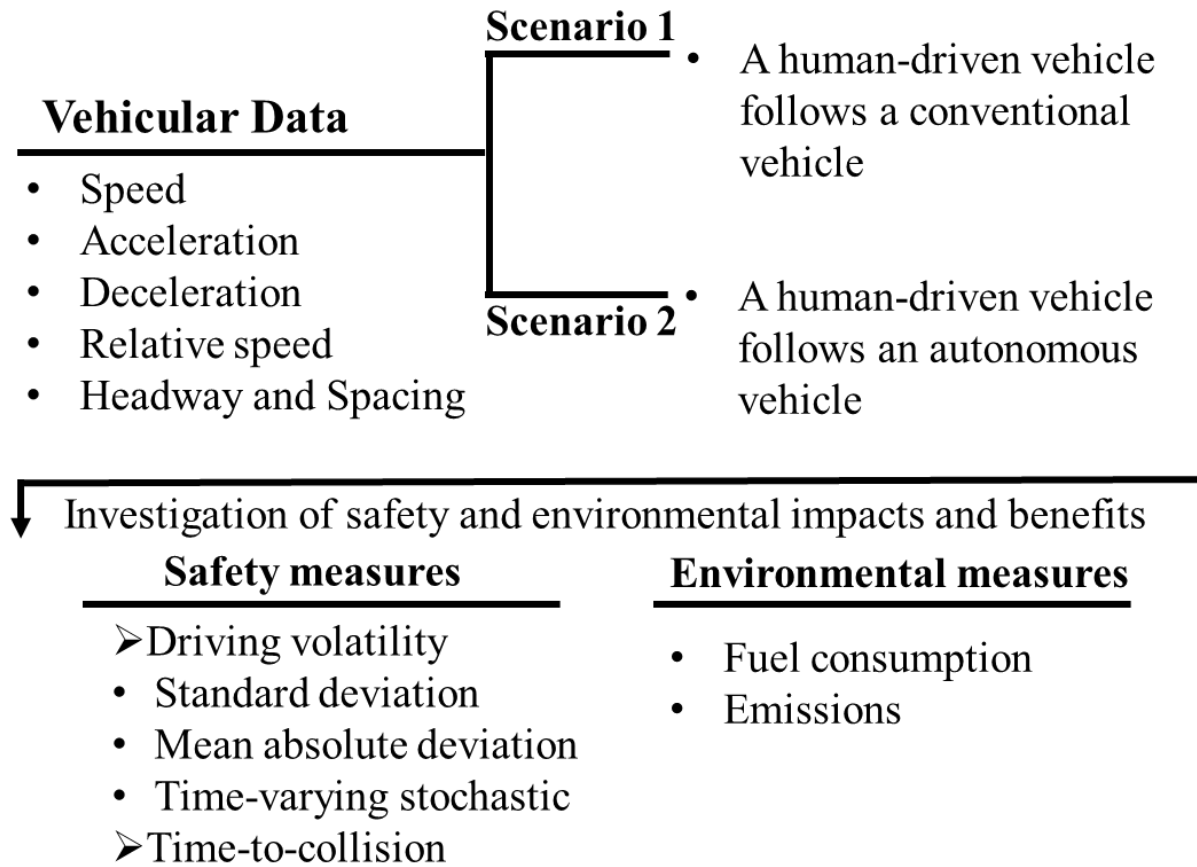


Figure 4-1 Framework of study

Safety Indicator

In the literature, several surrogate safety measures are introduced to quantify safety performance both at the macro or micro level. In terms of micro-level analysis, driving volatility is used to quantify traffic safety (126; 179) and time-to-collision, TTC, is used to quantify longitudinal crash risk (49; 101). In this study, we utilized the concept of temporal driving volatility and TTC to quantify the safety performance of the vehicles under different scenarios. The safety measures are described in this section.

Driving Volatility Measures

Driving volatility is a surrogate measure of safety that quantifies the deviation of driving behavior from the norm. In other words, driving volatility captures alterations in the longitudinal control of a vehicle. The volatility can be represented by vehicular speed, acceleration, and jerk (164; 166). To capture volatility, several measures have been introduced (117; 164). Previous studies have shown that volatility indices (measured before a traffic collision) are highly correlated with not only crash probability (Kamrani et al. 2019), but also with crash severity (Arvin et al. 2019 a, c). These can be used as surrogate measures of safety, along with time to collision. Also, analysis of the segments and intersections revealed that higher volatility is associated with higher crash frequency (100; 180). Therefore, higher driving volatility in terms of the following measures represents a lower safety state. In this study, three measures are used to capture driving volatility which are discussed as follows:

1. Standard deviation (S_{dev}):

This measure uses the standard deviation to capture the data variation using Equation 1 . It can be applied to vehicular speed and acceleration (164). An increase in this measure represents higher speed/acceleration changes, indicating higher driving volatility and a lower safety state.

$$S_{dev} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

Where S_{dev} is the standard deviation, x_i is the observed value I , $\bar{x}$ is the mean, and n is the number of observations.

2. Mean Absolute Deviation (MAD):

This measure captures the mean distance between the observation of a variable and the central tendency of the data as shown in Equation 2. It can be applied to the vehicular speed and acceleration/deceleration. Note that it captures acceleration and deceleration simultaneously (164). Mean absolute deviation is similar to standard deviation, but simpler. Larger MAD represents higher driving volatility and indicates a lower safety state.

$$D_{mean} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (2)$$

Where D_{mean} is the mean absolute deviation, n denotes the number of observations and $\bar{x}$ is the mean.

3. Time-Varying Stochastic Volatility (VolFi):

This measure is calculated using Equations 3 and 4 (181; 182). It requires positive time-series observations (164). Thereby, it only applies to vehicular speeds of positive values above zero. Note

that in this study, vehicle speeds were always above zero. Higher VolFi shows higher speed fluctuations that indicate higher driving volatility and a lower safety state.

$$V_f = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (r_i - \bar{r})^2} \quad (3)$$

$$r_i = \ln \left(\frac{x_i}{x_{i-1}} \right) * 100 \quad (4)$$

Where x_{i-1} is the previous observation regarding the observation x_i ; $\ln$ is the natural logarithm; $\bar{r}$ denotes the mean of parameter "r".

Calculation of Volatility Measures

In this study, the driving volatility measures are calculated at two levels including trip-based volatility and temporal driving volatility. In trip-based volatility, the driving volatility measures are applied to the data of each vehicle during a trip. The outcome of this approach is an aggregated value of a volatility measure on a trip that represents the driving volatility of a vehicle during a trip.

Temporal driving volatility aims to capture alterations in instantaneous driving behavior and generate time-series data. The advantage of this concept is that it captures the time dependency between observations, which can identify the durations that a driver is showing volatile behavior (179). To calculate temporal driving volatility, similar to the concept of moving average, a 3-second time-window is considered and the measure is assigned to the subject time. It was determined that volatility measures calculated using a 3-second time-frame window provide the highest correlation of volatility measures with crash risk compared to 1, 2, and 5 second time windows (183). Figure 4-2 illustrates the calculation of temporal volatility utilizing the moving window. Note that the aggregation and average of temporal driving volatility belonging to a period is the total driving volatility of that period.

The algorithm to calculate temporal driving volatility is provided as follows (183):

For each driver, $i = (1, N)$

For each trip taken by driver i , $j = (1, M_i)$

For each second of trip j taken by driver i $t_{k_{i,j}} \in [30, T_{i,j}]$

Step 1: Subset three seconds (30 deci-seconds) of data $[t_{k_{i,j}} - 30, t_{k_{i,j}}]$

Step 2: Record kinematic information of vehicle

Step 3: Apply volatility functions

Step 4: Assign calculated volatility measures and extracted kinematic information to time $t_{k_{i,j}}$; $Volatility_{v,t_{k_{i,j}}}$

where i is index of drivers; j is index of trips; M_i is number of trips taken by driver I ; $T_{i,j}$ is total travel time of trip j taken by driver I ; $t_{k_{i,j}}$ is time k of trip j taken by driver I ; $(x_{ki,j}, y_{ki,j})$; $Volatility_{v,t_{k_{i,j}}}$ represents volatility measure v at time k of trip j taken by driver i

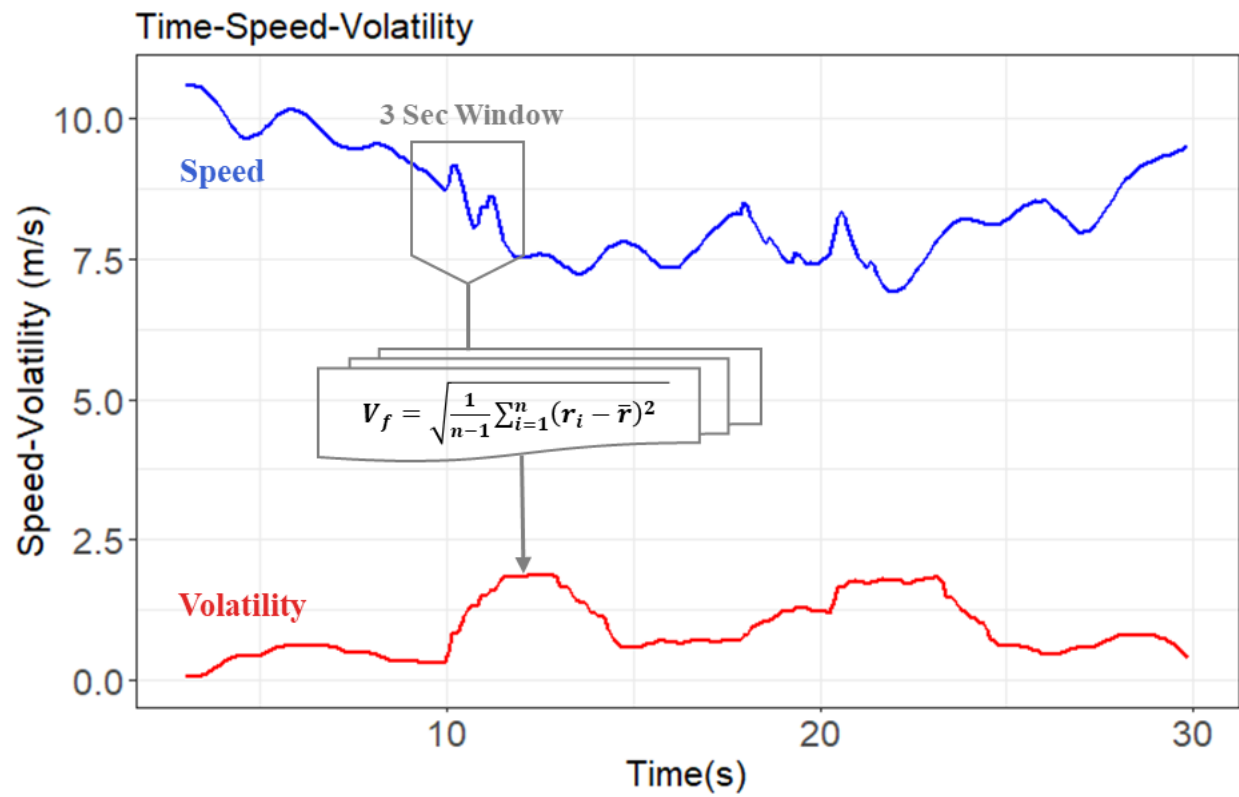


Figure 4-2 Temporal driving volatility calculation

Time-to-Collision

Time-to-collision (TTC) is a surrogate safety measure (26) which was introduced by (184) as a rear-end crash risk index (26; 49). This measure is applied to set a relationship between microscopic traffic data and longitudinal safety of vehicles (49). If two successive vehicles continue at their present speed while occupying the same lane, TTC represents the time required for the vehicles to collide if the following vehicle moves faster than the leading vehicle. TTC can be defined as follows:

$$TTC(t) = \frac{X_L(t) - X_F(t) - L}{V_F(t) - V_L(t)}, \quad \forall V_F(t) > V_L(t) \quad (5)$$

where $X_F(t)$ and $X_L(t)$ are the positions of the following and leading vehicles respectively with respect to time; $V_F(t)$ and $V_L(t)$ are the speed of following and leading vehicle with respect to time and L is the length of the leading vehicles.

Environmental Measure

Fuel Consumption

The measure of fuel consumption in a car equipped with an internal combustion engine is a function of different factors including speed, acceleration, fuel type, atmospheric pressure and temperature, and gear and engine efficiency. Therefore, a complicated model is required to estimate vehicle's fuel consumption accurately. However, a simplified model can approximate the engine functioning and provide a fair estimation of actual fuel consumption (129). In this study, regarding the similarity of the vehicles and the environmental condition, it is assumed that the only parameters contributing to the fuel consumption are the vehicle's speed and acceleration. This assumption is reasonable because this study aims to relatively compare the fuel consumption of similar vehicles in different scenarios. To this end, the model proposed by Kamal et al. is used to calculate the fuel consumption as shown in Eq. (6) (130; 131). The model takes advantage of the relation between speed, acceleration, and fuel consumption.

$$F_v = (c_0 + c_1 v_t + c_2 v_t^2 + c_3 v_t^3 + a_t(d_0 + d_1 v_t + d_2 v_t^2)) \quad (6)$$

Where F_v is the estimated fuel consumption (milliliter per second), v_t is the vehicle speed at time t (m/s), a_t is the acceleration of the vehicle at time t (m/s²), $c_0 = 0.1569$, $c_1 = 2.450 \times 10^{-2}$, $c_2 = -7.415 \times 10^{-4}$, $c_3 = 5.975 \times 10^{-5}$, $d_0 = 0.07224$, $d_1 = 9.681 \times 10^{-2}$, and $d_2 = 1.075 \times 10^{-3}$. It is assumed that the vehicle consumes no fuel during deceleration where $a_t < 0$. The consumption parameters used in Eq. 6 were estimated through a curve-fitting process developed by (131). Moreover, the following parameters are the default values for a representative conventional passenger vehicle in this study. The equivalent mass for the vehicles is $M_v = 1200$ kg, the drag coefficient is $C_D = 0.32$, the air density is $\rho_a = 1.184$ kg/m³, the frontal area is $A_f = 2.5$ m², and the rolling resistance coefficient is $\mu = 0.015$ (130; 131).

Emissions

The emissions measure is calculated based on the vehicle-specific power (VSP) microscopic model. The VSP model is regression-based (185) and accounts for vehicular second-by-second speed, acceleration, and terrain gradient (132). The VSP is calculated using Equation 7. Given VSP, total emissions for each vehicle can be estimated using Equation 8.

$$VSP = v[1.1a + 9.81 \sin(\arctan(\text{Grade})) + 0.132] + 0.000302v^3 \quad (7)$$

Where VSP is the vehicle specific power [kW/metric ton]; v is vehicle speed [m/s]; a is acceleration (+) or deceleration (-) [m/s^2]; and grade is terrain gradient [$\pm\%$].

$$\text{Total Emissions} = \text{CO}_2^{\text{VSP Mode}} * t + \text{CO}^{\text{VSP Mode}} * t + \text{NO}_x^{\text{VSP Mode}} * t + \text{HC}^{\text{VSP Mode}} * t \quad (8)$$

Where VSP Mode can be derived from Table 4-1; t is the duration vehicle drives in the VSP Mode; CO_2 is Carbon Dioxide (g/s); CO is Carbon Monoxide (g/s); NO_x is Nitrogen Oxide (g/s); HC is Hydrocarbon (g/s).

Data

The data used in this study consist of two scenarios with a fleet including a lead vehicle and a follower (Figure 4-3). In addition, a control vehicle is considered for making similar experimental setups and preserving the constancy of other variables in the scenarios. In other words, the control vehicle is used to control the trajectories of the other two vehicles in the fleet and to make sure they are following the exact set speed profiles in both scenarios. The AV used in the study is Texas A&M University's Automated Chevrolet Bolt, which is programable and able to run any speed time-series. Accordingly, during the studied scenarios, the speed profiles of the lead vehicle are implemented by the AV. In scenario 1, a conventional human-driven vehicle follows an AV which is programmed to perform the speed profiles of a human driver. In the second scenario, the same conventional vehicle follows the AV that performs the speed profiles of an AV (Rahmati et al. 2019). The possibility of human errors from executing the predetermined speed profiles was eliminated by utilizing the AV to emulate human driver behavior in scenario 1 and AV behavior in scenario 2.

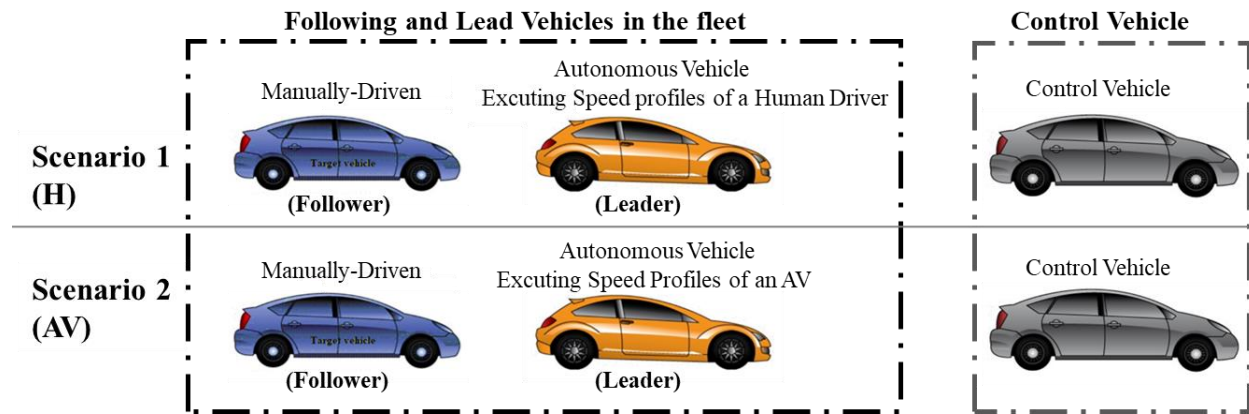
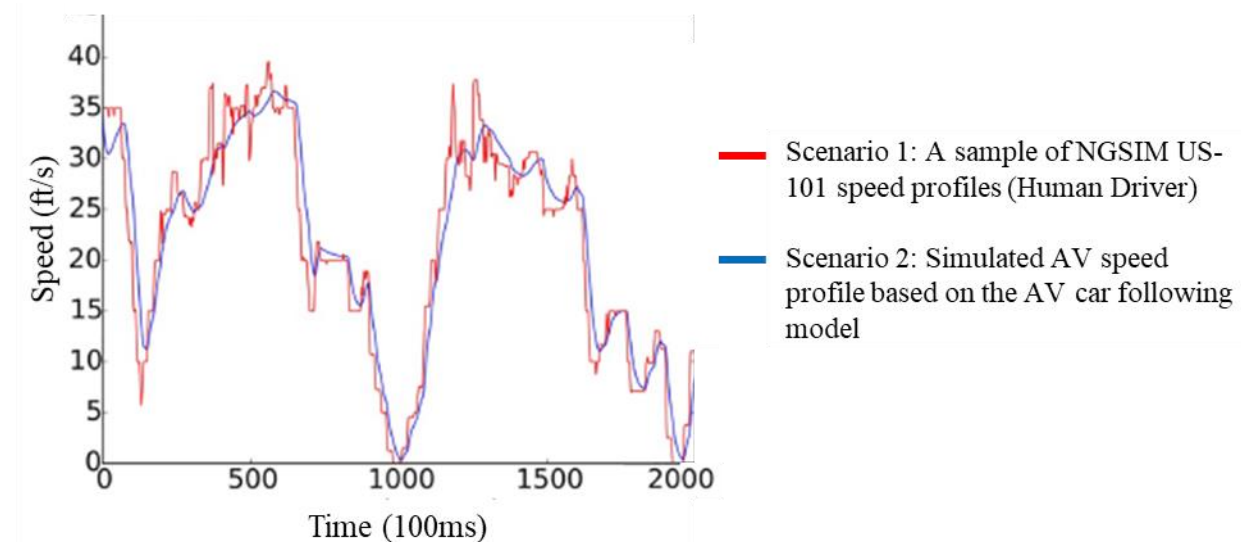
Data was collected from nine drivers in a testbed (AV testing track on the RELIS Campus, Texas A&M University). The drivers were asked to follow the lead vehicle individually during the experiment. It is worth mentioning that to avoid potential biases in the behavior of the drivers, they were not given any information about the lead vehicle and the test scenarios before and during the experiment. As a result, the drivers did not know whether they were following a human driver or an AV speed profile. The vehicular speed, acceleration, position, relative distance, and relative speed of the lead and following vehicle are recorded with a resolution of 10 Hertz.

The speed profiles used by the AV in scenario 1 are selected based on real-world observations from NGSIM data (186) to represent human driver behavior in real-world traffic conditions. The selected speed profiles from the NGSIM dataset were collected during morning peak over a 2,100-ft length segment on U.S. 101 in Los Angeles, California on June 15, 2005 (172). In scenario 2, the selected speed profiles from NGSIM data were given to a microscopic simulation model based on the AV car-following model for calibration and generating speed profiles representing AV behavior. According to (172), the trajectories of five platoons from the NGSIM dataset are selected and the associated speed profiles are extracted. Figure 4-4 illustrates a sample of the speed profiles. The red line shows the speed profile of a human driver from the NGSIM dataset and the blue line illustrates the simulated speed profile using the AV car-following model. As shown in Figure 4-4, a similar pattern can be observed in both scenarios while the simulated speed profile of the AV appears much smoother than the speed profile of a human driver (172).

Table 4-1 VSP Modes and Average Emissions Rate (132)

VSP Mode	Definition (kw/t)	CO ₂ (g/s)	CO (mg/s)	NO _x (mg/s)	HC (mg/s)
1	VSP < -2	1.671	7.807	0.901	0.450
2	-2 ≤ VSP < 0	1.458	3.908	0.628	0.257
3	0 ≤ VSP < 1	1.135	3.347	0.346	0.406
4	1 ≤ VSP < 4	2.233	8.335	1.173	0.432
5	4 ≤ VSP < 7	2.920	10.959	1.706	0.530
6	7 ≤ VSP < 10	3.525	17.013	2.368	0.705
7	10 ≤ VSP < 13	4.107	20.026	3.103	0.822
8	13 ≤ VSP < 16	4.635	29.222	4.234	0.976
9	16 ≤ VSP < 19	5.161	35.531	5.069	1.112
10	19 ≤ VSP < 23	5.633	55.068	5.865	1.443
11	23 ≤ VSP < 28	6.535	113.824	7.623	2.061
12	28 ≤ VSP < 33	7.585	207.586	12.149	3.373
13	33 ≤ VSP < 39	9.024	441.775	15.456	4.857
14	39 ≤ VSP	10.088	882.300	17.863	10.948

Notes: The rates are for Tier 1 light-duty gas vehicles with engine displacement <3.5L and odometer < 50,000 miles

**Figure 4-3 Vehicles Platooning Scenarios****Figure 4-4 A Sample of the leading vehicle (AV) Speed Profile (172)**

The controller of the AV is Adaptive Cruise Control (ACC). The acceleration term of the algorithm is the major contributor to the AV's behavior in terms of controlling the longitudinal movement through the throttle and brake. The use of AVs with low levels of automation is intentional because such situations are now being encountered in traffic flow. Since there is no connectivity between conventional vehicles and the AV, and the longitudinal movement of the vehicles is the focus of this study, a deterministic acceleration modeling framework is appropriate for modeling the AV's car-following behavior. The acceleration of the AV was determined based on a model developed by (187) and is presented in Equations 9-15.

$$a_n^d(t) = P_a a_{n-1}(t - \tau) + P_v(v_{n-1}(t - \tau) - v_n(t - \tau)) + P_d(s_n(t - \tau) - d_{ref}) \quad (9)$$

where a_n^d is the acceleration of vehicle i ; P_a , P_v , and P_d are the parameters of the model, $P_a = 1.0$, $P_v = 0.58$, and $P_d = 0.1$, that were developed in a study done by Van Aren et al., based on the smooth and fast reaction of the AV controller without leading to unsafe situations (187); v_n is the speed of vehicle n ; τ is the reaction time; s_n denotes the spacing; d_{ref} is the maximum of the minimum distance (d_{min}), following distance based on the reaction time (d_{system}), and the safe following distance (d_{safe}). The minimum distance was set at 2.0 meter (172) and d_{system} and d_{safe} are calculated using Equations 10 and 11 (187).

$$d_{safe} = \frac{v_{n-1}^2}{2} \left(\frac{1}{a_n^d} - \frac{1}{a_{n-1}^d} \right) \quad (10)$$

$$d_{system} = v_n \tau \quad (11)$$

Based on the maximum feasible deceleration of the AV, the maximum safe speed is calculated as follows (172):

$$\Delta x_n = (x_{n-1} - x_n - l_{n-1}) + v_n \tau + \frac{v_{n-1}^2}{2a_{n-1}^{decc}} \quad (12)$$

$$\Delta x = \min[\text{Sensor Detection Range}, \Delta x_n] \quad (13)$$

$$v_{max} = \sqrt{-2a_i^d \Delta x} \quad (14)$$

where n and $n-1$ respectively indicate the AV and the lead vehicle; v_n is the speed of vehicle n ; x_n denotes the location of vehicle n ; l_n is the length of vehicle n ; and a_n^d is the maximum deceleration of vehicle n . It was assumed that the sensor detection range based on the AV radar type is 90 m $\pm$ 2.5%. Finally, the acceleration controller of the AV is defined as follows:

$$a_n(t) = \min(a_n^d(t), K(v_{max} - v_n(t))) \quad (15)$$

where k is a model parameter which equals 1.0.

For more information about the data and the AV controllers please refer to (172).

Results

As mentioned, the objective of this study is to explore the impacts of automated vehicles on the driving behavior of conventional human-driven vehicles and estimates the effects on safety and the environment. After preprocessing the raw data collected from the following human-driven

vehicle in the scenarios, the speed data of driver number 5 in scenarios 1 and driver number 8 in scenario 2 in speed profile 4, and driver number 9 in the first scenario of speed profile 3, show some missing values and lags in time-speed diagrams with respect to their lead vehicle. Therefore, these observations were excluded from the data. Consequently, the speeds of the vehicles with respect to time show no outliers or errors. Table 4-2 presents the descriptive statistics of the following human-driven vehicle speeds and the spacing between the lead and following vehicle. In the second scenario (where the lead Vehicle's speed profile represents AV behavior), the mean values of speed for all drivers at a confidence level of 0.95% are significantly higher than the first scenario (where the lead vehicle's speed profile represents human driver behavior). Additionally, the standard deviation of speed in scenario 2 shows slightly lower values than the first scenario for most of the speed profiles (except speed profile 1). This can be observed from the time-speed diagrams illustrated in Figure 4-5 and Figure 4-6, where vehicle speeds have more alterations in scenario 1 than scenario 2. This will be discussed more using the volatility measures in this section. Concerning the spacing between the following vehicle and the lead vehicle, scenario 2, with a mean of 16.6, is lower than scenario 1 with a mean of 17.0; however, the difference is not statistically significant at a confidence level of 0.95%. Moreover, Time-Headway (TWH) is calculated for the following vehicle. The result shows that the percentage of TWH values lower than 2 seconds (selected as the critical THW) is 22.2% for scenario 1, and 23.3% for scenario 2; however, the difference is not very noticeable. Also, no statistically significant difference was found between the minimum THW values for the scenarios at a confidence level of 95%.

Safety Evaluation

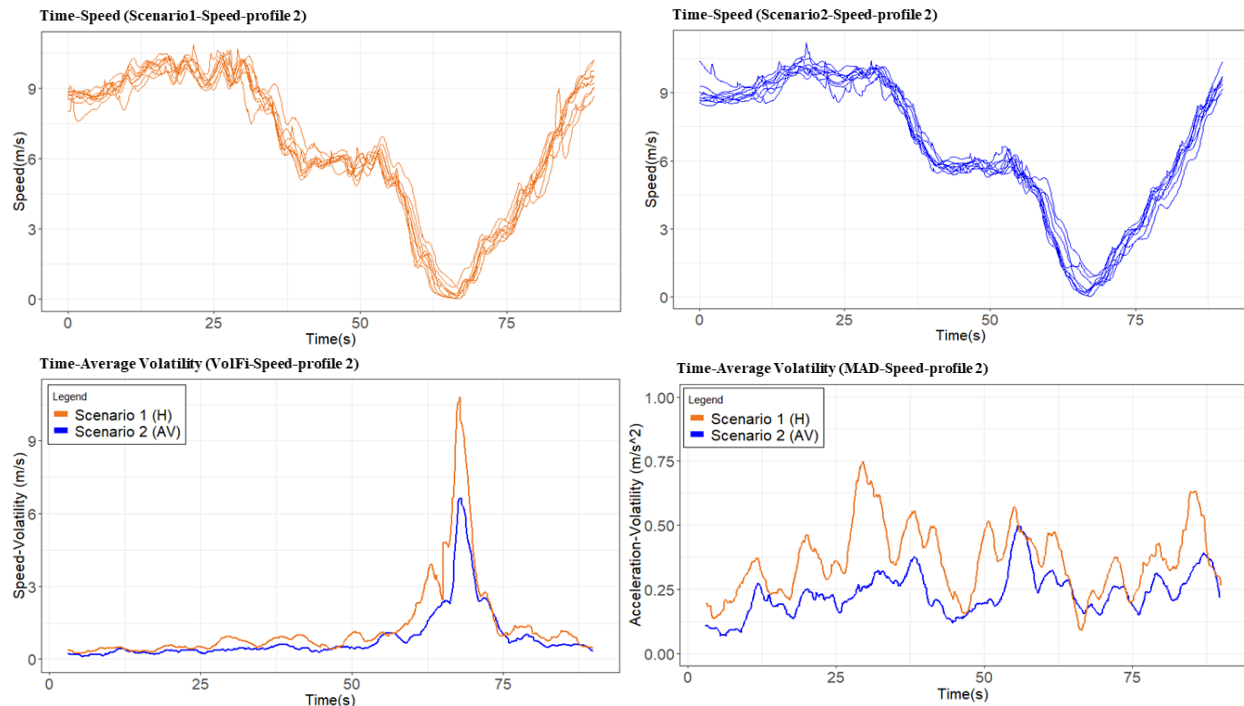
Driving Volatility

The volatility measures are computed for the vehicular speed and acceleration for each of the two scenarios. The results of the speed and acceleration volatility measures are presented in Table 4-3. Note that the values are the aggregation and average of temporal driving volatility of all drivers belonging to each speed-profile that represent trip-based volatility. Note that “%Difference” rows represent the relative volatility. The negative values of %Difference represent the relative decrease in the volatility of each speed-profile, and the positive values show the relative increase in the volatility. On average, the results indicate that drivers mostly show lower speed and acceleration volatility while following an AV in scenario 2 compared with scenario 1. Furthermore, the temporal driving volatility is illustrated in Figure 4-5 and Figure 4-6 just for a demonstration of the results of some volatility measures in speed-profiles 2 and 3. The orange lines represent scenario 1 and the blue lines represent scenario 2. Note that the values of the volatility measures are the average of all drivers in each speed-profile and scenario. The figures indicate a higher level of volatility of speed (VolFi) and acceleration (MAD) for drivers following an AV compared to following a human driver.

Table 4-3 presents the mean and confidence interval of driving volatility reduction when following an AV in a confidence level of 95% with respect to each volatility measure. For instance, based on the time-varying stochastic volatility (VolFi) measure we can say on average an $18.8\% \pm 6.8\%$ reduction in the speed volatility can be concluded when a human-driven vehicle follows an AV. Likewise, the mean absolute deviation (MAD) measure, on average shows a $23.5\% \pm 5.3\%$ reduction in the acceleration volatility.

Table 4-2 Descriptive Statistics of Speed (m/s) and Spacing (m), Average of All Drivers

Speed Profile	Scenario	Lead Vehicle's speed profile	No. of Observation	No. of Drivers	Duration (s)	Min	Max	Mean	SD
Speed (m/s)									
1	Scenario 1	Human	898	9	90	0.076	11.068	6.075	3.013
	Scenario 2	AV	898	9	90	0.011	11.156	6.208	3.030
2	Scenario 1	Human	898	9	90	0.027	10.864	6.512	3.003
	Scenario 2	AV	898	9	90	0.032	11.206	6.584	2.968
3	Scenario 1	Human	799	8	80	0.125	9.784	4.372	2.678
	Scenario 2	AV	799	9	80	0.203	9.365	4.552	2.597
4	Scenario 1	Human	399	8	40	1.613	10.654	7.420	2.480
	Scenario 2	AV	399	8	40	1.959	10.470	7.530	2.464
5	Scenario 1	Human	299	9	30	6.838	10.893	9.015	0.994
	Scenario 2	AV	299	9	30	6.769	11.101	9.231	0.874
Spacing (m)									
1	Scenario 1	Human	898	9	90	6.867	39.871	16.575	5.984
	Scenario 2	AV	898	9	90	8.293	33.017	16.701	5.324
2	Scenario 1	Human	899	9	90	7.795	39.410	17.303	6.740
	Scenario 2	AV	899	9	90	8.187	36.327	16.754	5.799
3	Scenario 1	Human	799	8	80	7.159	32.533	15.013	5.131
	Scenario 2	AV	799	9	80	7.911	28.816	14.616	4.483
4	Scenario 1	Human	399	8	40	9.213	32.628	17.880	5.060
	Scenario 2	AV	399	8	40	8.380	29.910	16.900	5.232
5	Scenario 1	Human	299	9	30	11.140	39.250	20.830	5.717
	Scenario 2	AV	299	9	30	10.620	30.530	20.430	5.087

**Figure 4-5 Speed-profile 2: Time-Speed and Time-Volatility Diagrams in Scenario 1 (Orange Lines) and Scenario 2 (Blue Lines) Including All Nine Drivers**

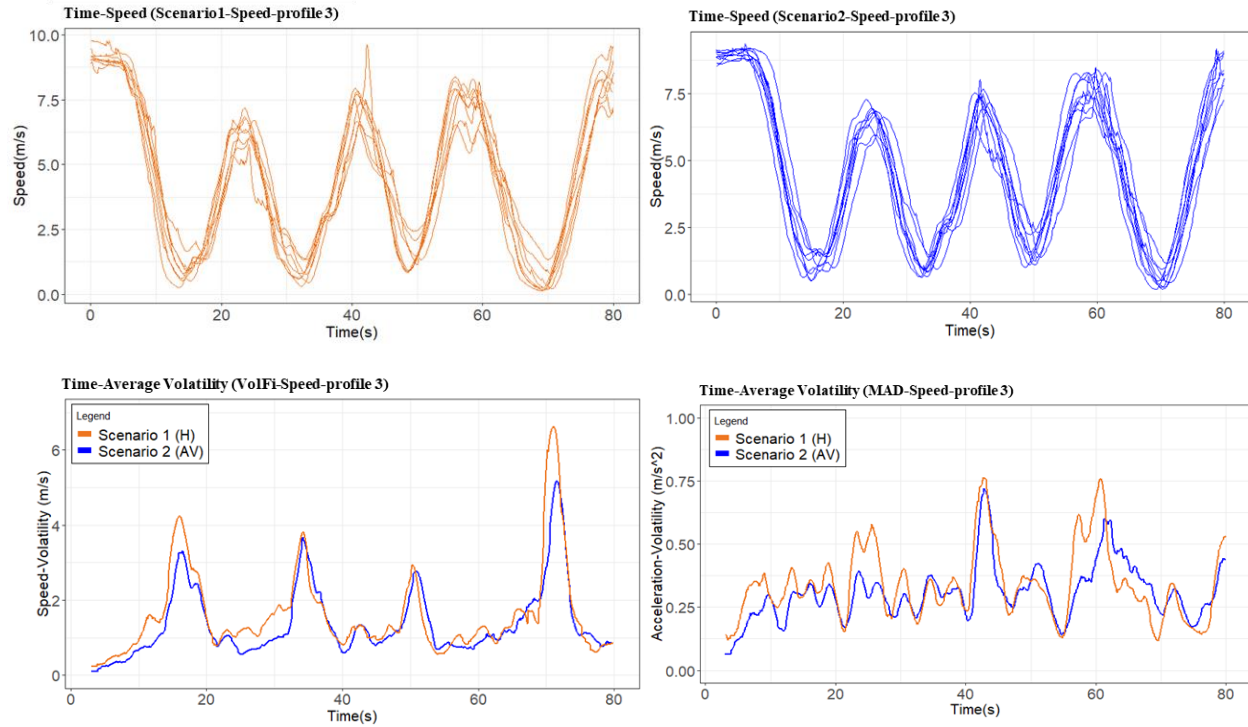


Figure 4-6 Speed-profile 3: Time-Speed and Time-Volatility Diagrams in Scenario 1 (Orange Lines) and Scenario 2 (Blue Lines) Including All Nine Drivers

Table 4-3 Driving Volatility Values, Average of All Drivers

Speed Profile		1		2		3		4		5	
Measure	Scenario-Mode	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speed Volatility											
S _{dev} (m/s)	1-H	0.719	0.056	0.714	0.042	1.401	1.401	1.007	0.042	0.654	0.059
	2-AV	0.568	0.048	0.568	0.021	1.352	1.352	0.925	0.058	0.515	0.037
	% Difference	-21.0%		-20.4%		-3.5%		-8.2%		-21.2%	
MAD (m/s)	1- H	0.609	0.044	0.606	0.035	1.191	0.140	0.855	0.037	0.553	0.050
	2-AV	0.480	0.042	0.481	0.017	1.150	0.096	0.784	0.051	0.433	0.031
	% Difference	-21.2%		-20.7%		-3.5%		-8.2%		-21.7%	
VolFi (m/s)	1-H	1.125	0.423	1.316	0.435	1.615	0.241	0.720	0.166	0.613	0.154
	2-AV	0.873	0.211	0.858	0.154	1.381	0.111	0.626	0.110	0.478	0.121
	% Difference	-22.5%		-34.8%		-14.4%		-13.1%		-22.0%	
Acceleration Volatility											
S _{dev} (m/s ²)	1- H	0.439	0.156	0.471	0.085	0.435	0.083	0.442	0.117	0.539	0.127
	2-AV	0.312	0.056	0.317	0.056	0.395	0.067	0.384	0.079	0.437	0.107
	% Difference	-28.8%		-32.6%		-9.1%		-13.2%		-19.0%	
MAD (m/s ²)	1- H	0.340	0.106	0.365	0.058	0.341	0.053	0.340	0.071	0.435	0.090
	2-AV	0.223	0.036	0.234	0.033	0.307	0.041	0.285	0.049	0.318	0.067
	% Difference	-34.3%		-35.9%		-10.0%		-16.2%		-26.9%	

Time to Collision

Time-to-Collision (TTC) is calculated for the following vehicle for each driver in each scenario. The average of the minimum and mean values of TTC for all drivers are presented in Table 4-4. The results show that the minimum and mean values of TTC for the second scenario are higher than the first scenario, indicating that drivers who follow an AV have relatively safer following behavior compared to those who follow a human-driven vehicle in a similar condition.

As mentioned before, the spacing between the vehicles is slightly lower in scenario 2 than in scenario 1, but the difference is not statistically significant. Closer vehicle spacing can cause lower TTC when we have a fixed or higher speed difference between the two following vehicles. In other words, closer spacing indicates a lower safety margin if the relative speed is not close to zero. Therefore, the higher values of TTC in scenario 2 resulted from the better speed adjustments (lower speed difference) between the following vehicles. The mean of relative speed for all drivers in scenarios 1 and 2 are 0.778 and 0.669 respectively. The lower value in scenario 2 indicates that there is an increased similarity in the speed of the following vehicle and the lead vehicle. This can be seen in Figure 4-7 where the black and red lines represent the average relative speed (difference) of the following vehicle and the leader in the first and second scenarios, respectively. Generally, the red line is closer to the zero axis, where the relative speed equals zero, demonstrating that there is an increased similarity in the speed of the following vehicle and the lead vehicle.

Environmental Indicators Evaluation

Table 4-5 exhibits the fuel consumption of each driver in each scenario. On average, it can be deduced that in similar conditions, human-driven vehicles consume less fuel while following an AV. The negative values of %Difference represent the relative decrease in fuel consumption in scenario 2 compared to scenario 1. The results reveal that on average with a confidence level of 95%, a human driver consumes $10.6\% \pm 3.5\%$ less fuel when follows an AV compared to a human driver. Similarly, the total emissions of the following vehicle in each scenario are calculated. Note that the terrain gradient is assumed to be zero. The results presented in Table 4-6 almost show a lower amount of emissions in the second scenario. On average, with a confidence level of 95%, a $4\% \pm 1.4\%$ reduction in emissions can be concluded while a human-driven vehicle follows an AV.

Discussion and Limitations

This study explores the impacts of automated vehicles on driver behavior in conventional human-driven vehicles and estimates the effects on safety and the environment. Figure 4-8 illustrates a summary of the key results. The results show that the behavior of conventional vehicle drivers is different when following an AV compared to following a human driver and the difference has considerable impacts on safety and the environment. This study reveals that an AV in a mixed platoon can significantly reduce the driving volatility of human-driven vehicles in terms of speed and acceleration. Lower driving volatility indicates smoother and safer driving behavior (117; 165). This finding is consistent with previous studies that have shown that an AV in a platoon of human-driven vehicles can dampen stop-and-go waves (174) stabilize the flow (175), and enhance safety (183; 188). Moreover, the results show that as AVs drive smoother than human drivers (172), AVs influence the following human-driven vehicles. According to the results, there is an increased similarity in the speed of the following human-driven vehicle and the lead AV. This is confirmed in the literature, as human drivers appeared to modify their driving behavior due to less speed alterations in vehicle platoons involving CAVs (9). Less speed alterations and better speed adjustments of human drivers when following AVs can potentially improve safety.

Table 4-4 Descriptive Statistics of TTC (sec), Average of All Drivers

Speed Profile	1		2		3		4		5	
Scenario	1 (H)	2 (AV)	1 (H)	2 (AV)	1 (H)	2 (AV)	1 (H)	2 (AV)	1 (H)	2 (AV)
Minimum	5.07	5.96	5.25	6.29	3.60	4.01	4.64	4.72	11.30	15.16
Mean	164.1	276.5	187.3	249	36.6	78.1	89.8	208.4	126	208.5

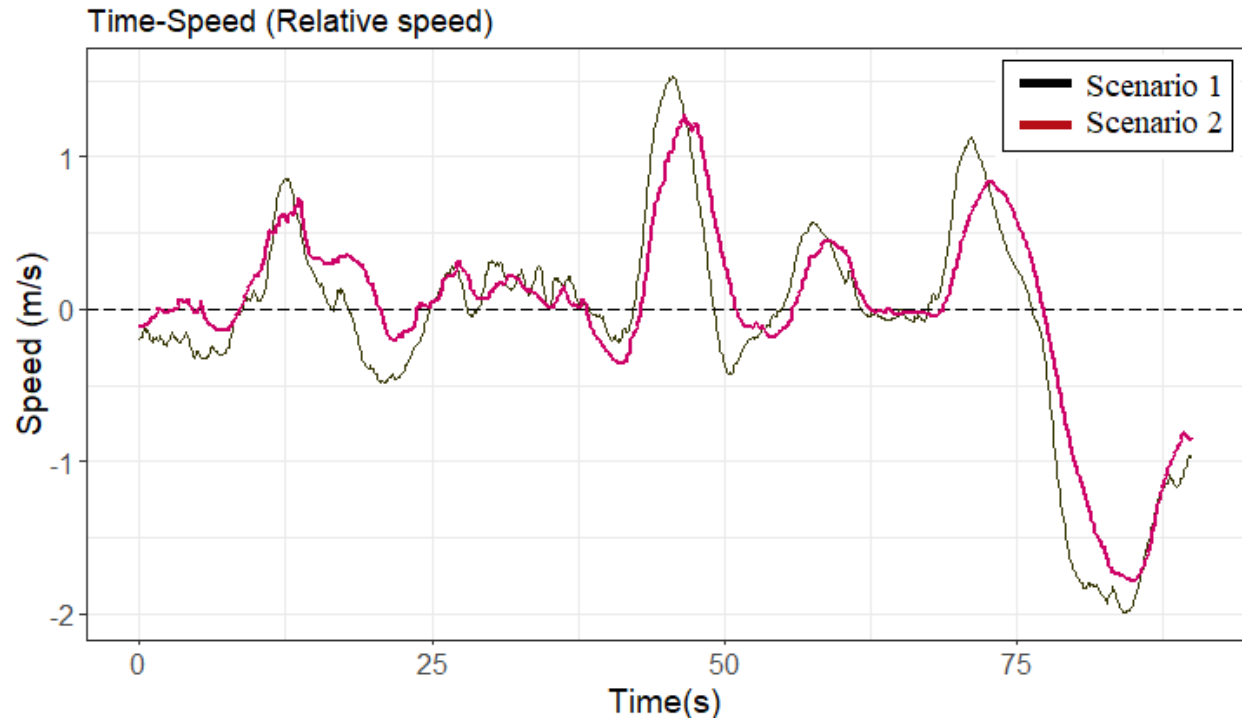


Figure 4-7 A Sample of the Average Relative Speed-Time Diagram for Scenario 1 and 2

Table 4-5 Vehicle's Fuel Consumption (in ml), Based on Kamal et al. Model (131)

Speed Profile	1			2			3			4			5		
Scenario	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference
Driver 1	30.87	25.28	-18.11%	25.44	22.45	-11.75%	26.62	24.35	-8.53%	14.66	13.74	-6.28%	11.04	9.60	-13.04%
Driver 2	32.37	27.94	-13.69%	31.94	25.15	-21.26%	28.36	24.60	-13.26%	18.92	17.02	-10.04%	15.07	11.10	-26.34%
Driver 3	29.71	24.95	-16.02%	22.45	25.28	12.61%	24.53	25.08	2.24%	16.13	14.79	-8.31%	10.62	9.28	-12.62%
Driver 4	30.09	24.67	-18.01%	23.96	21.01	-12.31%	25.47	27.06	6.24%	15.66	15.37	-1.85%	10.65	8.45	-20.66%
Driver 5	34.75	25.96	-25.29%	31.94	20.61	-35.47%	19.80	24.43	23.38%	15.46	NA	NA	13.72	11.75	-14.36%
Driver 6	28.55	26.90	-5.78%	26.46	22.10	-16.48%	26.41	26.05	-1.36%	16.63	15.17	-8.78%	11.32	10.39	-8.22%
Driver 7	29.81	23.62	-20.76%	25.55	19.74	-22.74%	21.43	23.81	11.11%	16.87	15.25	-9.60%	10.78	10.73	-0.46%
Driver 8	29.06	25.17	-13.39%	26.72	21.68	-18.86%	24.76	23.78	-3.96%	NA	13.67	NA	11.04	9.31	-15.67%
Driver 9	29.16	24.40	-16.32%	25.59	21.36	-16.53%	NA	26.13	NA	15.41	15.53	0.78%	10.47	8.83	-15.66%
Average	30.48	25.43	-16.38%	26.67	22.15	-15.87%	24.67	25.03	1.98%	16.22	15.07	-6.30%	11.64	9.94	-14.11%

Note: NA denoting the vehicles that either has missing values or error in the data

Table 4-6 Vehicle's Total Emissions (in grams) Based on the VSP Model

Speed Profile	1			2			3			4			5		
Scenario	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference	1 (H)	2 (AV)	%Difference
Driver 1	187.2	174.5	-6.8%	176.7	168.7	-4.5%	166.6	162.1	-2.7%	87.1	86.5	-0.7%	65.2	63.4	-2.8%
Driver 2	190.2	179.7	-5.8%	189.1	171.0	-9.6%	169.1	161.5	-4.5%	95.7	91.1	-4.8%	74.2	65.4	-11.9%
Driver 3	179.1	171.8	-4.1%	160.2	167.1	4.3%	156.7	160.1	2.2%	86.8	83.3	-4.1%	62.8	61.6	-1.9%
Driver 4	181.9	167.3	-8.0%	166.7	160.8	-3.5%	159.0	160.9	1.2%	87.1	86.0	-1.3%	63.3	56.9	-10.2%
Driver 5	191.5	170.9	-10.8%	186.5	158.9	-14.8%	143.4	156.6	9.2%	86.3	NA	NA	69.8	64.4	-7.7%
Driver 6	178.4	176.3	-1.1%	171.2	164.9	-3.7%	162.0	161.0	-0.6%	89.4	85.9	-3.9%	64.7	61.9	-4.4%
Driver 7	182.3	169.7	-6.9%	168.1	157.4	-6.4%	149.5	155.0	3.6%	88.3	86.7	-1.8%	64.0	62.9	-1.8%
Driver 8	183.5	177.6	-3.2%	178.2	167.0	-6.2%	163.6	160.8	-1.7%	NA	85.8	NA	65.9	63.0	-4.5%
Driver 9	179.2	166.9	-6.8%	169.7	159.4	-6.1%	NA	160.3	NA	88.1	88.9	0.9%	62.3	58.1	-6.8%
Average	183.7	172.8	-5.9%	174.0	163.9	-5.6%	158.7	159.8	0.9%	88.6	86.8	-2.2%	65.8	61.9	-5.8%

Note: NA denoting the vehicles that either has missing values or error in the data

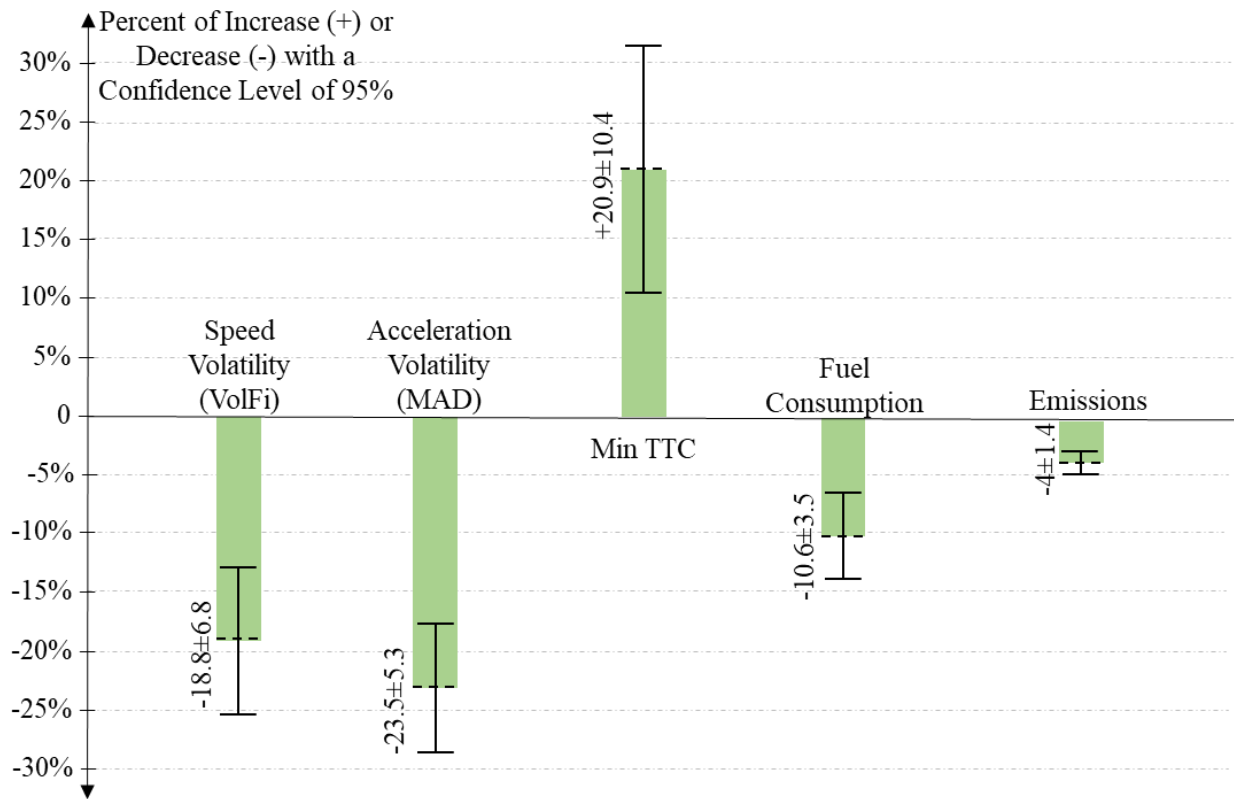


Figure 4-8 Comparison Between Scenario 2 (Human-driven vehicle and AV) and Scenario 1 (Two Human-driven vehicles)

Moreover, the analysis showed a remarkable improvement in TTC in scenario 2 that resulted from the better speed adjustments (lower speed difference) between the vehicles. Increasing TTC for a vehicle can potentially reduce conflicts between consecutive vehicles and reduce the risk of rear-end crashes. Based on different simulation-based studies in the literature, CAVs in mixed traffic improve road safety by increasing the minimum value of TTC (176) and mitigating traffic conflicts (9; 177). In contrast to this finding, Sinha et al. found that an increase in the CAV penetration rate in mixed traffic mitigates the magnitude of TTC between conventional vehicles and CAVs. The difference in their findings could be due to different CAV operational characteristics or different simulation assumptions that potentially have a direct impact on the evaluation of safety effect on mixed traffic flow (176). However, in the real world, the diversity of human driving behavior in a mixed traffic environment might result in expecting different impacts of CAVs.

From an environmental perspective, higher driving volatility is a major contributor to fuel consumption and emissions (163). Given that human drivers show lower speed and acceleration volatility when following an AV, the results of this study show that in similar conditions, human-driven vehicles consume significantly less fuel and produce fewer emissions when following an AV. This finding is consistent with previous studies that concluded AVs in a mixed fleet with conventional vehicles can reduce overall fuel consumption (189) and emissions (144).

The key limitations of this study are the restrictions of scenarios, the limited number of drivers, speed profiles, and the platoon size used in this study. It is essential in the future to consider more complicated scenarios where vehicular speed has higher fluctuations and consider interactions with other traffic in mixed traffic environments to conduct more generalizable conclusions. As this study utilized a limited number of test iterations, there might be randomness and error in the data. Furthermore, the findings of this study might be sensitive to the AV controller. In other words, human drivers may behave differently when following AVs with different algorithms since the AV performance is strongly dependent on its controller. For future studies, it is suggested to repeat this study on different AV algorithms and compare the results. Also, it is suggested for future studies to investigate the change in human drivers' behavior by informing them whether they are following an AV or a conventionally driven vehicle.

Conclusions

This study aims to quantify uncertainties in the interaction of human-driven vehicles and AVs in mixed traffic. It strives to create new knowledge about the impacts of AVs on the driving behavior of following vehicles in a mixed platoon. While the potential impacts of AVs on different users of a transportation system are expected, their direction and extent are not clear. This study quantifies key impacts and points to how AVs may impact the performance of conventional vehicles. To that end, this study explores the impacts of AVs on driver behavior of conventional human-driven vehicles and estimates the resulting effects on safety and the environment. In order to determine the impacts of AVs on human drivers, test-bed data from two sets of scenarios are utilized. The experiment comprises 9 drivers that attempt to follow 5 speed-profiles with two scenarios per profile. In scenario one, a human-driven vehicle follows an AV that implements the speed profiles of human drivers (base), and in scenario two, a human-driven vehicle follows an AV that executes the speed profiles of AVs. It should be noted that the drivers did not know whether they are following a human driver speed profile or an AV speed profile to influence their behaviors. Safety is evaluated through the comparison of several volatility measures and time-to-collision. Likewise,

fuel consumption and emissions are used to investigate environmental impacts. Note that the method used in this study is repeatable and the dataset is publicly available. Overall, the results show that due to their greater stability, lower speed fluctuation, and lower driving volatility, AVs in mixed platoons can cause substantial behavioral changes in following conventional vehicle drivers, with beneficial effects on safety, energy, and the environment.

The results show that the behavior of conventional vehicle drivers is different while following an AV compared with following a conventional vehicle, and this difference has considerable impacts on safety and the environment. On average, based on different volatility measures, a driver following an AV demonstrated lower driving volatility in terms of speed and acceleration. For instance, on average at a confidence level of 95%, an $18.8\% \pm 6.8\%$ reduction in speed volatility and a $23.5\% \pm 5.3\%$ reduction in acceleration volatility can be observed when a human-driven vehicle follows an AV. Since an increase in speed and acceleration volatility significantly increases the frequency of crashes (117; 164) and has a significant association with rear-end, sideswipe, angle, and head-on crashes (117), it is possible that AVs in mixed traffic can reduce the frequency and risk of different types of crashes for the following human-driven vehicles. Furthermore, due to an enhancement in the following behavior of the conventional human-driven vehicles in terms of speed adjustment when following an AV, a larger value for minimum time-to-collision (TTC) is observed that indicates safer following behavior and a lower rear-end crash risk.

From the environmental perspective, on average, a human-driven vehicle consumes up to $10.6\% \pm 3.5\%$ less fuel while following an AV and produces $4\% \pm 1.4\%$ fewer emissions at a confidence level of 95%. These findings indicate that the stability of the AV performance could be adjusted in a way that maximizes the AV performance and also the performance of conventionally-driven following vehicles in mixed traffic. The findings of this study can be used by engineers and practitioners to consider AV's impacts on other road users in a big picture to adjust, enhance, and optimize the AV performance based on the possible interaction between AVs and non-AVs and change in driving behavior of following vehicles.

This study quantifies the expected impacts of mixed vehicle platoons, uncovering the positive impacts that automated vehicles can have on the following vehicles. The results reduce the uncertainty of vehicle automation in terms of their impacts on mixed traffic. To further improve the confidence in the results, it will be necessary to collect even larger samples of data. Additionally, the results are obtained from analyzing a limited set of scenarios and field tests in specific circumstances. Therefore, the results are limited to the discussed scenarios and might not be generalizable to other conditions. To have more generalizable results, more data can be collected expanding the scenarios examined in different mixed traffic environments and conditions, e.g., adverse weather, and different roadway classifications.

CHAPTER 5: HOW EFFECTIVE ARE PEDESTRIAN CRASH PREVENTION SYSTEMS IN IMPROVING PEDESTRIAN SAFETY?

A version of this chapter is accepted for presentation in the 101st Transportation Research Board Conference and is under review in the Journal of Accident Analysis and Prevention.

Abstract

Over the past few years, the number of fatalities and severe injuries of vulnerable road users, particularly pedestrians, has risen substantially. Clearly, the safe mobility of pedestrians is critical in our transportation system. Technology can help reduce vehicle-pedestrian crashes, fatalities, and injuries. Emerging technologies such as pedestrian crash prevention (PCP) systems utilized in on-road vehicles have the potential to mitigate pedestrian crash severity or prevent crashes. However, the reliability and effectiveness of these technologies have remained uncertain. This study contributes toward understanding the effectiveness of PCP systems utilized in on-road vehicles with a low level of automation by investigating different pedestrian crossing scenarios. The Insurance Institute for Highway Safety field test data from 2018 to 2021 is harnessed, where several on-road vehicles and their PCP systems are evaluated in terms of safety. The sensing of pedestrians is done through a vehicle-based vision system without connectivity with external devices located on the infrastructure or carried by the pedestrian. The large-scale experimental dataset is comprised of 3125 tests of 92 vehicles with different sizes, makes, and models. The empirical results indicate that in hazardous pedestrian-vehicle conflict situations, the performance of PCP systems has been improved during recent years. The test data shows that on average in 70% of the tests, the PCP systems avoid pedestrian crashes. However, for the occurred crashes, PCP systems on average were able to mitigate impact speeds more than 50%. In real-life situations, this could translate to substantial reductions in injury risk. Through rigorous analysis, the associations of key factors in different pedestrian crossing scenarios and the performance of PCP systems are explored and discussed in this paper.

Keywords: Vehicle Pedestrian Crashes; Pedestrian Crash Prevention System; Pedestrian Autonomous Emergency Braking System; Pedestrian Collision Avoidance System.

Introduction

More people are walking as a mode of transportation in recent years. Over the past few years, the fatality of vulnerable road users has been rising significantly due to increasing transportation demand. According to the United States National Highway Transportation Safety Administration, 36,096 people were killed in traffic accidents in 2019, with 6,205 of them being pedestrians. An alarming 62% increase in pedestrian fatalities in urban areas between the years 2010 and 2019 raises concern about the safety of pedestrians (190). In addition to the loss of life, about 6,000 annual pedestrian deaths in the U.S. cost society about \$6 billion. Pedestrians have been recognized as the most at-risk road users because of their low level of protection. In recent decades, various efforts have been made to decrease pedestrian traffic fatalities and severe injuries. One popular effort is to recognize factors affecting vehicle-pedestrian crash severity and apply different countermeasures to improve those factors (191-194). In addition, technology can be part of the solution to reduce the number of vehicle-pedestrian crashes, fatalities, and injuries (195-197).

In recent years, some emerging technologies such as adaptive signal control technology (196) and Advanced Driving Assistance Systems (ADAS) have the potential to mitigate crash severity or prevent crashing. The Pedestrian Crash Prevention (PCP) system, also known as the “Pedestrian Automatic Emergency Braking System”, “Pedestrian Collision Avoidance System”, and “frontal pedestrian impact mitigation braking”, is an emerging safety technology utilized in on-road vehicles with a low level of automation. This technology delivers automatic braking for vehicles when pedestrians are in the forward path of the vehicle’s travel and the driver has taken insufficient action to avoid an imminent crash. It is widely expected that autonomous emergency braking systems equipped with a pedestrian detection system have the potential to prevent pedestrian crashes or mitigate their severity. However, the reliability and effectiveness of these technologies have remained uncertain. Many existing on-road vehicles equipped with pedestrian autonomous emergency braking systems had a high failure rate to completely avoid a collision or to significantly mitigate the impact speed (198).

This study aims to contribute toward understanding the reliability and effectiveness of PCP systems utilized in on-road vehicles with a low level of automation. Moreover, this study aims to explore factors affecting the reliability and effectiveness of the technology and to investigate situations in which the PCP system may or may not perform well. To this end, this study analyzes the Insurance Institute for Highway Safety (IIHS) dataset from 2018 to 2021 where several on-road vehicles and their technologies have been evaluated in terms of safety. The findings of this study are based on field test data, which are unique. The rest of the paper is organized as follows: a literature review, the study methodology, data description, results, and finally, the conclusion.

Literature Review

In recent years, the number of on-road vehicles equipped with automatic emergency braking has been increased. It is expected that PCP systems mitigate vehicle-pedestrian severity by either preventing the collision or reducing the impact speed (199). It is shown that reducing the speed at impact in pedestrian crashes from 50 to 25 km/h mitigates fatality risk by about 85% (200). The literature shows that if the market penetration of PCP systems becomes 100% and the systems perform flawlessly, it is expected that pedestrian fatality and severe injury would be reduced by about 15% and 38% each year, respectively (201). However, this is not a present-day reality, and not all pedestrian crashes can be avoided by PCP systems (199). The effectiveness and reliability of PCP systems rely on many factors such as sensor accuracy and cover area (202), system activation time and latency (199), and maximum deceleration (202). Previous studies found that increasing maximum deceleration (202; 203), increasing sensor field of view, and reducing braking initiation time of PCP systems can significantly reduce pedestrian collisions and pedestrian fatalities (203).

While it is clear that the pedestrian crash prevention systems are beneficial to pedestrians and vulnerable road users, there are still some uncertainties about the reliability and effectiveness of the technology. Furthermore, a limited number of factors affecting PCP system performance have been studied in the literature.

Methodology

In this study, a regression model for estimating vehicle-pedestrian impact speed facilitates the exploration of factors affecting the performance and effectiveness of the PCP system to prevent

vehicle-pedestrian crashes or reduce the impact speeds. However, because vehicles can safely stop via the PCP system before hitting the pedestrian, the model may suffer from sample-selection bias. Therefore, the model should first estimate whether the PCP system can stop safely with no collision with the pedestrian, and then if the vehicle hits the pedestrian, the model estimates the speed at impact with the pedestrian. To perform such a model, Heckman's sample selection approach is utilized. Heckman in 1976 (204; 205) introduced a two-stage model to address the bias from a non-random sample selection. Heckman's sample selection approach fits regression models by using either Heckman's two-step consistent estimator or full maximum likelihood. In this study, the full maximum likelihood approach (204) is utilized. The regression equation is shown in Equation 1 and the selection equation is illustrated in Equation 2.

$$y_i = x_i\beta + u_{1i} \quad (1)$$

$$z_i\gamma + u_{2i} > 0 \quad (2)$$

where $u_1 \sim N(0, \sigma)$, $u_2 \sim N(0, 1)$, and y_i is the speed at impact for observation i . In this study, a dummy variable is defined to be used as the selection equation of the Heckman sample selection approach. The binary variable is illustrated in Equation 3. The PCP system is labeled as successful collision avoidance when it can successfully detect a pedestrian in imminent danger of being hit by the vehicle, engage the automatic emergency braking system, and prevent the impact. Otherwise, it is labeled as failure collision avoidance. The log-likelihood for observation j can be calculated using Equation 4.

$$\text{Success/Failure of the PCP system} = \begin{cases} 0: \text{successful Collision Avoidance} \\ 1: \text{Failure Collision Avoidance} \end{cases} \quad (3)$$

$$LL_j = \begin{cases} \omega_j \ln \Phi\left(\frac{z_j\gamma + (y_i - x_j\beta)\rho/\sigma}{\sqrt{1-\sigma^2}}\right) - \frac{\omega_j}{2} \left(\frac{y_i - x_i\beta}{\sigma}\right)^2 - \omega_j \ln(\sqrt{2\pi}\sigma), & \text{if } y_i \text{ observed } (y_i > 0) \\ \omega_j \ln \Phi(-z_i\gamma), & \text{if } y_i \text{ not observed } (y_i = 0) \end{cases} \quad (4)$$

where Φ is the standard cumulative normal and ω_j is an optimal weight for observation j and $\rho = \text{corr}(u_1, u_2)$ which is the correlation between the unobserved determinants and σ is the standard deviation of the outcome. When $\rho \neq 0$, standard regression methods provide biased results but the Heckman selection model delivers consistent and asymptotically efficient estimates (206; 207).

Data

To explore factors affecting the reliability and performance of the PCP system, this study analyzes the Insurance Institute for Highway Safety (IIHS) dataset from 2018 to 2021, where several on-road vehicles and their technologies have been evaluated in terms of safety. The dataset is comprised of 3,125 tests of 92 vehicles with different sizes, makes, and models from 2018 to 2021 that have been tested in 6 scenarios with three different test speeds including 20, 40, and 60 km/h (12, 25, and 37 mph). The test scenarios are illustrated in Figure 5-1. In the test scenarios, using dummies, illustrated in Figure 5-2, that move across or stand in the roadway, pedestrian crash prevention capabilities are evaluated. In the 1st and 2nd scenarios, a vehicle is traveling at speeds

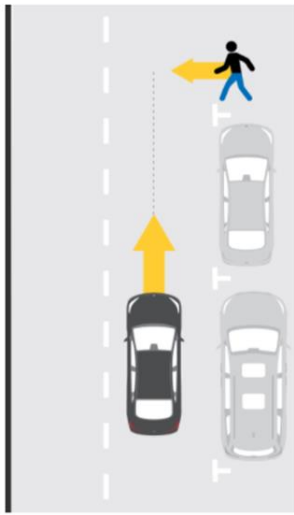
of 20 km/h and 40 km/h while a child darts into the street from behind two parked vehicles in the path of the oncoming vehicle with an unobstructed view. In the 3rd and 4th scenarios, a vehicle is driving at 20 km/h and 40 km/h, and an adult pedestrian steps into the street. Finally, in the 5th and 6th scenarios, a vehicle is driving at 40 km/h and 60 km/h while an adult pedestrian walks along the road in the same direction as the vehicle moves, facing away from traffic. After putting together the data, preprocessing the data, and performing an error checking, the descriptive statistics are reasonable and the data seem valid and reliable. The descriptive statistics and frequencies of the key variables are presented in Table 5-1 and Table 5-2.

In this study, the speed at impact is the dependent variable in the proposed model. The histogram of the speed at impact is illustrated in Figure 5-3. It can be seen that the distribution of impact speeds larger than zero is quite normal.

Results

The test data shows that in 30% of the tests, crashes did occur. However, for those crashes, PCP systems were, on average, able to mitigate impact speeds substantially. For instance, at the test speeds of 60, 40, and 20 km/h at which crashes occurred, PCP systems were able to mitigate the average impact speeds to about 31.94, 16.34, and 5.33 km/h. This potentially can mitigate the severity of pedestrian crashes. According to a review and meta-analysis study conducted by Hussain et al. (208), the overall association between impact speed (km/h) and risk of fatality for pedestrians is illustrated in Figure 5-4. The red curve represents the results from synthesizing 15 relevant studies (black curves) using a multivariate meta-regression model and shows the risk of fatality for pedestrians (208). According to Figure 5-4, pedestrian fatality risks for the average impact speeds are calculated and presented in Table 5-3. It can be observed that PCP systems can substantially mitigate the risk of fatality for pedestrians if drivers do not engage to brake earlier. For instance, at an impact speed of 60 km/h, the risk of fatality is 54%. Risk of fatality is reduced to 5% in the tests in which the PCP systems did not avoid collisions, and to 1.5% considering all tests.

Table 5-4 presents examples of the performance of PCP systems of some of the vehicles available in the IIHS data. It can be seen that vehicles with different brands exhibited different PCP system performances. For example, in 63% of the studied pedestrian crossing scenarios, the PCP system of Tesla Model 3 successfully avoided pedestrian crashes while in other vehicles such as Ford Fusion, Audi A4, Volvo S60, and Lexus ES350, this percentage is 17%, 83%, 83%, and 80%, respectively. Surprisingly, in test scenarios with higher speeds (60 km/h), Tesla Model 3 has a better performance with 60% successful collision avoidance while in other similar vehicles it is much lower.



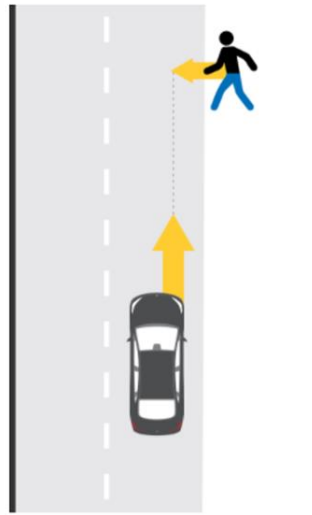
Perpendicular child:

Scenario 1: (CPNC_50)

Child runs into road;
Parked vehicles obstruct view;
Tests run at 20 km/h (12 mph)

Scenario 2:

Child runs into road;
Parked vehicles obstruct view;
Tests run at 40 km/h (25 mph)



Perpendicular adult:

(CPNA_25)

Scenario 3:

Adult walks across road
Tests run at 20 km/h (12 mph)

Scenario 4:

Adult walks across road
Tests run at 40 km/h (25 mph)



Parallel adult: (CPLA_25)

Scenario 5:

Adult in right lane near edge of
road, facing away from traffic;
Tests run at 40 km/h (25 mph)

Scenario 6:

Adult in right lane near edge of
road, facing away from traffic;
Tests run at 60 km/h (37 mph)

Figure 5-1 Test Scenarios in the Insurance Institute for Highway Safety Data (209)



Figure 5-2 IIHS Pedestrian Crash Prevention System Testing, IIHS Images (209)

Table 5-1 Descriptive Frequencies of Key Variables

Variable (N=92 Vehicles and 3125 Tests)		Frequency of Tests	Percent
Success/Failure of PCP system	No collision (speed at impact = 0)	2192	70
	Collisions (speed at impact > 0)	933	30
Test Speed (km/h)	20	1043	33
	40	1562	50
	60	520	17
Scenario	1-Perpendicular Child 20 km/h (12 mph)	522	16.7
	2-Perpendicular Child 40 km/h (25 mph)	520	16.6
	3-Perpendicular Adult 20 km/h (12 mph)	521	16.7
	4-Perpendicular Adult 40 km/h (25 mph)	522	16.7
	5-Parallel Adult 40 km/h (25 mph)	520	16.6
	6-Parallel Adult 60 km/h (37 mph)	520	16.6
Vehicle Model Year	2018	120	4
	2019	1353	43
	2020	1141	37
	2021	511	16
Vehicle Body Type	Hatchback-2 Vehicles	60	2
	Compact sedan-17	601	19
	Mid-size sedan-20	663	21
	Compact SUV-30	960	31
	Mid-size SUV-17	661	21
	Minivan-2	60	2
	Full-size sedan-1	30	1
	Pickup Truck-3	90	3
Vehicle Manufacturer's Reported Weight (base model)	≤1370 kg. (≤ 3,000 lbs.)	360	12
	1370 – 1588 kg. (3,001 – 3,500 lbs.)	1204	38
	1589 – 1814 kg. (3,501 – 4,000 lbs.)	840	27
	1815 – 2041 kg. (4,001 – 4,500 lbs.)	481	15
	>2041 kg. (>4,500 lbs.)	240	8
Vehicle Manufacturer's Suggested Retail Price (base model)	≤ \$30,000	1684	54
	\$30,001 - \$35,000	360	12
	\$35,001 - \$40,000	450	14
	\$40,001 - \$50,000	391	12
	> \$50,000	240	8

Table 5-2 Descriptive Statistics of Variables

Variable (N=92 Vehicles and 3125 Tests)		Min	Max	Mean	SD
Speed at Impact with Pedestrians (km/h) N=3125 Tests	All Years and Test Speeds	0	60.78	7.08	13.30
	2018	0	60.75	12.43	16.37
	2019	0	60.78	9.96	15.64
	2020	0	60.28	4.60	10.20
	2021	0	60.00	3.71	9.23
	20 km/h	0	20.71	2.05	5.45
	40 km/h	0	40.89	6.68	12.23
	60 km/h	0	60.78	18.34	19.40
Conditional Speed at Impact with Pedestrians (Given that Collisions Occurred, N=933 Tests)	All Test Speeds	0.03	60.78	23.70	14.07
	20 km/h	0.50	20.71	14.53	5.53
	40 km/h	0.03	40.89	23.41	11.49
	60 km/h	2.70	60.78	28.06	17.40
Speed Reduction by PCP System (km/h) N=3125 Tests	All Years and Test Speeds	0	60.88	29.49	14.96
	2018	0	60.31	24.56	16.42
	2019	0	60.79	26.61	15.80
	2020	0	60.88	31.95	13.30
	2021	0	60.41	32.87	14.04
	20 km/h	0	20.78	17.83	5.46
	40 km/h	0	40.87	33.25	12.21
	60 km/h	0	60.88	41.43	19.37
Conditional Speed Reduction by PCP System (Given that Collisions Occurred, N=933 Tests)	All Test Speeds	0	56.77	20.38	16.39
	20 km/h	0	19.54	5.33	5.37
	40 km/h	0	39.88	16.34	11.47
	60 km/h	0	56.77	31.94	17.37
Maximum Deceleration by PCP System (m/s ²) N=3125 Tests	All years	0	14.86	8.97	2.52
	2018	0.08	11.13	7.48	3.28
	2019	0	14.86	8.53	3.10
	2020	0	11.79	9.47	1.78
	2021	0	11.25	9.36	1.41
Autonomous Emergency Braking Time-to-collision N=3125 Tests	All years	0	4.53	0.93	0.50
	2018	0	1.58	0.79	0.38
	2019	0	4.53	0.89	0.59
	2020	0	4.51	0.99	0.41
	2021	0	4.53	0.96	0.41
Speed at Initial Braking by PCP System (km/h), N=3125 Tests		19.05	60.88	36.56	13.78
Vehicle Weight (lbs.), N=3125 Tests		2584	5754	3755	610.8

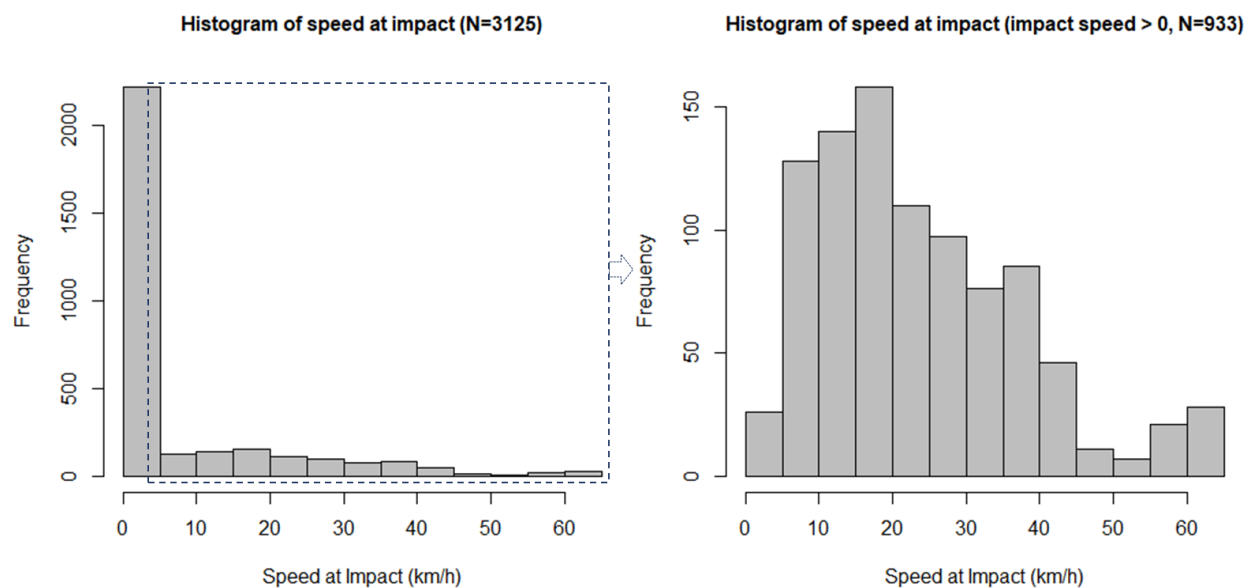


Figure 5-3 Histogram of the Dependent Variable (Speed at Impact km/h). The Right Histogram Is an Expansion of the Distribution on the Left after Removing Zeros

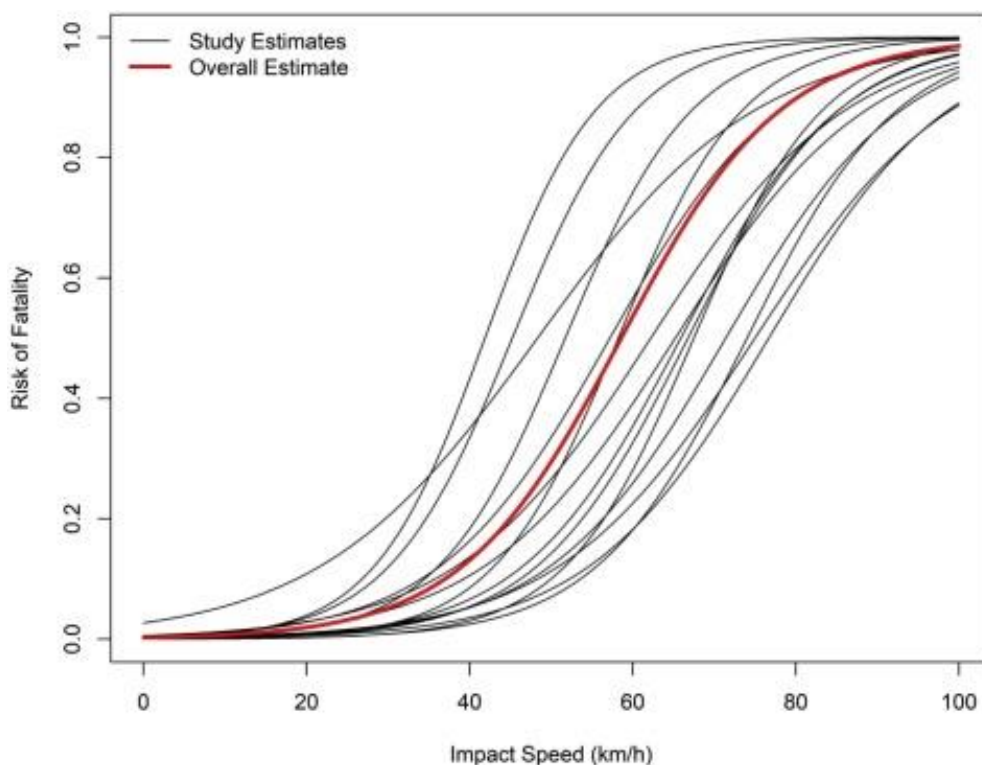


Figure 5-4 “Plot for S-shaped Curves for Pedestrian Fatality Risk by Impact Speed” (208)

Table 5-3 Estimation of the Risk of Pedestrian Fatality by Impact Speed

No PCP system		PCP system (only tests with crashes)		PCP system (all tests)	
Impact speed	Risk of fatality	Average impact speed	Risk of fatality	Average impact speed	Risk of fatality
20 km/h	2%	14.5 km/h	1%	2.1 km/h	~0%
40 km/h	14%	23.4 km/h	3%	6.7 km/h	~0%
60 km/h	54%	28.1 km/h	5%	18.3 km/h	1.5%

Table 5-4 Examples of the Performance of PCP Systems

Vehicle Brand	Test Speed (km/h)	No. of Tests	Average Max Decel. (m/s ²)	Success Rate of PCP system	Average speed at impact (km/h)
Tesla Model3 (2019)	20	15	8.625	67%	31.910
	40	10	7.393	60%	15.340
	60	5	9.530	60%	24.910
	Mean	30	8.365	63%	21.975
Ford Fusion (2019)	20	10	1.717	0%	18.690
	40	15	5.335	33%	36.846
	60	5	9.652	0%	34.430
	Mean	30	4.848	17%	29.100
Audi A4 (2019)	20	10	9.947	100%	0
	40	15	10.224	100%	0
	60	5	10.116	0%	9.871
	Mean	30	10.113	83%	9.871
Volvo S60 (2019)	20	10	9.056	100%	0
	40	15	9.812	87%	24.221
	60	5	9.566	40%	27.201
	Mean	30	9.529	83%	26.009
Lexus ES350 (2019)	20	10	8.512	90%	9.831
	40	15	9.180	100%	0
	60	5	8.005	0%	33.515
	Mean	30	8.762	80%	29.568
Honda Civic (2019)	20	10	9.285	100%	0
	40	15	8.799	80%	19.577
	60	5	8.537	20%	11.344
	Mean	30	8.918	77%	14.872
Toyota Prius (2021)	20	10	9.001	100%	0
	40	15	9.235	100%	0
	60	5	9.899	0%	19.601
	Mean	30	9.267	83%	19.601
Acura TLX (2021)	20	10	9.296	93%	18.516
	40	15	10.096	100%	0
	60	5	8.648	20%	11.433
	Mean	30	9.455	83%	12.850

Figure 5-5 illustrates the percentage of successful collision avoidance of the PCP system from the year 2018 to 2021 for the 3 test speeds. As shown, the percentage of success has increased over time. The PCP system is successful when it can detect a pedestrian in imminent danger of being hit by the vehicle, engage the automatic emergency braking system, and avoid the impact. For example, on average, only 20% of the tested vehicles with the initial speed of 60 km/h in 2018 were able to stop before colliding with the pedestrian. This value has increased to 32% in 2019, 33% in 2020, and 49% in 2021. In addition, based on the frequencies of total tests that resulted in collisions with pedestrians presented in Table 5-5, it can be observed that the percentage of tests that PCP systems failed to avoid pedestrian collisions is continuously reduced during recent years. Figure 5-6 demonstrates the density plot (violin plot) of speed at impact versus car model year. It can be observed that the speeds at impact with pedestrians have been reduced over time from the years 2018 to 2021. In addition, Figure 5-6 illustrates the box plots of the Autonomous Emergency Braking (AEB) time-to-collision versus car model year. It can be noted that the mean for time-to-collision has been increased over time from 2018 to 2021 indicating a better performance of the PCP system.

Heckman Sample Selection Model

A Heckman sample selection regression is developed to estimate the speed at impact with the pedestrian during the test scenarios. Table 5-6 presents the results of the Heckman sample selection model. The model has converged to a value of $\rho = 0.827$, which indicates that the Heckman selection model delivers consistent and asymptotically efficient estimates. Also, the likelihood-ratio test is 68.05, which justifies the Heckman selection model with the data. According to the Wald Chi-Square test (1031.16), the general goodness of fit of the model is acceptable.

The coefficients on the variables (maximum deceleration, scenario, vehicle model year, vehicle manufacturer's suggested retail price, and vehicle manufacturer's suggested weight) represent the estimated marginal effects of the independent variables in the regression model. According to the empirical results obtained from the analysis of the on-road vehicles equipped with PCP systems, it can be seen that a one-unit increase in the maximum deceleration of PCP systems reduces the speed at impact with pedestrians by about 3.8 units. Likewise, previous studies found that increasing maximum deceleration improves the performance of the PCP system (202; 203). According to the modeling results, the speed at impact in scenario 2, scenario 4, scenario 5, and scenario 6 are relatively significantly higher than in scenario 1. Furthermore, the speed at impact decreases in the newer generation of car model years 2020 and 2021, compared to 2018 and 2019. The reason might be related to the improvement in PCP technology and the system algorithm over the years. It can be seen in Table 5-2 that the average maximum deceleration of the PCP system also has been increased over time from 7.48 m/s² in 2018 to 9.36 m/s² in 2021. Interestingly, it can be seen in the IIHS tests that cars with higher Manufacturer's Suggested Retail Price are associated with lower speeds at impact with pedestrians. One possible reason to explain this may be that luxury and expensive cars may be equipped with better high-end and superior quality features and technologies compared to cheaper cars. The results also show that lighter vehicles are significantly associated with a lower speed at impact with pedestrians. In other words, pedestrian crash prevention systems can mitigate the speed at impact more intensely in lighter vehicles.

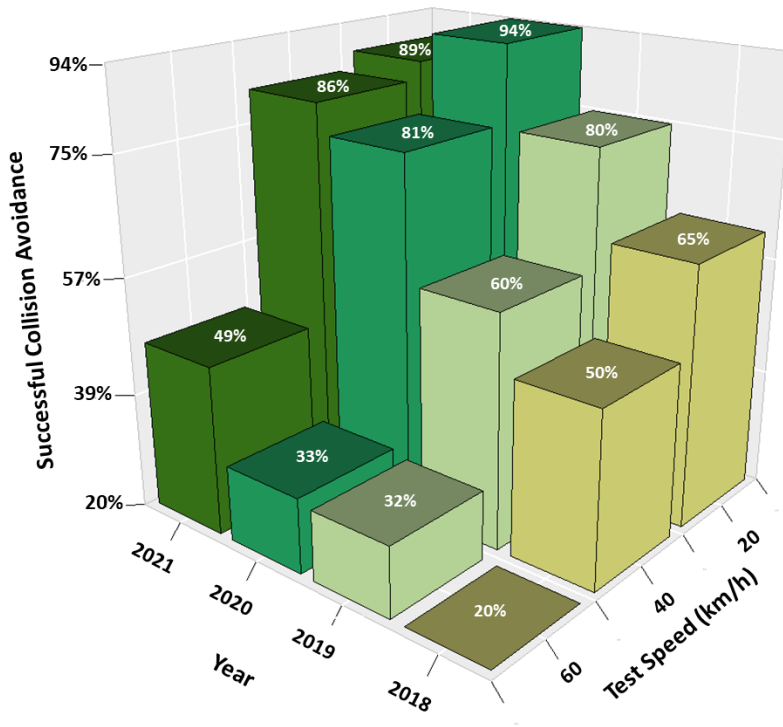


Figure 5-5 Percentage of Successful Collision Avoiding by the PCP Systems

Table 5-5 Frequency of Tests Resulted in Collisions

Year (N=3125)	Number of Total Tests	Frequency of Tests Resulted in Collisions (Speed at Impact > 0)	Percent of Failed Collision Avoidance
2018	120	60	50%
2019	1353	514	38%
2020	1141	261	23%
2021	511	98	19%
All	3125	933	30%

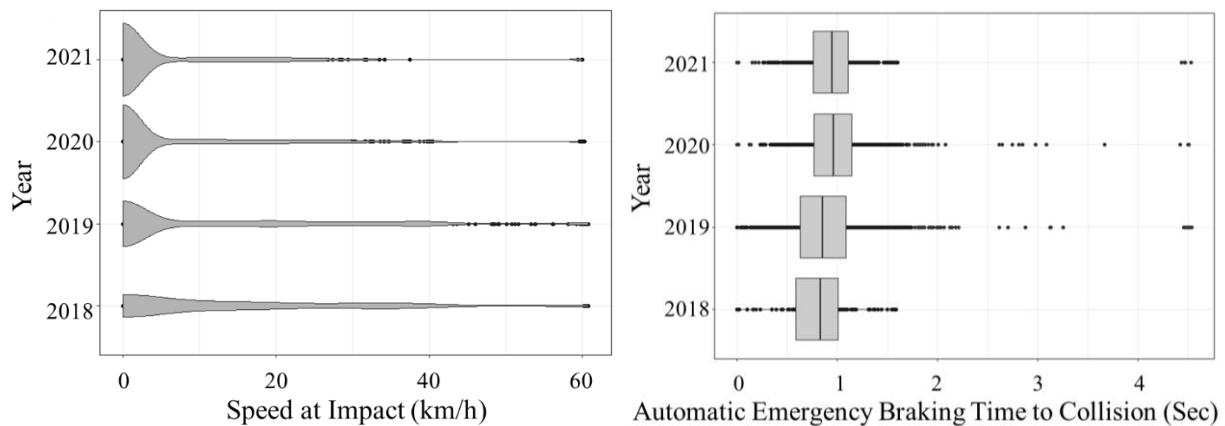


Figure 5-6 Density plot of speed at impact (left side) and Box plots of time-to-collision (right side) for different car model years from 2018 to 2021

Table 5-6 Heckman Sample Selection Regression

Y: Speed at Impact (N=3125)				
Variables		β	Z-statistic	P-value
Constant		18.650	7.50	0.000
Maximum Deceleration (m/s ²)		-3.809	-24.31	0.000
Scenario	1-Perpendicular Child 20 km/h (base)			
	2-Perpendicular Child 40 km/h	20.474	17.60	0.000
	3-Perpendicular Adult 20 km/h	-4.582	-2.95	0.003
	4-Perpendicular Adult 40 km/h	10.680	8.57	0.000
	5-Parallel Adult 40 km/h	6.313	4.28	0.000
	6-Parallel Adult 60 km/h	26.300	22.83	0.000
Vehicle Model Year	2018	2.854	1.69	0.091
	2019	5.303	4.97	0.000
	2020	-1.006	-0.88	0.379
	2021 (base)			
Vehicle	≤\$30,000	4.578	2.73	0.006
Manufacturer's	\$30,001 - \$35,000	1.671	1.00	0.320
Suggested Retail	\$35,001 - \$40,000	5.030	2.95	0.003
Price (base model)	\$40,001 - \$50,000	3.333	1.94	0.052
	>\$50,000 (base)			
Vehicle	≤3,000 lbs. (base)			
Manufacturer's	3,001 – 3,500 lbs.	2.003	1.95	0.051
Reported Weight	3,501 – 4,000 lbs.	3.361	2.88	0.004
(base model)	4,001 – 4,500 lbs.	7.108	4.73	0.000
	> 4,500 lbs.	7.955	3.85	0.000
Y: 0 = No Impact, 1= Impact				
Constant		1.179	5.79	0.000
Maximum Deceleration (m/s ²)		-0.326	-17.31	0.000
Scenario	1-Perpendicular Child 20 km/h (base)			
	2-Perpendicular Child 40 km/h	1.351	13.63	0.000
	3-Perpendicular Adult 20 km/h	-0.373	-3.00	0.003
	4-Perpendicular Adult 40 km/h	0.578	5.56	0.000
	5-Parallel Adult 40 km/h	-0.029	-0.24	0.809
	6-Parallel Adult 60 km/h	2.028	19.92	0.000
Vehicle Model Year	2018	0.774	4.65	0.000
	2019	0.631	6.92	0.000
	2020	0.166	1.73	0.083
	2021 (base)			
Vehicle Body Type	Compact Sedan (base)			
	Compact Hatchback	0.070	0.32	0.752
	Mid-size Sedan	0.133	1.54	0.123
	Compact SUV	0.018	0.24	0.813
	Mid-size SUV	0.277	3.12	0.002
	Minivan	0.198	0.95	0.341
	Full-size Sedan	0.802	3.32	0.001
	Pickup Truck	-0.025	-0.13	0.899
athrho (inverse hyperbolic tangent of rho)		1.177	7.30	0.000
Ln sigma		2.346	51.92	0.000
rho		0.827		
Sigma (standard error of residual)		10.443		
lambda		8.633		
Log Likelihood = -4474.855; AIC = 9023.71; BIC = 9247.46				
Chi2 = 68.05; Prob>Chi2 = 0.000; Wald Chi2 = 1031.16; Prob>Wald Chi2 = 0.000				

Table 5-7 presents the marginal effects of the probability of being selected by the Heckman sample selection model. The results indicate that the PCP systems in vehicles model years 2021 and 2020 have a higher probability of preventing vehicle-pedestrian crashes. Likewise, it can be observed that pedestrians in scenarios 2 and 6 have a higher probability of being hit by a vehicle equipped with the PCP system. For varying vehicle body types, the PCP systems utilized in pickup trucks, compact sedans, compact SUVs, and mid-size sedans have a higher probability to prevent pedestrian collisions.

Limitations

This study is not without limitations. The key limitations of this study are the restrictions of variables and the restrictions and simplification of scenarios. Moreover, the results are obtained from examining a limited number of scenarios and field tests in specific circumstances. Therefore, the results are restricted to the analyzed scenarios and might not be generalizable to other conditions. Future studies can consider more complicated scenarios with more test speeds, different weather conditions, and different traffic environments to conduct more generalizable conclusions.

Conclusions

In recent years, emerging technologies such as pedestrian crash prevention systems utilized in on-road vehicles with low levels of automation have the potential to mitigate pedestrian crash severity or prevent crashes. However, the effectiveness of these technologies has remained uncertain. This study contributes toward understanding the effectiveness of PCP systems utilized in on-road vehicles with a low level of automation by harnessing field test data collected by the Insurance Institute for Highway Safety data from 2018 to 2021.

The empirical results indicate that in hazardous pedestrian-vehicle conflict situations, the performance of PCP systems has been improved during recent years. This can potentially reduce the number of pedestrian crashes, fatalities, and severe injuries by increasing the market penetration of the technologies. Furthermore, the PCP systems can substantially mitigate the impact speeds. In real-life situations, this would translate to substantial reductions in fatality risk.

Based on the developed model, the results show that increasing the maximum deceleration rate of the PCP system and lower weight of vehicles can significantly improve the performance of the PCP system by decreasing the speed at impact with pedestrians. Furthermore, based on the studied pedestrian crossing scenarios, the results show that the PCP system in scenario 6 (an adult walking in the right lane near the edge of a road facing away from traffic and a vehicle traveling at 60 km/h) and scenario 2 (a vehicle traveling at 40 km/h and a child darting into the road while some parked vehicles obstruct the view of the vehicle) have a less effective performance and a higher probability of colliding with pedestrians.

The findings of this study are based on the large-scale field test data, which are unique. The findings can provide auto manufacturers with some insights to improve the performance of PCP systems and can help transportation practitioners better understand the impact of PCP technology on traffic safety.

Table 5-7 Marginal Effects of the Probability of being Selected (Probit Model)

Y: 0 = No Impact, 1= Impact (N=3125)				
		dy/dx	Z-statistic	P-value
Maximum Deceleration (m/s ²)		-0.067	-19.94	0.000
Scenario	1-Perpendicular Child 20 km/h (base)			
	2-Perpendicular Child 40 km/h	0.334	14.96	0.000
	3-Perpendicular Adult 20 km/h	-0.043	-3.04	0.002
	4-Perpendicular Adult 40 km/h	0.108	5.58	0.000
	5-Parallel Adult 40 km/h	-0.004	-0.24	0.808
	6-Parallel Adult 60 km/h	0.556	25.27	0.000
Vehicle Model	2018	0.165	4.28	0.000
Year	2019	0.131	7.39	0.000
	2020	0.031	1.76	0.079
	2021 (base)			
Vehicle Body Type	Compact Sedan (base)			
	Compact Hatchback	0.014	0.31	0.755
	Mid-size Sedan	0.027	1.54	0.123
	Compact SUV	0.004	0.24	0.813
	Mid-size SUV	0.058	3.13	0.002
	Minivan	0.041	0.92	0.357
	Full-size Sedan	0.185	3.02	0.002
	Pickup Truck	-0.005	-0.13	0.898

ACKNOWLEDGMENT

The authors wish to thank Transportation Engineering & Science Program (TESP), CEE, UTK, and the Insurance Institute for Highway Safety organization. The author also would like to thank the following financial support for the research, authorship, and publication/presentation of this article: This study was partially supported by the Collaborative Sciences Center for Road Safety, CSCRS, a U.S. DOT funded National Transportation Center, under the Grant No. 69A3551747113, and the Tennessee Department of Transportation, TDOT.

CHAPTER 6: UNDERSTANDING THE ROLE OF FASTER EMERGENCY MEDICAL SERVICE RESPONSE IN SURVIVAL TIME OF PEDESTRIANS IN FATAL CRASHES

A version of this chapter is under review in the journal of Accident Analysis and Prevention.

Mahdinia, I., Mohammadnazar, A., & Khattak, A. J. (2021). Understanding the Role of Faster Emergency Medical Service Response in Survival Time of Pedestrians in Fatal Crashes. Under review in the journal of Accident Analysis & Prevention.

Abstract

Fatalities and severe injuries among vulnerable road users, particularly pedestrians, are rising. In addition to the loss of life, about 6,000 annual pedestrian deaths in the U.S. cost society about \$6 billion. Contrary to the assumption that all fatal pedestrian-involved crashes are similar, instantaneous death is substantially more severe than death that occurs several days after the crash. Instead of homogenizing all pedestrian fatal crashes, this study takes into account the severity of fatal injury crashes as a timeline based on the survival time of pedestrians. This study extracts valuable information from fatal crashes by examining pedestrians' survival time ranging from early death to death within 30 days of the crash. The Fatality Analysis Reporting System dataset is utilized from 2015 to 2018. The emergency medical service (EMS) response time is the key post-crash measure, while controlling for pedestrian, driver, roadway, and environmental characteristics. Notably, the response time and survival time can cause endogeneity, i.e., the response times may be shorter for more severe crashes. Due to the spatial and temporal nature of traffic crashes, to extract the association of different variables with pedestrians' survival time, a geographically and temporally weighted truncated regression with a two-stage residual inclusion treatment (local model) is estimated. The local model can overcome the endogeneity limitation (between EMS response time and survival time) and uncover the potentially spatially and temporally varying correlates of pedestrians' survival time with associated factors to account for unobserved heterogeneity. Moreover, to verify the variations are noticeable, a truncated regression with the two-stage residual inclusion treatment is developed (global model). The modeling results indicate that while capturing the unobserved heterogeneity, the local model outperformed the global model. The empirical results show that EMS response time, speeding, and some pedestrian behaviors are the most important factors that affect pedestrians' survival time in fatal injury crashes. However, the effect of factors on pedestrians' survival time is noticeably varied spatially and temporally. The results and their implications are discussed in detail in the paper.

Keywords: Pedestrian Time-to-Death; Survival Time; Emergency Medical Service Response Time; endogeneity; heterogeneity.

Introduction

Walking as a mode of transportation has risen in recent years. An increasing number of fatal traffic crashes is one of the adverse impacts of increasing transportation demand (87). Over the past few years, the fatality of vulnerable road users has been rising significantly. According to the United States National Highway Transportation Safety Administration, 36,560 people were killed in traffic accidents in 2018, with 6,283 of them being pedestrians. An alarming 3.4% increase in pedestrian fatalities between the years 2017 and 2018 raises concern about the safety of pedestrians (210). Pedestrians have been recognized as the most at-risk road users because of their low level

of protection. In recent decades, various efforts have been made to decrease pedestrian traffic fatalities. One popular effort is to recognize factors affecting vehicle-pedestrian crash severity and apply different countermeasures to improve those factors in a way to decrease the number of pedestrian fatalities and severe injuries.

To discover contributing factors and extract their complex association with pedestrian fatalities, many previous studies utilized different modeling approaches (191-194; 211; 212). Most of the models were developed to estimate stationary relationships between injury severity and associated factors while some studies utilized heterogeneity-based models to capture unobserved heterogeneity (213) and variation of the relationships between injury severity and associated factors across space and time (193; 214). Researchers analyzing crash severity usually assume that all fatal crashes are similar. However, instant death in a fatal crash is generally substantially more severe than death several days afterward. Instead of homogenizing all fatal pedestrian crashes, this study takes into account the severity of fatal injury crashes as a timeline based on the survival time of pedestrians ranging from early death to death within 30 days of the crash. To that end, the Fatality Analysis Reporting System (FARS) dataset from 2015 to 2018 is utilized. The findings of this study provide insights into pre-crash factors in terms of pedestrian, driver, roadway, and environmental characteristics, and more importantly, the post-crash factor of Emergency Medical Service (EMS) response times that can positively or negatively affect the survival time (time-to-death) of a pedestrian victim in a car crash.

In this study, the EMS response time is defined as the difference between crash time and the arrival time of EMS personnel. The literature discusses this response time as a key important variable that significantly contributes to the survival time of vehicle occupants in a crash (215; 216). Notably, response time and survival time can have endogeneity, i.e., the response times may be shorter for more severe crashes. Also, it is possible that the same set of observed and unobserved factors simultaneously influence EMS response time and pedestrians' survival time (215). For instance, consider a crash that occurred during off-peak-hours in an urban area. Usually, lower traffic demand and higher vehicular speeds may increase the severity of pedestrian-car crashes. However, these conditions may allow EMS personnel to respond faster and arrive at the crash scene in a shorter period of time. Therefore, it is crucial to deal with this endogeneity when modeling pedestrian survival time to have theoretically plausible relationships between endogenous variables and the variable of interest. To that end, the Two-stage Residual Inclusion (2SRI) treatment is applied as it is utilized to overcome the endogeneity limitation (215). Finally, to extract the associations of different variables with pedestrians' survival time and address the potential existence of unobserved heterogeneity, a geographically and temporally weighted truncated regression with 2SRI treatment is estimated. The rest of the paper is organized as follows: a literature review, the study methodology, data and variable description, test results, discussion, and finally, the conclusion.

Literature Review

Several past studies have focused on the severity of pedestrian injuries in traffic crashes (191-194). Researchers have found many pedestrian, environmental, and roadway variables that contribute to pedestrian-vehicle crash severity. Pedestrian characteristics include pedestrian behavior and

violation of traffic laws (193; 217), pedestrian age (218), drug intoxication (193), and alcohol consumption (193; 219). Environmental characteristics consist of different weather conditions (220; 221), and roadway characteristics encompass land use (222; 223), crash location (224; 225), and impact speed (200; 208; 225). Most of the previous studies offer useful findings on what factors influence vehicle-pedestrian crash fatality. Most of the previous studies considered stationary relationships between injury severity and associated factors. However, a few studies captured the unobserved heterogeneity across the pedestrian crash observations (193; 213; 214).

Studies that categorized vehicle-pedestrian crashes into fatal vs. non-fatal categories assume that all fatal crashes are similar in terms of severity. However, instant death in a fatal crash is substantially more severe than a death several days after a crash (215). Very few studies have considered the severity differences among fatal crashes. The time to death of pedestrians killed in traffic accidents is associated with different factors such as impact speed and pedestrian age (226-228). Yasmin et al. analyzed the driver fatal-injury severity in vehicle crashes and showed that the severity of a fatal crash varies based on the survival time of the victims after the crash time. They categorized the survival timeline to seven fatal-injury severity categories as an ordered variable and developed a generalized ordered logit model to estimate the fatal-injury severity of car crashes. Their findings indicated that many factors can affect the time-to-death of a driver victim in a car crash (215). Ju and Sohn (2014) used the Weibull regression approach and investigated the factors that are correlated with the victim's survival time (229). One of the most important factors found in previous studies is the EMS response time. It has been revealed that an early EMS response can potentially increase the survival time of a traffic crash victim (215; 228; 230; 231). Previous studies found that a minute reduction in EMS response time and a ten-minute reduction in EMS response time can decrease the risk and probability of death for a car crash victim by 4.08% (232) and 33% (233), respectively.

While previous studies provided significant understandings of the factors associated with injury severity, there is much to be further examined. The most important gaps found in the literature are that few studies have investigated the factors affecting the survival time of pedestrians in fatal injury crashes, specifically pedestrian behavioral factors and EMS response time. The majority of studies considered all fatal crashes the same in terms of severity. Furthermore, most of the studies assumed stationary associations between pedestrian survival time and associated factors and did not capture the potential existence of unobserved heterogeneity. This study contributes to continuing research on pedestrian fatal crashes. Specifically, instead of homogenizing all pedestrian fatal crashes, this study takes into account the severity of fatal injury crashes as a timeline based on the survival time of pedestrians and extracts a large range of factors, particularly pedestrian behavioral factors and EMS response time that may influence pedestrian time-to-death. Furthermore, this study utilizes a spatiotemporal method to model pedestrians' survival time to address the potential unobserved heterogeneity. The findings of this study provide insights into pre-crash and post-crash factors in terms of pedestrian, driver, roadway, and accident characteristics as pre-crash factors, and EMS response times as a post-crash factor that can affect the survival time of a pedestrian victim in a fatal car crash.

Methods

This study focuses on analyzing the survival time of pedestrians involved in fatal injury crashes and extracts associations of several pre-crash and post-crash variables. To address unobserved heterogeneity across crash observations in space and time, a geographically and temporally weighted truncated regression with 2SRI treatment is developed (local model). The model can overcome the endogeneity limitation (between EMS response time and time to death) and uncover the potentially spatially and temporally varying correlates of pedestrians' time-to-death with associated factors. Moreover, to verify the variations are noticeable, a truncated regression as a common statistical approach for analyzing survival time (234) with the two-stage residual inclusion treatment is developed (global model). A model with 2SRI endogenous treatment comprises a two-step procedure (235) which is described in this chapter. Figure 6-1 presents the flow diagram of the models that will be described in this chapter. A model with 2SRI endogenous treatment comprises a two-step procedure (Terza et al. 2008). To develop both the global and local models, in the first stage, an instrumental variable is developed to overcome the endogeneity bias. Then, in the second stage, the instrumental variables alongside other variables are used to develop the local and global models. The local and global models are compared to select the best fit.

First Stage: 2SRI Endogenous Treatment (global model)

The two-stage residual inclusion treatment method was introduced by Hausman (236). In the first stage, the residuals are estimated from an auxiliary linear regression estimation of the endogenous variables (i.e., EMS Response time). In the second step, a model is computed by including the first-stage residuals as an additional regressor alongside the endogenous variables in estimating the outcome of interest variables (215; 235). It is notable that the actual values for the endogenous variables are kept in the second-stage; and at the same time the residuals from the first stage have replaced the unobserved confounders to control for endogeneity (235). Statistically significant coefficients for the residual and the endogenous variable, EMS response time, and the expected impact of EMS response time on the response variable are the evidence for controlling the endogeneity bias. The coefficient of the residual term is an indication of the non-intuitive impact of EMS response time due to the correlation between unobserved confounders and EMS response time (215). Equations 1 and 2 are used to conduct the first stage. Let i be the indicators to represent pedestrians, x_i be a column vector of the observable exogenous variables, x_e represent a set of e ($e = 1, 2, 3, \dots, E$) endogenous variables, and x_u is defined as unobserved confounding variables correlated with the variable of interest, Y (i.e., pedestrian time-to-death), and the endogenous variables (i.e., EMS Response Time). x_u can generate endogeneity bias in the model (215). In this study, it is hypothesized that EMS Response Time may be correlated with the unobserved determinants of pedestrian time-to-death. Therefore, in this study, EMS Response Time is the only endogenous variable, $e = 1$. Using Equation 1, the endogenous variable, x_e , can be estimated through an ordinary least squares model (235).

$$x_e = \rho w_i + x_u \quad (1)$$

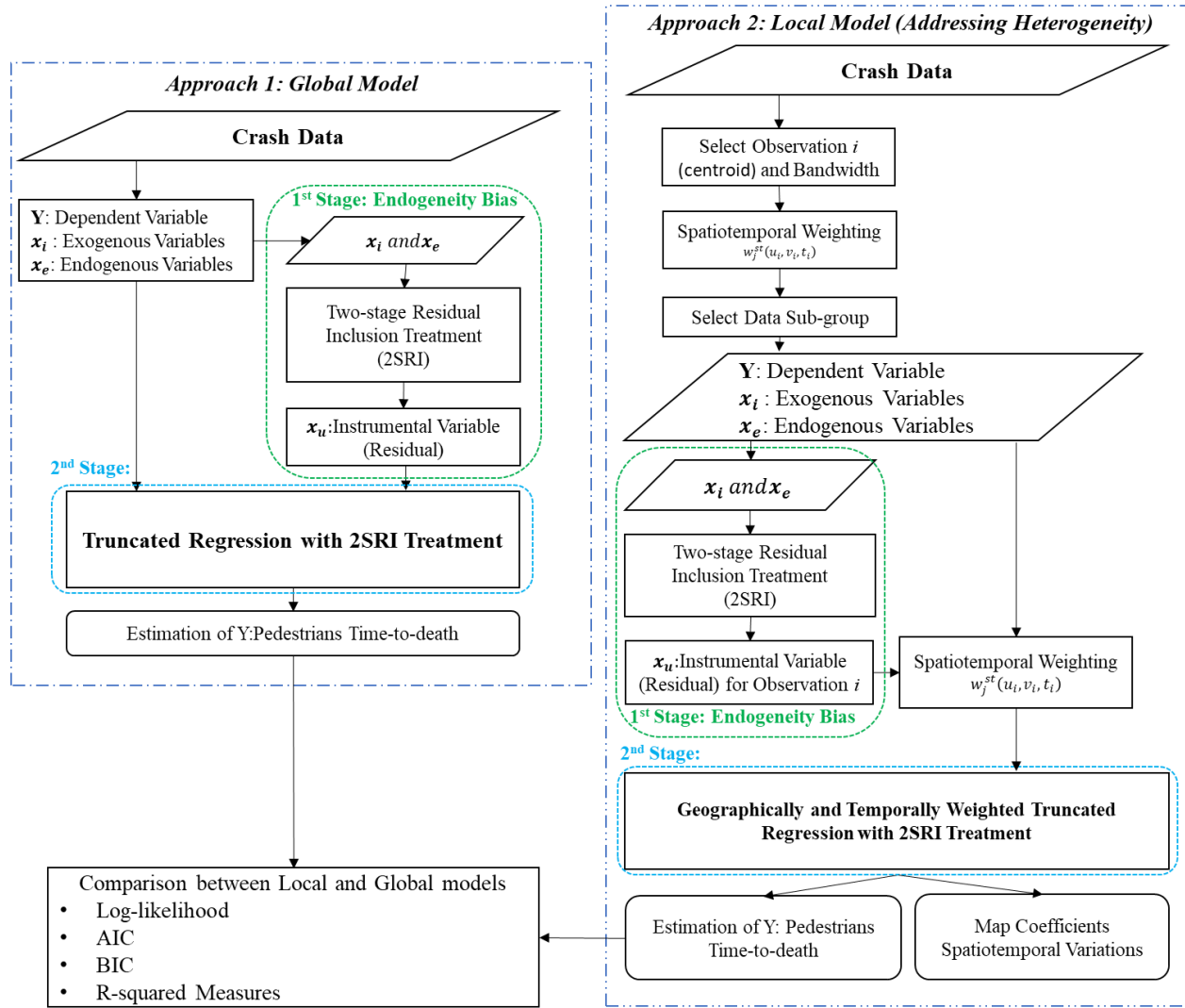


Figure 6-1 Flow diagram of the models

where ρ is a corresponding row vector of parameter estimates and $w_i = [x_i \ b_i]$ is a vector of the observed exogenous variables (e.g., crash time, lighting condition, roadway functional classification) and a set of instrumental variables b_i . x_u is residuals and can be calculated using Equation 2.

$$x_u = x_e - \rho w_i \quad (2)$$

In this study, x_e is the EMS Response Time variable, w_i represents the variables that correlate with the EMS Response Time variable including Crash Time, Lighting Condition, and Roadway Functional Classification. The residual (instrumental variable) is added to the outcome model (Y) to overcome endogeneity bias.

Second Stage: Truncated Regression (global model)

In the second stage, pedestrian time-to-death (survival time) is the predictive variable, which is a non-negative continuous variable and considered left-truncated. Truncated regression fits a model of a response variable on regressors from a restricted part of a population. Therefore, a truncated regression model with a lower and/or upper truncation limit is utilized. If x has a normal distribution with mean μ and standard deviation σ , the density of the truncated normal distribution is as follows (237):

$$f(x|a < x < b) = \frac{\frac{\phi}{\sigma}(\frac{x-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \quad (3)$$

where Φ and ϕ are the distribution and density functions of the standard normal distribution, in which a is the lower limit and b is the upper limit. The truncated regression model with a maximum likelihood function can be summarized as follows:

$$Y = \beta X + \epsilon, \text{ and } \epsilon \sim N(0, \sigma^2 I) \quad (4)$$

$$\ln L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - x_j\beta)^2 - \sum_{j=1}^n \log(\Phi(\frac{b-x_j\beta}{\sigma}) - \Phi(\frac{a-x_j\beta}{\sigma})) \quad (5)$$

where Y represents the continuous dependent variable, ϵ represents the error component that follows a truncated normal distribution, and $\ln L$ represents the log-likelihood. In this study, as the pedestrians' time-to-death values are greater than zero and have a right-skewed distribution, the lower limit of truncation is considered zero ($a = 0$). Likewise, as the data only include deaths within 30 days of the crashes, the upper limit is considered $b = 10.674$ which is the logarithm of 43200 (minutes).

Geographically and Temporally Weighted Truncated Regression (local model)

As heterogeneity in the crash data among vast areas and long periods can result in a different association between the response variable and explanatory variables, Geographically and Temporally Weighted Regression (GTWR) has been utilized to simultaneously deal with the spatial and temporal variations of model coefficients (193; 214). GTWR takes a sub-group of observations with respect to a target location and time and assigns weights to the observations

based on the geographical and temporal distance to the target observation as a centroid of the sub-group. The centroid location moves from one observation to another. The observations in a sub-group are selected from the adjacent observations to the centroid that are located within a certain distance called bandwidth. Several sizes of bandwidth should be tested to determine the optimal bandwidth based on model performance. Model performance is measured by the Akaike Information Criterion (AIC) to obtain for the optimal bandwidth (193; 238). In this study, the optimal bandwidth is 550 observations, which was selected by considering models with smaller AICs, and with a reasonable sign for the average of each coefficient.

To weight observations, two terms of weights are calculated. First, the spatial weights of the observations in a sub-group are computed based on the geographical distance to the corresponding centroid. The Bi-square Kernel Weighting function is utilized in this study to determine the spatial weights (214; 239).

$$w_j^{spatial}(u_i, v_i) = \left[1 - \left(\frac{d_{ij}}{m}\right)^2\right]^2 \quad (6)$$

where u_i and v_i are the longitude and latitude coordinates of centroid i , $w_j^{spatial}(u_i, v_i)$ is a spatial weight for the observation j with centroid i of the group; d_{ij} is the geographical distance between observation j and the centroid i ; and m is the maximum distance between the centroid j and the farthest observation in the group. Similarly, the observations are weighted based on their temporal distance as follows using the similar Bi-square Kernel Weighting function (193; 214):

$$w_j^{temporal}(t_i) = \left[1 - \left(\frac{t_{ij}}{m}\right)^2\right]^2 \quad (7)$$

where $w_j^{temporal}(t_i)$ is the temporal weight for observation j of a sub-group; t_{ij} is the temporal distance between observation j and the centroid i in the group; and m is the maximum distance between the centroid j and the farthest observation in the group. Note that (t_i) represents any time information and in this study, the basis of the temporal variation is year. Finally, the spatial and temporal weights are multiplied to calculate the spatiotemporal weights using equation (8) (193; 214):

$$w_j^{st}(u_i, v_i, t_i) = w_j^{spatial}(u_i, v_i) \times w_j^{temporal}(t_i) \quad (8)$$

where w_j^{st} is the spatiotemporal weight of the observation j located in coordination (u_j, v_j) and time (t_i) .

Given the spatiotemporal weights for all observations, those with larger weights will be selected to form a sub-group for a target observation (centroid) with respect to a given bandwidth. Afterward, observations belong to a sub-group are re-weighted only based on the sub-group samples. Eventually, a local truncated regression (second stage) with 2SRI treatment (first stage) by incorporating the assigned spatiotemporal weights is estimated for all sub-groups. GTWR generates a local set of coefficients for each target observation. In order to examine whether the

spatiotemporal variations of coefficients for all variables are noticeable, the outputs from GTWR and the global model are compared by a Non-stationary Test. The Non-stationary Test was defined as follows (193; 214):

$$\Delta = \beta^{3^{rd}Q} - \beta^{1^{st}Q} \begin{cases} > 1.96(S.E.) \text{ and } \max|z| > 1.96, & \text{Non-stationary Test is passed} \\ \text{otherwise,} & \text{Non-stationary Test is failed} \end{cases} \quad (9)$$

where Δ is an interquartile range (the difference between the 75th and 25th percentile) of coefficient β , Q is quantile, $S.E.$ represents the standard error of the coefficients from the global model, and $|z|$ is the z-value of the coefficients in the local model. A passed Non-stationary Test for a coefficient indicates that the variation of the coefficient over space and time is substantial, otherwise, the coefficient is stationary. For more information about the GTWR method, please refer to (193; 214). STATA 16.1 and R packages were used to carry out the analysis in this study.

Model Selection Approach

To draw a comparison between local and global models, the log-likelihood of the estimated models, Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), and McFadden Pseudo R-squared are used. These measures can be calculated using Equations 5, 10, 11, and 12 respectively.

$$AIC = 2K - 2L \quad (10)$$

$$BIC = K \cdot \ln(n) - 2\ln L \quad (11)$$

$$McFadden's \text{ Pseudo } R^2 = 1 - \frac{\ln L}{\ln L_{intercept}} \quad (12)$$

where n is the number of observations, k is the number of regressors, L is the log-likelihood of the estimated model, and $\ln L_{intercept}$ is the likelihood value of the model without regressors.

Data

This study utilizes the Fatality Analysis Reporting System (FARS) data from 2015 to 2018 for all U.S. states (240). FARS dataset collects crashes from the US Department of Transportation National Highway Traffic Safety Administration's National Center for Statistics and Analysis in which at least one person dies within 30 days from the time of the crash. The FARS dataset typically provides deeper crash investigations of fatal crashes. This study focuses on the fatality outcomes of one-pedestrian one-vehicle crashes, and other types of crashes are excluded from the dataset. Additionally, the instant death at a crash scene where pedestrian time-to-death equals zero is excluded from the dataset because these observations do not correlate with EMS response time. The total number of crashes after cleaning the data from missing values and error-checking the data by examining descriptive statistics is 4,983 pedestrian fatalities in 4 years. Figure 6-2 demonstrates the number of observations in each U.S. state. Note that Indiana is excluded from the dataset as all the observations of this state have missing values. Among the variables available in the data, those that were frequently used in previous studies were selected and are categorized into pre-crash and post-crash characteristics, which will be discussed in this section.

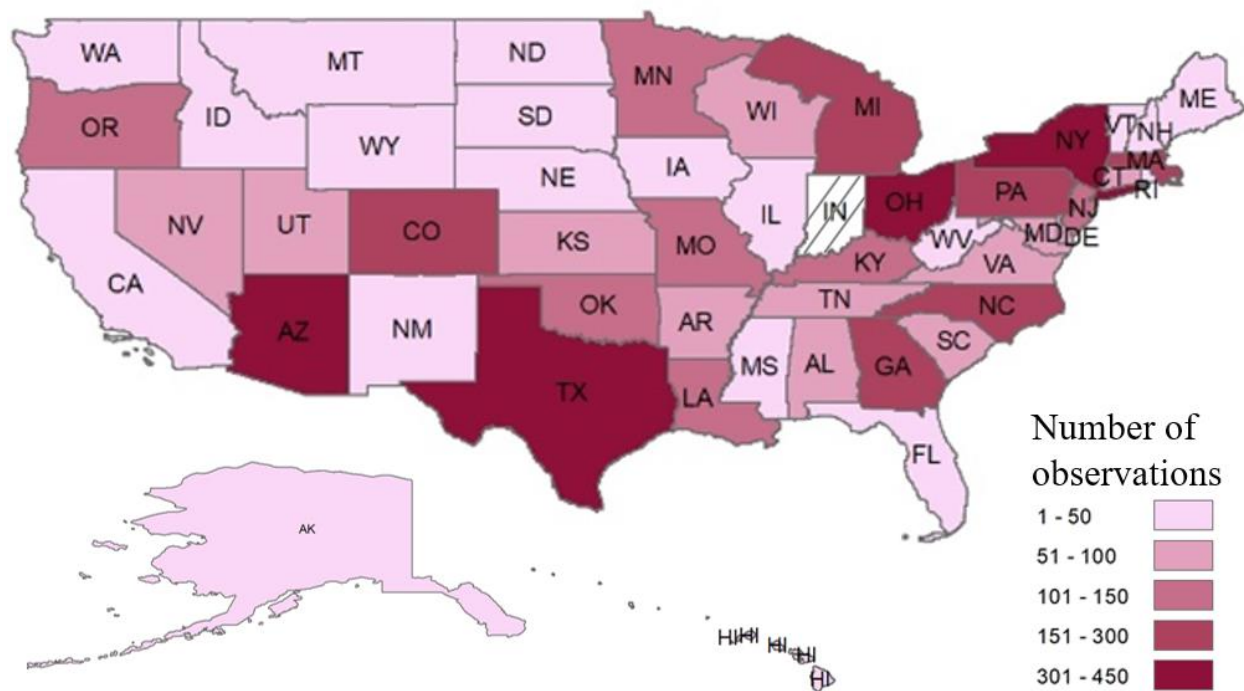


Figure 6-2 Number of observations in each U.S. state

Pre-crash Variables

Pre-crash variables are categorized into environmental, driver, pedestrian, and roadway characteristics. The descriptive frequencies of the pre-crash variables are presented in Table 6-1. For environmental characteristics, crash time of day, lighting conditions, and visibility conditions are selected. Crash time is divided into 5 categories: morning peak, mid-day, afternoon peak, evening, and late night. Lighting conditions are divided into daylight, dark with no lighting, and dark with lighting. Finally, two types of pedestrian visibility conditions are available: good visibility and poor visibility. Two variables are selected to represent driver behavior characteristics: driver hit and run action and driver intoxication.

Pedestrian characteristics are categorized into 4 variables: age, gender, pedestrian intoxication, and behavior. Pedestrian age is divided into 3 categories: under 30, between 30 and 60, and above 60. Finally, 8 pedestrian behavior categories were available in the data: working or playing in a roadway, walking or running along a roadway, dash and/or dart-out, crossing a roadway while the vehicle is turning, crossing a roadway while the vehicle is not turning, crossing an expressway, waiting to cross, and other behavior.

In terms of roadway characteristics, roadway functional classification, speed limit, and crash location are the selected variables. The roadway functional classification includes interstates, principal arterials, minor arterials, collectors, and local roadways. Speed limits are categorized into 3 categories: under 30 mph, 35 and 40, and above 45 mph. Crash location includes crashes at an intersection, not-intersection, and unknown locations.

Post-crash Variables

EMS response time and pedestrian time-to-death are the post-crash variables. EMS response time is the difference between the crash time and the arrival time of the EMS personnel at the crash scene. This variable is an integer variable with a minimum of one minute and a maximum of 713 minutes (11 hours and 53 minutes). Likewise, pedestrian time-to-death is an integer variable that is determined based on the difference between the crash time and the pedestrian time of death. This variable has a minimum of 2 and a maximum of 43,119 minutes. The descriptive statistic of the EMS response time and the pedestrian time-to-death are presented in Table 6-2.

Data Transformation

The EMS response time and pedestrian time-to-death (TTD) variables exhibit a large range of values. These variables also have a right-skewed distribution (Figure 6-3). A natural logarithm transformation is applied to normalize the EMS response time and pedestrian time-to-death variables. Since a truncated regression is utilized to estimate the pedestrian time-to-death, the TTD should have a truncated normal distribution. Table 6-2 presents the descriptive statistics of the transformed variables. The average response time is 10.70 minutes but it is 41% and 39% higher in a rural area (14.07) compared to an urban area (10.01), and for a hit and run crash (13.93), respectively. Figure 6-4 illustrates the scatterplot of EMS response times vs pedestrians' time-to-death. As can be seen in some of the extreme cases, it took hours for the EMS personnel to respond.

Table 6-1 Variable statistics-FARS dataset

Variable (N = 4983 crashes with a death in each crash)		Frequency	Percent
(Environmental characteristics)			
Crash Time of Day	Morning peak	547	11
	Mid-day	592	12
	Afternoon peak	850	17
	Evening	1886	38
	Late-night	1108	22
Lighting Condition	Daylight	1208	24
	Dark	1575	32
	Dark lighted	2200	44
Visibility	Good visibility	4120	83
	Poor visibility	863	17
(Driver characteristics)			
Hit and Run	No	4157	83
	Yes	826	17
Driver Intoxication	No	4639	93
	Yes	344	7
(Pedestrian characteristics)			
Pedestrian Age	<30	1114	22
	30-60	2330	47
	>60	1539	31
Pedestrian Gender	Male	3457	69
	Female	1526	31
Pedestrian Intoxication	No	1518	31
	Yes	1066	21
	Unknown	2399	48
Pedestrian Behavior	Working or playing in roadways	56	1
	Walking/running along roadways	600	12
	Dash/dart-out	539	11
	Crossing roadway-vehicle not turning	2108	42
	Crossing roadway-vehicle turning	270	6
	Crossing expressway	151	3
	Waiting to cross	308	6
	Others	951	19
(Roadway characteristics)			
Crash Location	At intersections	1531	30
	Not at intersections	3433	69
	Unknown	19	1
Functional Roadway System	Interstate	314	7
	Principal arterial	2140	43
	Minor arterial	1165	23
	Collector	509	10
	Local road	855	17
Speed Limit (MPH)	<=30	1076	22
	35-40	1709	34
	>=45	2198	44

Table 6-2 Descriptive statistics of transformed variables

Variable (N=4983)	Min	Mean	Max	Standard Deviation
Time-to-Death (Minute)	2	2117	43119	5648.284
Transformed Time-to-Death ($\text{Log}_e\text{-TTD}$)	0.693	4.865	10.672	2.381
EMS Response Time (Minute)	1	10.700	713.0	27.139
Transformed EMS Response Time ($\text{Log}_e\text{-EMS}$)	0	1.945	6.569	0.764

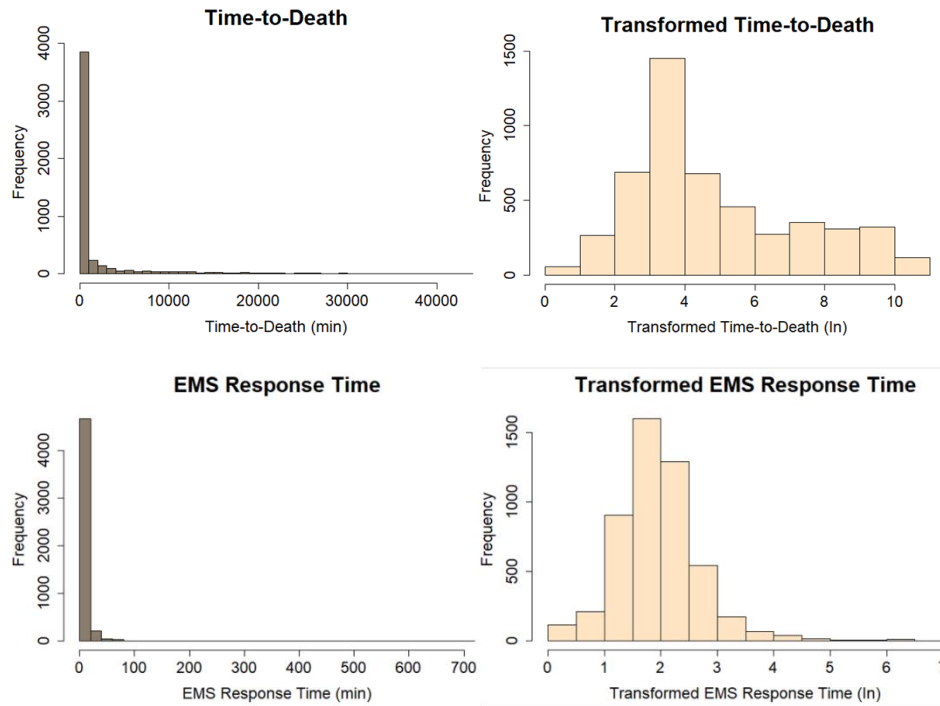


Figure 6-3 Distribution of TTD and EMS response time before and after the transformation

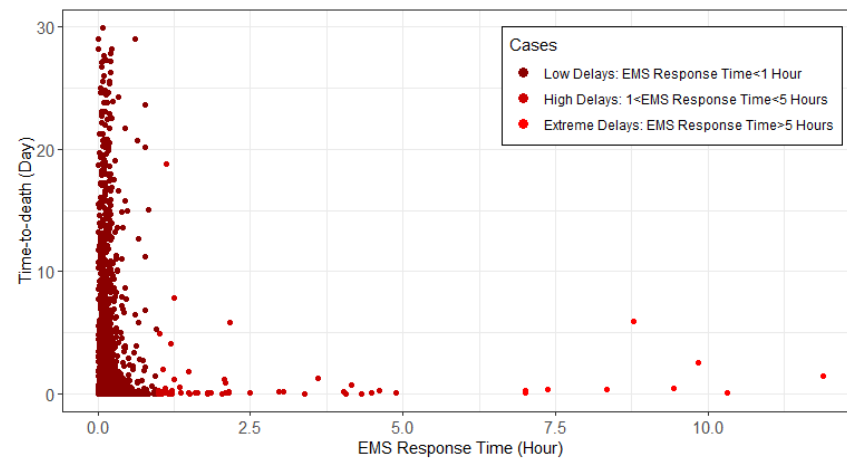


Figure 6-4 Scatterplot of pedestrian time-to-death vs EMS response time

Results

First Stage: 2SRI Endogenous Treatment Results

Table 6-3 presents the linear regression results for the endogenous variable (EMS response time) along with the descriptive statistics of the residuals in stage 1. The model shows that log-EMS response time is significantly correlated with crash time, lighting conditions, and the roadway functional classification. The results also show that a fatal pedestrian crash in dark and not-lighted conditions compared to daylight and dark lighted conditions, and interstates compared to other roadway functional classifications are expected to have larger EMS response times. Likewise, late-night hours and morning peak hours have higher EMS response times compared to other daytime. EMS response times for crashes on local roads, collectors, minor arterials, and principal arterials respectively decrease by about 25%, 19%, 29%, and 31% compared to interstates that usually have the highest posted speed limits.

Second Stage: Truncated Regression Results (global model)

Table 6-4 presents the results of the truncated regression model (the global model in stage 2). The results in the table can be interpreted as a change, with a positive sign for increasing and a negative sign for decreasing the natural logarithm of pedestrian time-to-death, with respect to the base categories for the categorical exogenous variables. The model shows that a pedestrian time-to-death after a fatal injury crash is significantly correlated with the EMS response time with a reasonable negative sign. Note that the sign was positive without the 2SRI treatment due to the endogeneity limitation. The residual from the EMS response time model (stage 1) is used as a regressor in the model to control for the endogeneity bias for the effect of variables on the pedestrians' survival time. Note that the significance of residual term shows evidence for endogeneity and different specifications were tested to inspect the coefficient on the residuals regressor. The coefficients were relatively stable across these specifications. The model with the best fit was selected. It can be deduced that for any 1% increase in EMS response time, pedestrian TTD reduces by 0.85%. The marginal effect of EMS response time at means implies that a 1-minute increase in the mean of EMS response time reduces pedestrians' survival time by 4.4 minutes. In other words, higher EMS response times are correlated with an increased likelihood of earlier death for pedestrians in car crashes.

Regarding pre-crash variables in terms of pedestrian characteristics, the results show that the likelihood of early death is increased for pedestrians above the age of 30. Likewise, intoxicated pedestrians compared to unintoxicated pedestrians and female pedestrians compared to male pedestrians experience reduced time-to-death. More remarkably, in terms of pedestrian behavior, it can be deduced that waiting to cross, working or playing in roadways, and crossing expressways are relatively the most detrimental and lethal behavior.

Considering roadways characteristics, crashes on roadways with higher speed limits have significantly reduced pedestrian survival time. Similarly, crashes at intersections have a lower likelihood of early death compared to other locations. For driver and environmental characteristics, the model indicates that intoxicated drivers and pedestrians in poor visibility conditions decrease pedestrian survival time. Additionally, hit and run behavior by drivers is correlated with significant reductions in the survival time of pedestrians.

Table 6-3 Linear regression estimates for log-EMS response time (Stage 1)

Variable (N=4983)	β	Standard Error	t-value	P-value	
Constant	2.524	0.046	54.813	0.000	
Crash Time (base: Late Night)					
Morning Peak	-0.086	0.043	-1.988	0.047	
Mid-day	-0.121	0.054	-2.264	0.024	
Afternoon Peak	-0.128	0.037	-3.497	0.000	
Late-evening	-0.173	0.028	-6.082	0.000	
Lighting Condition (base: Dark Not Lighted)					
Daylight	-0.196	0.041	-4.757	0.000	
Dark-lighted	-0.257	0.025	-10.332	0.000	
Roadway Functional classification					
(base: Interstate)					
Principal Arterial	-0.365	0.045	-8.053	0.000	
Minor Arterial	-0.338	0.048	-7.081	0.000	
Collector	-0.207	0.054	-3.846	0.000	
Local	-0.292	0.050	-5.877	0.000	
F-statistic: 26.3, p-value: < 2.2e-16					
Multiple R-squared: 0.050, Adjusted R-squared: 0.048					
Residuals:	Min	Mean	Median	Max	Standard Deviation
	-2.396	0.000	-0.030	4.780	0.745

Table 6-4 Truncated with 2SRI treatment regression model

Variables (N = 4983)	β	t-value	P-value	Exp(β)
Constant	9.400	13.066	0.000	12088.38
Logarithm of EMS	-0.858	-2.482	0.013	0.424
Residual	1.151	3.304	0.001	3.161
<i>Pedestrian Characteristics</i>				
Age (base: <30)				
30-60	-0.404	-3.577	0.000	0.668
>60	-0.287	-2.297	0.022	0.751
Gender (base: Male) Female	-0.390	-4.090	0.000	0.677
Intoxication (base: No)				
Yes	-0.391	-3.086	0.002	0.676
Unknown	-0.111	-1.112	0.266	0.895
Behavior (base: Crossing, vehicle turning)				
Working or playing in roadway	-1.633	-3.567	0.000	0.195
Walking/running along roadway	-1.322	-5.271	0.000	0.267
Dash/dart-out	-0.992	-4.082	0.000	0.371
Crossing, vehicle not turning	-1.319	-6.201	0.000	0.267
Crossing expressway	-1.329	-3.646	0.000	0.265
Waiting to cross	-1.694	-6.172	0.000	0.184
Others	-1.395	-6.108	0.000	0.248
<i>Roadway Characteristics</i>				
Speed limit (base: <=30 mph)				
35-40	-0.420	-3.373	0.001	0.657
>=45	-1.115	-8.568	0.000	0.328
Crash Locations (base: At intersections)				
Not at intersections	-0.215	-2.079	0.038	0.807
Unknown	-0.414	-0.583	0.560	0.661
Functional Classification (base: Local and collector)				
Interstate	-0.592	-2.376	0.018	0.553
Arterial	-0.441	-0.402	0.688	0.643
<i>Driver Characteristics</i>				
Hit and Run (base: No) Yes	-0.417	-3.471	0.000	0.659
Driver Intoxication (base No) Yes	-0.342	-1.979	0.048	0.710
<i>Environmental Characteristics</i>				
Visibility (base: Good) Poor	-0.346	-2.914	0.004	0.708
Crash Time (base: off-peak midday) Other	-0.265	-2.300	0.021	0.767
Sigma	2.631	63.094	0.000	13.888
Log-Likelihood	-10991.36			
AIC	22034.73			
R-squared	0.067			
Adjusted R-squared	0.063			
Pseudo R-squared	0.029			

Table 6-5 presents the results of the geographically and temporally weighted truncated regression, the GTWTR model (the local model in stage 2), and the results of the non-stationary test. Non-stationarity tests revealed that all the variables passed the test and have a substantial spatiotemporal variation. In general, the sign and magnitude of the mean of each coefficient are as expected and are consistent with the global model. Model performance measures including R-squared measures, Log-likelihood, and AIC indicate that the GTWTR model outperformed the global truncated regression.

Second Stage: Geographically and Temporally Weighted Truncated Regression Results (local model)

Figure 6-5 illustrates the spatiotemporal variations of the coefficient of the EMS response time variable. The colors in each sub-figure represent the magnitude of the coefficient. The center map shows the local coefficient averaged across all years (2015-2018) for each State and the sub-figures show the local coefficient within one specific year. According to the map in the center, a 1% delay in EMS response time in some states in the Northeastern, Midwest, and West Coast states reduces pedestrian time-to-death more drastically. The spatiotemporal variations of the coefficient of the EMS response time variable illustrated in Figure 6-5 indicate in which state and in which year, a delay in EMS response time was even more likely to associate with early death in crashes. In addition to the spatial variation, it can be observed that there is a substantial temporal variation of the association between EMS response times and pedestrian survival times between 2015 and 2018 across the states. For instance, in 2015 and 2016, the strongest associations between EMS response times and pedestrian survival times are in some states in the east, while in 2017, the strong associations are shifted toward states in the west and central part of the U.S. However, these variations can be due to some unobserved factors related to socio-economic, traffic, and built environment conditions. Although this study reveals these temporal variations, further studies need to be carried out to investigate the reasons behind the temporal variations to adopt better strategies for the safety improvement of pedestrians in the future.

Figure 6-6 demonstrates the spatial variation of the coefficients averaged across all years associated with each pedestrian behavior. Likewise, Figure 6-7 illustrates some of the other variables. Considering the spatial variation of the coefficients of pedestrian behaviors, it can be observed that the greater values for the estimates of variable Crossing Expressway are concentrated in the West Coast, West (Rocky Mountains), and Southwest of the U.S., while in the central and eastern areas, it has a smaller impact on pedestrians' survival time. Moreover, whether pedestrians walk or run along roadways has a stronger impact on pedestrians' survival time in Southwest and South. Furthermore, the association of Dash/Dart-out behavior with pedestrians' survival time is stronger in the West Coast, Southwest, and South. Similarly, based on Figure 6-7, the variable Driver Hit and Run has a higher correlation with the survival time of pedestrians in the West and Northeast of the U.S. It can be because of some unobserved factors related to the cultural characteristics of these regions. Furthermore, in some states in the East Coast area and the West of the U.S., Pedestrian Intoxication has a stronger impact on pedestrians' survival time.

Table 6-5 Geographically and temporally weighted truncated with 2SRI treatment regression model

Variable (N=4983)	Meanβ	Min β	1stq. β	3rdq. β	Max β	Max Z 	Δ	Test
Constant	8.605	2.053	7.529	9.524	14.217	8.079	1.994	Yes
Logarithm of EMS	-0.711	-3.787	-1.186	-0.150	2.452	3.439	1.036	Yes
Residual	0.962	-2.327	0.481	1.410	3.991	3.668	0.929	Yes
<i>Pedestrian Characteristics</i>								
Age (base: <30)								
30-60	-0.204	-0.981	-0.393	0.009	0.628	3.032	0.402	Yes
>60	-0.224	-0.975	-0.494	-0.002	0.657	3.066	0.491	Yes
Gender (base: Male) Female	-0.320	-1.281	-0.521	-0.131	0.504	3.704	0.389	Yes
Intoxication (base: No)								
Yes	-0.272	-1.199	-0.392	-0.104	0.456	3.508	0.288	Yes
Unknown	-0.256	-2.231	-0.526	0.017	0.678	3.417	0.544	Yes
Behavior (base: Crossing, vehicle turning)								
Working or playing in roadway	-0.909	-3.250	-1.566	-0.339	1.594	3.866	1.227	Yes
Walking/running along roadway	-0.976	-2.410	-1.349	-0.539	0.186	3.960	0.811	Yes
Dash/dart-out	-0.743	-2.114	-1.196	-0.290	0.760	4.099	0.907	Yes
Crossing, vehicle not turning	-1.045	-3.247	-1.498	-0.520	0.781	3.805	0.978	Yes
Crossing expressway	-1.218	-3.399	-1.922	-0.488	1.342	4.123	1.434	Yes
Waiting to cross	-1.278	-3.401	-1.852	-0.690	0.368	4.818	1.162	Yes
Others	-0.124	-2.432	-1.056	2.426	2.623	27.176	3.482	Yes
<i>Roadway Characteristics</i>								
Speed limit (base: <=30 mph)								
35-40	-0.451	-1.649	-0.856	-0.137	0.759	4.705	0.719	Yes
>=45	-1.299	-3.391	-1.581	-0.810	0.087	5.865	0.771	Yes
Crash Locations (base: At intersections)								
Not at intersections	-0.270	-0.901	-0.427	-0.103	0.351	3.223	0.324	Yes
Unknown	-0.201	12.282	-0.836	0.775	3.118	2.106	1.610	Yes
Functional Classification (base: Local and collector)								
Interstate	-0.382	-2.228	-0.849	0.123	1.867	3.545	0.972	Yes
Arterial	-0.224	-1.034	-0.458	-0.011	0.453	3.187	0.447	Yes
<i>Driver Characteristics</i>								
Hit and Run (base: No) Yes	-0.270	-0.955	-0.474	-0.067	0.412	3.135	0.407	Yes
Driver Intoxication (base No) Yes	-0.290	-1.931	-0.455	-0.077	0.635	3.541	0.378	Yes
<i>Environmental Characteristics</i>								
Visibility (base: Good) Poor	-0.349	-1.090	-0.607	-0.119	0.324	3.380	0.488	Yes
Crash Time (base: off-peak midday) Other	-0.326	-1.108	-0.565	-0.065	0.529	3.653	0.500	Yes
Sigma	4.108	2.210	2.371	7.450	11.178	27.827	5.079	Yes
Log-Likelihood	-10939.4							
AIC	21928.84							
R-squared	0.114							
Adjusted R-squared	0.110							
Pseudo R-squared	0.040							

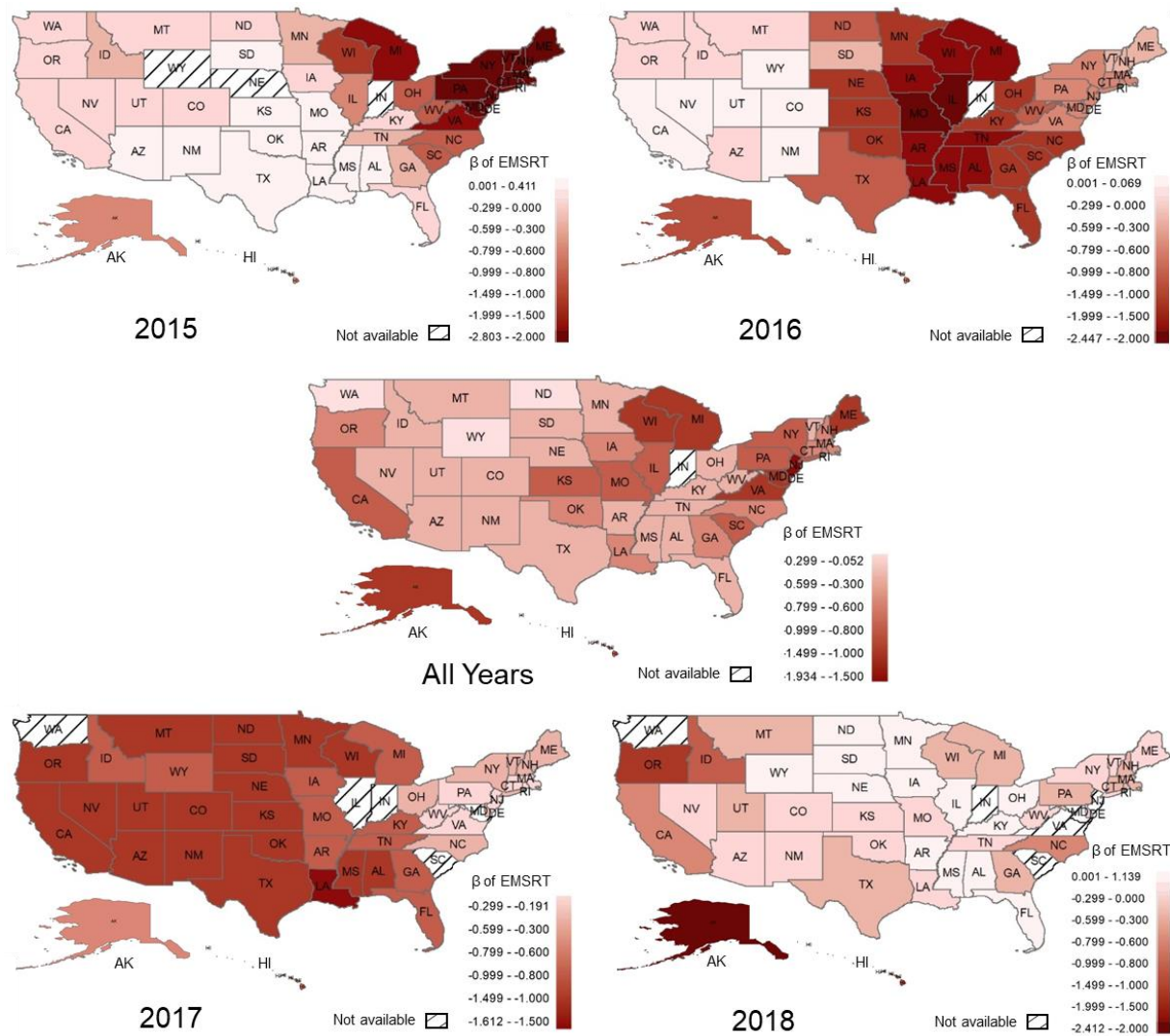


Figure 6-5 Spatiotemporal variation of the coefficient of EMS response time estimated by the GTWTR model.

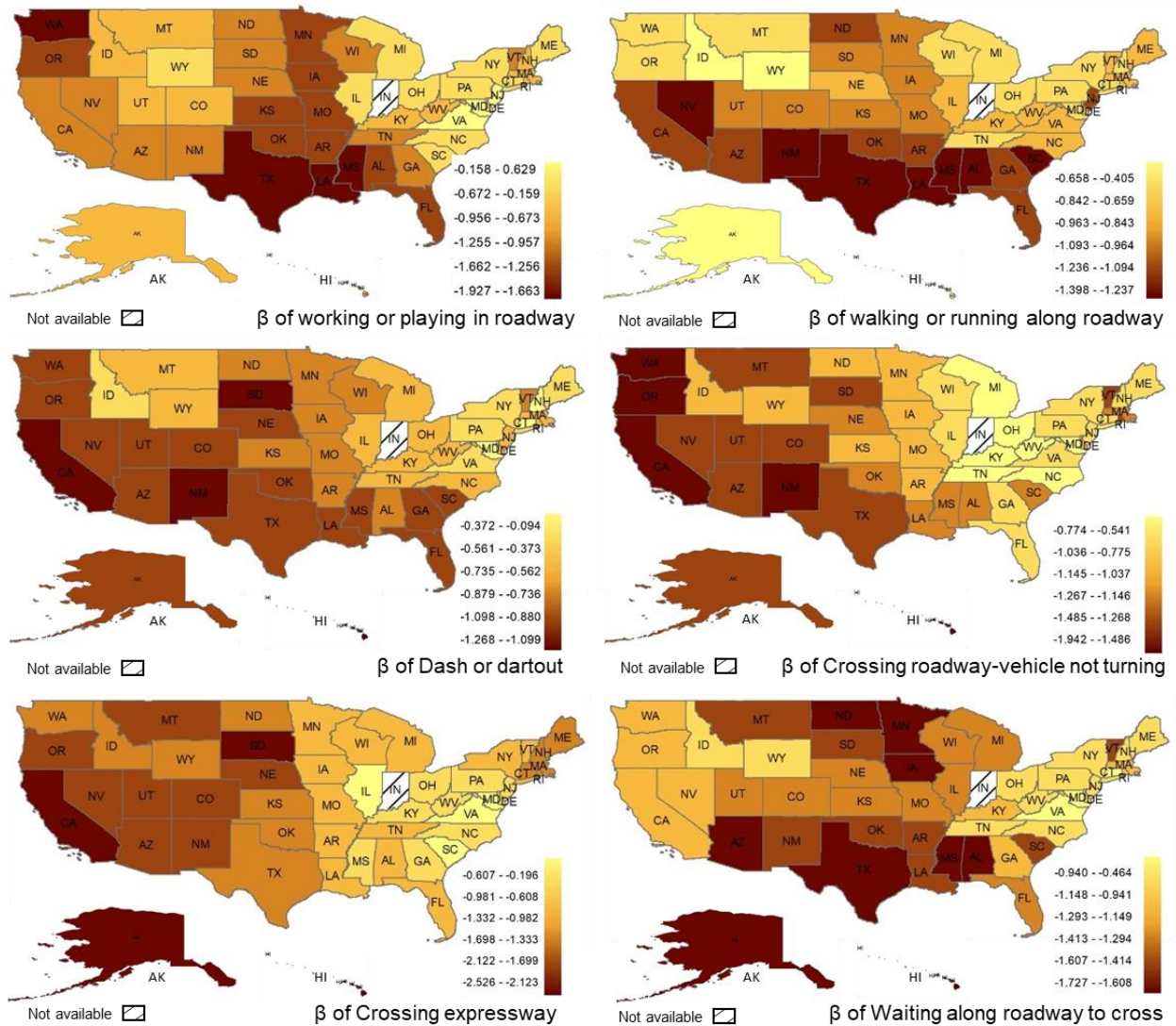


Figure 6-6 Spatial variation of the coefficient of pedestrian behaviors averaged across all years estimated by the GTWTR model.

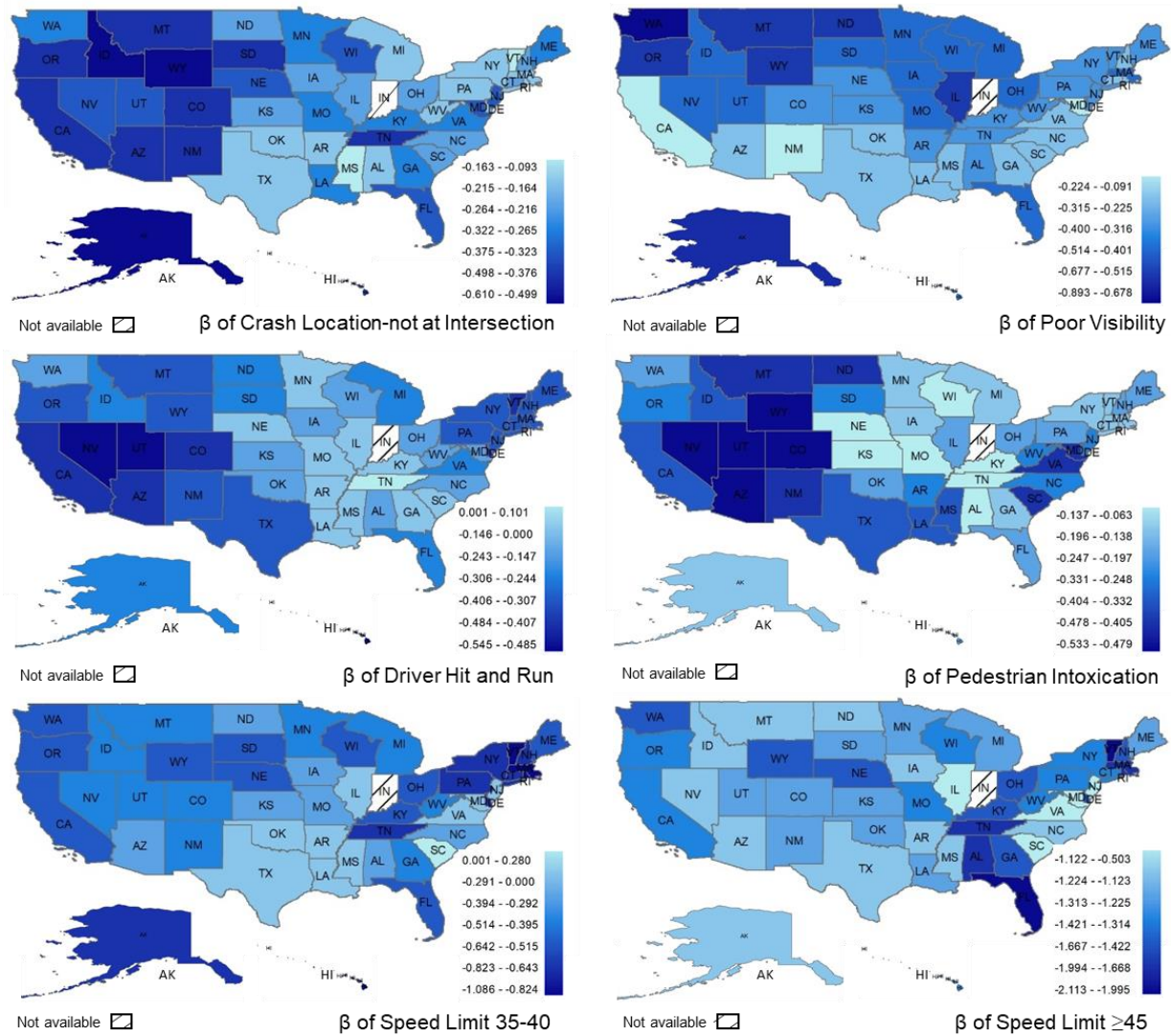


Figure 6-7 Spatial variation of some coefficients averaged across all years estimated by GTWTR model.

The spatial variation of the local parameter estimates at a state-level delivers valuable information about which areas interventions may be needed regarding the studied factors and their associations with pedestrians' survival time. Safety practitioners can identify states and regions to prioritize safety improvements based on critical variables contributing to pedestrians' survival time. Furthermore, future studies can focus on specific variables in order to find out the reason behind the stronger impact of some variables on pedestrians' survival time in some regions. This can help practitioners to identify and apply appropriate countermeasures to improve the safety of pedestrians in these regions.

Discussion and Limitations

This study explores the survival time (time-to-death) of pedestrians involved in fatal injury crashes and extracts associations with different pre-crash and post-crash factors, specifically pedestrian behavior and EMS response time with pedestrian survival time. The empirical results from the estimated model show that a delay in an EMS response time significantly reduces the survival time of pedestrians in fatal crashes. This finding is consistent with previous studies in which EMS response times with an early response were shown to potentially increase the survival time of car passengers in a traffic crash (215; 216; 228; 230). EMS response times for crashes on local roads, collectors, minor arterials, and principal arterials respectively decrease by about 25%, 179%, 29%, and 31% compared to interstates, which usually have the highest mobility and posted speed limits. Reasons for discrepancies in EMS response times may be related to a greater travel time of EMS personnel to arrive at crash scenes on interstates due to greater distances between emergency medical centers and interstates. Also, EMS response times are significantly higher during late-night hours compared to other hours, most likely due to longer notification times during the late night. The average response time is 41% and 39% higher in a rural area compared to an urban area and for a hit-and-run crash.

Moreover, the results show that many pre-crash factors in different categories including pedestrian, roadway, environmental, and driver characteristics contribute to pedestrian survival time. The findings regarding pedestrian characteristics are consistent with previous studies in which pedestrian behaviors and violations (193), increasing pedestrian age (218), and intoxication from drugs (193) and alcohol (193; 219) can adversely increase crash severity and result in reductions of pedestrian survival time. Likewise, considering roadway characteristics, previous studies have found that crash location significantly influences crash severity (215; 224; 241). The results of this study show that crash locations at intersections increase the time-to-death of pedestrians compared with non-intersections and other locations. Previous studies have also shown that pedestrians involved in crashes at intersections seem to be less at risk of severe injury or death compared to crashes at non-intersection locations (193; 242). One possible explanation may be the behaviors of pedestrians and vehicle drivers when approaching intersections. Motorists usually behave more cautiously at intersections as they expect pedestrian and vehicle crossings and are more prone to slow down. Likewise, modeling results show that some pedestrian behaviors are relatively more detrimental and lethal, such as working or playing in roadways and crossing expressways. These behaviors bring pedestrians into conflict with vehicles in situations where motorists may not expect pedestrians. Concerning roadway characteristics, in previous studies, it was found that crashes with higher impact speeds (200; 208) are associated with a higher

propensity for fatal crashes, and this study reveals that these factors reduce pedestrian survival times in fatal injury crashes.

The results from the GTWTR model indicate that the model coefficients vary spatially and temporally. The EMS response time coefficient varies across time and space from -3.79 to 2.45. This variation could be due to unobserved heterogeneity that could be captured by unobserved factors such as pedestrian pre-existing medical conditions or types of pedestrian damage in car crashes that are not available in the data.

This study is not without limitations. One limitation of this study is the low number of observations with emergency medical system information, which reduced the total number of observations in the dataset from 17,000 observations to 4,983 observations. Additionally, FARS data does not have some relevant post-crash variables that can affect pedestrian survival time such as type of damage, underlying pre-existing medical conditions and medical information, distance from crash scenes to nearby hospitals, and treatment offered to pedestrians.

Conclusions

Fatalities and severe injuries among vulnerable road users, particularly pedestrians, are rising. While pedestrian-involved fatal crashes are sometimes assumed to be similar in terms of severity, crashes that cause instant death are substantially more severe than crashes resulting in death several days afterward. Instead of homogenizing all pedestrian fatal crashes as the same, this study extracts valuable information from fatal crashes by examining pedestrians' survival time ranging from early death to death within 30 days. To that end, this study utilizes a spatiotemporal method with a two-stage residual inclusion treatment to model pedestrians' survival time to overcome the endogeneity limitation and account for unobserved heterogeneity. The results indicate that while capturing the unobserved heterogeneity, the local GTWTR model outperformed the conventional global truncated regression model based on improving the log-likelihood, AIC, BIC, and different estimated R-squared measures.

This study shows that the associations of variables with pedestrian survival time substantially vary across space and over time, due to the unobserved heterogeneity that associates with location and time. The results from the GTWTR model deliver valuable information about which areas and at what time interventions may be needed regarding the role of EMS response time or other studied factors. Empirical results show that a delay in EMS response time significantly decreases pedestrian survival time in fatal injury crashes. It can be deduced that for a 10% increase in EMS response time, on average, pedestrian time-to-death reduces by 7.9%. However, this effect varied spatially and temporally. For example, one minute delay in EMS response time in some states in the North East, Midwest, and on the West Coast reduces pedestrian time-to-death more intensely compared to the other states. The reasons for that may relate to several factors such as different socioeconomic factors, racial groups, and spatial patterns of health conditions that can be studied in the future. Also, the results show that the average response times and the likelihood of early death are higher in hit and run crashes. It is shown that the variable Driver Hit and Run has a higher correlation with the survival time of pedestrians in the West and Northeast of the U.S. It can be because of some unobserved factors related to the cultural characteristics of these regions that can be studied in the future. The findings of this study can

provide safety practitioners with some key factors that have the potential to save pedestrian lives, specifically through faster pedestrian crash detection and emergency response. Safety practitioners can identify states and regions to prioritize safety improvements based on critical variables contributing to pedestrians' survival time. This can help practitioners to identify and apply appropriate countermeasures to improve the safety of pedestrians in these regions.

Consistent with the literature, this study shows that the effect of the key factors on the survival time of pedestrians in fatal injury crashes is consistent with their effect on the severity of pedestrian crashes particularly in terms of fatal vs. non-fatal outcomes. By understanding which factors influence the survival time of pedestrians in fatal injury crashes and applying proper countermeasures, it is expected that a pedestrian victim has a higher chance to avoid instantaneous death at the crash scene and has a higher chance of survivability when provided proper medical services. However, due to the limitation of the data that only includes fatal injury crashes, these results cannot be directly extrapolated to non-fatal crashes or provide insights on how to reduce pedestrian crash severity in terms of converting fatal crashes to non-fatal crashes. To understand the direct influence of the studied factors on both pedestrian survival time and survivability, future studies should create a more comprehensive dataset that encompasses both fatal and non-fatal pedestrian crashes. Moreover, the findings of this study cannot be generalized widely because of the distinctive characteristics of the United States. Future studies should also consider more post-crash factors such as medical information, type of damage, and the treatment offered to pedestrians. They could also consider other statistical instrumental approaches for the endogeneity treatment. Furthermore, analyzing different temporal aspects such as the time of day, month, or season remains for future studies.

ACKNOWLEDGMENT

The authors would like to thank the following entities for financial support for the research, authorship, and/or publication of this article: the Collaborative Sciences Center for Road Safety, the U.S. Department of Transportation National University Transportation Center promoting safety, and the Tennessee Department of Transportation, TDOT.

CHAPTER 7: CAN FASTER CRASH NOTIFICATION TIME INCREASE SURVIVAL TIME OF BICYCLISTS IN FATAL CRASHES: APPLICATION OF EXPLAINABLE MACHINE LEARNING

A version of this chapter is presented at the 100th Transportation Research Board Conference.

Mahdinia, I., Mohammadnazar, A., & Khattak, A. J. (2021). Can faster crash notification time increase survival time of bicyclists in fatal crashes: application of explainable machine learning. Under review in the Journal of Safety Research.

Abstract

Introduction: Over recent years, the fatality of vulnerable road users has been rising considerably. Bicyclists have been recognized as one of the most at-risk road users. A better understanding of fatal bicyclist crashes will help safety practitioners to reduce fatalities. While bicyclist-involved fatal crashes are often assumed to be similar in studies, instant death in a fatal crash is substantially more severe than death caused by a crash several days afterward. This study extracts valuable information from fatal bicyclist crashes by analyzing bicyclist time-to-death ranging from instant death to death within 30 days using the Fatality Analysis Reporting System (FARS) dataset between 2015 and 2019. This study categorizes variables based on the Haddon Matrix in terms of pre-crash, during-crash, and post-crash and uses crash notification time as the key post-crash measure that can affect bicyclist survival time. **Methods:** To extract associations of several variables with bicyclist time-to-death, an explainable XGBoost regression model is developed and a novel application of the SHAP technique is demonstrated to interpret the results. **Results:** The SHAP results show that a large delay in crash notification time, as a fundamental part of Emergency Medical Service (EMS) response time, considerably reduces bicyclist time-to-death and increases the likelihood of early death in fatal crashes. **Conclusions:** the average crash notification time for hit-and-run crashes, for crashes in rural areas, and crashes during late-night hours are significantly higher than non-hit-and-run crashes, urban areas, and other hours respectively. These situations are not rare and they may increase crash notification delay if there are no witnesses or survivors capable of notifying emergency responders. **Practical Applications:** To save cyclists' lives and reduce delays in crash notification times, on-road vehicle technologies such as the advanced automatic collision notification system need to be widely utilized to inform emergency responders immediately after a crash.

Keywords: Bicyclist Time-to-Death, Crash Notification Time, Explainable Machine Learning, XGBoost, SHAP Value.

Introduction

Bicyclists have been recognized as one of the most at-risk road users because of their low level of protection. According to the United State National Highway Transportation Safety Administration, 33,654 people were killed in traffic accidents in 2018 with 857 of them being pedal cyclists. Pedal cyclist fatalities have been increased by about 38 percent between 2010 and 2018 (210). This statistic raises concerns about the safety of bicyclists. In recent years, many studies have been conducted to recognize and understand factors contributing to bicyclist-vehicle crash severity. But researchers and safety practitioners usually categorize crashes into fatal and non-fatal and assume that all fatal crashes are similar. However, instant death in a fatal crash is generally substantially

more severe than death several days afterward. In this study, instead of homogenizing all fatal bicyclist crashes as the same, this study takes into account the severity of fatal crashes as a timeline based on the time-to-death (survival time) of bicyclists ranging from instant death to death within 30 days of the crash. To that end, the Fatality Analysis Reporting System (FARS) dataset from 2015 to 2019 is utilized.

The emergency medical service (EMS) response time is a key post-crash factor that can affect the traffic crash victim's survival time (215; 216) and the crash injury severity (243). The EMS response time is defined as the difference between crash time and the arrival time of EMS personnel. It comprises the period between crash time and the time that emergency medical services are notified, the notification time, and the period between notification time and the arrival time of EMS personnel at the crash scene. Some specific situations may increase the notification time such as hit-and-run crashes, crashes at late night hours, or crashes in rural areas, which can significantly delay the arrival of EMS personnel. That said, it is imperative to investigate such situations that potentially can increase crash notification times. Several systems have been introduced and investigated to help promptly identify crashes and reduce the notification time and EMS response time such as the automatic crash notification (ACN) system (230; 244; 245), mobile phone proliferation (246), and Connected and Automated Vehicles (247; 248).

In recent years, different machine learning techniques have been introduced. These methods have been broadly utilized in traffic safety due to their great predictive performance, and their ability to consider multiple independent variables, particularly in analyzing crash severity. For example, different Decision Tree methods (249-252), Neural Network (253; 254), Support Vector Machine models (255-257), and the XGBoost model (258) have been used. The machine learning methods have been compared with statistical methods in crash severity analysis (257; 259), and it is revealed that many machine learning methods perform better than statistical models in terms of prediction (257; 259-261). The models can also consider multiple predictors by eliminating possible multicollinearity (262; 263). However, these methods are generally criticized as they mostly perform as a black-box (259), and consequently, they are less explainable and interpretable than traditional statistical methods. However, over the last few years, to explain predictions from complex methods and seek causality inference, several novel techniques have been introduced such as variable importance plots (264; 265), and SHAP value (266). These techniques recently have been widely used in traffic safety (263; 267).

To analyze the bicyclists' time-to-death involved in fatal crashes and extract associations of various factors with bicyclist survival time, an Extreme Gradient Boosting (XGBoost) tree is developed and the SHAP (SHapley Additive exPlanations) value is estimated to interpret the results. XGBoost is a relatively new machine learning approach that was introduced by Chen and Guestrin in 2016 (268). The SHAP value is introduced by Lundberg and Lee in 2017, which captures the importance of a variable by comparing what a complex machine learning model predicts with and without the variable (266). Finally, associations of several pre-crash variables and crash notification time as a key post-crash factor are explored and explained. The implication of this study assists practitioners in traffic safety by recognizing the key factors that can potentially reduce bicyclist fatalities. The rest of the paper is organized as follows: a review of the literature

followed by the methodology. Subsequently, the description of data and variables are presented, followed by the results and discussion. In the final section, the conclusion is presented.

Literature Review

The severity of bicyclists-vehicle crashes has been the focus of many studies in the past (269-272). Many variables contribute to the severity of a bicyclists-vehicle crash in different categories such as accident characteristics including the crash time of day (271), and weekdays or weekend (273), bicyclists characteristics including bicyclists age (271), bicyclists intoxication (274), environmental characteristics such as different weather conditions (271), lighting conditions (269; 271), and roadway characteristics such as land use (275), crash location (274), and speed limit (269).

Most of the previous studies offer useful findings on what factors influence vehicle-bicyclist crash fatality, particularly in the context of fatal vs. non-fatal. The studies that categorized vehicle-bicyclist crashes into fatal vs. non-fatal categories assume that all fatal crashes are similar in terms of severity. However, instant death in a fatal crash is substantially more severe than a death caused by a crash after several days of the crash time (215). In the literature, very few studies consider the fact of the difference between the severity of fatal crashes. Few studies have analyzed the fatal-injury severity of vehicle crashes and showed that the severity of a fatal crash varies based on the survival time of the car passengers after the crash time (215; 229). One of the most important factors found in previous studies is the Emergency Medical Service (EMS) response time. It is revealed that an early response of EMS can potentially increase the survival time of a car crash victim (215; 228; 230). It is found in previous studies that a minute reduction in EMS response time and ten minutes reduction in it can decrease the risk and probability of death for a car crash victim by 4.08% (232) and 33% (233) respectively.

The most important gap found in the literature is that neither of the previous studies has investigated the survival time of bicyclists in fatal-injury crashes. In contrast, most studies considered all fatal crashes the same in terms of severity. This study contributes to continuing research on bicyclist fatal crashes. Specifically, instead of homogenizing all bicyclist fatal crashes as the same, this study takes into account fatal injury crashes as a continuous severity level based on the survival time of bicyclists ranging from instant death to death within 30 days as it is available in the FARS data. Moreover, previous studies utilized statistical models to investigate the causality of factors in crash severity, while this study explores the application of a machine learning method (i.e., XGBoost) and interprets the results by utilizing a special technique (i.e., SHAP value). The findings of this study provide insights into the pre-crash and post-crash factors in terms of the bicyclist, driver, roadway, and accident characteristics as the pre-crash factors, and more importantly, the crash notification time as a post-crash factor that can affect the survival time of a bicyclist victim in a fatal car crash.

Methods

This study focuses on analyzing the survival time of bicyclists involved in fatal injury crashes and extracting associations of several variables with bicyclists time-to-death that are categorized based on the Haddon Matrix. To that end, as demonstrated in Figure 7-1, an XGBoost tree model is developed and the SHAP (SHapley Additive exPlanations) technique is utilized to explain the

results. These techniques are explained in this section. A machine learning approach is used in this study first because it is evident that many machine learning methods perform better than statistical models in terms of prediction (257; 259-261), and can consider multiple predictors by eliminating possible multicollinearity (262; 263). Second, the method used in this study can deal with nonlinearity in the data and can learn more deeply from the data (268; 276).

Haddon Matrix

The Haddon matrix is a common technique developed by William Haddon to combine the epidemiologic triangle of human factors, vehicle and equipment factors, and environmental factors with 3 stages of pre-crash, during, and post-crash phases (277; 278). Based on this tool, the studied factors can be classified and special strategies can be designed for road traffic fatality and injury prevention (279).

Machine Learning Model: XGBoost

XGBoost is a decision-tree-based machine learning method that utilizes the Gradient Boosting (GB) framework (280) and can be used to solve classification, ranking, and regression problems. In comparison with GB, generally, XGBoost utilizes a more regularized model formalization. This feature provides better over-fitting control and results in better performance. The algorithm of XGBoost requires several parameters to be determined based on the best model performance (268). Therefore, considering using SHAP values in this study, the following hyperparameters should be tuned based on the best model performance and reasonable SHAP values:

- The number of iterations or the number of trees (nrounds)
- The learning rate (eta). Selecting a small value for eta helps the model to capture deep interactions in the data, but the model will need so much time to train
- Gamma specifies the minimum loss reduction required to make a split;
- The maximum number of rules separating the top of the tree from any leaf (max_depth)
- The minimum number of individuals allowed to be in a partition (min_child_weight)
- The fraction of predictors a tree is allowed to consider when making a new rule (colsample_bytree)
- The fraction of instances used to build a tree (subsample)
- The learning objective that is determined based on the output distribution (objective)
- The evaluation metrics; “eval_metric” is Root Mean Square Error (RMSE) for regression.

For more information about the XGBoost algorithm and the parameters, please refer to (268).

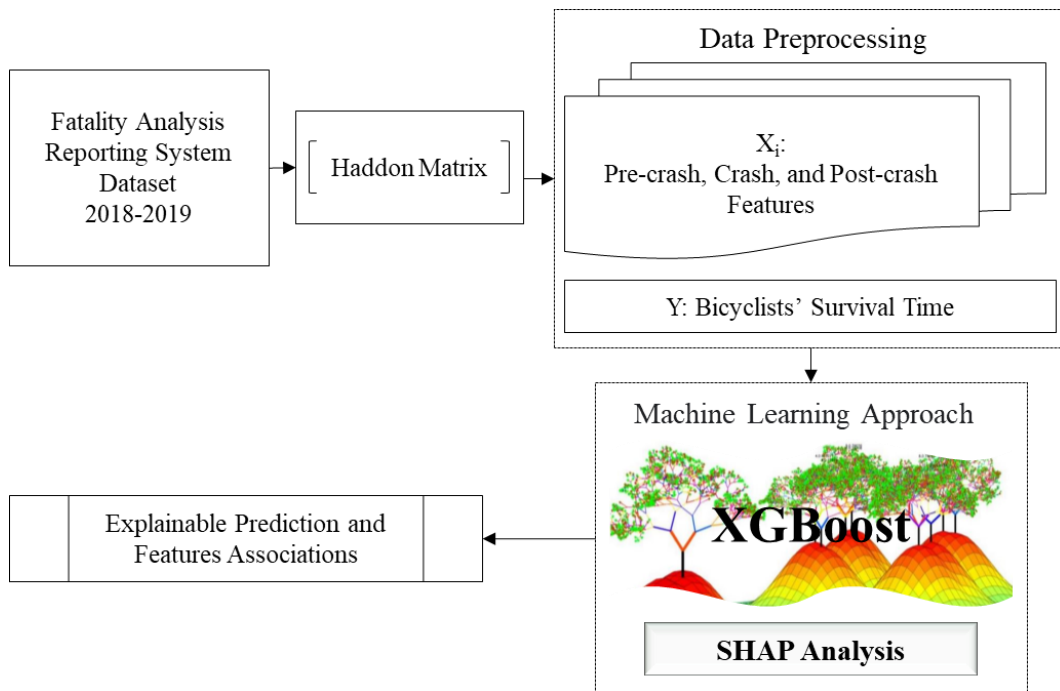


Figure 7-1 Study Framework

SHAP Value

SHAP value is a new technique that can be used to interpret an XGBoost model. This technique was introduced by Lundberg and Lee in 2017, which captures the importance of a variable by comparing what a complex machine learning model predicts with and without the variable (266). Equations 5 and 6 determine SHAP values based on the contribution of a variable.

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} [\nu(S \cup \{i\}) - \nu(S)] \quad (5)$$

where φ_i is the contribution of variable i , ν is model output with a group N with n variables.

$$g(z') = \varphi_0 + \sum_{i=1}^M \varphi_i z'_i \quad (6)$$

where $g(z')$ is a linear function of binary variables g based on the additive variable attribution method; $z' \in [0,1]^M$ which is one when a variable is observed and 0 otherwise; M is the number of input variables. For more information please refer to (266; 281).

Data

This study utilizes the Fatality Analysis Reporting System (FARS) data from 2015 to 2019 for all U.S. states. FARS dataset collects crashes from the US Department of Transportation National Highway Traffic Safety Administration's National Center for Statistics and Analysis in the U.S. where at least one person dies within 30 days from the time of the crash. The FARS dataset typically provides deeper crash investigations of fatal crashes. This study focuses on the fatality outcomes of one-bicyclist one-vehicle crashes. Other types of crashes are excluded from the dataset. The total number of crashes after cleaning the data from missing values and error-checking the data by examining descriptive statistics is 1,001 bicyclist fatalities in 5 years. Among the variables available in the data, those that were frequently used in previous studies were selected which will be discussed in this section. The variables are classified based on the Haddon Matrix and are presented in Table 7-1.

Variables

The selected variables are categorized into environmental, Human, and vehicle or bike characteristics. The descriptive frequencies of the variables are presented in Table 7-2. In this study, environmental characteristics are comprised of crash time of day, crash day, lighting conditions, land use, speed limit, crash location, sidewalk presence, and crosswalk presence. Crash time is divided into 3 categories: peak hours, late-night hours, and other hours. The crash day variable includes weekday or weekend. Lighting conditions are divided into daylight or dark. Speed limits are categorized into 3 categories: under 30 mph, 35 and 40 mph, and above 45 mph. Crash location includes crashes at intersections or not-intersections. Land use comprises urban areas and rural areas. Human characteristics are categorized into 4 variables including age, gender, driver intoxication, and driver hit and run action. Hit-and-run crashes represent driver behavior characteristics. Bicyclist age is divided into 3 categories: under 30, between 30 and 60, and above 60. Gender includes male and female bicyclists. Finally, in terms of vehicle/bike characteristics, bicyclist direction of travel consists of 3 categories including traveling with traffic, facing traffic, and unknown directions.

Table 7-1 Classification of the variables based on the Haddon Matrix

PHASES	FACTORS		
	<i>Human</i>	<i>Vehicle/Bike/Vector</i>	<i>Environment</i>
<i>Pre-crash</i>	Intoxicated Driver	Bicyclist direction of travel	- Sidewalk - Crosswalk - Speed limit
<i>Crash</i>			- Land use - Crash time of day - Crash day - Lighting - Crash location
<i>Post-crash</i>	- Driver: Hit and run - Bicyclist age - Bicyclist gender	Crash notification time (Automated crash notification systems)	Crash notification time

Table 7-2 Variable Statistics

Variable (No. of observation = 1001)		Frequency	Percent
<i>(Environmental characteristic)</i>			
Crash Time of Day	Peak hours	353	35
	Late-night hours	270	27
	Other hours	378	38
Crash Day	Weekend	241	24
	Weekday	760	76
Lighting Condition	Daylight	537	54
	Dark	464	46
<i>(Driver characteristic)</i>			
Hit and Run	No	854	85
	Yes	147	15
<i>(Bicyclist characteristics)</i>			
Age	<30	280	28
	30-60	457	46
	>60	264	26
Gender	Male	851	85
	Female	150	15
Bicyclist Direction of Travel	With Traffic	651	65
	Facing Traffic	160	16
	Unknown	190	19
<i>(Roadway characteristics)</i>			
Crash Location	At intersection	399	40
	Not at intersection	602	60
Land Use	Rural	266	27
	Urban	735	73
Crosswalk Present	No	802	80
	Yes	182	18
	Unknown	17	2
Sidewalk Present	No	630	63
	Yes	353	35
	Unknown	18	2
Speed Limit (MPH)	≤30	238	24
	35-40	306	30
	≥45	457	46

Crash notification time and bicyclist time-to-death are the post-crash variables. The crash notification time variable is an integer variable with a minimum of zero and a maximum of 241 minutes. Note that after error-checking the data, a few observations had very large crash notification times because the time for EMS notification was before the time of the crash. These observations were excluded from the dataset due to the error. Another integer post-crash variable is bicyclist time-to-death, which is determined based on the difference between the crash time and the bicyclist's time of death. This variable has a minimum of one minute and a maximum of 43,467 minutes. It should be pointed out that the instant death at a crash scene where bicyclist time-to-death equals zero is excluded from the dataset because it does not correlate with crash notification time and EMS response time. The descriptive statistic of these variables is presented in Table 7-3.

Data Transformation

The notification time and bicyclist time-to-death (TTD) variables exhibit a large range of values. These variables also have a right-skewed distribution. A natural log transformation is applied to reduce the skewness of the crash notification time variable and reduce the range of values. Note that one unit is added to the values of crash notification time so that there are no zero values for doing the logarithm transformation. Similarly, since the range of bicyclist time-to-death values is so large, a quad root transformation is selected to represent the learning objective of the XGBoost model as the Gamma distribution. Table 7-3 presents the descriptive statistics of the transformed variables.

Results and Discussion

As this study aims to investigate the survival time of bicyclists involved in fatal injury crashes, the XGBoost model was trained on 85% of the data based on randomly selected observations. The remaining 15% data was used to test the model performance. Besides, a 10-fold cross-validation process was applied to the training data to test the stability of the model performance. Finally, the model with the best fit and reasonable SHAP values was selected. Figure 7-2 illustrates the selected hyper-parameters of the XGBoost model and the estimated SHAP summary plot for the model. The plot orders the features based on their importance to estimate bicyclist time-to-death. The plot exhibits that crash notification time is the most important factor affecting a bicyclist's survival time.

Based on the high and low feature values (i.e., high values are shown with purple dots and low values are shown with yellow dots), overall relationships between the features and bicyclist time-to-death can be extracted. A positive sign for SHAP value stands for increasing and a negative sign stands for decreasing the transformed bicyclist time-to-death. According to the estimated SHAP values, bicyclists involved in fatal-injury crashes in roadways with a posted speed limit higher than 45 mph had shorter survival times.

Table 7-3 Descriptive Statistics of Transformed Variables

Variable (No. of observations = 1001)	Min	Mean	Max	Standard Deviation
Time-to-Death (Minute)	1	2470	40467	5467.7
Transformed Time-to-Death (Quad Root)	1	4.613	14.183	3.043
Crash Notification Time (Minute)	0	2.738	241	10.032
Transformed Crash Notification Time (Ln)	0	0.782	5.489	0.839

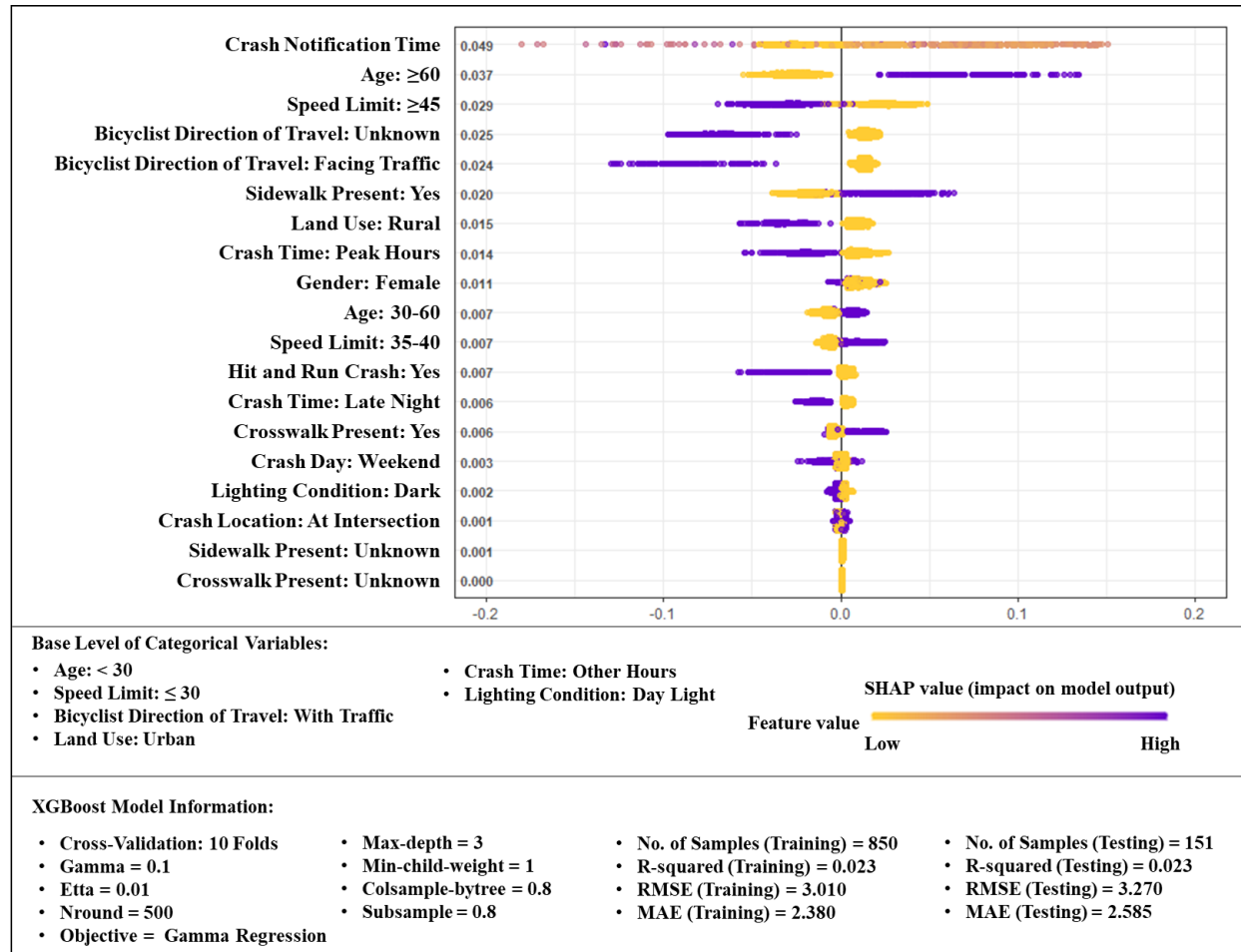


Figure 7-2 Shap Summary Plot of XGBoost Model for Bicyclist Time-to-death

On the other hand, the presence of a sidewalk and the presence of a crosswalk increase cyclists' time-to-death. Furthermore, it can be seen that facing traffic as the bicyclist direction of travel is more lethal than the direction of travel with traffic. Based on the estimated SHAP values, it can also be observed that younger bicyclists had surprisingly shorter survival times compared to older cyclists. One reason for it is that older bicyclists have different preferences and behavior than younger bicyclists. Bernhoft and Carstensen (2008) state that “the older road users seem to be more influenced by the fact that an action is illegal than the younger road users are” (282). Therefore, older bicyclists seem to be more cautious (283) and behave in a way that they feel safer. Another reason is that older cyclists may be involved in less severe crashes, but due to their weak physical and health conditions, they may die several days after the crash during hospitalization. However, in the same situation, a younger victim has a higher chance to survive.

More importantly, vehicle-bicycle fatal crashes in rural areas, in hit-and-run crashes, and during peak hours and late-night hours had noticeably reduced survival times of bicyclists. Figure 7-3Figure 7-3 demonstrates SHAP value Plots of the natural logarithm of crash notification time in which it presents the association between the transformed crash notification time and bicyclist time-to-death. The plots demonstrate the value of crash notification time on the x-axis and the SHAP value on the y-axis to determine the impact of different crash notification times on the model output. Overall, as illustrated in Figure 7-3-A larger crash notification times (higher feature values) have negative SHAP values. This indicates that a large delay in the crash notification time increases the likelihood of early death in cyclist's fatal crashes. It can also be deduced that very severe crashes usually receive very prompt notification times as the SHAP values for some of the very quick notification times are slightly below zero. Figure 7-3-B, C, and D demonstrate SHAP values of the transformed crash notification time and differentiate crashes between late-night hours and other hours, between hit-and-run and non-hit-and-run crashes, and between crashes in rural and urban areas, respectively. According to these plots and as presented in Table 7-4, the average crash notification time during late-night hours is significantly higher than other hours at a 90% confidence level. Likewise, at a 95% confidence level, the average crash notification time for hit-and-run crashes and crashes in rural areas are significantly higher than non-hit-and-run crashes and urban areas, respectively.

The SHAP results show that a large delay in crash notification time as a fundamental part of EMS response time considerably reduces bicyclist time-to-death and increases the likelihood of early death in fatal vehicle-bicyclist crashes. This is consistent with several studies where EMS response times with an early response were shown to potentially increase the victim survival time in a car crash (215; 216; 228; 230). The average crash notification time shows that it takes longer in rural areas compared to urban areas to notify EMS personnel. This finding is consistent with previous studies where EMS response time is higher in rural areas compared to urban areas (284-286). Additionally, hit-and-run crashes may increase a crash notification delay if there are no witnesses or survivors capable of calling emergency services, which can cause a substantial delay in EMS response time (287). The same logic can be used for explaining the longer crash notification time during late-night hours compared to daytime.

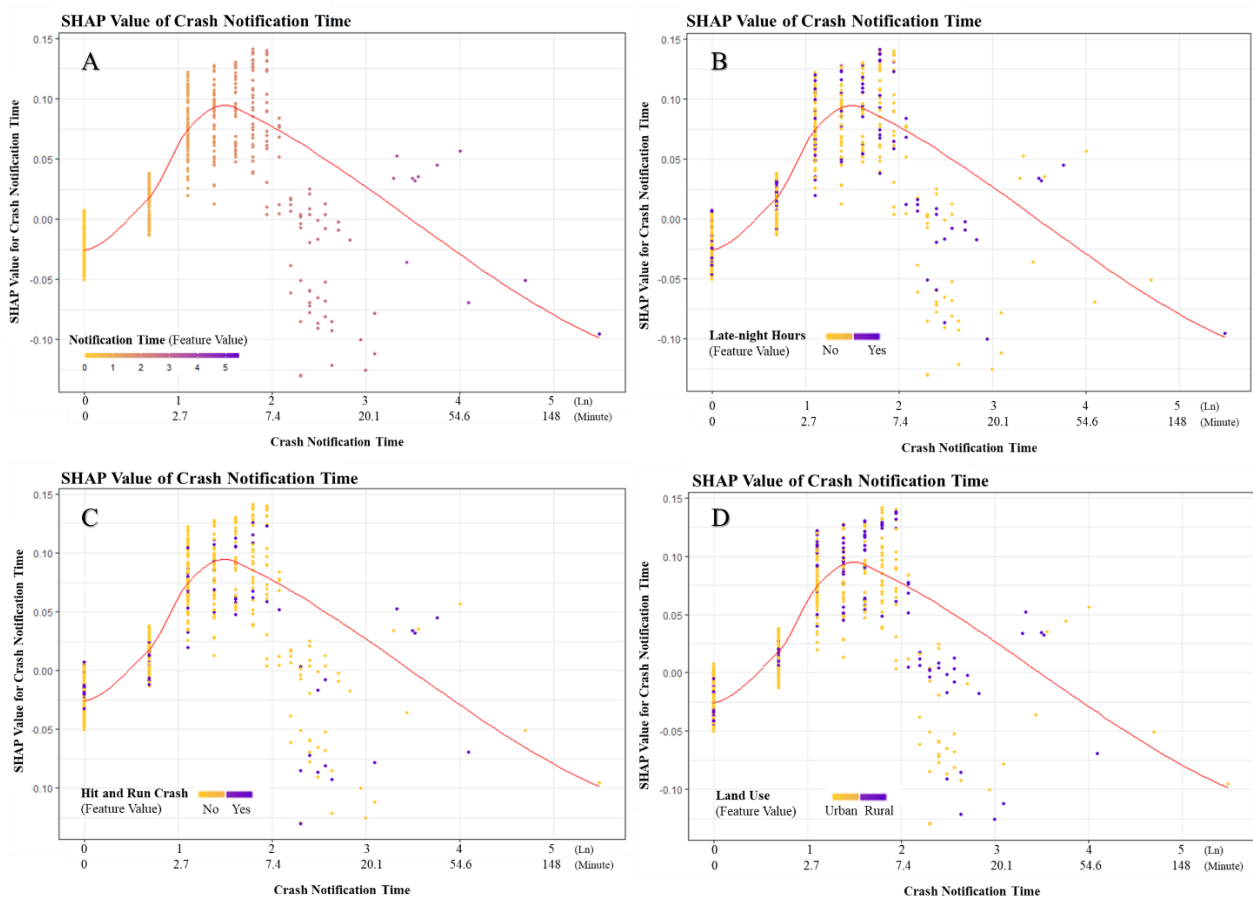


Figure 7-3 SHAP value Plot of Natural Logarithm of Crash Notification Time

Table 7-4 Difference in the mean of crash notification time in late-night hours, hit and run crashes, and urban or rural areas

Variable	Late-night Hours		Hit and Run Crashes		Land Use	
	No	Yes	No	Yes	Urban	Rural
Mean of Crash Notification Time (minute)	2.393	3.674	2.530	3.946	2.473	3.470
P-value of one-tailed t-test	0.096		0.048		0.049	
Confidence Level	90%		95%		95%	

Conclusions

The objective of this study is to analyze the bicyclist time-to-death involved in fatal crashes and extract associations of various factors with bicyclist time-to-death and more importantly, crash notification time as an important post-crash factor. To that end, variables are selected and categorized based on the Haddon Matrix in terms of pre-crash, during-crash, and post-crash factors and an Extreme Gradient Boosting (XGBoost) tree is developed and SHAP (SHapley Additive exPlanations) technique is utilized to interpret the results. A key contribution of the paper is the novel application of these explainable machine learning techniques to improve bicyclist safety as these techniques perform better than statistical models in terms of prediction and can learn more deeply from the data by dealing with nonlinearity in the data.

The SHAP results show that a large delay in crash notification time as a fundamental part of EMS response time considerably reduces bicyclist time-to-death and increases the likelihood of early death in fatal vehicle-bicyclist crashes. Additionally, the average crash notification time for hit-and-run crashes, for crashes in rural areas, and crashes during late-night hours are significantly higher than non-hit-and-run crashes, urban areas, and other hours respectively. These situations are not rare and may increase a crash notification delay if there are no witnesses or survivors capable of notifying emergency responders that can cause a substantial delay in EMS response time. Recent on-road vehicle technologies have the potential to alleviate these situations. Studies indicate that utilizing an advanced automatic collision notification system can deliver an average crash notification time of 1 minute (244). The system is an automatic safety technology aimed to inform emergency responders immediately after a crash. Other technologies that need further investigation include improved mobile phone coverage especially in rural areas, automatic location capability, and the use of GPS technology to improve response times.

The results also indicate that posted speed limits, bicyclist direction of travel, cyclist age, lighting conditions, crash locations, land use, availability of crosswalks and sidewalk, hit and run crashes, and certain crash times, are associated with bicyclist time-to-death. The findings of this study can impact practice by providing safety practitioners with key factors that have the potential to save bicyclist lives, specifically through faster bicyclist crash detection and emergency responders' notification. Future studies should also consider more post-crash factors such as medical information, type of damage, and the treatment offered to cyclist victims.

This study is not without limitations. One limitation of this study is the low number of observations with emergency medical system information, which noticeably reduced the total number of observations in the dataset. Additionally, the FARS dataset does not have some relevant post-crash variables that can affect victims' survival times such as medical information, type of damage, victim underlying medical conditions, and treatment offered to victims.

ACKNOWLEDGMENT

The author(s) would like to thank the following financial support for the research, authorship, and/or publication of this article: This project was partially supported by the Tennessee Department of Transportation, TDOT and the U.S. DOT through the Collaborative Sciences Center for Road Safety (Grant No. 69A3551747113). The views in this paper are solely those of the authors who are responsible for the content of this publication.

CHAPTER 8: CONCLUSIONS

With the emergence of connected and automated vehicles (CAVs) equipped with wireless communication technologies, it is important to consider the large amount of information generated by electronic sensors in the combined ecosystems of CAVs. Generated large-scale empirical data has significant potential to facilitate a deeper understanding of what can be achieved from deploying CAVs on roadways and what transportation-related problems can be solved. Connected and automated vehicle technologies have the potential to significantly improve transportation system performance. But due to the limitation in the empirical CAV data, knowledge about their impacts is very limited. As such, this study attempts to fill a portion of the existing gap. The overall intellectual merit and contribution of this study is a robust analysis of the impacts of vehicles' automation and cooperation technologies using emerging large-scale data obtained from CAV field tests. While the benefits of these technologies are expected, this study quantifies and more realistically assesses the expected benefits of the technologies based on field experimental data. Furthermore, this dissertation explores the potential for CAV technologies to address some of the important transportation problems related to vulnerable road users.

The main focus of this dissertation is on CAVs with a low level of automation. The key hypothesis of the dissertation is that the technologies utilized in the CAVs with a low level of automation (i.e., SAE automation levels 1 and 2) can benefit the environment, traffic safety, and the safety of vulnerable road users. In this dissertation, some of the potential impacts of the advanced driver-assistance systems (ADAS), including adaptive cruise control (ACC), cooperative adaptive cruise control (CACC), pedestrian crash prevention (PCP) system, and advanced automated collision notification (ACN) system are investigated. The observed benefits from real-life testing show that ACC and CACC technologies substantially improve traffic safety and benefit the environment. Furthermore, this dissertation shows that AVs in mixed traffic streams can induce behavioral changes in conventional vehicle drivers, with some beneficial effects on safety and the environment. On average, a driver that follows an AV exhibits lower driving volatility in terms of speed and acceleration, which represents more stable traffic flow behavior and lower crash risk. The analysis showed a remarkable improvement in TTC as a result of the notably better speed adjustments of the following vehicle (i.e., lower differences in speeds between the lead and following vehicles) in the second scenario. Furthermore, human-driven vehicles were found to consume less fuel and produce fewer emissions on average when following an AV. Regarding the PCP systems the test data shows that on average in 70% of the tests, the pedestrian crash prevention systems avoid pedestrian crashes. However, for the occurred crashes, PCP systems on average were able to mitigate impact speeds more than 50%. In real-life situations, this could translate to substantial reductions in injury risk. Furthermore, it is shown that the CAN system potentially can save vulnerable road users' lives by reducing delays in crash notification times and EMS response times. The overall benefits of the studied CAV technologies from the Haddon's Matrix concept in terms of pre-crash, during-crash, and post-crash versus human, vehicle, and environmental factors are presented in Table 8-1. The overall implication of this dissertation assists practitioners in traffic safety and CAV technology developers by having more empirical evidence regarding the impacts of connectivity and automation for wide-scale deployment. To have more generalizable results, it is recommended for future studies to collect more experimental or real-world data and expand the scenarios examined in different mixed traffic environments and conditions, e.g., adverse weather, different roadway classifications, and edge

cases. Furthermore, for future studies, it is suggested to repeat this study on different AV algorithms and compare the results. Also, it is suggested for future studies to investigate the change in human drivers' behavior by informing them whether they are following an AV or a conventionally driven vehicle. Future studies can also investigate the potential impact of CAVs on other aspects of transportation such as mobility, accessibility, and sustainability.

Table 8-1 Classifying CAV technologies' benefits from Haddon's Matrix concept

	Human Factors:	Vehicle Factors:	Environmental Factors:
Pre-crash	ACC System: <i>Reduces driving volatility and increases TTC → Improves Safety</i> CACC System: <i>Reduces driving volatility and increases TTC → Improves Safety</i> PCP System: <i>Reduces human error → Improves Safety</i>	PCP System: <i>Avoids crashes → Improves Safety</i>	
Crash		PCP System: <i>Reduces impact speeds → Improves Safety</i>	
Post-Crash			ACN System: <i>Reduces EMS response times → Improves Safety</i>

REFERENCES

- [1] Berry, M. W., and M. Castellanos. Survey of text mining. *Computing Reviews*, Vol. 45, No. 9, 2004, p. 548.
- [2] Kim, S. M., and E. Hovy. Extracting opinions, opinion holders, and topics expressed in online news media text. *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, Association for Computational Linguistics, 2006. pp. 1-8.
- [3] Pang, B., and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 2004. p. 271.
- [4] Marletto, G. Who will drive the transition to self-driving? A socio-technical analysis of the future impact of automated vehicles. *Technological Forecasting and Social Change*, Vol. 139, 2019, pp. 221-234.
- [5] Ghiasi, A., O. Hussain, Z. S. Qian, and X. Li. A mixed traffic capacity analysis and lane management model for connected automated vehicles: A Markov chain method. *Transportation Research Part B: Methodological*, Vol. 106, 2017, pp. 266-292.
- [6] Lee, S., E. Jeong, M. Oh, and C. Oh. Driving aggressiveness management policy to enhance the performance of mixed traffic conditions in automated driving environments. *Transportation Research Part A: Policy and Practice*, Vol. 121, 2019, pp. 136-146.
- [7] Mirheli, A., L. Hajibabai, and A. Hajbabaie. Development of a signal-head-free intersection control logic in a fully connected and autonomous vehicle environment. *Transportation Research Part C: Emerging Technologies*, Vol. 92, 2018, pp. 412-425.
- [8] Talavera, E., A. Díaz-Álvarez, F. Jiménez, and J. Naranjo. Impact on congestion and fuel consumption of a cooperative adaptive cruise control system with lane-level position estimation. *Energies*, Vol. 11, No. 1, 2018, p. 194.
- [9] Papadoulis, A., M. Quddus, and M. Imprialou. Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 12-22.
- [10] Morando, M. M., Q. Tian, L. T. Truong, and H. L. Vu. Studying the safety impact of autonomous vehicles using simulation-based surrogate safety measures. *Journal of Advanced Transportation*, Vol. 2018, 2018.
- [11] Zhao, D., H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa, and C. S. Pan. Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 3, 2017, pp. 595-607.
- [12] Eriksson, A., S. M. Petermeijer, M. Zimmermann, J. C. de Winter, K. J. Bengler, and N. A. Stanton. Rolling out the red (and green) carpet: supporting driver decision making in automation-to-manual transitions. *IEEE Transactions on Human-Machine Systems*, Vol. 49, No. 1, 2019, pp. 20-31.
- [13] Yang, D., S. Zheng, C. Wen, P. J. Jin, and B. Ran. A dynamic lane-changing trajectory planning model for automated vehicles. *Transportation Research Part C: Emerging Technologies*, Vol. 95, 2018, pp. 228-247.
- [14] Hu, J., L. Kong, W. Shu, and M.-Y. Wu. Scheduling of connected autonomous vehicles on highway lanes. *2012 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2012. pp. 5556-5561.
- [15] Pek, C., P. Zahn, and M. Althoff. Verifying the safety of lane change maneuvers of self-driving vehicles based on formalized traffic rules. *2017 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2017. pp. 1477-1483.

- [16] Bevely, D., X. Cao, M. Gordon, G. Ozbilgin, D. Kari, B. Nelson, J. Woodruff, M. Barth, C. Murray, and A. Kurt. Lane change and merge maneuvers for connected and automated vehicles: A survey. *IEEE Transactions on Intelligent Vehicles*, Vol. 1, No. 1, 2016, pp. 105-120.
- [17] Bengler, K., K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner. Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, Vol. 6, No. 4, 2014, pp. 6-22.
- [18] Zheng, N. N., S. Tang, H. Cheng, Q. Li, G. Lai, and F. Y. Wang. Toward intelligent driver-assistance and safety warning systems. 2004.
- [19] Tang, A., and A. Yip. Collision avoidance timing analysis of DSRC-based vehicles. *Accident Analysis & Prevention*, Vol. 42, No. 1, 2010, pp. 182-195.
- [20] Golestan, K., R. Soua, F. Karray, and M. S. Kamel. Situation awareness within the context of connected cars: A comprehensive review and recent trends. *Information Fusion*, Vol. 29, 2016, pp. 68-83.
- [21] Noy, I. Y., D. Shinar, and W. J. Horrey. Automated driving: Safety blind spots. *Safety Science*, Vol. 102, 2018, pp. 68-78.
- [22] Harper, C. D., C. T. Hendrickson, and C. Samaras. Cost and benefit estimates of partially-automated vehicle collision avoidance technologies. *Accident Analysis & Prevention*, Vol. 95, 2016, pp. 104-115.
- [23] Khazraeian, S., M. Hadi, and Y. Xiao. Safety impacts of queue warning in a connected vehicle environment. *Transportation Research Record*, No. 2621, 2017, pp. 31-37.
- [24] Jeong, E., C. Oh, and S. Lee. Is vehicle automation enough to prevent crashes? Role of traffic operations in automated driving environments for traffic safety. *Accident Analysis & Prevention*, Vol. 104, 2017, pp. 115-124.
- [25] Zhang, Y., C. Wu, C. Qiao, and Y. Hou. The effects of warning characteristics on driver behavior in connected vehicles systems with missed warnings. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 138-145.
- [26] Li, Y., H. Wang, W. Wang, L. Xing, S. Liu, and X. Wei. Evaluation of the impacts of cooperative adaptive cruise control on reducing rear-end collision risks on freeways. *Accident Analysis & Prevention*, Vol. 98, 2017, pp. 87-95.
- [27] Li, Y., C. C. Xu, L. Xing, and W. Wang. Integrated cooperative adaptive cruise and variable speed limit controls for reducing rear-end collision risks near freeway bottlenecks based on micro-simulations. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 11, 2017, pp. 3157-3167.
- [28] Li, Y., Z. Li, H. Wang, W. Wang, and L. Xing. Evaluating the safety impact of adaptive cruise control in traffic oscillations on freeways. *Accident Analysis & Prevention*, Vol. 104, 2017, pp. 137-145.
- [29] Kianfar, R., B. Augusto, A. Ebadighajari, U. Hakeem, J. Nilsson, A. Raza, R. S. Tabar, N. V. Irukulapati, C. Englund, and P. Falcone. Design and experimental validation of a cooperative driving system in the grand cooperative driving challenge. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 3, 2012, pp. 994-1007.
- [30] Robinson, C. L., L. Caminiti, D. Caveney, and K. Laberteaux. Efficient coordination and transmission of data for cooperative vehicular safety applications. *Proceedings of the 3rd International Workshop on Vehicular AD HOC Networks*, Association for Computing Machinery (ACM), 2006. pp. 10-19.

- [31] Jeong, E., C. Oh, G. Lee, and H. Cho. Safety impacts of intervehicle warning information systems for moving hazards in connected vehicle environments. *Transportation Research Record*, No. 2424, 2014, pp. 11-19.
- [32] Kusano, K. D., and H. C. Gabler. Safety benefits of forward collision warning, brake assist, and autonomous braking systems in rear-end collisions. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 4, 2012, pp. 1546-1555.
- [33] Lee, K., and H. Peng. Evaluation of automotive forward collision warning and collision avoidance algorithms. *Vehicle System Dynamics*, Vol. 43, No. 10, 2005, pp. 735-751.
- [34] Qin, Y., and H. Wang. Influence of the feedback links of connected and automated vehicle on rear-end collision risks with vehicle-to-vehicle communication. *Traffic injury prevention*, 2019, pp. 1-5.
- [35] Caveney, D. Cooperative vehicular safety applications. *IEEE Control Systems Magazine*, Vol. 30, No. 4, 2010, pp. 38-53.
- [36] Rau, P., M. Yanagisawa, and W. G. Najm. Target crash population of automated vehicles. *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 2015. pp. 1-11.
- [37] Eidehall, A., J. Pohl, F. Gustafsson, and J. Ekmark. Toward autonomous collision avoidance by steering. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 1, 2007, pp. 84-94.
- [38] Wardzinski, A. The role of situation awareness in assuring safety of autonomous vehicles. In *Computer Safety, Reliability, and Security, Proceedings*, No. 4166, Springer-Verlag Berlin, Berlin, 2006. pp. 205-218.
- [39] Moon, S., I. Moon, and K. Yi. Design, tuning, and evaluation of a full-range adaptive cruise control system with collision avoidance. *Control Engineering Practice*, Vol. 17, No. 4, 2009, pp. 442-455.
- [40] Chira-Chavala, T., and S. M. Yoo. Potential safety benefits of intelligent cruise control systems. *Accident Analysis & Prevention*, Vol. 26, No. 2, 1994, pp. 135-146.
- [41] Yue, L., M. Abdel-Aty, Y. Wu, and L. Wang. Assessment of the safety benefits of vehicles' advanced driver assistance, connectivity and low level automation systems. *Accident Analysis & Prevention*, Vol. 117, 2018, pp. 55-64.
- [42] Bagloee, S. A., M. Tavana, M. Asadi, and T. Oliver. Autonomous vehicles: Challenges, opportunities, and future implications for transportation policies. *Journal of modern transportation*, Vol. 24, No. 4, 2016, pp. 284-303.
- [43] Rahman, M. S., M. Abdel-Aty, J. Lee, and M. H. Rahman. Safety benefits of arterials' crash risk under connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, Vol. 100, 2019, pp. 354-371.
- [44] Akella, M. R., C. Bang, R. Beutner, E. M. Delmelle, R. Batta, A. Blatt, P. A. Rogerson, and G. Wilson. Evaluating the reliability of automated collision notification systems. *Accident Analysis & Prevention*, Vol. 35, No. 3, 2003, pp. 349-360.
- [45] Baker, R. T., J. Wagner, and Ieee. Policy Pathways to Vehicle Automation Industry Perspectives on the Role of Public Policy in Autonomous Vehicle Development. In *2013 International Conference on Connected Vehicles and Expo*, Ieee, New York, 2013. pp. 431-436.
- [46] Yang, H., Z. Wang, and K. Xie. Impact of connected vehicles on mitigating secondary crash risk. *International Journal of Transportation Science and Technology*, Vol. 6, No. 3, 2017, pp. 196-207.

- [47] Qin, Y. Y., H. Wang, and B. Ran. Impact of connected and automated vehicles on passenger comfort of traffic flow with vehicle-to-vehicle communications. *Ksce Journal of Civil Engineering*, Vol. 23, No. 2, 2019, pp. 821-832.
- [48] Elbanhawi, M., M. Simic, and R. Jazar. In the passenger seat: Investigating ride comfort measures in autonomous cars. *IEEE Intelligent Transportation Systems Magazine*, Vol. 7, No. 3, 2015, pp. 4-17.
- [49] Rahman, M. S., and M. Abdel-Aty. Longitudinal safety evaluation of connected vehicles' platooning on expressways. *Accident Analysis & Prevention*, Vol. 117, 2018, pp. 381-391.
- [50] Magdici, S., and M. Althoff. Fail-safe motion planning of autonomous vehicles. Presented at 19th International Conference on Intelligent Transportation Systems, New York, 2016.
- [51] Kianfar, R., P. Falcone, and J. Fredriksson. A distributed model predictive control approach to active steering control of string stable cooperative vehicle platoon. *IFAC Proceedings Volumes*, Vol. 46, No. 21, 2013, pp. 750-755.
- [52] Kianfar, R., M. Ali, P. Falcone, and J. Fredriksson. Combined longitudinal and lateral control design for string stable vehicle platooning within a designated lane. *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2014. pp. 1003-1008.
- [53] Alam, A., A. Gattami, K. H. Johansson, and C. J. Tomlin. Guaranteeing safety for heavy duty vehicle platooning: Safe set computations and experimental evaluations. *Control Engineering Practice*, Vol. 24, 2014, pp. 33-41.
- [54] Zhang, W., S. Guhathakurta, and E. B. Khalil. The impact of private autonomous vehicles on vehicle ownership and unoccupied VMT generation. *Transportation Research Part C: Emerging Technologies*, Vol. 90, 2018, pp. 156-165.
- [55] Cui, J., L. S. Liew, G. Sabaliauskaite, and F. Zhou. A review on safety failures, security attacks, and available countermeasures for autonomous vehicles. *Ad Hoc Networks*, 2018.
- [56] Dixit, V., Z. Xiong, S. Jian, and N. Saxena. Risk of automated driving: Implications on safety acceptability and productivity. *Accident Analysis & Prevention*, Vol. 125, 2019, pp. 257-266.
- [57] 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.
- [58] Kyriakidis, M., R. Happee, and J. C. F. de Winter. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 32, 2015, pp. 127-140.
- [59] Schoitsch, E. Autonomous vehicles and automated driving status, perspectives and societal impact. *Information technology, society and economy strategic cross-influences (IDIMT-2016). 24th Interdisciplinary Information Management Talks*, Vol. 45, No. 1, 2016, pp. 405-424.
- [60] Li, Y., Y. Tu, Q. Fan, C. Dong, and W. Wang. Influence of cyber-attacks on longitudinal safety of connected and automated vehicles. *Accident Analysis & Prevention*, Vol. 121, 2018, pp. 148-156.
- [61] Petit, J., and S. E. Shladover. Potential cyberattacks on automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 2, 2015, pp. 546-556.
- [62] Taeihagh, A., and H. S. M. Lim. Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, Vol. 39, No. 1, 2019, pp. 103-128.
- [63] Skarbek-Żabkin, A., and M. Szczepanek. Autonomous vehicles and their impact on road infrastructure and user safety. *2018 XI International Science-Technical Conference Automotive Safety*, IEEE, 2018. pp. 1-4.

- [64] Cui, L., J. Hu, B. B. Park, and P. Bujanovic. Development of a simulation platform for safety impact analysis considering vehicle dynamics, sensor errors, and communication latencies: Assessing cooperative adaptive cruise control under cyber attack. *Transportation Research Part C: Emerging Technologies*, Vol. 97, 2018, pp. 1-22.
- [65] Winkle, T. Safety benefits of automated vehicles: Extended findings from accident research for development, validation and testing. In *Autonomous Driving*, Springer, 2016. pp. 335-364.
- [66] Ge, J. I., S. S. Avedisov, C. R. He, W. B. Qin, M. Sadeghpour, and G. Orosz. Experimental validation of connected automated vehicle design among human-driven vehicles. *Transportation Research Part C: Emerging Technologies*, Vol. 91, 2018, pp. 335-352.
- [67] Vahidi, A., and A. Eskandarian. Research advances in intelligent collision avoidance and adaptive cruise control. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4, No. 3, 2003, pp. 143-153.
- [68] Al Najada, H., and I. Mahgoub. Autonomous vehicles safe-optimal trajectory selection based on big data analysis and predefined user preferences. *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE, 2016. pp. 1-6.
- [69] Weiß, C. V2X communication in Europe – From research projects towards standardization and field testing of vehicle communication technology. *Computer Networks*, Vol. 55, No. 14, 2011, pp. 3103-3119.
- [70] Clark, H., and J. Feng. Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. *Accident Analysis & Prevention*, Vol. 106, 2017, pp. 468-479.
- [71] Matthews, G., C. Neubauer, D. J. Saxby, R. W. Wohleber, and J. Lin. Dangerous intersections? A review of studies of fatigue and distraction in the automated vehicle. *Accident Analysis & Prevention*, 2018.
- [72] Vogelpohl, T., M. Kühn, T. Hummel, and M. Vollrath. Asleep at the automated wheel: Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*, 2018.
- [73] Wiedemann, K., F. Naujoks, J. Wörle, R. Kenntner-Mabiala, Y. Kaussner, and A. Neukum. Effect of different alcohol levels on take-over performance in conditionally automated driving. *Accident Analysis & Prevention*, Vol. 115, 2018, pp. 89-97.
- [74] Wu, Y., K. Kihara, Y. Takeda, T. Sato, M. Akamatsu, and S. Kitazaki. Effects of scheduled manual driving on drowsiness and response to take over request: A simulator study towards understanding drivers in automated driving. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 202-209.
- [75] Petermeijer, S., P. Bazilinskyy, K. Bengler, and J. De Winter. Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics*, Vol. 62, 2017, pp. 204-215.
- [76] Dixit, V. V., S. Chand, and D. J. Nair. Autonomous vehicles: Disengagements, accidents and reaction times. *Plos One*, Vol. 11, No. 12, 2016, p. 14.
- [77] Favarò, F., S. Eurich, and N. Nader. Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. *Accident Analysis & Prevention*, Vol. 110, 2018, pp. 136-148.
- [78] Forster, Y., F. Naujoks, A. Neukum, and L. Huestegge. Driver compliance to take-over requests with different auditory outputs in conditional automation. *Accident Analysis & Prevention*, Vol. 109, 2017, pp. 18-28.
- [79] Fraade-Blanar, L., and N. Kalra. Autonomous vehicles and federal safety standards: An exemption to the rule? *RAND*, 2017.

- [80] Serban, A. C., E. Poll, and J. Visser. Tactical safety reasoning: A case for autonomous vehicles. *87th Vehicular Technology Conference (VTC Spring)*, IEEE, 2018. pp. 1-5.
- [81] Rodríguez Palmeiro, A., S. van der Kint, L. Vissers, H. Farah, J. C. F. de Winter, and M. Hagenzieker. Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 58, 2018, pp. 1005-1020.
- [82] Gupta, S., M. Vasardani, B. Lohani, and S. Winter. Pedestrian's risk-based negotiation model for self-driving vehicles to get the right of way. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 163-173.
- [83] Goel, S., and Y. Yuan. Emerging Research in Connected Vehicles [Guest Editorial]. *IEEE Intelligent Transportation Systems Magazine*, Vol. 7, No. 2, 2015, pp. 6-9.
- [84] O'callaghan, D., D. Greene, J. Carthy, and P. Cunningham. An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, Vol. 42, No. 13, 2015, pp. 5645-5657.
- [85] Jolliffe, I. Principal component analysis. *Technometrics*, Vol. 45, No. 3, 2003, p. 276.
- [86] Mahdinia, I., and M. Habibian. Evaluating the transportation system performance based on efficiency, effectiveness, and efficacy: a case study of the US states. Presented at Transportation Research Board Annual Meeting, Washington DC, 2017.
- [87] Mahdinia, I., M. Habibian, Y. Hatamzadeh, and H. Gudmundsson. An indicator-based algorithm to measure transportation sustainability: A case study of the US states. *Ecological indicators*, Vol. 89, 2018, pp. 738-754.
- [88] 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.
- [89] Nickkar, A., and Y.-J. Lee. Evaluation of Dedicated Lanes for Automated vehicles at Roundabouts with Various Flow Patterns. *arXiv preprint arXiv:1907.07025*, 2019.
- [90] USDOT. Preparing for the Future of Transportation: Automated Vehicles 3.0. U.S. Department of Transportation, 2018.
- [91] NHTSA. National Highway Traffic Safety Administration. Automated driving systems 2.0: A vision for safety. No. 812, Washington, DC: US Department of Transportation, 2017. p. 442.
- [92] CARMA, Cooperative Automation Research Mobility Applications. US Department of Transportation, Federal Highway Administration. <https://cms7.fhwa.dot.gov/research/research-programs/operations/carma-overview>. Accessed April, 2019.
- [93] Mahmoudzadeh, A., H. Razi-Ardakani, and M. Kermanshah. Studying crash avoidance maneuvers prior to an impact considering different types of driver's distractions. *Transportation research procedia*, Vol. 37, 2019, pp. 203-210.
- [94] Li, M., K. Boriboonsomsin, G. Wu, W.-B. Zhang, and M. Barth. Traffic energy and emission reductions at signalized intersections: a study of the benefits of advanced driver information. *International Journal of Intelligent Transportation Systems Research*, Vol. 7, No. 1, 2009, pp. 49-58.
- [95] Petit, J., and S. E. Shladover. Potential cyberattacks on automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 2, 2014, pp. 546-556.
- [96] Kelarestaghi, K. B., K. Heaslip, M. Khalilikhah, A. Fuentes, and V. Fessmann. Intelligent transportation system security: hacked message signs. *SAE International Journal of Transportation Cybersecurity and Privacy*, Vol. 1, No. 11-01-02-0004, 2018, pp. 75-90.

- [97] Kidando, E., R. Moses, M. Ghorbanzadeh, and E. E. Ozguven. Traffic Operation and Safety Analysis on an Arterial Highway: Implications for Connected Vehicle Applications. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2018. pp. 2753-2758.
- [98] Mahdavian, A., A. Shojaei, and A. Oloufa. Assessing the Long-and Mid-Term Effects of Connected and Automated Vehicles on Highways' Traffic Flow and Capacity. *International Conference on Sustainable Infrastructure*, 2019. p. 263.
- [99] Shladover, S. E., D. Su, and X.-Y. Lu. Impacts of cooperative adaptive cruise control on freeway traffic flow. *Transportation Research Record*, Vol. 2324, No. 1, 2012, pp. 63-70.
- [100] Arvin, R., M. Kamrani, and A. J. Khattak. Examining the Role of Speed and Driving Stability on Crash Severity Using SHRP2 Naturalistic Driving Study Data. Presented at Transportation Research Board 98th Annual Meeting, Washington DC, 2019.
- [101] Essa, M., and T. Sayed. Traffic conflict models to evaluate the safety of signalized intersections at the cycle level. *Transportation Research Part C: Emerging Technologies*, Vol. 89, 2018, pp. 289-302.
- [102] Bibeka, A., P. Songchitruksa, and Y. Zhang. Assessing environmental impacts of ad-hoc truck platooning on multilane freeways. *Journal of Intelligent Transportation Systems*, 2019, pp. 1-12.
- [103] Sohrabi, S., H. Khreis, and D. Lord. Beauty or the Beast? Impacts of Autonomous Vehicles on Public Health. *Journal of Transport & Health*, Vol. 14, 2019, p. 100765.
- [104] Seiler, P., A. Pant, and K. Hedrick. Disturbance propagation in vehicle strings. *Transactions on automatic control*, Vol. 49, No. 10, 2004, pp. 1835-1842.
- [105] Ma, J., E. Leslie, A. Ghiasi, Z. Huang, and Y. Guo. Empirical analysis and modeling of a bundled connected automated vehicle application using experimental data. Presented at Transportation Research Board, Washington DC, 2019.
- [106] Kikuchi, S., N. Uno, and M. Tanaka. Impacts of shorter perception-reaction time of adapted cruise controlled vehicles on traffic flow and safety. *Journal of transportation engineering*, Vol. 129, No. 2, 2003, pp. 146-154.
- [107] Li, Y., H. Wang, W. Wang, S. Liu, and Y. Xiang. Reducing the risk of rear-end collisions with infrastructure-to-vehicle (I2V) integration of variable speed limit control and adaptive cruise control system. *Traffic injury prevention*, Vol. 17, No. 6, 2016, pp. 597-603.
- [108] Mamouei, M., I. Kaparias, and G. Halikias. A framework for user-and system-oriented optimisation of fuel efficiency and traffic flow in Adaptive Cruise Control. *Transportation Research Part C: Emerging Technologies*, Vol. 92, 2018, pp. 27-41.
- [109] He, Y., B. Ciuffo, Q. Zhou, M. Makridis, K. Mattas, J. Li, Z. Li, F. Yan, and H. Xu. Adaptive cruise control strategies implemented on experimental vehicles: A review. *IFAC-PapersOnLine*, Vol. 52, No. 5, 2019, pp. 21-27.
- [110] Milanés, V., and S. E. Shladover. Handling cut-in vehicles in strings of cooperative adaptive cruise control vehicles. *Journal of Intelligent Transportation Systems*, Vol. 20, No. 2, 2016, pp. 178-191.
- [111] Milanés, V., S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura. Cooperative adaptive cruise control in real traffic situations. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 1, 2013, pp. 296-305.
- [112] Almannaa, M. H., H. Chen, H. A. Rakha, A. Loulizi, and I. El-Shawarby. Field implementation and testing of an automated eco-cooperative adaptive cruise control system in the

- vicinity of signalized intersections. *Transportation Research Part D: Transport and Environment*, Vol. 67, 2019, pp. 244-262.
- [113] Shladover, S. E., C. Nowakowski, J. O'Connell, and D. Cody. Cooperative adaptive cruise control: Driver selection of car-following gaps. *17th ITS World Congress*, 2010.
- [114] Parsa, A. B., H. Taghipour, S. Derrible, and A. K. Mohammadian. Real-time accident detection: coping with imbalanced data. *Accident Analysis & Prevention*, Vol. 129, 2019, pp. 202-210.
- [115] Shi, Q., and M. Abdel-Aty. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 380-394.
- [116] NHTSA. National motor vehicle crash causation survey: Report to Congress. No. 811, National Highway Traffic Safety Administration Technical Report DOT HS, 2008. p. 059.
- [117] 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.
- [118] 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.
- [119] Kamrani, M., B. Wali, and A. J. Khattak. Can data generated by connected vehicles enhance safety? Proactive approach to intersection safety management. *Transportation Research Record*, No. 2659, 2017, pp. 80-90.
- [120] Kamrani, M., R. Arvin, and A. J. Khattak. Analyzing Highly Volatile Driving Trips Taken by Alternative Fuel Vehicles. Presented at Transportation Research Board 97th Annual Meeting Washington DC, United States, 2018.
- [121] Hoseinzadeh, N., R. Arvin, A. J. Khattak, and L. D. Han. Integrating safety and mobility for pathfinding using big data generated by connected vehicles. *Journal of Intelligent Transportation Systems*, 2020, pp. 1-17.
- [122] Ghasemzadeh, A., and M. M. Ahmed. Drivers' lane-keeping ability in heavy rain: preliminary investigation using SHRP 2 naturalistic driving study data. *Transportation Research Record*, Vol. 2663, No. 1, 2017, pp. 99-108.
- [123] Mousavi, S. M., Z. Zhang, S. A. Parr, A. Pande, and B. Wolshon. Identifying High Crash Risk Highway Segments Using Jerk-Cluster Analysis. *Journal of Transportation engineering*, 2008.
- [124] Liu, H., X. Kan, S. E. Shladover, and X.-Y. Lu. Using Cooperative Adaptive Cruise Control (CACC) to Form High-Performance Vehicle Streams: Simulation Results Analysis. UC Berkeley: California Partners for Advanced Transportation Technology. Retrieved from <https://escholarship.org/uc/item/31w2f555>, 2018.
- [125] Bujanovic, P. Developing vehicle platoons and predicting their impacts. Doctor of Philosophy Thesis. The University of Texas at Austin, 2018.
- [126] Arvin, R., M. Kamrani, and A. Khattak. The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. *Accident Analysis & Prevention*, Vol. 132, 2019.
- [127] Arvin, R., and A. j. Khattak. Harnessing big data generated by connected vehicles to monitor safety performance: Application of Geographically Weighted Negative Binomial Regression. Presented at Transportation Research Board, Washington DC, 2020.

- [128] Saffarzadeh, M., N. Nadimi, S. Naseralavi, and A. R. Mamdoohi. A general formulation for time-to-collision safety indicator. *Proceedings of the Institution of Civil Engineers-Transport*, No. 166, Thomas Telford Ltd, 2013. pp. 294-304.
- [129] Guzzella, L., and C. Onder. *Introduction to modeling and control of internal combustion engine systems*. Springer Science & Business Media, 2009.
- [130] Kamal, M. A. S., M. Mukai, J. Murata, and T. Kawabe. Model predictive control of vehicles on urban roads for improved fuel economy. *IEEE Transactions on control systems technology*, Vol. 21, No. 3, 2012, pp. 831-841.
- [131] Kamal, M. A. S., M. Mukai, J. Murata, and T. Kawabe. Ecological vehicle control on roads with up-down slopes. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 3, 2011, pp. 783-794.
- [132] Frey, H., A. Unal, J. Chen, S. Li, and C. Xuan. Methodology for developing modal emission rates for EPA's multi-scale motor vehicle & equipment emission system. US Environmental Protection Agency, Ann Arbor, Michigan, 2002.
- [133] Tiernan, T., N. Richardson, P. Azeredo, W. G. Najm, and T. Lochrane. Test and evaluation of vehicle platooning proof-of-concept based on cooperative adaptive cruise control. John A. Volpe National Transportation Systems Center (US), 2017.
- [134] Savitzky, A., and M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, Vol. 36, No. 8, 1964, pp. 1627-1639.
- [135] Bujanovic, T., P. Bujanovic, T. Lochrane, J. Hu, and C. M. Walton. Cooperative Adaptive Cruise Control Algorithm with Priority Weights Assigned to Downstream Vehicles for Increased Safety. *Transportation Research Board 97th Annual Meeting*, 2018.
- [136] Weißmann, A., D. Görges, and X. Lin. Energy-optimal adaptive cruise control combining model predictive control and dynamic programming. *Control Engineering Practice*, Vol. 72, 2018, pp. 125-137.
- [137] Zlocki, A., and P. Themann. Methodology for quantification of fuel reduction potential for adaptive cruise control relevant driving strategies. *IET Intelligent Transport Systems*, Vol. 8, No. 1, 2014, pp. 68-75.
- [138] Mahdinia, I., and M. Habibian. Evaluating the Transportation System Performance Based on Efficiency, Effectiveness and Efficacy: A Case Study of the US States. 2017.
- [139] NHTSA. National Highway Traffic Safety Administration. Fatal motor vehicle crashes: overview. Washington, DC: US Department of Transportation, 2017.
- [140] 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.
- [141] Wang, L., H. Zhong, W. Ma, M. Abdel-Aty, and J. Park. How many crashes can connected vehicle and automated vehicle technologies prevent: A meta-analysis. *Accident Analysis & Prevention*, Vol. 136, 2020, p. 105299.
- [142] Wadud, Z., D. MacKenzie, and P. Leiby. Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practice*, Vol. 86, 2016, pp. 1-18.
- [143] Greenblatt, J. B., and S. Shaheen. Automated vehicles, on-demand mobility, and environmental impacts. *Current sustainable/renewable energy reports*, Vol. 2, No. 3, 2015, pp. 74-81.
- [144] Stern, R. E., Y. Chen, M. Churchill, F. Wu, M. L. Delle Monache, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work. Quantifying air quality benefits resulting from few autonomous vehicles

- stabilizing traffic. *Transportation Research Part D: Transport and Environment*, Vol. 67, 2019, pp. 351-365.
- [145] Patella, S., F. Scrucca, F. Asdrubali, and S. Carrese. Carbon Footprint of autonomous vehicles at the urban mobility system level: A traffic simulation-based approach. *Transportation Research Part D: Transport and Environment*, Vol. 74, 2019, pp. 189-200.
- [146] Stogios, C., D. Kasraian, M. J. Roorda, and M. Hatzopoulou. Simulating impacts of automated driving behavior and traffic conditions on vehicle emissions. *Transportation Research Part D: Transport and Environment*, Vol. 76, 2019, pp. 176-192.
- [147] Brown, K. E., and R. Dodder. Energy and emissions implications of automated vehicles in the US energy system. *Transportation Research Part D: Transport and Environment*, Vol. 77, 2019, pp. 132-147.
- [148] Litman, T. *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, Canada, 2017.
- [149] Sinha, A., S. Chand, K. P. Wijayaratna, N. Virdi, and V. Dixit. Comprehensive safety assessment in mixed fleets with connected and automated vehicles: A crash severity and rate evaluation of conventional vehicles. *Accident Analysis & Prevention*, Vol. 142, 2020, p. 105567.
- [150] Hoseinzadeh, N., Y. Liu, L. D. Han, C. Brakewood, and A. Mohammadnazar. Quality of location-based crowdsourced speed data on surface streets: A case study of Waze and Bluetooth speed data in Sevierville, TN. *Computers, Environment and Urban Systems*, Vol. 83, 2020, p. 101518.
- [151] Khattak, A. J., I. Mahdinia, S. Mohammadi, A. Mohammadnazar, and B. Wali. Big Data Generated by Connected and Automated Vehicles for Safety Monitoring, Assessment and Improvement, Final Report (Year 3). *arXiv preprint arXiv:2101.06106*, 2021.
- [152] Mohammadnazar, A., R. Arvin, and A. J. Khattak. Classifying travelers' driving style using basic safety messages generated by connected vehicles: application of unsupervised machine learning. *Transportation Research Part C: Emerging Technologies*, Vol. 122, 2021, p. 102917.
- [153] Sadeghinassr, B., A. Akhavan, and Q. Wang. Estimating Commuting Patterns from High Resolution Phone GPS Data. In *Computing in Civil Engineering 2019: Data, Sensing, and Analytics*, American Society of Civil Engineers Reston, VA, 2019. pp. 9-16.
- [154] Karacasu, M., and A. Er. An analysis on distribution of traffic faults in accidents, based on driver's age and gender: Eskisehir case. *Procedia-Social and Behavioral Sciences*, Vol. 20, 2011, pp. 776-785.
- [155] Wang, W., J. Xi, A. Chong, and L. Li. Driving style classification using a semisupervised support vector machine. *IEEE Transactions on Human-Machine Systems*, Vol. 47, No. 5, 2017, pp. 650-660.
- [156] Deng, C., C. Wu, N. Lyu, and Z. Huang. Driving style recognition method using braking characteristics based on hidden Markov model. *Plos One*, Vol. 12, No. 8, 2017.
- [157] Castignani, G., T. Derrmann, R. Frank, and T. Engel. Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. *IEEE Intelligent Transportation Systems Magazine*, Vol. 7, No. 1, 2015, pp. 91-102.
- [158] Aljaafreh, A., N. Alshabat, and M. S. N. Al-Din. Driving style recognition using fuzzy logic. Presented at 2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012), 2012.
- [159] Lemarié, L., F. Bellavance, and J.-C. Chebat. Regulatory focus, time perspective, locus of control and sensation seeking as predictors of risky driving behaviors. *Accident Analysis & Prevention*, Vol. 127, 2019, pp. 19-27.

- [160] Richer, I., and J. Bergeron. Differentiating risky and aggressive driving: Further support of the internal validity of the Dula Dangerous Driving Index. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 620-627.
- [161] Azimi, G., A. Rahimi, H. Asgari, and X. Jin. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. *Accident Analysis & Prevention*, Vol. 135, 2020, p. 105355.
- [162] Faria, M. V., G. O. Duarte, R. A. Varella, T. L. Farias, and P. Baptista. Driving for decarbonization: Assessing the energy, environmental, and economic benefits of less aggressive driving in Lisbon, Portugal. *Energy Research Social Science*, Vol. 47, 2019, pp. 113-127.
- [163] Rios-Torres, J., J. Liu, and A. Khattak. Fuel consumption for various driving styles in conventional and hybrid electric vehicles: Integrating driving cycle predictions with fuel consumption optimization. *International Journal of Sustainable Transportation*, Vol. 13, No. 2, 2019, pp. 123-137.
- [164] 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, pp. 290-301.
- [165] ---. The Role of Aggressive Driving and Speeding in Road Safety: Insights from SHRP2 Naturalistic Driving Study Data. 2019.
- [166] Wang, X., A. J. Khattak, J. Liu, G. Masghati-Amoli, and S. Son. What is the level of volatility in instantaneous driving decisions? *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 413-427.
- [167] Sharma, A., Y. Ali, M. Saifuzzaman, Z. Zheng, and M. M. Haque. Human factors in modelling mixed traffic of traditional, connected, and automated vehicles. *International Conference on Applied Human Factors and Ergonomics*, Springer, 2017. pp. 262-273.
- [168] Ali, Y., Z. Zheng, and M. M. Haque. Connectivity's impact on mandatory lane-changing behaviour: Evidences from a driving simulator study. *Transportation Research Part C: Emerging Technologies*, Vol. 93, 2018, pp. 292-309.
- [169] Ali, Y., Z. Zheng, M. M. Haque, M. Yildirimoglu, and S. Washington. Understanding the discretionary lane-changing behaviour in the connected environment. *Accident Analysis & Prevention*, Vol. 137, 2020, p. 105463.
- [170] Ali, Y., M. M. Haque, Z. Zheng, S. Washington, and M. Yildirimoglu. A hazard-based duration model to quantify the impact of connected driving environment on safety during mandatory lane-changing. *Transportation Research Part C: Emerging Technologies*, Vol. 106, 2019, pp. 113-131.
- [171] Tanveer, M., F. A. Kashmiri, H. Naeem, H. Yan, X. Qi, S. M. A. Rizvi, T. Wang, and H. Lu. An Assessment of Age and Gender Characteristics of Mixed Traffic with Autonomous and Manual Vehicles: A Cellular Automata Approach. *Sustainability*, Vol. 12, No. 7, 2020, p. 2922.
- [172] Rahmati, Y., M. Khajeh Hosseini, A. Talebpour, B. Swain, and C. Nelson. Influence of Autonomous Vehicles on Car-Following Behavior of Human Drivers. *Transportation Research Record*, 2019, pp. 1-13.
- [173] Zheng, F., C. Liu, X. Liu, S. E. Jabari, and L. Lu. Analyzing the impact of automated vehicles on uncertainty and stability of the mixed traffic flow. *Transportation Research Part C: Emerging Technologies*, Vol. 112, 2020, pp. 203-219.
- [174] Stern, R. E., S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. m. Haulcy, H. Pohlmann, F. Wu, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work.

- Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. *Transportation Research Part C: Emerging Technologies*, Vol. 89, 2018, pp. 205-221.
- [175] Cui, S., B. Seibold, R. Stern, and D. B. Work. Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations. Presented at 2017 IEEE Intelligent Vehicles Symposium (IV), 2017.
- [176] Ye, L., and T. Yamamoto. Evaluating the impact of connected and autonomous vehicles on traffic safety. *Physica A: Statistical Mechanics and its Applications*, Vol. 526, 2019, p. 121009.
- [177] Viridi, N., H. Grzybowska, S. T. Waller, and V. Dixit. A safety assessment of mixed fleets with connected and autonomous vehicles using the Surrogate Safety Assessment Module. *Accident Analysis & Prevention*, Vol. 131, 2019, pp. 95-111.
- [178] Pan, T., W. H. Lam, A. Sumalee, and R. Zhong. Multiclass multilane model for freeway traffic mixed with connected automated vehicles and regular human-piloted vehicles. *Transportmetrica A: Transport Science*, 2019, pp. 1-29.
- [179] Mahdinia, I., R. Arvin, A. J. Khattak, and A. Ghiasi. Safety, Energy, and Emissions Impacts of Adaptive Cruise Control and Cooperative Adaptive Cruise Control. *Transportation Research Record*, 2020, p. 0361198120918572.
- [180] Hoseinzadeh, N., R. Arvin, A. J. Khattak, and L. D. Han. Integrating safety and mobility for pathfinding using big data generated by connected vehicles. *Journal of Intelligent Transportation Systems*, Vol. 24, No. 4, 2020, pp. 404-420.
- [181] Figlewski, S. Forecasting volatility using historical data. 1994.
- [182] Kamrani, M., R. Arvin, and A. J. Khattak. Analyzing highly volatile driving trips taken by alternative fuel vehicles. *arXiv preprint arXiv:1807.03861*, 2018.
- [183] Arvin, R., A. J. Khattak, M. Kamrani, and J. Rio-Torres. Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections. *Journal of Intelligent Transportation Systems*, 2020, pp. 1-18.
- [184] Hayward, J. C. Near miss determination through use of a scale of danger. 1972.
- [185] Bandeira, J., T. G. Almeida, A. J. Khattak, N. M. Rouphail, and M. C. Coelho. Generating emissions information for route selection: Experimental monitoring and routes characterization. *Journal of Intelligent Transportation Systems*, Vol. 17, No. 1, 2013, pp. 3-17.
- [186] FHWA. Federal Highway Administration. Next Generation Simulation: US101 Freeway Dataset., 2007.
- [187] Van Arem, B., C. J. Van Driel, and R. Visser. The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7, No. 4, 2006, pp. 429-436.
- [188] Mousavi, S. M., O. A. Osman, D. Lord, K. K. Dixon, and B. Dadashova. Investigating the safety and operational benefits of mixed traffic environments with different automated vehicle market penetration rates in the proximity of a driveway on an urban arterial. *Accident Analysis & Prevention*, Vol. 152, 2021, p. 105982.
- [189] Delle Monache, M. L., T. Liard, A. Rat, R. Stern, R. Bhadani, B. Seibold, J. Sprinkle, D. B. Work, and B. Piccoli. Feedback control algorithms for the dissipation of traffic waves with autonomous vehicles. In *Computational Intelligence and Optimization Methods for Control Engineering*, Springer, 2019. pp. 275-299.
- [190] NHTSA. Traffic Safety Facts: Overview of Motor Vehicle Crashes in 2019. National Highway Traffic Safety Administration, 2020.

- [191] Zajac, S. S., and J. N. Ivan. Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accident Analysis & Prevention*, Vol. 35, No. 3, 2003, pp. 369-379.
- [192] Wang, J., H. Huang, P. Xu, S. Xie, and S. Wong. Random parameter probit models to analyze pedestrian red-light violations and injury severity in pedestrian–motor vehicle crashes at signalized crossings. *Journal of Transportation Safety & Security*, 2019, pp. 1-20.
- [193] 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.
- [194] Jahangeer, A. A., S. S. Anjana, and V. R. Das. A Hierarchical Modeling Approach to Predict Pedestrian Crash Severity. In *Transportation Research*, Springer, 2020. pp. 355-366.
- [195] Yue, L., M. Abdel-Aty, Y. Wu, J. Yuan, and M. Morris. Influence of pedestrian-to-vehicle technology on drivers' response and safety benefits considering pre-crash conditions. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 73, 2020, pp. 50-65.
- [196] Khattak, Z. H., M. D. Fontaine, B. L. Smith, and J. Ma. Crash severity effects of adaptive signal control technology: An empirical assessment with insights from Pennsylvania and Virginia. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 151-162.
- [197] Jannat, M., S. M. Roldan, S. A. Balk, and K. Timpone. Assessing potential safety benefits of advanced pedestrian technologies through a Pedestrian Technology Test Bed. *Journal of Intelligent Transportation Systems*, Vol. 25, No. 2, 2020, pp. 139-156.
- [198] American-Automobile-Association. Automatic emergency braking with pedestrian detection. No. 20, October 2019.
- [199] Haus, S. H., R. Sherony, and H. C. Gabler. Estimated benefit of automated emergency braking systems for vehicle–pedestrian crashes in the United States. *Traffic injury prevention*, Vol. 20, No. sup1, 2019, pp. S171-S176.
- [200] Rosén, E., and U. Sander. Pedestrian fatality risk as a function of car impact speed. *Accident Analysis & Prevention*, Vol. 41, No. 3, 2009, pp. 536-542.
- [201] Chauvel, C., Y. Page, B. Fildes, and J. Lahaussé. Automatic emergency braking for pedestrians effective target population and expected safety benefits. *23rd international technical conference on the enhanced safety of vehicles (ESV 2013)*, 2013.
- [202] Jeppsson, H., M. Östling, and N. Lubbe. Real life safety benefits of increasing brake deceleration in car-to-pedestrian accidents: Simulation of Vacuum Emergency Braking. *Accident Analysis & Prevention*, Vol. 111, 2018, pp. 311-320.
- [203] Rosén, E., J.-E. Källhammer, D. Eriksson, M. Nentwich, R. Fredriksson, and K. Smith. Pedestrian injury mitigation by autonomous braking. *Accident Analysis & Prevention*, Vol. 42, No. 6, 2010, pp. 1949-1957.
- [204] Heckman, J. J. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 1979, pp. 153-161.
- [205] Heckman, J. J. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, NBER, 1976. pp. 475-492.
- [206] Abbasi, S., J. Ko, and J. Kim. Carsharing station location and demand: Identification of associated factors through Heckman selection models. *Journal of Cleaner Production*, Vol. 279, 2021, p. 123846.
- [207] STATA. User's Guide, StataCorp. Stata: Release 17. Statistical Software. College Station, TX: StataCorp LLC., 2021.

- [208] Hussain, Q., H. Feng, R. Grzebieta, T. Brijs, and J. Olivier. The relationship between impact speed and the probability of pedestrian fatality during a vehicle-pedestrian crash: A systematic review and meta-analysis. *Accident Analysis & Prevention*, Vol. 129, 2019, pp. 241-249.
- [209] IIHS. *Insurance Institute for Highway Safety*, <https://www.iihs.org/>. Accessed 4/9/2021, 2021.
- [210] NHTSA. 2018 Traffic Safety Facts A Compilation of Motor Vehicle Crash Data (Annual Report) 2020.
- [211] Batouli, G., M. Guo, B. Janson, and W. Marshall. Analysis of pedestrian-vehicle crash injury severity factors in Colorado 2006–2016. *Accident Analysis & Prevention*, Vol. 148, 2020, p. 105782.
- [212] Li, Y., and W. Fan. Modelling the severity of pedestrian injury in pedestrian—vehicle crashes in North Carolina: A partial proportional odds logit model approach. *Journal of Transportation Safety & Security*, Vol. 12, No. 3, 2020, pp. 358-379.
- [213] Wali, B., A. J. Khattak, and N. Ahmad. Injury severity analysis of pedestrian and bicyclist trespassing crashes at non-crossings: A hybrid predictive text analytics and heterogeneity-based statistical modeling approach. *Accident Analysis & Prevention*, Vol. 150, p. 105835.
- [214] Liu, J., X. Li, and A. J. Khattak. An integrated spatio-temporal approach to examine the consequences of driving under the influence (DUI) in crashes. *Accident Analysis & Prevention*, Vol. 146, 2020, p. 105742.
- [215] Yasmin, S., N. Eluru, and A. R. Pinjari. Analyzing the continuum of fatal crashes: A generalized ordered approach. *Analytic methods in accident research*, Vol. 7, 2015, pp. 1-15.
- [216] Ma, L., H. Zhang, X. Yan, J. Wang, Z. Song, and H. Xiong. Smooth associations between the emergency medical services response time and the risk of death in road traffic crashes. *Journal of Transport & Health*, Vol. 12, 2019, pp. 379-391.
- [217] Mukherjee, D., and S. Mitra. A comparative study of safe and unsafe signalized intersections from the view point of pedestrian behavior and perception. *Accident Analysis & Prevention*, Vol. 132, 2019, p. 105218.
- [218] Kim, J.-K., G. F. Ulfarsson, V. N. Shankar, and S. Kim. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention*, Vol. 40, No. 5, 2008, pp. 1695-1702.
- [219] Hezaveh, A. M., and C. R. Cherry. Walking under the influence of the alcohol: A case study of pedestrian crashes in Tennessee. *Accident Analysis & Prevention*, Vol. 121, 2018, pp. 64-70.
- [220] Zhai, X., H. Huang, N. Sze, Z. Song, and K. K. Hon. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. *Accident Analysis & Prevention*, Vol. 122, 2019, pp. 318-324.
- [221] Mokhtarimousavi, S., J. C. Anderson, A. Azizinamini, and M. Hadi. Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. *International Journal of Transportation Science and Technology*, Vol. 9, No. 2, 2020, pp. 100-115.
- [222] Chen, Z., and W. Fan. Modeling Pedestrian Injury Severity in Pedestrian-Vehicle Crashes in Rural and Urban Areas: Mixed Logit Model Approach. *Transportation Research Record*, 2019, p. 0361198119842825.
- [223] Tay, R., J. Choi, L. Kattan, and A. Khan. A multinomial logit model of pedestrian–vehicle crash severity. *International Journal of Sustainable Transportation*, Vol. 5, No. 4, 2011, pp. 233-249.

- [224] Haleem, K., P. Alluri, and A. Gan. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accident Analysis & Prevention*, Vol. 81, 2015, pp. 14-23.
- [225] Rahimi, A., G. Azimi, H. Asgari, and X. Jin. Injury Severity of Pedestrian and Bicyclist Crashes Involving Large Trucks. *International Conference on Transportation and Development 2020*, American Society of Civil Engineers Reston, VA, 2020. pp. 110-122.
- [226] Hutchinson, T. P. Factors affecting the times till death of pedestrians killed in road accidents. *Injury*, Vol. 6, No. 3, 1975, pp. 208-212.
- [227] Fife, D. Time from injury to death (survival time) among fatally injured pedestrians. *Injury*, Vol. 18, No. 5, 1987, pp. 315-318.
- [228] Clark, D. E., R. J. Winchell, and R. A. Betensky. Estimating the effect of emergency care on early survival after traffic crashes. *Accident Analysis & Prevention*, Vol. 60, 2013, pp. 141-147.
- [229] Ju, Y., and S. Y. Sohn. Time to death analysis of road traffic accidents in relation to delta V, drunk driving, and restraint systems. *Traffic injury prevention*, Vol. 15, No. 8, 2014, pp. 771-777.
- [230] Clark, D. E., and B. M. Cushing. Predicted effect of automatic crash notification on traffic mortality. *Accident Analysis & Prevention*, Vol. 34, No. 4, 2002, pp. 507-513.
- [231] Lee, J., M. Abdel-Aty, Q. Cai, and L. Wang. Effects of emergency medical services times on traffic injury severity: A random effects ordered probit approach. *Traffic injury prevention*, Vol. 19, No. 6, 2018, pp. 577-581.
- [232] Meng, Q., and J. Weng. Uncertainty analysis of accident notification time and emergency medical service response time in work zone traffic accidents. *Traffic injury prevention*, Vol. 14, No. 2, 2013, pp. 150-158.
- [233] Sánchez-Mangas, R., A. García-Ferrer, A. De Juan, and A. M. Arroyo. The probability of death in road traffic accidents. How important is a quick medical response? *Accident Analysis & Prevention*, Vol. 42, No. 4, 2010, pp. 1048-1056.
- [234] Klein, J. P., and M. L. Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2006.
- [235] Terza, J. V., A. Basu, and P. J. Rathouz. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, Vol. 27, No. 3, 2008, pp. 531-543.
- [236] Hausman, J. A. Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1978, pp. 1251-1271.
- [237] Davidson, R., and J. G. MacKinnon. Estimation and inference in econometrics. *OUP Catalogue*, 1993.
- [238] Fotheringham, A. S., C. Brunsdon, and M. Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.
- [239] Fotheringham, A. S., M. E. Charlton, and C. Brunsdon. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A*, Vol. 30, No. 11, 1998, pp. 1905-1927.
- [240] FARS. Fatality Analysis Reporting System, United States Department of Transportation, National Highway Traffic Safety Administration. <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>.
- [241] 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.
- [242] Eluru, N., C. R. Bhat, and D. A. Hensher. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis & Prevention*, Vol. 40, No. 3, 2008, pp. 1033-1054.

- [243] Hosseinzadeh, A., and R. Kluger. Do EMS times associate with injury severity? *Accident Analysis & Prevention*, Vol. 153, 2021, p. 106053.
- [244] Lahaussé, J. A., B. N. Fildes, Y. Page, and M. P. Fitzharris. The potential for automatic crash notification systems to reduce road fatalities. *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, No. 52, Association for the Advancement of Automotive Medicine, 2008. p. 85.
- [245] Plevin, R. E., R. Kaufman, L. Fraade-Blanar, and E. M. Bulger. Evaluating the potential benefits of advanced automatic crash notification. *Prehospital and disaster medicine*, Vol. 32, No. 2, 2017, pp. 156-164.
- [246] Stickles, J. L., J. M. Kempema, and L. H. Brown. Effect of mobile phone proliferation on crash notification times and fatality rates. *The American journal of emergency medicine*, Vol. 36, No. 1, 2018, pp. 24-26.
- [247] Obenauf, A. W. CONNECTED AND AUTONOMOUS VEHICLES EFFECTS ON EMERGENCY RESPONSE TIMES. 2019.
- [248] Obenauf, A. W., R. R. Souleyrette, R. M. Kluger, and A. Pratelli. IMPACT OF SELF-DRIVING AND CONNECTED VEHICLES ON EMERGENCY RESPONSE: THE CASE OF THE USA AND IMPLICATIONS FOR ITALY. *WIT Transactions on The Built Environment*, Vol. 189, 2019, pp. 101-112.
- [249] Toran Pour, A., S. Moridpour, R. Tay, and A. Rajabifard. Modelling pedestrian crash severity at mid-blocks. *Transportmetrica A: Transport Science*, Vol. 13, No. 3, 2017, pp. 273-297.
- [250] Kwon, O. H., W. Rhee, and Y. Yoon. Application of classification algorithms for analysis of road safety risk factor dependencies. *Accident Analysis & Prevention*, Vol. 75, 2015, pp. 1-15.
- [251] Kashani, A. T., and A. S. Mohaymany. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety Science*, Vol. 49, No. 10, 2011, pp. 1314-1320.
- [252] Lee, C., and X. Li. Predicting driver injury severity in single-vehicle and two-vehicle crashes with boosted regression trees. *Transportation Research Record*, Vol. 2514, No. 1, 2015, pp. 138-148.
- [253] Abdelwahab, H. T., and M. A. Abdel-Aty. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. *Transportation Research Record*, Vol. 1746, No. 1, 2001, pp. 6-13.
- [254] Delen, D., R. Sharda, and M. Bessonov. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis & Prevention*, Vol. 38, No. 3, 2006, pp. 434-444.
- [255] Yu, R., and M. Abdel-Aty. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, Vol. 63, 2014, pp. 50-56.
- [256] Li, Z., P. Liu, W. Wang, and C. Xu. Using support vector machine models for crash injury severity analysis. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 478-486.
- [257] Ahmadi, A., A. Jahangiri, V. Berardi, and S. G. Machiani. Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods. *Journal of Transportation Safety & Security*, Vol. 12, No. 4, 2020, pp. 522-546.
- [258] Mokoatle, M., D. Vukosi Marivate, and P. Michael Esiefarienrhe Bukohwo. Predicting Road Traffic Accident Severity using Accident Report Data in South Africa. *Proceedings of the 20th Annual International Conference on Digital Government Research*, 2019. pp. 11-17.

- [259] Zhang, J., Z. Li, Z. Pu, and C. Xu. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEE Access*, Vol. 6, 2018, pp. 60079-60087.
- [260] Wahab, L., and H. Jiang. A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. *Plos One*, Vol. 14, No. 4, 2019, p. e0214966.
- [261] Iranitalab, A., and A. Khattak. Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 27-36.
- [262] Garg, A., and K. Tai. Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *International Journal of Modelling, Identification and Control*, Vol. 18, No. 4, 2013, pp. 295-312.
- [263] Parsa, A. B., A. Movahedi, H. Taghipour, S. Derrible, and A. K. Mohammadian. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, Vol. 136, 2020, p. 105405.
- [264] Molnar, C. *Interpretable Machine Learning*. Lulu. com, 2020.
- [265] Altmann, A., L. Tološi, O. Sander, and T. Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, Vol. 26, No. 10, 2010, pp. 1340-1347.
- [266] Lundberg, S. M., and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 2017. pp. 4765-4774.
- [267] Mihaita, A.-S., Z. Liu, C. Cai, and M.-A. Rizoio. Arterial incident duration prediction using a bi-level framework of extreme gradient-tree boosting. *arXiv preprint arXiv:1905.12254*, 2019.
- [268] Chen, T., and C. Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016. pp. 785-794.
- [269] Yan, X., M. Ma, H. Huang, M. Abdel-Aty, and C. Wu. Motor vehicle–bicycle crashes in Beijing: Irregular maneuvers, crash patterns, and injury severity. *Accident Analysis & Prevention*, Vol. 43, No. 5, 2011, pp. 1751-1758.
- [270] Salon, D., and A. McIntyre. Determinants of pedestrian and bicyclist crash severity by party at fault in San Francisco, CA. *Accident Analysis & Prevention*, Vol. 110, 2018, pp. 149-160.
- [271] Kim, J.-K., S. Kim, G. F. Ulfarsson, and L. A. Porrello. Bicyclist injury severities in bicycle–motor vehicle accidents. *Accident Analysis & Prevention*, Vol. 39, No. 2, 2007, pp. 238-251.
- [272] Behnood, A., and F. Mannering. Determinants of bicyclist injury severities in bicycle–vehicle crashes: a random parameters approach with heterogeneity in means and variances. *Analytic methods in accident research*, Vol. 16, 2017, pp. 35-47.
- [273] Liu, S., and W. Fan. Investigating Factors Affecting Injury Severity in Bicycle-Vehicle Crashes: A Day-of-Week Analysis with Partial Proportional Odds Logit Models. *Canadian Journal of Civil Engineering*, No. ja, 2020.
- [274] Moore, D. N., W. H. Schneider IV, P. T. Savolainen, and M. Farzaneh. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. *Accident Analysis & Prevention*, Vol. 43, No. 3, 2011, pp. 621-630.
- [275] Meuleners, L. B., A. H. Lee, and C. Haworth. Road environment, crash type and hospitalisation of bicyclists and motorcyclists presented to emergency departments in Western Australia. *Accident Analysis & Prevention*, Vol. 39, No. 6, 2007, pp. 1222-1225.
- [276] He, J., and P. R. Hahn. Stochastic tree ensembles for regularized nonlinear regression. *arXiv preprint arXiv:2002.03375*, 2020.
- [277] Peden, M., R. Scurfield, D. Sleet, A. A. Hyder, C. Mathers, E. Jarawan, A. Hyder, D. Mohan, and E. Jarawan. *World report on road traffic injury prevention*. World Health Organization, 2004.

- [278] Haddon Jr, W. The changing approach to the epidemiology, prevention, and amelioration of trauma: the transition to approaches etiologically rather than descriptively based. *American journal of public health and the Nations health*, Vol. 58, No. 8, 1968, pp. 1431-1438.
- [279] Masoumi, K., A. Forouzan, H. Barzegari, A. A. Darian, F. Rahim, B. Zohrevandi, and S. Nabi. Effective factors in severity of traffic accident-related traumas; an epidemiologic study based on the Haddon matrix. *Emergency*, Vol. 4, No. 2, 2016, p. 78.
- [280] Friedman, J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, Vol. 38, No. 4, 2002, pp. 367-378.
- [281] Shapley, L. S. A value for n-person games. *Contributions to the Theory of Games*, Vol. 2, No. 28, 1953, pp. 307-317.
- [282] Bernhoft, I. M., and G. Carstensen. Preferences and behaviour of pedestrians and cyclists by age and gender. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 11, No. 2, 2008, pp. 83-95.
- [283] Liu, J., S. Jones, E. K. Adanu, and X. Li. Behavioral pathways in bicycle-motor vehicle crashes: From contributing factors, pre-crash actions, to injury severities. *Journal of Safety Research*, 2021.
- [284] Mayer, J. D. Emergency medical service: delays, response time and survival. *Medical care*, 1979, pp. 818-827.
- [285] Gonzalez, R. P., G. R. Cummings, H. A. Phelan, M. S. Mulekar, and C. B. Rodning. Does increased emergency medical services prehospital time affect patient mortality in rural motor vehicle crashes? A statewide analysis. *The American journal of surgery*, Vol. 197, No. 1, 2009, pp. 30-34.
- [286] Doggett, S., D. R. Ragland, and G. Felschundneff. Prehospital Response Time and Traumatic Injury—A Review. 2018.
- [287] Tay, R., U. Barua, and L. Kattan. Factors contributing to hit-and-run in fatal crashes. *Accident Analysis & Prevention*, Vol. 41, No. 2, 2009, pp. 227-233.

VITA

Originally from Iran, Iman Mahdinia grew up in Mashhad, the second populated city in the northeast of Iran. After high school, he attended the Islamic Azad University of Mashhad in Iran and received a Bachelor of Science degree in Civil Engineering. Before graduating with his undergraduate degree, he knew he wanted to attend graduate school. He chose to attend the Amirkabir University of Technology in Tehran, Iran, to pursue a Master of Science degree in Transportation Engineering. He started his Doctor of Philosophy program in Civil Engineering with a concentration in Transportation Engineering at Penn State in August 2017 and joined the University of Tennessee in August 2018 to continue and pursue a doctoral degree in Transportation Engineering and a master's degree in Statistics. His research interest includes connected and automated vehicles, traffic safety, and transportation sustainability. After graduation, he will begin his new position as a postdoctoral position at UTK. He is incredibly grateful for all the support from his family, friends, advisors, and colleagues as he begins his new career.