Understanding Fixed Object Crashes with SHRP2 Naturalistic Driving Study Data

Haiyan Hao

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Civil Engineering

Linbing Wang, Chair
Montasir M. Abbas
Yingfeng Li

June 29th, 2018
Blacksburg, Virginia

Keywords: Fixed-object crash, Naturalistic driving study, Logistic regression analysis, Support vector machine, Artificial neural network

Understanding Fixed Object Crashes with SHRP2 Naturalistic Driving Study Data

Haiyan Hao

ABSTRACT

Fixed-object crashes have long time been considered as major roadway safety concerns. While previous relevant studies tended to address such crashes in the context of roadway departures, and heavily relied on police-reported accidents data, this study integrated the SHRP2 NDS and RID data for analyses, which fully depicted the prior to, during, and after crash scenarios. A total of 1,639 crash, near-crash events, and 1,050 baseline events were acquired. Three analysis methods: logistic regression, support vector machine (SVM) and artificial neural network (ANN) were employed for two responses: crash occurrence and severity level. Logistic regression analyses identified 16 and 10 significant variables with significance levels of 0.1, relevant to driver, roadway, environment, etc. for two responses respectively. The logistic regression analyses led to a series of findings regarding the effects of explanatory variables on fixed-object event occurrence and associated severity level. SVM classifiers and ANN models were also constructed to predict these two responses. Sensitivity analyses were performed for SVM classifiers to infer the contributing effects of input variables. All three methods obtained satisfactory prediction performance, that was around 88% for fixed-object event occurrence and 75% for event severity level, which indicated the effectiveness of NDS event data on depicting crash scenarios and roadway safety analyses.

Understanding Fixed Object Crashes with SHRP2 Naturalistic Driving Study Data

Haiyan Hao

GENERAL AUDIENCE ABSTRACT

Fixed-object crashes happen when a single vehicle strikes a roadway feature such as a curb or a median, or runs off the road and hits a roadside feature such as a tree or utility pole. They have long time been considered as major highway safety concerns due to their high frequency, fatality rate, and associated property cost. Previous studies relevant to fixed-object crashes tended to address such crashes in the contexture of roadway departures, and heavily relied on police-reported accident data. However, many fixed-object crashes involved objects in roadway such as traffic control devices, roadway debris, etc. The police-reported accident data were found to be weak in depicting scenarios prior to, during crashes. Also, many minor crashes were often kept unreported.

The Second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study (NDS) is the largest NDS project launched across the country till now, aimed to study driver behavior or, performance-related safety problems under real-world scenarios. The data acquisition systems (DASs) equipped on participated vehicles collect vehicle kinematics, roadway, traffic, environment, and driver behavior data continuously, which enable researchers to address such crash scenarios closely. This study integrated SHRP2 NDS and roadway information database (RID) data to conduct a comprehensive analysis of fixed-object crashes. A total of 1,639 crash, near-crash events relevant to fixed objects and animals, and 1,050 baseline events were used. Three analysis methods: logistic regression, support vector machine (SVM) and artificial neural network (ANN) were employed for two responses: crash occurrence and severity level.

The logistic regression analyses identified 16 and 10 variables with significance levels of 0.1 for fixed-object event occurrence and severity level models respectively. The influence of explanatory variables was discussed in detail. SVM classifiers and ANN models were also constructed to predict the fixed-object crash occurrence and severity level. Sensitivity analyses were performed for SVM classifiers to infer the contributing effects of input variables. All three methods achieved satisfactory prediction accuracies of around 88% for crash occurrence prediction and 75% for crash severity level prediction, which suggested the effectiveness of NDS event data on depicting crash scenarios and roadway safety analyses.

ACKNOWLEDGEMENT

I am extremely grateful to my advisor, Dr. Linbing Wang for his guidance, inspiration, and encouragement throughout my Master program with his patience and knowledge. I would also like to extend my acknowledgement to Dr. Montasir M. Abbas and Dr. Yingfeng Li for serving as my advisory committee, their support and cooperation is much appreciated. In addition, I want to thank Dr. Hao Chen from VTTI for his effort in reviewing this thesis and giving valuable suggestions. I appreciate Dr. Miguel A. Perez, Lisa Goodell and all other staff at VTTI for their general assistance to make this research running smoothly. Last but not the least, I am extending my sincere thanks to Megha Rao, Yufeng Liu, Wei Liu, Hui Han and all my other friends for their hearty support and priceless friendship. It is so lucky to meet you guys.

DEDICATION

This work is dedicated to my parents for their endless love, support and encouragement.

TABLE OF CONTENTS

ABSTRACT .... **ii**
GENERAL AUDIENCE ABSTRACT .... **iii**
ACKNOWLEDGEMENT .... **iv**
DEDICATION .... **v**
TABLE OF CONTENTS .... **vi**
LIST OF FIGURES .... **ix**
LIST OF TABLES .... **x**
ABBREVIATIONS .... **xi**
Chapter 1 Introduction **1**
1.1 Background 1
1.2 Thesis Objective 3
1.3 Thesis Contribution 3
1.4 Thesis Layout 4
Chapter 2: Literature review **6**
2.1 The Influence of Contributing Factors 6
2.2 The SHRP2 NDS and RID Database 7
2.3 The Analysis Methods 10
2.4 Comparison Between this Study and Previous Fixed-object Crash Studies 11
Chapter 3. Data **14**
3.1 Data Collection 14
3.2 Additional Event Data Collection and Pre-processing 16
3.3 Explanatory Variables 23
3.4 Response Variables 23
Chapter 4. Methodology **25**
4.1 Analysis Scenarios 25
4.2 Odds Ratio 25

4.3 Logistic Regression 26
4.4 SVM Analysis 28
4.5 ANN Analysis 31
4.6 Feature Vector Composition 32
Chapter 5. Logistic Regression Analysis Results 36
5.1 Binary Logistic Regression Analysis Results for SCE Probability 36
5.1.1 Modelling Results 36
5.1.2 Model Goodness-of-Fit Statistics 39
5.1.3 SCE Probability Modelling Findings 40
5.2 Ordinal Logistic Regression Results for SCE Severity Level 44
5.2.1 Modelling Results 44
5.2.2 Model Goodness-of-Fit Statistics 47
5.2.3 SCE Severity Level Modelling Findings 48
5.3 Conclusion 50
Chapter 6. SVM and ANN Analysis Results 51
6.1 Results of SVM Analysis 51
6.1.1 Classification Results 51
6.1.2 Sensitivity Analysis Results 54
6.2 Results of ANN Classification 55
6.3 Conclusion 58
Chapter 7. Fixed-object Crash Case Studies 59
7.1 Fixed-object Crash Case Studies 59
7.2 Conclusion 70
Chapter 8: Conclusions and Recommendations 72
8.1 Effects of Influential Factors on Fixed-object Crashes 72
8.1.1 Effects of Driver Behavior Factors on Fixed-object Crashes 72
8.1.2 Effects of Roadway and Traffic-Related Factors on Fixed-object Crashes 73
8.1.3 Effects of Environment-related Factors on Fixed-object Crashes 75

8.2 Comparison between Logistic Regression and Machine Learning Methods 76

8.3 Recommendations 77

REFERENCE 80

APPENDIX A. Detail Table of Explanatory Variables 84

APPENDIX B: SAS Output 88

SAS output for binomial logit regression model.. 88

SAS output for ordinal logit regression model 93

LIST OF FIGURES

Figure 1. Traffic fatalities in fixed-object crashes, 2007-2016 2
Figure 2. Mean cost for fixed-object Crashes 2
Figure 3. Crash Involvement and Fatality Caused by Struck Object Type 3
Figure 4. A Schematic Illustration of Thesis Organization 5
Figure 5. SHRP2 Participants by Gender and Age Group 8
Figure 6. An Example of Miss Matching Event due to Densely Located Streets. 18
Figure 7. An Example of Matching Erratic GPS Points to Correct Route 19
Figure 8. Examples of Events Occurred on Roadways with No RID Data. 20
Figure 9. Example of Events Occurred at Intersection 21
Figure 10. Distribution of Events for Fixed-object Crashes Analysis 24
Figure 11. Hyperplane and Margins for an SVM Trained with Two Classes 30
Figure 12. ANN Basic Structure 31
Figure 13. A diagram of k-fold cross validation, with k=5. 34
Figure 14. ROC Curves for Binary SVM Classifiers with Quadratic Kernel 53
Figure 15. ROC Curves for Multi-Class SVM Classifier with Linear Kernel 53
Figure 16. Sensitivity Analysis Results of Binary SVM Classifier 54
Figure 17. Sensitivity Analysis Results of Multi-class SVM Classifier 55
Figure 18. An Example of Neural Network Structure with Two Hidden Layer 56
Figure 19. ROC Curves for ANN models 57

LIST OF TABLES

Table 1. Data Categories Collected in the SHRP2 Project 9
Table 2. SCE Distributed by Event Nature 15
Table 3. Typical 2x2 Frequency Table for Calculating Odds Ratios. 26
Table 4. Significant Explanatory Variables for SCE Probability Modelling............................... 37
Table 5. Parameters and Odds Ratios of Significant Variables for SCE Probability Modelling.. 38
Table 6. Goodness of Fit Measures for Binary Logistic Regression Model................................ 40
Table 7. Testing Global Null Hypothesis: β=0 for the Binary Logistic Regression Model. 40
Table 8. Model Prediction Results with Original Data for Binary Logistic Regression Model... 40
Table 9. Significant Explanatory Variables for SCE Severity Level Modelling. 45
Table 10. Parameters and Odds Ratios of Significant Variables for SCE Severity Level Modelling 46
Table 11. Model Fitness Measures for the Ordinal Logistic Regression Model. 47
Table 12. Testing Global Null Hypothesis: β=0 for the Ordinal Logistic Regression Model. 47
Table 13. Confusion Matrix of Ordinal Event Severity Prediction. 48
Table 14. Classification Accuracy of SVM Classifiers with Different Kernels 51
Table 15. Confusion Matrix of Binary SVM Classifier with Quadratic Kernel 52
Table 16. Confusion Matrix of Multi-Class SVM Classifier with Linear Kernel 52
Table 17. Classification Accuracy of ANN Classifiers with Different Structures 56
Table 18. Confusion Matrix of ANN Classifier for SCE, Baseline Classification....................... 57
Table 19. Confusion Matrix of ANN Classifier for SCE Severity Classification 57
Table 20. Prediction Accuracy for Different Methods 76

ABBREVIATIONS

AADT: Annual Average Daily Traffic
AIC: Akaike's Information Criterion
ANN: Artificial Neural Network
AUC: Area Under Curve
CART: Classification and Regression Tree
DAS: Data Acquisition System
DOT: Department of Transportation
FARS: Fatality Analysis Reporting System
FPR: False Positive Rate
GES: General Estimate System
GPS: Global Positioning System
IRI: International Roughness Index
LOS: Level of Service
MCA: Multiple Correspondence Analysis
MLE: Maximum Likelihood Estimate
NDS: Naturalistic Driving Study
OR: Odds Ratio
PII: Personal Identifying Information
RID: Roadway Information Database
ROC: Receiver Operating Characteristic
SC: Schwarz Criterion
SCE: Safety Critical Event
SHRP2: The Second Strategic Highway Research Program
SVM: Support Vector Machine
TBI: Traumatic Brain Injury
TPR: True Positive Rate

Chapter 1 Introduction

1.1 Background

Fixed-object crashes happen when a single vehicle strikes a roadway feature such as a curb or a median or runs off the road and hits a roadside feature such as a tree or utility pole. They have long time been considered as major highway safety concerns for state transportation agencies, insurance companies, and the automotive industries due to below characteristics:

- Severe outcome

  According to FARS (Fatality Analysis Reporting System)'s 2015 report, fixed-object crashes accounted for 14.7% (or 92600) of all crashes in nationwide but resulted in 30.9% (or 9939) fatal crashes (*1*). In addition to the high fatality rate, fixed-object crashes are also more likely to cause other severe diseases such as traumatic brain injury (TBI) and disability (*2*).

- High associated cost

  Compared to other crash types, fixed-object crashes cause large amounts of monetary losses including property damages and hospital expenses. As illustrated in Figure 2, the average costs of colliding with trees and utility poles can reach as high as $90,000, while colliding with other objects are also relatively expensive. The high associated costs of fixed-object crashes bring significant burden to the insurance industry (*3*).

- High frequency

  The fixed-object crash is among the most common crash types in the nation. Figure 1 shows the number and percentage of traffic fatalities involved in fixed-object crashes between 2007 and 2016 (*4*, *5*). It clearly illustrates that the fatalities due to fixed-object crashes as a percentage of total fatalities have not decreased much over the last 10 years.

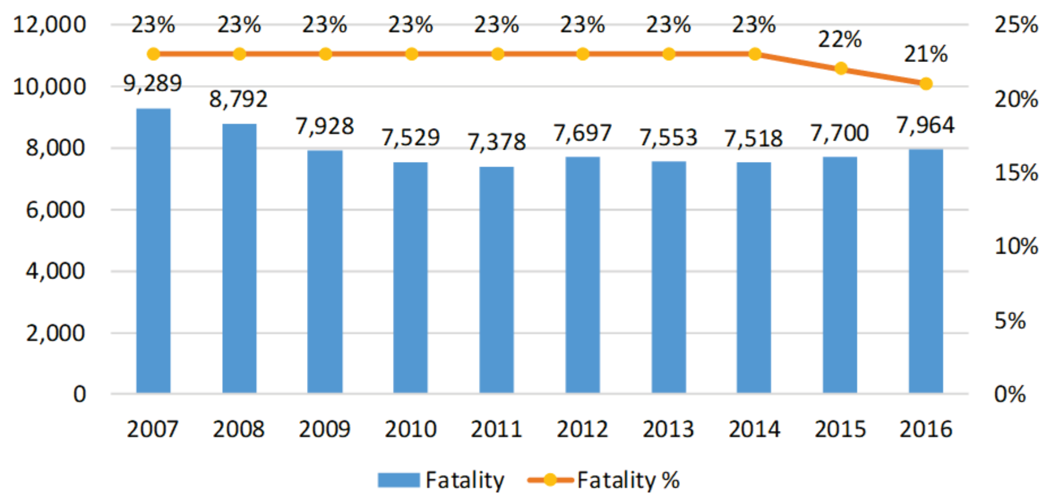


Figure 1. Traffic fatalities in fixed-object crashes, 2007-2016 (*4*)

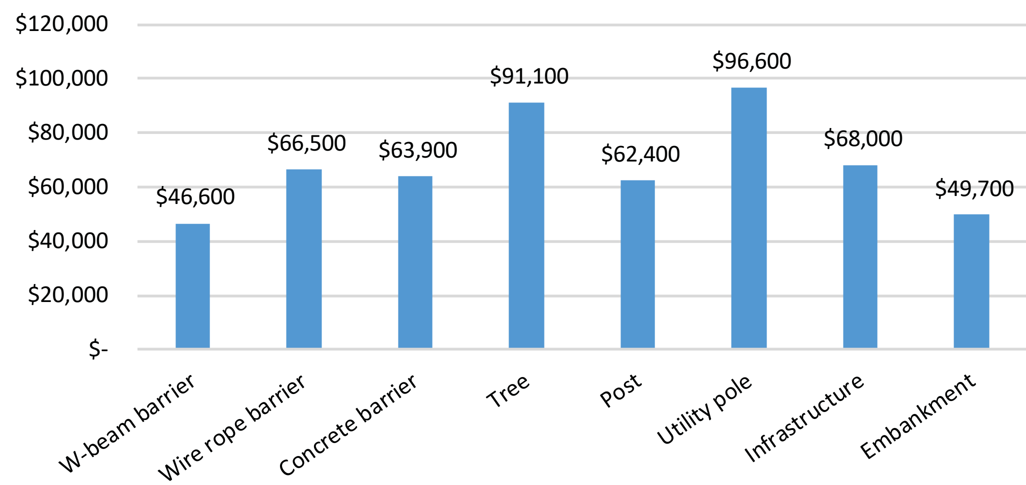


Figure 2. Mean cost for fixed-object Crashes (*3*)

Fixed-object crashes involve a variety of object types on roadway and roadside. The consequences of colliding with different objects are also different. According to 2015 FARS report, culverts/curbs/ditches, poles/posts, and trees were among the most frequently struck object types, and they were responsible for over 60% of fatalities when combined (Figure 3) (*1*).

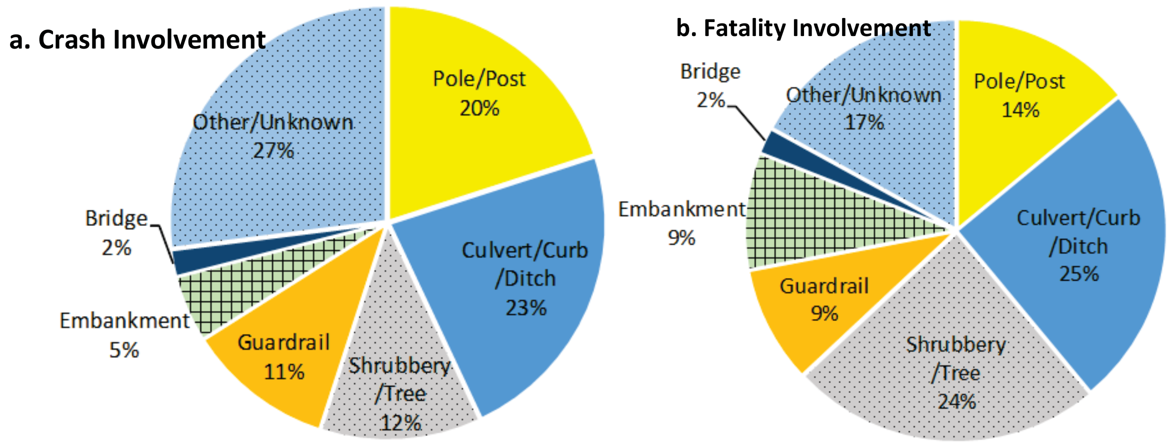


Figure 3. Crash Involvement and Fatality Caused by Struck Object Type (*1*)

1.2 Thesis Objective

The objective of this thesis is to perform a comprehensive analysis to understand the effects of risk factors relevant to environment, roadway, traffic, driver, etc., on fixed-object crashes. In order to achieve this objective, SHRP2 NDS and RID database are integrated for analysis, and three analysis methods: logistic regression, support vector machine (SVM), and artificial neural network (ANN) are employed for two responses: fixed-object event occurrence and severity level.

1.3 Thesis Contribution

The findings of this thesis will probably contribute to below aspects:

- The detailed risk factors of fixed-object crashes can provide information for driver educators, roadway designers, and policymakers regarding what exposure conditions contribute to fixed-object crash occurrence and more severe outcome.

- The identified types of fixed objects and the roadway characteristics commonly involved in fixed-object crashes can help transportation planners and agencies to optimize the roadway design, traffic control, improve public transportation safety and minimize the hardware lifecycle costs.

- A comparison will be drawn between the results of traditional logit regression methods and machine learning methods used in this study, which may help other transportation safety analysts on method selection.

1.4 Thesis Layout

This thesis consists of eight chapters. Chapter 1 introduces what is fixed-object crash and why it should be studied, followed by thesis objectives, contributions, and layout. In Chapter 2, a thorough literature review is conducted including contributing factors influencing fixed-object crash occurrence or severity level, the background of SHRP2 NDS project and developed databases, the analysis methods commonly employed for transportation safety problem, and a comparison between the characteristics of this study and previous studies relevant to fixed-object crashes. Chapter 3 documents detailed steps of event selection, data acquisition, and preprocessing steps before the analyses. Chapter 4 records the three analysis methods and tools involved in this study. Chapter 5 documents the analysis results and corresponding findings of logit regression analyses, and Chapter 6 documents the analysis results and findings of two machine learning analyses. In Chapter 7, the author conducts several case studies to illustrates typical fixed-object event scenarios identified in this study, how events happen and how the drivers perform during the event. In Chapter 8, a conclusion of this study is drawn, followed by recommendations based on the analysis findings.

Figure 4 schematically illustrates the analysis work conducted in this thesis, which involves two databases: SHRP2 NDS and RID databases, three analysis methods: logit regression, SVM and ANN, and two responses: event occurrence and severity level.

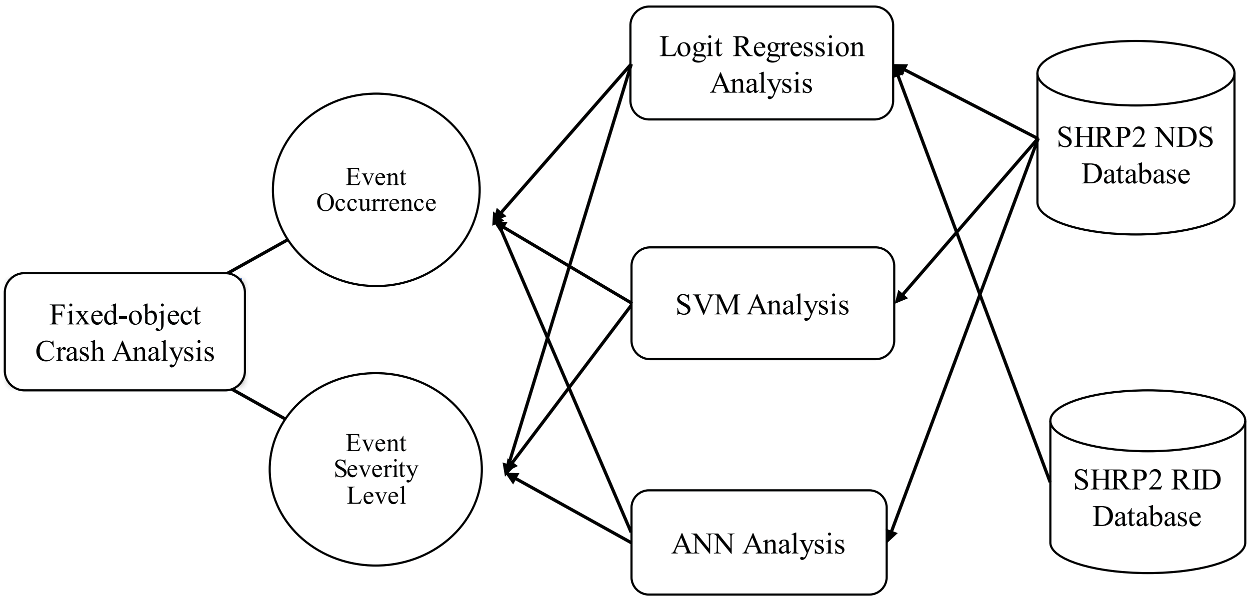


Figure 4. A Schematic Illustration of Thesis Organization

Chapter 2: Literature review

2.1 The Influence of Contributing Factors

Many researches have been carried out to understand fixed-object crashes considering their importance to transportation safety. Previous fixed-object crash analysis works addressed the effect of influential factors including environment, roadway, driver, vehicle, etc., on crash frequency and/or severity level. The identified influence was as following:

- Driver-related factors. Fixed-object crashes can be attributed to a variety of factors, but driver performance error related factors should be the most influential ones. Driver impairments such as alcohol involvement, drowsiness, and distraction were widely studied to associate with increased risk levels of fixed-object crashes. Improper driver maneuvers such as inappropriate braking, steering, and speeding were also identified to increase the probability of fixed-object crashes (*5, 6*, *7*, *8*). Seatbelt usage, however, was found to decrease the driver injury severity in fixed-object crashes (*9*). Other driver-related factors such as driver age and gender were also found to influence fixed-object crash severity level.

- Roadway-related factors. It was found that fixed-object crashes were more likely to occur on roadway segments with lower functional classifications, flat and rolling terrains (as against hilly terrain), asphalt pavement (as against concrete pavement), unpaved shoulder, and no median barriers (*10*). Undivided rural roadways with curved alignment and high posted speed limits were particularly prone to roadway departures that could lead to crashes involving fixed objects (*6*). In addition, intersections and narrower clear zones also increased the probability of more severe fixed-object crashes (*11*).

- Environment-related factors. Some studies showed that the installation of roadway lighting was significantly correlated with decreases in fixed-object crash severity. Adverse weather conditions such as foggy, snowy, and rainy were often correlated with increased crash severity level (*11*). Other studies, however, showed little association between crash severity and weather, lighting conditions (*9*).

- Vehicle-related factors. Studies showed that heavy trucks (as against passenger cars) and vehicle function failures (e.g. ineffective brakes, blown tires) were frequently associated with severe fixed-object crash outcome, while activation of airbags tended to reduce crash severity (*9, 11, 12)*.

- Struck object type. Many works identified trees and utility poles as the most hazardous fixed objects involved in fixed-object crashes (*5*, *13*). The varied mean cost showed in Figure 2 also suggested the different outcome when conflicting with different object types. Studies found that colliding with objects such as guardrail ends, bridge rails, trees and utility poles tended to increase the probability of fatal injury while colliding with guardrail face, concrete barrier associated with lower driver injury severity (*11*).

2.2 The SHRP2 NDS and RID Database

The Second Strategic Highway Research Program (SHRP2) launched the Naturalistic Driving Study (NDS) between 2010 and 2013 to investigate the driving conditions under real-world scenarios (*14*). With more than 3,500 participant drivers recruited from all age groups (see Figure 5), the study collected a massive amount of naturalistic driving and related data at six sites across the country (see Table 1 for detailed data categories):

- Bloomington, Indiana;
- Buffalo, New York;
- Central Pennsylvania;
- Durham, North Carolina;
- Seattle, Washington; and
- Tampa Bay, Florida.

Figure 5. SHRP2 Participants by Gender and Age Group (*14*)

SHRP2 NDS used an onboard data acquisition system (DAS) for roadway condition, vehicle kinematic, and driver behavior data collection. The DAS consists of a forward radar, four video cameras (monitoring driver's face, hand, passenger side, forward and rear roadway respectively), accelerometers, Global Positioning System (GPS) receivers, computer vision lane-tracking capability, and data storage equipment (*15*).

The final NDS database contained information of more than 5 million vehicle miles and 41,000 events. SHRP2 defined three event types including crashes, near-crashes, and baseline events. Baseline events were sample epochs of trips that represented normal driving conditions and typical driver behaviors. Crashes and near-crashes in the NDS database are also collectively referred to as safety critical events (SCEs). At the time when the study was conducted, the NDS database contained 8,758 SCEs, 19,998 balanced baseline events stratified by participants and proportion of time that the driver drove above 5 mph, and 12,583 additional baseline events. Researchers can access the front view videos and event detail table of all these events through SHRP2 Insight website (https://insight.shrp2nds.us) with IRB training certification (*14*).

In conjunction with the NDS database, SHRP2 also developed the Roadway Information Database (RID) with relatively detailed traffic and roadway information for the six NDS sites (*16*). The SHRP2 RID incorporated both data collected by instrumented vehicles on selected routes, and existing roadway data from government, public, and private sources. The linkages between the

SHRP2 driving and roadway data allow researchers to effectively identify driving data on particular road segments of interest (*17*).

Table 1. Data Categories Collected in the SHRP2 Project (*18*)

Participant Assessments • Demographic Questionnaire • Driving History • Driving Knowledge • Medical Conditions and Meds • ADHD Screening • Risk Perception • Frequency of Risky Behavior • Sensation Seeking Behavior • Sleep Habits • Visual, Physical, and Cognitive Test Results • Exit Interview
Vehicle Information • Make, Model, Year, Body, Style • Vehicle Condition (tires, battery, etc.) • Safety and Entertainment Systems.
Continuous Data • Face, Forward, Rear, and Instrument Panel Video • Vehicle Network Data • Accelerometers/Gyros, Forward RADAR, GPS • Additional Sensor Data.
Trip Summary Data • Characterization of Trip Content • Start Time and Duration of Trip • Min, Max, Mean Sensor Data • Time and Distance Driven at Various Speeds, Headways • Vehicle Systems Usage
Event Data • Crashes, Near Crashes, Baseline Events • 30s Events with Classifications • Post-Crash Interviews • Other Crash Data
Cell Phone Records • Subset of Participant Drivers • Call Time and Duration • Call Type (Text, Call, Pic, etc.)
Roadway Data • Matching trip GPS to roadway database • Roadway classifications • Other roadway data.

2.3 The Analysis Methods

Statistical analysis methods have long time played a predominant role in transportation safety analyses. According to the different variable types and data structures, different analysis methods can be employed. Recently, machine learning methods such as SVM and ANN gain increased attention of transportation safety analysts. Many studies have been conducted to evaluate the application of machine learning methods on crash analysis.

Descriptive analyses quantitatively summarize the dataset. Souleyrette et al. in his work calculated the mean crash rate (number of crashes per million vehicle mile) of roadway segments with different facilities such as barrier, shoulder, curvature alignment, and grade. Tukey's t test was conducted to find facilities causing significant differences in the mean crash rate (*10*). Takemoto et al. used descriptive analysis to identify factors attributing to fixed-object crashes, where the author briefly summarized the crash counts for different casual factors (*8*). Many state agency reports used descriptive statistic to present the crash data for its excellent ability to collect, organize and compare data in a manageable form (*1*).

Different from descriptive analyses, logistic regression analyses are inferential statistical methods and are frequently employed for transportation safety analysis. Considering the different explanatory and response variable types, they can take the form of linear logistic regression for continuous responses (*6*), binary logistic regression for dichotomous responses (*9, 12,* 19), multinomial logistic regression for multiple, discrete response variables (7), ordered logistic regression when the response variables are discrete and numerically ordered, and nested logistic regression when some response variables should be nested for analysis (*11*). Some works also used probit regression instead of logit regression. For example, Yamamoto and Shankar presented an ordered-response probit model for fixed-object collision severity level analysis (*20*).

Odds ratio is also an effective method to identify causal factors by comparing the crash odds with and without the particular exposure condition (*21*). Some previous studies involving SHRP2 NDS database used odds ratio for its simplicity and effectiveness (*22*, *23*).

Both logit regression and odds ratio are parametric analysis methods. However, the parametric methods can suffer from many limitations. For example, parametric analyses make assumptions about the population distribution and the relationship between response and independent variables, require the independence of independent variables, and usually only involve a small number of variables. Once some of these basic assumptions are violated, erroneous estimations and incorrect inferences could be produced. Hence, in recent transportation analysis works, machine learning methods such as CART (*24*), MCA (multiple correspondence analysis) (*25*) and SVM (*26*, *27*) have gained increased attention from transportation safety analysts. Delen et al. compared four methods including ANN, SVM, decision tree and logistic regression for an injury severity analysis and showed that the machine learning methods had comparable or even better prediction accuracy than traditional parametric analysis methods (*28*).

2.4 Comparison Between this Study and Previous Fixed-object Crash Studies

When reviewing previous analysis works relevant to fixed-object crashes, it has been found that most of them share one or more following characteristics:

- Most previous fixed-object crash analyses tended to address such crashes in the context of roadway departure. Thus the analysis results would lead to recommendations of roadway treatment such as preventing vehicles from running off the road, or installation of roadside barriers, advanced warning signs and wider clear zones. Although a large proportion of fixed-object crashes did involve roadside objects, many fixed-object crashes involved with objects on the roadway such as sign structures, traffic control devices, and debris. Debris on roadways, for example, contributed to over 200,000 police-reported crashes and resulted in nearly 39,000 injuries and 500 deaths between 2011–2014 (*29*).

- Most previous transportation safety analyses heavily relied on police-reported crash data, which were usually provided by state Department of Transportation (DOT), the nationwide databases such as General Estimate System (GES), FARS, local transportation agencies, insurance companies, etc. However, many minor crashes were often kept unreported. Such minor crashes, although did not cause significant injuries or property damages, could still provide important insights on how fixed-object crashes occur. In addition, the analyses based on police reports or insurance reports were also limited to post-crash information, leaving out critical information of real-time driver behavior and reaction prior to and during the crashes.

- The traditional data sources only provided crash data, hence previous works often used descriptive analysis, or drew comparisons between crash severity levels or different crash types due to the unknown normal driving conditions. Some studies used crash rate (number of crashes versus traveled miles) as the response variable. However, this type of analyses mainly addressed roadway characteristics factor, i.e. it was possible to acquire the accumulated traveled mileage for a road segment, but it was difficult to acquire such accumulated traveled mileage in different environmental conditions or with different driver behavior.

In this thesis, the SHRP2 NDS and RID database were integrated to study the fixed-object crashes, which perfectly solved above stated problems. Firstly, the struck objects were identified from real-time videos and driver narratives, which were not limited to roadside objects, but also involved objects lying on roadways. It should be noted that though animals were often categorized as "not fixed" objects in transportation reports, we also took them into account considering that many fixed-object crashes or roadway departures happen after the driver maneuver to avoid colliding with animals. Note that all the "fixed object crashes" mentioned in the following content referred to collisions with either fixed objects or animals. Secondly, the DASs could capture every minor crash occurrence, and led to a large number of low-level conflictions and even near crashes. Besides, the video cameras monitored outside roadway and inside cabin continuously, which could adequately depict the prior to, during and post-crash scenarios. Moreover, the sampled baseline events provided researchers an insight of normal driving conditions. Researchers were able to understand the prevalence of particular exposure factors on crash occurrence by comparing their frequencies in the SCEs and baseline events.

The author also noticed that there were other naturalistic driving studies conducted other than the SHRP2 NDS project such as 100 car naturalistic driving studies, the precursor of SHRP2 NDS project (*30*), PROLOGUE and UDRIVE projects launched in Europe. But most of them used machines to simulate the real driving environment or only captured a small number of events. None of them was comparable with the SHRP2 NDS project in data variety, participant numbers, and project length. Even so, the number of collected crash, near-crash events was still far less than the number of accidents used in previous studies, which the data were provided by traditional databases such as state DOT, GES, FARS, etc. Also, the crash severity level distribution in this study was different from what previous studies defined. In addition, as SHRP2 data were reduced by trained reductionists, biases may exist regarding event causes, severities.

Chapter 3. Data

3.1 Data Collection

SHRP2 NDS is the largest naturalist driving study project launched across the country till now, aimed to study the driver behavior or performance-related safety problem under real-world scenarios. As of today, over 3,500 participants recruited from six states were involved in the project, and more than 5 million vehicle miles, 41,000 crash, near-crash, and baseline events have been collected (*14, 15*).

This study used SHRP2 NDS and RID data, firstly SCEs relevant to fixed objects were filtered using below criteria:

- Event Severity = Crash, near-crash and
- Event Nature = Conflict with parked vehicle; conflict with animal; conflict with obstacle/object in roadway and single vehicle conflict

The initial data request resulted in 2,304 crashes and near crashes. Then further the author eliminated the following events from analysis:

- Events where subject vehicles were parking or starting in a parking lot. These events typically involved slow moving vehicles colliding with parked cars and parking lot appurtenances such as parking blocks, columns/walls, and bollards. These events represent a different causation nature than that of fixed-object crashes involving moving vehicles.

- Events when the subject vehicle was backing. Most of these events occurred within parking lots, with the rest occurring at driveways or streets with on-street parking spots.

- Single vehicle conflicts that did not involve fixed objects. For example, events involving vehicles losing control, rotating in the road, or stopping due to vehicle failure.

At the end of this process, the author kept a total of 1,639 relevant events as the study group. Although the focus of this project was fixed-object crashes, the author included crashes involving animals in the study as well. Animal-related crashes frequently involved high-speed vehicles and could cause roadway departures leading to secondary crashes involving roadside objects. The distribution of SCE samples by selected event natures was recorded in Table 2. In addition, 1050 participant balanced baseline events were also acquired as the control group.

Table 2. SCE Distributed by Event Nature

Event Nature	Crash	Near Crash	Total
Single vehicle conflict*	901	192	1093
Conflict with obstacle/object in roadway	58	65	123
Conflict with animal	65	301	366
Conflict with parked vehicle	5	52	57
Total	**1029**	**610**	**1639**

*: Single vehicle conflict here referred to events involving a single vehicle running off the road, or a single vehicle conflicting with roadside objects.

For each selected SCE and baseline event, the author acquired the following datasets in the SHRP2 NDS database:

- SHRP2 event data, including event detail table and the associated video files. The event detail table contained detailed and comprehensive information about each event collected by SHRP2 data reductionists based on the original DAS sampled data.

- Time series data, including variables depicting the travel speed, GPS location, acceleration of each analyzed SHRP2 event. SHRP2 team provided time series data of 30 seconds for SCEs and 10 seconds for baseline events. Most time series data were sampled at a 1 Hz frequency, excluding the network speed sampled at 10 Hz frequency.

- Driver age from the SHRP2 driver demographic questionnaire data for all analyzed SHRP2 events.

- Driving history questionnaire data including average annual mileage, years of driving, and the number of previous violations for drivers involved in all analysis events.

- RID data including roadway alignment, speed limit, number of lanes, IRI, and AADT for locations where the analyzed SHRP2 events took place. The RID data were provided in GIS format.

3.2 Additional Event Data Collection and Pre-processing

For the purpose of this project, the author collected a number of additional data elements that were not previously available in the SHRP2 NDS database, and matched RID data with the event data using a GIS-based process.

Critical speed

As the vehicle speed varied throughout each event, critical speed was used to gauge the pre-event traveling speed of the analyzed event. SHRP2 NDS database includes approximately 30-second epochs for SCEs and 10-second for baseline events. The critical speed was taken as the average speed for 10 seconds baseline events, and the maximum speed for the following three timestamps of each SCE (*14*):

- Event start: the timestamp that was identified as the point of time in the video defining the beginning of crash, near-crash events.
- Subject reaction start: the timestamp at which point the driver was seen to recognize and begin to react to the event, defined as the first change in facial expression or the first movement of a body part in a way indicating the awareness and/or the start of an evasive maneuver, whichever occurred first. The reaction time was often coded after the impact proximity for low-level tire strikes and coded as "-1" when the driver took no evasive maneuver or no reaction to the collision.
- Impact proximity: the timestamp at which point the subject vehicle physically contact with the object of conflict. In the case of a near crash, it is the timestamp when the subject vehicle is at the closest distance to the object of conflict. If more than one incident type occurs, the impact proximity timestamp is coded for the most severe or the first incident type if both are of the same severity.

Driver reaction time

Driver reaction time was the time spent by the subject driver to make an evasive maneuver, which was determined as the time difference between impact proximity timestamp and subject reaction start timestamp. In the cases that drivers reacted after striking the object, or did not react to the events, the driver reaction time was coded as 0.

Struck object type

Front view videos capture lots of information of event scenarios. Each video of the 1639 SCEs was viewed carefully by the author, information such as struck object type, size, distance, etc. was retrieved from videos.

GIS matching

The integration of NDS and RID database was achieved by linking the location variables (longitudes and latitudes) provided in time series data with the roadway feature classes of six states provided in GIS format. During the matching process, the author faced the following challenges and considerations.

- Events were linear instead of point features. The author obtained latitudinal and longitudinal coordinates for the study events along with the associated time series data. The coordinates, however, were provided for the entire lengths of events at a frequency of 1 Hz. The lengths of events varied significantly, with many lasting approximately 30 seconds. At a speed of 30 mph, a 30-second epoch translates to a distance traveled of 1,320 ft. Through this entire distance, subject vehicles frequently changed roadways (especially for events at intersections) or traversed multiple roadway sections with different traffic and roadway characteristics.

- Some events took place at locations with densely located streets. Urban areas frequently had densely located roadways and side streets. Divided roadways, in addition, were represented by two closely located lines due to the two different directions of traffic. Multiple roadways closely located in the vicinity of an event resulted in challenges to correctly locate the event on the right roadway solely based on spatial relationships. In the case presented in Figure 6, the trip located on a divided roadway. The trip was supposed to be matched to the top street as the vehicle traveling from east to west. However, as there was no corresponding RID data collected for top street, the trip was automatically matched to the bot street, which was a missing matching.

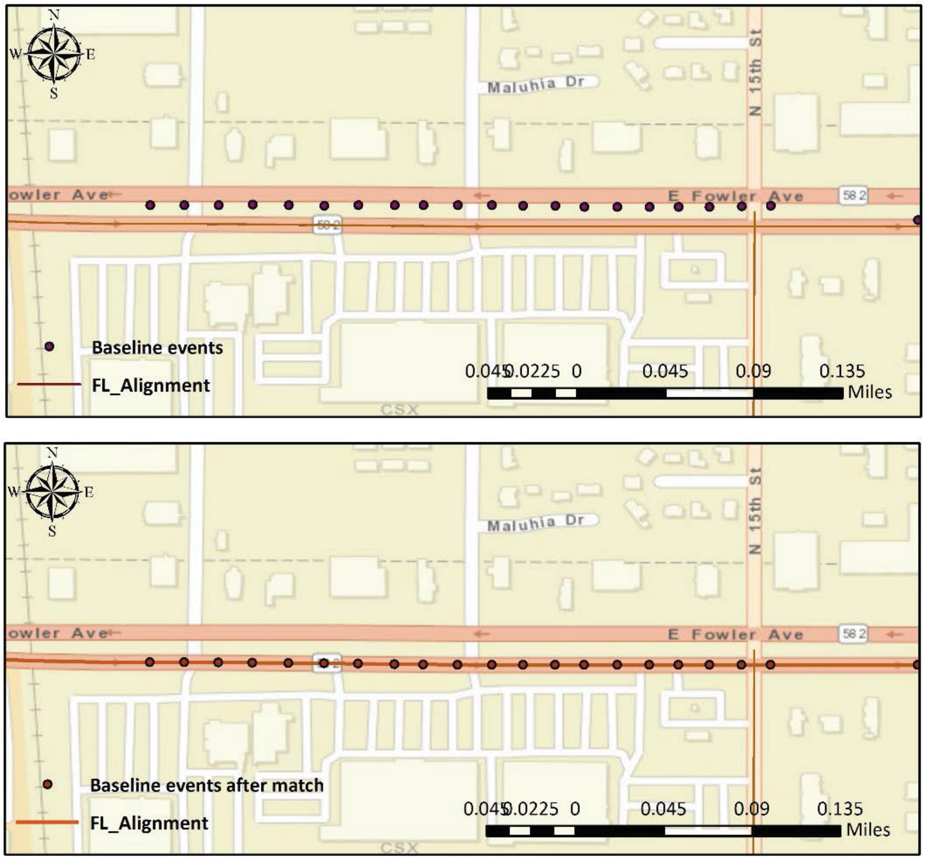


Figure 6. An Example of Miss Matching Event due to Densely Located Streets.

- The NDS GPS sensors in some cases performed erratically near bridges, tunnels, or highrise buildings; or when traveling at low speeds. The erratic GPS coordinates for certain data points of some events added further challenges for the data matching. For example, Figure 7 exhibited a trip with erratic sampled GPS points, that were matched to the correct route with the GIS-aided approach.

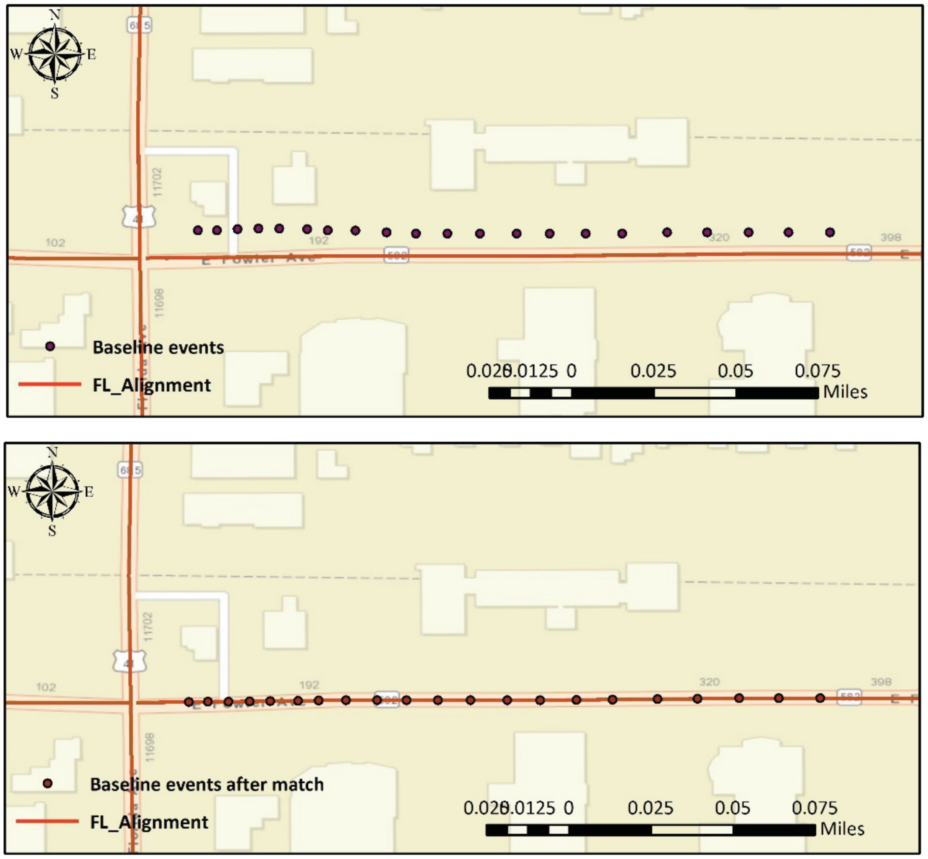


Figure 7. An Example of Matching Erratic GPS Points to Correct Route

- RID did not include data for many rural and local roadways. RID included data from two sources: state transportation agency-generated data (e.g., AADT and speed limit) and SHRP2-collected data (e.g., curvature). State generated data only included detailed information on state-maintained roadways. Roadway data collected by the NDS team only covered a relatively small mileage of roadways with focuses on roadways with higher NDS trip densities (e.g., freeways and major arterials) in areas in and surrounding major metropolitan regions. Figure 8 showed two events that occurred in urban areas, but on roadways where RID data were not collected.

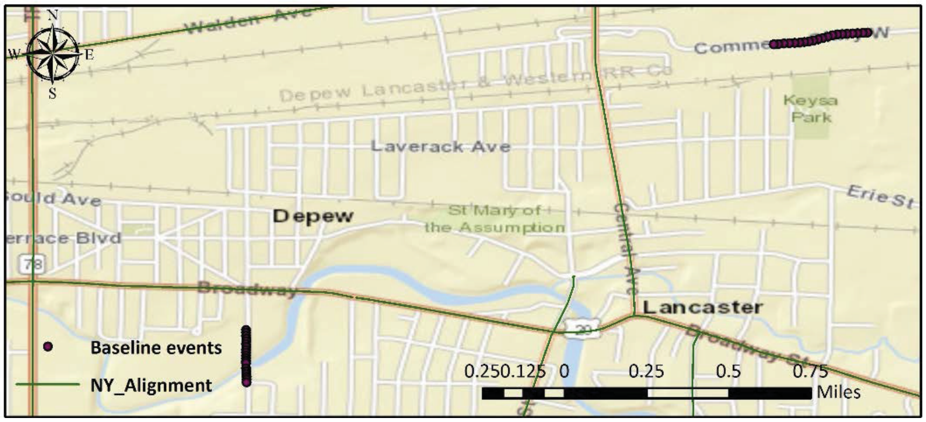


Figure 8. Examples of Events Occurred on Roadways with No RID Data.

- Many events happened at or near intersections. Events occurring at or near intersections traverse multiple roadways with different traffic and roadway characteristics. Different portions of the event epochs would be located on different roadways strictly following spatial relationships. Figure 9 presented a trip that vehicle traversed at an intersection, where the GPS points were matched in two separate routes. In this case, we used the route that corresponded to more GPS points as the determined route, as it explained the prior crash driving conditions more.

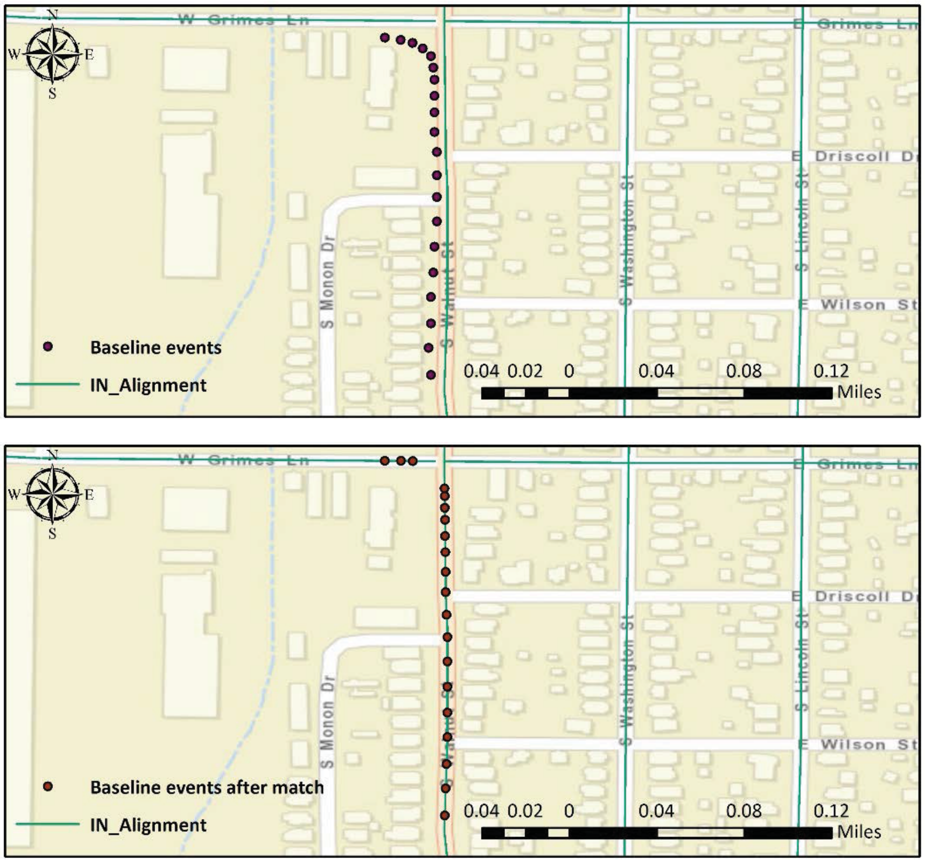


Figure 9. Example of Events Occurred at Intersection

To address the aforementioned challenges, the author used the following GIS-based approach to match the data:

- Identify data points representing event locations. For crash and near-crash events, the author used the coordinates between the last data point before the event start timestamp and the first data point after the impact proximity timestamp, which in most cases corresponded to a period between 1 to 5 seconds (2 to 6 GPS data points at the frequency of 1 Hz). Considering the low data sampling rate, 1 Hz, of GPS devices in DAS, linear interpolation is used to make the vehicle trajectory more clearly when displaced in GIS software. The majority of baseline events were approximately 10 seconds long and therefore all data points for each baseline events were used to represent their locations.

- Insert intermediate location points for crash and near-crash events. To improve the number of data points representing the location of each crash or near crash, the author inserted nine intermediate points between each pair of adjacent GPS points based on a linear relationship and equal distance. The larger number of points helped the spatial join process to better determine the roadway segment corresponding to the largest portion of each event.

- Map event data points into feature classes. The author used the Display XY data tool in the ArcGIS® software to perform this task.

- Join events with RID data. The author performed this task using the ArcGIS Near tool which spatially matched each event data point with the nearest roadway feature within a 75ft. search radius. This process was performed for the multiple RID data layers containing the needed roadway and traffic information and by state. Note that the same data layers in RID for the six NDS sites contained inconsistent data formats, including in particular different attribute names for the same variables. This inconsistency required the data matching process to be performed separately by state.

- Determine correct matches of roadway segments for events. For each event, the author selected the roadway segment to which the most number of data points were matched. This segment was considered as the location where the event took place (e.g., the roadway segment corresponding to the largest proportion of the event). This step ensured that each event was only matched to one link segment.

- Perform quality assurance/quality control (QA/QC). After the automated data matching process, the author performed a QA/QC by manually reviewing the matching results on the ArcGIS platform with original event locations and roadway feature classes overlaid. This process particularly focused on events at locations that were more likely to be subject to the previously described challenges.

At the end of this process, the author was able to match RID roadway and traffic information with 1,538 events including 694 SCEs and 844 baseline events. The RID database did not contain the needed data for the remaining events.

Personal Identification Information (PII) removal

In order to protect study participants, all data that can be used to identify or potentially identify the participants should be reduced or removed from analyses. GPS locations are such PII data, which are viewed in secure data enclave and reduced for analysis (e.g. AADT of 13,121 is reduced to an interval: 10,000 to 15,000).

3.3 Explanatory Variables

With the above data acquisition and pre-processing, 33 explanatory variables in 3 major categories including environment, driver, roadway were obtained from different datasets. See Appendix A. for the detail table of explanatory variables regarding their definitions, variable values, and frequencies.

3.4 Response Variables

For fixed-object crash, near-crash occurrence analysis, the response variable was a binary value with the 1639 SCEs coded as "1" and 1050 baseline events coded as "0". For event severity level analysis, the severity level categorization of SHRP2 NDS project was adopted, that was four severity levels for crash events. In addition, near-crash events were treated as level 5 events (*14*):

- Level 1 most severe crashes: crashes that involve airbag deployments, vehicle rollover, high Delta V, vehicle towing or that cause any injury of drivers or other road users. If the injury present, it should be sufficient to warrant a doctor's visit, including those self-reported and those observed from video files. A high Delta V is defined as changes in speed of the subject vehicle in any direction during impact greater than 20 mph (excluding curb strikes) or acceleration on any axis greater than +/-2g (excluding curb strikes).

- Level 2 police reportable crashes: crashes that do not meet for Level 1 crash criteria, but cause sufficient property damage that is police reportable (minimum of $1500 worth of damage, as estimated from video), or crashes that reach an acceleration on any axis greater than +/-1.3g (excluding curb strikes), or there is a police report noted for the crashes. Most large animal strikes and signage strikes are considered Level 2.

- Level 3 minor crashes: crashes which physically impact with another object, but with minimal damage that do not meet the Level 1 and Level 2 requirement. Most roadway departures, small animal strikes, all curb and tire strikes potentially in conflict with oncoming traffic, and other curb strikes with an increased risk element (e.g., would have resulted in worse had curb not been there, usually related to some driver behavior or state.)

- Level 4 low-level tire strikes: crashes that only involve minimal tire strikes with little/no risk element (e.g., clipping a curb during a tight turn). The distinction between level 3 and level 4 is that level 3 crashes would lead to worse conditions if the curb not been there, but level 4 crashes would not due to limited risk present.

- Level 5 near crashes: events that require a rapid evasive maneuver by the subject driver or any other road users. The maneuver performed should not be pre-meditated and during the evasion, no physical contact is made with other objects, and also the maneuver must not result in a roadway departure.

The distribution of the events used in this study is presented in Figure 10.

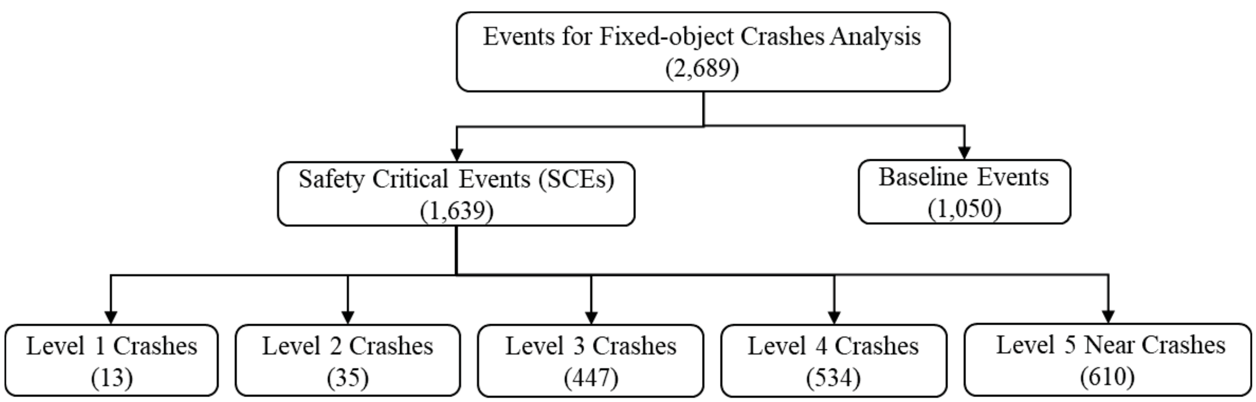


Figure 10. Distribution of Events for Fixed-object Crashes Analysis

Chapter 4. Methodology

4.1 Analysis Scenarios

To understand factors that significantly contribute to fixed-object crashes and quantify risks due to such factors, the author considered the following scenarios for the data analysis:

- Logistic regression analyses
  - Binary logistic regression for SCE occurrence.
  - Ordinal logistic regression for SCE severity.

- SVM analyses
  - Binary SVM classifier for SCE and baseline classification.
  - Multi-class SVM classifier for five event severity classification.

- ANN analyses
  - ANN model with two output nodes for SCEs and baseline events prediction.
  - ANN model with five output nodes for five event severity prediction.

This study performed logistic regression analyses using SAS® Studio software package (*31*), and performed the SVM and ANN analyses with the MATLAB® Classification Learner Toolbox (*32*) MATLAB® Neural Network Toolbox ™ (*33*), respectively. The following sections in this chapter introduced the methodology of these methods in a more detailed level.

4.2 Odds Ratio

Odds ratios measure the association between explanatory variables and responses, which are calculated by comparing the odds of event occurrence with and without the particular exposure condition. An odds ratio greater than one suggests that the factor contributes to the event occurrence. Odds ratios are calculated with below two-by-two frequency table (Table 3) (*21*).

Table 3. Typical 2x2 Frequency Table for Calculating Odds Ratios.

	Exposure Status		
Outcome Status		With	Without
	Yes	a	b
	No	c	d

$$\mathrm{OR} = \frac{a/c}{b/d}$$

Where:

a = Number of exposed cases;

b = Number of exposed non-cases;

c = Number of unexposed cases; and

d = Number of unexposed non-cases.

The 95% confidence interval for a given set of observations can be calculated as:

$$95\%\mathrm{CI} = \exp(\mathrm{In}(\mathrm{OR}) \pm 1.96 \times \mathrm{sd})$$

Where sd is the standard deviation and calculated with the following equation:

$$\mathrm{sd} = (\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d})^{0.5}$$

In the context of logistic regression, the odds ratio can be correlated with parameter estimates (β_i) as following:

$$\mathrm{OR} = e^{\beta_i}$$

4.3 Logistic Regression

Binary logit regression models the log odds of SCE probability as a response of a variety of explanatory variables:

$$\mathrm{logit}(P(\mathrm{Y} = 1)) = \log\left(\frac{P(\mathrm{Y} = 1)}{1 - P(\mathrm{Y} = 1)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

or

$$P(\mathrm{Y}=1)=\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)}}$$

Where:

P is the probability of SCE occurrence;

β_0 is the intercept; and

β_i is the estimated parameters for explanatory factor X_i.

Similarly, ordinal logistic regression takes the following form for SCE severity outcome:

$$\mathrm{logit}(P_j)=\log\left(\frac{P_j}{1-P_j}\right)=\beta_{0j}+\beta_1 X_1+\cdots+\beta_k X_k, j=1,2,3,4,5$$

or

$$P_j=\frac{1}{1+e^{-(\beta_{0j}+\beta_1 X_1+\cdots+\beta_k X_k)}}, j=1,2,3,4,5$$

Where:

j is the event severity level;

P_j is the probability of SCE occurrence with severity level $\leq j$;

β_{0j} is the intercept for SCE severity level $\leq j$; and

β_i is the estimated parameters for explanatory factor X_i.

Hence the probability for each response level is:

$$Prob(\mathrm{Y}=1)=P_1$$

$$Prob(\mathrm{Y}=2)=P_2-P_1$$

$$Prob(\mathrm{Y}=3)=P_3-P_2$$

$$Prob(\mathrm{Y}=4)=P_4-P_3$$

$$Prob(\mathrm{Y}=5)=1-P_4$$

This study uses SAS® Studio software to conduct the above statistical analysis. All significant models and variables were selected at a 0.1 level of significance. Note that SAS also outputs the odds ratios for logistic regression models. However, odds ratios are calculated by making a contrast of the investigated value to reference level.

SAS use maximum likelihood estimation (MLE) to calculate the model parameters, and give the following statistics evaluating the goodness of model fitness:

- -2 log likelihood, AIC (Akaike's Information Criterion), SC (Schwarz Criterion). The log likelihood statistic was used to test the global null hypothesis that all parameters (β) associated with covariates were zero. The AIC and SC statistics adjusted the -2 log likelihood statistic for the number of terms in the model and the number of observations used. These statistics are used when comparing different models developed for the same data and lower values of these statistics indicate a model with better goodness-of-fit (*31*, *34*).

- Likelihood ratio test, Score test, Wald test. These three statistics test the global null hypothesis that $\beta=0$. The likelihood ratio test compares the deviation of the log likelihood function of models with estimated parameters to the global null hypothesis that all parameters (β) associated with covariates were zero. The Wald test is based on normal distribution theory. The score test is obtained as the value of first score function (first derivative of the likelihood function) at β. Higher values of these statistics indicate less probability of null hypothesis, hence better model goodness-of-fit. It should be mentioned that Wald test is also used in testing the significance of single parameters, where the comparison is made between maximum likelihood estimate of slope parameter β_i and standard error (*31*, *34*).

It should be noted that effort was also paid to use multinomial (non-ordinal) logistic regression to model the event severity outcomes. For this study, multinomial models had evidently less satisfactory goodness-of-fit statistics, and calculated parameters were of less significance, which may be due to the limited sample size of severe crash events.

4.4 SVM Analysis

SVM is a machine learning method primarily developed to perform binary classification tasks with hyperplanes constructed in a multidimensional space that separate cases of different class labels

(*35*). It can be easily extended to multi-class classifiers using methods such as one-against-all and one-against-one (*36)*:

- One-against-all methods construct k SVM classifiers for k categorical classes, where the mth SVM is trained with all the example in mth class with positive labels and all other examples with negative labels.
- One-against-one method constructs $\frac{k(k-1)}{2}$ classifiers, where each classifier is trained on data from two different classes (*36*).

In this study, two SVM classifiers, a binary classifier and multi-class classifier, are constructed for fixed-object crashes occurrence and severity level classification respectively, the following describes the basic theories and procedures for the SVM analyses.

In a binary SVM classifier, each selected event is represented as a vector $\{(x_i, y_i), i = 1, \ldots, N\}$. Where y_i is the class label and x_i is the feature vector used to describe the event. A hyperplane $w \cdot x + b = 0$ is learned from the training data so that:

- Samples with labels $y = +1$ and $y = -1$ are located on different sides of the hyperplane;
- The distance of the closest samples to the hyperplane on each side is maximized.

These closest events are also called support vectors, and the distance between support vectors to the hyperplane is the optimal margin, as Figure 11 shown. The optimal margin can be easily determined, that is $\frac{2}{\|w\|}$. Hence the problem of maximizing optimal margin can be turned into:

$$\text{Minimization}\|w\|, \text{ subject to } y_i(w \cdot x_i - b) \geq 1.$$

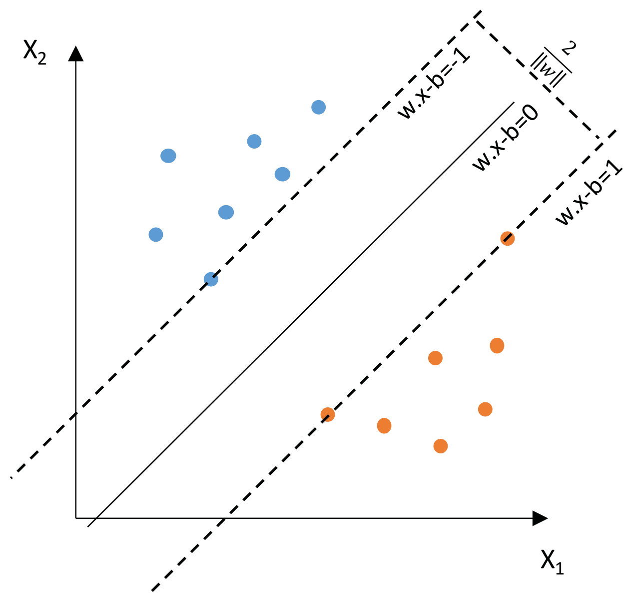


Figure 11. Hyperplane and Margins for an SVM Trained with Two Classes (*37*)

In many cases, samples cannot be linearly separated, or it is hard to find a hyperplane to perfectly separate all samples in different sides. The first problem can be solved by adopting transform kernels (*36*). Some of the most common adopted kernels include:

- Linear: $K(x_i, x_j) = x_i^T x_j$.
- Polynomial: $K(x_i, x_j) = (x_i^T x_j + 1)^d, d > 0$, where d is the order of the polynomial.
- Radial basis function (RBF): $K(x_i, x_j) = exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0$, where γ is a parameter controlling the influence of a single training example.
- Gaussian: $K(x_i, x_j) = exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j\|^2\right), \sigma > 0$. Notice that the Gaussian kernel is a special form of RBF kernel with γ replaced by $\frac{1}{2\sigma^2}$, and σ is also referred as kernel width parameter.

For the second problem, a slack variable ξ_i can be adopted to allow a certain degree of misclassification, Then the hyperplane for two classes become $w \cdot x + \mathrm{b} = \pm(1 - \xi_i)$. The optimized hyperplane can be determined based on a trade-off between maximizing the margin range and minimizing the misclassification rate by solving the following constrained optimization problem:

$$\text{Minimization: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}\xi_i, \text{ subject to } y_i(w\cdot x_i - b) \geq 1 - \xi_i, i = 1, \dots, k, \xi_i > 0.$$

Where C is a regularization parameter used to control the magnitude of allowing violation of the margin. It defines the trade-off between the rate of misclassification in the training data and the maximization of margin. In practice, the regularization parameter is selected by trial and error. Finally, when a new feature vector of testing event input, the classification result is defined as:

$$\mathrm{f(x) = sgn(w \cdot x + b)}$$

4.5 ANN Analysis

ANN is also a frequently used machine learning method, that can also be adopted for transportation safety analysis. A simple ANN typically comprises (Figure 12):

- An input layer inputs feature vectors;
- An output layer makes predictions for classes; and
- One or more hidden layers consisting of hidden nodes connecting nodes in other layers with weighted connections.

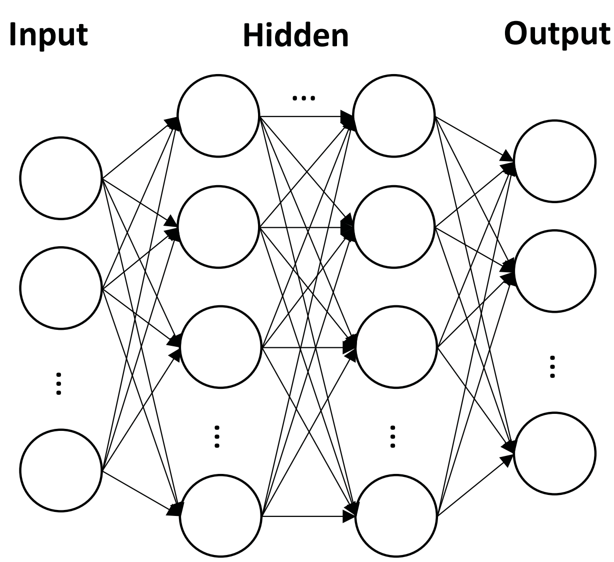


Figure 12. ANN Basic Structure

For each hidden node, the weighted sum of predecessor activations: $z = \sum weight \cdot input + bias$, will be transferred with an activation function, and then the outcome is connected to the node in next layer with corresponding weights. Some of the frequently used activation functions include:

- Step function: $if\ z > threshold, then\ A = 1, else\ A = 0$
- Linear function: $A = c \cdot z$
- Sigmoid function: $A = \frac{1}{1+e^{-z}}$
- Tanh function: $A = \frac{1}{1+e^{-2z}} - 1$
- Rectified Linear Unit (ReLU): $A = \max(0, z)$

During the training process, the weight and biases in the network will be updated in each iteration. The goal of training an ANN is to find weights and biases which minimize the error between desired output and actual output, as below cost function showed:

$$\text{Minimization: } C(w, b) = \frac{1}{2n} \sum \|y(x) - a\|^2$$

Where w is the weights, b is the biases, n is the number of training samples, $y(x)$ is the desired output, and a is the actual network output.

A common technique used to solve such minimization problems is gradient descent, which enforced the updated weights and biases could reduce the cost function for each iteration. And backpropagation method could further accelerate the ANN training process by subtracting or adding weights/biases with the amount proportional to error gradients (*38*).

4.6 Feature Vector Composition

Both SVM and ANN analyses need transferring variables into feature vectors for training and testing. In accordance with logistic regression analyses, the response variables for SVM analyses were:

- Binary classifier: -1 for baseline events and 1 for SCE.

- Multi-class classifier: from 1 to 5 for five crash severity levels defined in Chapter 3. Note that though numerical numbers were assigned, there was no ordinal relationship between the responses.

As for ANN classifiers, the response variables were dealt as vectors as following:

- ANN model for SCE occurrence: [1 0] for SCEs and [0 1] for baseline events.

- ANN model for SCE severity level
    - [1 0 0 0 0] for level 1 most severe crashes;
    - [0 1 0 0 0] for level 2 police-reportable crashes;
    - [0 0 1 0 0] for level 3 physical contact crashes;
    - [0 0 0 1 0] for level 4 low risk tire strikes crashes; and
    - [0 0 0 0 1] for level 5 near crashes.

The composition of feature vectors for explanatory variables follows below rules:

- Binary explanatory variables were coded as 0 and 1, e.g., Weather:
    - 1 representing driving in adverse weather; and
    - 0 representing driving in common weather.

- Ordinal explanatory variables were assigned with discrete numbers, e.g., Age group:
    - 1 representing the age of the driver is between 16-19,
    - 2 representing the age of the driver is between 20-29, and so on.

- Categorical values were assigned with a vector, e.g., Lighting
    - [1 0 0 0] representing daytime driving,
    - [0 1 0 0] representing driving in dawn or dusk,
    - [0 0 1 0] representing driving on an unlighted road in nighttime, while
    - [0 0 0 1] representing driving in a lighted road in nighttime.

- Continuous explanatory variables used its original value, e.g., speed: 70.12 km/h

The two machine learning analyses did not include RID data. This was because that the feature vector composition cannot deal with missing values. As aforementioned, the RID data elements such as AADT, speed limit, and alignment were missing for a large number of events. The problem of missing explanatory variables could be easily solved by using the "missing" statement in SAS software. The final constituted feature vector consisted of 85 and 99 numerical features for SCE occurrence and SCE severity level, respectively. The training, testing, and validation processes were fulfilled with MATLAB toolboxes.

During this study, the author used validation to avoid overfitting. Overfitting is a modelling error when the trained model is able to fit the limited training data closely, but would fail to predict new input data. A k-fold cross validation, for example, partitions the entire dataset into k roughly equal subsets. During each iteration, one subset is used for validation purpose, while the other $k-1$ subsets are used as the training sets. This process is repeated k times and the averaged estimation of the k repetitions will be taken as the final result (*39*). Figure 13 graphically illustrates the cross validation process with k =5.

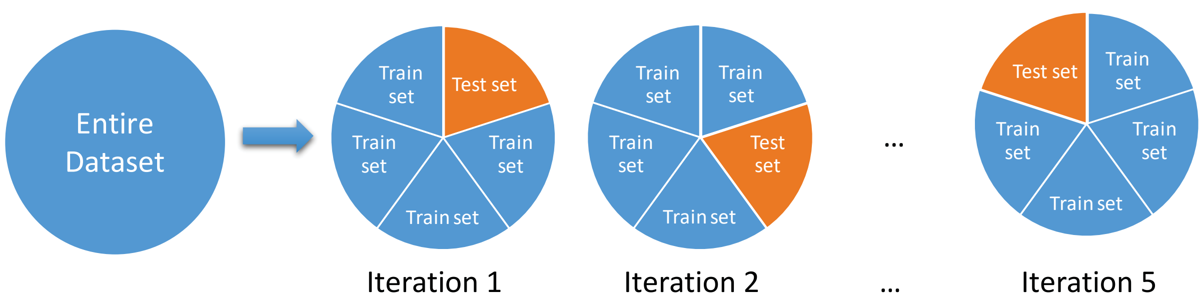


Figure 13. A diagram of k-fold cross validation, with k=5 (*39*).

Machine learning methods such as SVM and ANN are often criticized for their black-box like performance, which means that no interpretable parameters for each explanatory variables will be generated as compared to logit regression models. In order to mitigate this deficiency, some researchers (*28*, *40*, *41*) proposed sensitivity analysis for machine learning methods to infer the correlations between predictors and responses. The original sensitivity analysis proposed by those researchers evaluated the output variances with investigated variables added by a user-defined amount of changes, while other variables remained unchanged. This study adopted that sensitivity analysis concept to demonstrate the effects of explanatory variables. However, instead of introducing changes to existing features, the author removed individual input variables from the model and trained new classifiers. The changes in predictive performance of the new classifiers were then used to quantitatively measure the impacts of removed variables.

Chapter 5. Logistic Regression Analysis Results

5.1 Binary Logistic Regression Analysis Results for SCE Probability

5.1.1 Modelling Results

Binary logistic regression was used to model the effect of risk factors on the probability of an event being a SCE as against a baseline event. Table 4 listed the significant variables and associated statistical test results that were included in the binary logistic regression model for SCE probability. Table 5 further listed the estimated parameters and odds ratios of the significant variables and variable values for the model. In the modelling process, a reference level was assigned for each unordered categorical variable. Parameter estimates and odds ratios of other values were calculated by making the comparison to the reference level. The reference levels were primarily determined to reflect normal or optimal driving conditions. In addition, Table 5 only displayed values with a p-value less than 0.1 for unordered variables. Unordered variable values with a p-value larger than 0.1 were excluded here although the associated variables were tested to be significant. A detailed table of estimated parameters of all variable values could be found in Appendix B. In Table 5, the odds ratios and parameters for ordinal variables were calculated by making the comparison to the value that is one level lower, hence the interpretation of ordinal values should be accumulated over the lower levels. For continuous variables, the odds ratios and parameters reflect the influence of one unit increase of that variable on fixed-object crashes.

Table 4. Significant Explanatory Variables for SCE Probability Modelling

Effect	DF	Wald Chi-square	Pr > ChiSq
Pre-incident Maneuver	5	34.3704	<.0001
Maneuver Judgment	3	16.0199	0.0011
Driver Behavior	9	39.7489	<.0001
Passenger Existence	1	2.9601	0.0853
Secondary Task	10	53.0622	<.0001
Hands on the Wheel	2	10.1661	0.0062
Lighting	3	13.6342	0.0034
Surface Condition	1	3.4882	0.0618
Traffic Density	4	21.9477	0.0002
Contiguous Travel Lanes	5	13.7778	0.0171
Traffic Control	4	9.851	0.043
Grade	3	6.5556	0.0875
Locality	7	36.7197	<.0001
Critical Speed	1	60.5405	<.0001
AADT	8	76.1996	<.0001
IRI	6	46.0758	<.0001

DF = Degree of freedom;
AADT = Average annual daily traffic;
IRI = International roughness index.

Table 5. Parameters and Odds Ratios of Significant Variables for SCE Probability Modelling

Variable	Values	Parameter	Chi-square	Pr > ChiSq	Odds Ratio (95% CI)
Intercept		13.9110	0.0004	0.9846	
Driver Behavior Factors					
Pre-Incident Maneuver	Changing lanes	0.9423	7.1918	0.0073	2.566 (1.289-5.109)
	Going straight - unintentional "drifting"	1.3933	5.4315	0.0198	4.028 (1.248-13.002)
	Making a turn	2.2989	20.7979	<.0001	9.963 (3.709-26.759)
	Going straight (accelerate, decelerate, constant speed)	Reference			
Maneuver judgment	Unsafe and illegal	1.1902	4.4044	0.0358	3.288 (1.082-9.991)
	Unsafe but legal	2.4294	13.6215	0.0002	11.352 (3.124-41.243)
	Safe and legal	Reference			
Driver Behavior	Avoiding animal or other vehicle	3.6471	18.6589	<.0001	38.362 (7.332-200.715)
	Failed to signal, improper signal	1.8436	2.7283	0.0986	6.319 (0.709-56.33)
	Improper turn	4.4753	17.9293	<.0001	87.82 (11.065-697.012)
	Other	0.9841	3.5164	0.0608	2.675 (0.956-7.483)
	Sign, signal violation	1.671	2.8814	0.0896	5.318 (0.772-36.615)
	None	Reference			
Secondary Task	Adjusting/monitoring vehicle devices	0.878	4.1228	0.0423	2.406 (1.031-5.615)
	Personal hygiene	0.8268	4.1371	0.042	2.286 (1.031-5.071)
	Reaching, moving object in vehicle	3.3542	41.7081	<.0001	28.624 (10.343-79.218)
	No secondary tasks	Reference			
Hands on the Wheel	None or at least one hand off	-1.907	8.9612	0.0028	0.149 (0.043-0.518)
	Both hands	Reference			
Critical Speed	-	-0.0244	89.946	<.0001	0.976 (0.971-0.981)
Passenger Existence	Yes	-0.3237	2.9601	0.0853	0.723 (0.5-1.046)
	No	Reference			
Roadway and Traffic Variables					
Traffic Density	LOS A2	-0.6545	13.1572	0.0003	0.52 (0.365-0.74)
	LOS B	-0.3569	3.4302	0.064	0.7 (0.48-1.021)
	LOS D/E/F	-3.8469	10.1675	0.0014	0.021 (0.002-0.227)
	LOS A1	Reference			
Contiguous Travel Lanes	1	-13.2558	0.0003	0.9854	NA
	2	0.1449	0.1332	0.7151	1.156 (0.531-2.517)
	3	-0.6208	8.6733	0.0032	0.537 (0.356-0.812)
	4	0.0858	0.1052	0.7456	1.090 (0.649-1.830)
	5 and 5+	-0.2246	0.5641	0.4526	
	0	Lowest level			
Traffic Control	Other	1.757	5.5092	0.0189	5.795 (1.336-25.132)
	Sign control	0.5333	4.3504	0.037	1.705 (1.033-2.814)
	No traffic control	Reference			
Grade	Dip, hillcrest	1.3454	3.5741	0.0587	3.84 (0.952-15.492)
	Grade Up	0.3754	3.0712	0.0797	1.456 (0.957-2.215)
	Level	Reference			
Locality	Business/Industrial	0.8187	13.3666	0.0003	2.268 (1.462-3.517)
	Open Country	1.084	5.4653	0.0194	2.956 (1.191-7.336)
	Residential area	1.3483	35.9477	<.0001	3.851 (2.478-5.984)

Variable	Values	Parameter	Chi-square	Pr > ChiSq	Odds Ratio (95% CI)
	School/church/playground	0.7997	6.4612	0.011	2.225 (1.201-4.122)
	Interstate/Bypass/Divided-no signals	Reference			
Annual Average Daily Traffic (AADT)	≤5,000	-1.2394	21.66	<.0001	0.290 (0.172-0.488)
	5,000-10,000	0.0306	0.0088	0.9254	1.031 (0.543-1.956)
	10,000-15,000	-0.1253	0.1268	0.7218	0.882 (0.443-1.759)
	15,000-20,000	0.2026	0.2411	0.6234	1.225 (0.545-2.750)
	20,000-30,000	0.2493	0.3321	0.5644	1.283 (0.550-2.996)
	30,000-50,000	0.2295	0.3212	0.5709	1.258 (0.569-2.782)
	50,000-100,000	0.6368	2.1943	0.1385	1.890 (0.814-4.390)
	≥100,000	1.2827	8.9248	0.0028	3.605 (1.555–8.367)
	Missing	Lowest level			
International Roughness Index (IRI, in/mile)	≤20	1.4429	10.9066	0.001	4.233 (1.798-9.966)
	20-50	-2.6842	17.539	<.0001	0.068 (0.019-0.240)
	50-100	0.4296	0.6645	0.4150	1.537 (0.547-4.317)
	100-150	-0.00318	0.0001	0.9922	0.997 (0.527-1.887)
	150-200	0.9646	3.6997	0.0544	2.624 (0.982-7.011)
	>200	1.2233	3.4779	0.0622	3.398 (0.940-12.291)
	Missing	Lowest level			
Environmental Factors					
Lighting	Darkness, lighted	0.6621	11.0242	0.0009	1.939 (1.312-2.866)
	Darkness, not lighted	0.459	4.5099	0.0337	1.583 (1.036-2.417)
	Daylight	Reference			
Surface Condition	Icy/snowy/wet	0.3492	3.4882	0.0618	1.418 (0.983-2.046)
	Dry	Reference			

Note. LOS = level of service.
NA in odds ratio presented when one cell of the two-by-two calculation table is zero.

5.1.2 Model Goodness-of-Fit Statistics

The model optimization and variable selection in SAS were based on maximum likelihood estimation (MLE). SAS provided a number of model fitness statistics including -2 Log Likelihood, AIC, SC (Table 6). It should be mentioned that these goodness-of-fit measures are generally useful when comparing different models developed for the same dataset. The measured values of themselves may not provide significant information on how well the model described the original data. SAS also outputted statistics of likelihood ratio test, Wald test and score test indicating the significance of the constructed model (Table 7). To better understand how well the model described the data, as well as knowing the prediction performance of the constructed model, the author also reapplied the logistic regression model on the original dataset. Among 2,689 SCEs and baseline events used, the regression model correctly described the outcome of 89% of the events (Table 8).

Table 6. Goodness of Fit Measures for Binary Logistic Regression Model

Criterion	Intercept Only	Intercept and Covariates
AIC	3599.679	1588.039
SC	3605.576	2018.514
-2 Log L	3597.679	1442.039

Table 7. Testing Global Null Hypothesis: β=0 for the Binary Logistic Regression Model.

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2155.6398	72	<.0001
Score	1523.5849	72	<.0001
Wald	512.1695	72	<.0001

Table 8. Model Prediction Results with Original Data for Binary Logistic Regression Model

<table>
<tr><th colspan="3" rowspan="3">Total Events = 2,689</th><th colspan="2">Actual</th></tr>
<tr><th>SCE</th><th>Baseline</th></tr>
<tr><th>1,639</th><th>1,050</th></tr>
<tr><th rowspan="2">Predicted</th><th>SCE</th><th>1,559</th><td>1,455</td><td>114</td></tr>
<tr><th>Baseline</th><th>1,120</th><td>184</td><td>936</td></tr>
</table>

$$Accuracy = \frac{1455 + 936}{2689} \times 100\% = 88.9\%$$

5.1.3 SCE Probability Modelling Findings

Effects of Driver Behavior Factors

The results showed seven driver behavior-related factors with significant influence on SCE occurrence (see Table 5). The following summarizes the findings relevant to driver behavior-related variables:

- Pre-incident maneuver. Pre-incident maneuver describes the last action of driving maneuver that the subject driver took prior to the precipitating event. The results indicated that drivers were more likely to be involved in fixed-object crashes when or after making a turn, drifting unintentionally while driving straight, or changing lanes. The odds ratio associated with these three types of pre-incident maneuvers were approximately 10, 4, and 2.6, respectively, compared to going straight.

- Maneuver judgment. This variable depicts the SHRP2 data reductionists' judgment on the safety and legality of the pre-incident maneuver based on vehicle position and vehicle dynamics. An example of unsafe but legal maneuver is a vehicle traveling at speed limit on slippery pavement. Compared to safe and legal maneuvers, the results showed that unsafe maneuvers were significantly more likely to be a contributing factor for fixed-object crashes and near crashes. In particular, unsafe but legal maneuvers associated with an odds ratio of 11, indicating a risk level for fixed-object crashes and near crashes more than 11 times higher than that associated with safe and legal maneuvers.

- Driver behavior. The driver behavior variable depicts driver errors or maneuvers that led to or contributed to the occurrence of the event. As the results showed, a number of driver behaviors were identified as contributing factors for fixed-object crashes. Among them, avoiding animals or other vehicles (odds ratio = 38.4) and turning improperly (odds ratio = 87.8) were particularly risky in terms of causing fixed-object crashes.

- Secondary task. Generally, the engagement of secondary tasks while driving significantly increased the risk of fixed-object crashes and near crashes. In particular, reaching, moving objects in vehicles associated with an odds ratio of almost 30, indicating the significant contribution of such secondary task for fixed-object crashes. Other tasks such as personal hygiene related activities (e.g., putting on makeup or cutting nails), adjusting/monitoring vehicle devices also significantly increased fixed-object crash risks.

- Hands on the wheel. Note that the model suggested that drivers were less likely to involve in fixed-object crashes and near crashes when they had one hand or none on wheels. This outcome does not necessarily indicated driving with one or no hands on the wheel is safer. Driving with one hand on the wheel, however, was commonly considered to be a surrogate for less risky and relaxed driving environment (e.g., straight alignment with no or few traffic).

- Critical speed. The negative parameter for the critical speed variable suggested that fixed-object crashes were more likely to occur while traveling at lower speeds. This is plausible since fixed-object crashes are more common on local roads and/or at intersections compared to freeways and continuous segments on arterial roads.

- Passenger existence. The results showed that the presence of passengers in a subject vehicle reduced the risks of fixed-object crashes and near crashes. A possible explanation for this outcome is that passengers, particularly front seat passengers, could help reduce driver inattention and fatigue, and help drivers identify risky fixed objects. Interaction with passengers, identified as a type of secondary task, however, was found to contribute to fixed-object crashes and near crashes occurrence, though the contributing effect was less significant.

Effects of Roadway and Traffic-Related Factors

The results showed the following significant roadway and traffic-related variables relevant to fixed-object crashes and near crashes (see Table 5):

- Traffic density. This variable depicts the SHRP2 data team's perception of traffic density based on the event video data, with level of service (LOS) A1 being no leading traffic in any lanes and LOS F being stop and go or severe congestion condition. The regression model showed that the presence of leading traffic was generally beneficial for reducing the risk of fixed-object crashes. This is consistent with common observations that vehicles following other vehicles are less likely to involve in fixed-object crashes.

- Contiguous travel lanes. Contiguous travel lanes include all lanes on the roadway that the subject vehicle can travel into at the time of the precipitating event, including auxiliary lanes. The results showed that the presence of other lanes allowing subject vehicles to maneuver into generally reduced the risk of fixed-object crashes.

- Traffic control. The results suggested that traffic control devices, such as traffic signs and "other" devices, significantly increased the risks of fixed-object crashes except for lane markings. Traffic control devices, while attempting to improve traffic safety overall, result in fixed objects in the roadway environment and therefore could potentially increase the risks of fixed-object crashes. The “other” category under this variable includes traffic control measures that are not commonly seen on roadways, such as toll booths, parking gates, traffic circles, and roundabouts.

- Grade. The results showed that uphill, dips, or hillcrests increased risks of fixed object-crashes and near crashes.

- Locality. The locality variable describes the surroundings of the event roadway location. The results showed that more complex (possibly with more roadside objects or curbs/channelization devices) surroundings, such as business/industrial, residential, and school/church/playground areas were associated with significantly higher risks of fixed-object crashes. In particular, roadways characterized as open country had an odds ratio of about 3.

- AADT and IRI. AADT in this analysis did not reflect actual traffic condition during the events (see the traffic density variable). It is more considered as a surrogate for certain roadway types. Both AADT and IRI were obtained from RID for only a portion of the events. The analysis of AADT showed an increasing tendency of fixed-object crashes with increased AADT. The result of IRI showed that both smoother road surface ($IRI \leq 20$) or rougher road surface ($IRI > 150$) contributed to fixed-object crashes.

Effects of Environment-Related Factors

The binary logistic regression model suggested that lighting and road surface condition were significant variables contributing to fixed-object crashes and near crashes:

- Lighting conditions. Both darkness lighted and darkness not lighted conditions had an odds ratio higher than 1 compared to the daylight condition. The odds ratio for the former was found to be 1.9 and that for the latter was 1.6. Note that the odds ratio for darkness lighted was actually higher than that for darkness not lighted, which could be due to some of the following reasons:

- A large proportion of the events analyzed in this study occurred at or near intersections in urban areas. Street lighting is more likely to present in such areas and therefore the results reflect the roadway settings for which lighting can be considered as a surrogate.
- Roadway lighting, when designed/installed improperly, results in poor uniformity across the roadway surface. Brighter areas in some cases may attract driver attention from darker areas on the roadway surface and therefore help hide fixed objects if they happen to be located in those darker areas.
- When lighting levels are inadequate to light on-road/roadside objects, the presence of lighting becomes irrelevant to the risk levels of fixed-object crashes.
- Lighting infrastructure (i.e., light poles) can also create additional risks for fixed-object crashes if not properly placed.

- Surface condition. The analysis results suggested that icy/snowy/wet pavement surfaces had an odds ratio of 1.4 for fixed-object crashes compared to the dry pavement condition.

5.2 Ordinal Logistic Regression Results for SCE Severity Level

5.2.1 Modelling Results

Ordinal logistic regression was used to model the effects of risk factors on the probability of increased severity levels based on SHRP2 SCEs. The SCEs were divided into five severity levels with level 1 representing the most severe crashes and level 5 being less severe near crashes. In addition to the variables used in the binary logistic regression model, the SCE severity level modelling also included variables such as driver reaction time and struck object type. Table 9 Listed the significant variables and associated statistical test results. Table 10 listed the significant variable values, parameters, and odds ratios. Similarly, this analysis only involved variables with a significance level of 0.1. For unordered variables, only values with a p-value smaller than 0.1 were displayed. A detailed table of estimated parameters and odds ratios of all variable values could be found in Appendix B.

Table 9. Significant Explanatory Variables for SCE Severity Level Modelling.

Effect	DF	Wald Chi-square	Pr > ChiSq
Number of Violation	3	6.0748	0.108
Pre-incident Maneuver	5	13.0414	0.023
Driver Behavior	9	107.2417	<.0001
Weather	1	2.9861	0.084
Traffic Density	4	12.3309	0.0151
Locality	6	36.3792	<.0001
Critical Speed	1	3.7883	0.0516
Radius	6	21.7426	0.0013
Struck Object Type	12	266.8916	<.0001
Reaction Time	1	84.523	<.0001

Table 10. Parameters and Odds Ratios of Significant Variables for SCE Severity Level Modelling

Variable	Variable Value	Parameter	Chi-square	Pr > ChiSq	Odds ratio (95% CI)
Intercept 1	-	-5.4665	70.1777	<.0001	-
Intercept 2	-	-4.0579	44.9919	<.0001	-
Intercept 3	-	-0.8791	2.2621	0.1326	-
Intercept 4	-	1.0690	3.3350	0.0678	-
Driver Behavior Factors					
Number of violation	0	-0.9412	4.2593	0.039	0.390 (0.160-0.954)
	1	0.0429	0.1112	0.7388	1.044 (0.811-1.343)
	2 and 2+	0.1834	1.0579	0.3037	1.201 (0.847-1.704)
	Missing	Lowest Level			
Pre-incident Maneuver	Going straight but unintentionally drifted	0.4981	3.2443	0.0717	1.646 (0.957-2.83)
	Making a turn	0.3062	2.6188	0.1056	1.358 (0.937-1.968)
	Going straight (accelerating, decelerating, constant speed)	Reference			
Driver Behavior	Apparent unfamiliarity with roadway	1.3754	14.6353	0.0001	3.957 (1.956-8.005)
	Avoiding animal, or other vehicle	-1.1367	3.5268	0.0604	0.321 (0.098-1.051)
	Distracted	1.089	39.1993	<.0001	2.971 (2.113-4.178)
	Drowsy, sleepy, asleep, fatigued	1.2856	10.703	0.0011	3.617 (1.674-7.813)
	Exceeded safe speed, or speed limit	1.3257	40.4204	<.0001	3.765 (2.502-5.665)
	None	Reference			
Critical Speed (km/h)	-	0.00622	3.7883	0.0516	1.006 (1-1.013)
Reaction Time (s)	-	-0.5962	84.523	<.0001	0.551 (0.485-0.626)
Roadway and Traffic Factors					
Traffic Density	LOS A2	-0.3368	6.1728	0.013	0.714 (0.547-0.931)
	LOS B	-0.3219	4.5792	0.0324	0.725 (0.54-0.973)
	LOS C	-0.8883	6.7163	0.0096	0.411 (0.21-0.805)
	LOS A1	Reference			
Locality	Business/Industrial	0.8038	8.8657	0.0029	2.234 (1.316-3.792)
	Bypass/Divided Highway with traffic signals	0.8137	3.2428	0.0717	2.256 (0.931-5.47)
	Church/school/playground	1.0214	11.5393	0.0007	2.777 (1.54-5.007)
	Open Country	2.8109	33.7909	<.0001	16.624 (6.444-42.888)
	Residential area	0.8056	9.1334	0.0025	2.238 (1.327-3.774)
	Urban	0.642	2.7473	0.0974	1.9 (0.889-4.06)
	Interstate/Bypass/Divided Highway with no traffic signals	Reference			
Curve Radius (ft.)	≤500	0.3594	0.7040	0.4014	1.433 (0.619-3.317)
	500-1000	-0.1376	0.0672	0.7955	0.871 (0.308-2.467)
	1000-2000	0.0884	0.0295	0.8636	1.092 (0.398-2.995)
	2000-5000	-1.5421	7.0451	0.0079	0.214 (0.069-0.668)
	>5000	2.1959	15.2680	<.0001	8.988 (2.988-27.042)
	No curvature	-0.5585	2.1021	0.1471	0.572 (0.269-1.217)
	Missing	Lowest Level			
Struck Object Type	Animal	-2.2632	102.4069	<.0001	0.104 (0.067-0.161)
	Ditch	3.0768	32.6201	<.0001	21.688 (7.545-62.341)
	Roadway debris	0.919	10.1226	0.0015	2.507 (1.423-4.415)

Variable	Variable Value	Parameter	Chi-square	Pr > ChiSq	Odds ratio (95% CI)
	Stopped, backing, pulling out car	-3.2329	45.9126	<.0001	0.039 (0.015-0.1)
	Tree/shrub	1.4932	3.6178	0.0572	4.451 (0.956-20.737)
	Utility/light pole	2.1344	6.6982	0.0097	8.452 (1.679-42.551)
	Curb	Reference			
Environmental Factors					
Weather	Adverse Weather	0.2642	2.9861	0.084	1.302 (0.965-1.757)
	No adverse weather	Reference			

5.2.2 Model Goodness-of-Fit Statistics

Table 11 and Table 12 listed the model goodness-of-fit measures and global null hypothesis tests. Verification of the ordinal logistic regression model using the original data showed that the model was able to correctly describe 70% of the analyzed events (Table 13).

Table 11. Model Fitness Measures for the Ordinal Logistic Regression Model.

Criterion	Intercept Only	Intercept and Covariates
AIC	3968.104	3293.890
SC	3989.711	3585.590
-2 Log L	3960.104	3185.890

Table 12. Testing Global Null Hypothesis: β=0 for the Ordinal Logistic Regression Model.

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	774.2134	50	<.0001
Score	558.8899	50	<.0001
Wald	574.0044	50	<.0001

Table 13. Confusion Matrix of Ordinal Event Severity Prediction.

Total SCEs = 1,639			Counts by Actual Severity				
			1	2	3	4	5
			13	35	447	534	610
Counts by Predicted Severity	1	1	1	0	0	0	0
	2	0	0	0	0	0	0
	3	373	5	12	236	77	43
	4	698	3	9	140	440	106
	5	567	4	14	71	17	461

$$Accuracy = \frac{1 + 0 + 236 + 440 + 461}{1639} \times 100\% = 69.4\%$$

5.2.3 SCE Severity Level Modelling Findings

Impacts of Driver Behavior Factors on Event Outcome

The results showed five driver behavior-related factors with significant influence on SCE severity level (see Table 10). The following summarizes the findings relevant to driver behavior-related variables:

- Number of violation. This variable is defined as the violation times that the subjective driver holds in the last three years. This variable can be an indicator of the driver's driving habits to some extent. The results indicated that drivers with fewer violations in the past generally had a higher likelihood of having a less severe fixed-object confliction than those with more violations when a fixed-object event took place.

- Pre-incident maneuver. The results suggested that drivers were more likely to be involved in a more severe fixed-object crash during or after making a turn, or drifting unintentionally while going straight.

- Driver behavior. As the results showed, most identified driver behavior contributed to increased event severity relevant upon the occurrence of a fixed-object crash. The presence of factors, such as unfamiliarity with roadway, distraction, drowsiness/fatigue, and speeding, was associated with an odds ratio ranging from three to four, indicating significantly higher levels of risks for a more severe fixed-object crash. Avoiding animals or other vehicles, however, was found to be associated with lower odds of more severe fixed-object event outcomes, possibly because this driver action is an evasive maneuver and therefore is more likely to cause near crashes as compared to crashes.

- Critical speed. The positive parameter for the critical speed suggested that increased traveling speed was associated with higher severity fixed-object crashes. The results showed that an increase of 10 mph in speed increased the likelihood of a more severe fixed-object event by 1.1 times.

- Reaction time. The negative parameter for the reaction time indicated that drivers were often able to avoid severe fixed-object crash outcome effectively when given enough reaction time. An extra one second reaction time could reduce the probability of more severe fixed-object event outcomes by 1.8 times.

Effects of Roadway and Traffic-Related Factors

The results showed that the following significant roadway and traffic-related variables were associated with fixed-object SCEs severity (see Table 10):

- Traffic density. The ordinal logistic regression model showed that the presence of leading traffic was beneficial for reducing the risk of more severe fixed-object event outcomes. In particular, drivers were least likely to involve in a severe fixed-object crash when traveling in LOS C density (stable flow, restricted maneuverability and speed) with the least odds ratio of 0.4.

- Locality. The results showed that locations with more complex surroundings such as business/industrial, residential, urban, bypass/divided highway with traffic signals and school/church/playground areas were associated with higher probability of more severe fixed-object events, with odds ratios ranging from two to three. Note that roadways characterized with open country were particularly correlated with more severe fixed-object event outcomes, with an odds ratio as high as 16.6.

- Curve radius. This variable was obtained from RID for only a portion of the events. According to the result, roadways with curve radii ranging between 2000 and 5000 were associated with a less likelihood of severe fixed-object crash outcome, while roadways with curve radii larger than 5,000 were particularly correlated to more severe fixed-object event outcomes.

- Struck object type. Curbs were the most frequent type of struck objects and were often associated with low-risk tire strikes (used as the reference level for this variable). Comparatively, confliction with animals and vehicles showed a significantly lower likelihood of more severe fixed-object event outcomes judging from the low odds ratio (i.e., 0.1 and 0.04). This could be explained together with the negative parameter of avoiding animal/vehicle in driver behavior variable, which suggested that drivers tended to avoid these two objects more than curbs, and hence were more likely to make successful evasions resulting in near crashes. Colliding with other objects such as roadway debris, trees/shrubs, utility/light poles, and ditches generally led to more severe fixed-object event outcomes, with odds ratios being 2.5, 4.5, 8.5 and 21.7, respectively.

Effects of Environment-Related Factors

- The analysis results showed that driving in adverse weather conditions such as foggy, rainy and snowy weathers tended to be correlated with higher risks of more severe fixed-object event outcomes (i.e., an odds ratio of 1.3).

5.3 Conclusion

In this chapter, a binary logistic regression model and an ordinal logistic regression model were constructed for the occurrence and severity level of fixed-object events respectively. The model construction was fulfilled with SAS® Studio software package. With 0.1 significance level used as the criteria for variable selection, 16 and 10 variables in three major categories were identified for two models separately. The influence of variables on responses was illustrated with parameters and odds ratios, which were discussed in detail. The software also outputted goodness-of-fit statistics including AIC, SC, -2 Log L, and three global null hypothesis test results including likelihood ratio test, Wald test, and score test. These statistics combined showed that both constructed models had good fitness and significance. When the author reapplied the original datasets to constructed models, two models gave satisfactory correct prediction rate of 88.9% and 69.4%, respectively.

Chapter 6. SVM and ANN Analysis Results

6.1 Results of SVM Analysis

6.1.1 Classification Results

A binary SVM classifier and a multi-class SVM classifier were constructed for the fixed-object SCE occurrence and severity levels respectively. The dataset contains 2689 events (1639 SCE events distributed in five severity levels and 1050 baseline events). In the construction process, 5-fold cross validation was used to avoid potential overfitting, and six different kernel transformations were tested. Table 14 listed the classification accuracy for both SVM classifiers with different kernels. The classifier training and validation procedures were fulfilled with MATLAB® Classification Learner Application.

Table 14. Classification Accuracy of SVM Classifiers with Different Kernels

Kernel	Binary Classifier	Multi-class Classifier
Linear	86.3%	<u>74.7%</u>
Quadratic	<u>87.4%</u>	74.2%
Cubic	86.3%	72.8%
Fine Gaussian ($\sigma = 2.5$)	63.3%	39.4%
Medium Gaussian ($\sigma = 10$)	87.2%	73.6%
Coarse Gaussian ($\sigma = 40$)	85.6%	72.2%

It could be found from Table 14 that the binary SVM classifier with quadratic kernel and the multi-class classifier with linear kernel achieved the highest classification accuracy of 87.4% and 74.2%. Hence they were selected as the classifiers for fixed-object SCE occurrence and severity. It should be mentioned that these accuracy rates may vary slightly (<1%) for each training time. Table 15 and Table 16 showed the classification results of two SVM classifiers.

Table 15. Confusion Matrix of Binary SVM Classifier with Quadratic Kernel

Total Events = 2,689			**Actual**	
			SCE	**Baseline**
			1,639	**1,050**
Predicted	**SCE**	**1,520**	1,410	110
	Baseline	**1,169**	229	940

$$Accuracy = \frac{1410 + 940}{2689} \times 100\% = 87.4\%$$

Table 16. Confusion Matrix of Multi-Class SVM Classifier with Linear Kernel

Total SCEs = 1,639			**Counts by Actual Severity**				
			1	**2**	**3**	**4**	**5**
			13	**35**	**447**	**534**	**610**
Counts by Predicted Severity	**1**	**0**	0	0	0	0	0
	2	**5**	1	2	1	0	1
	3	**417**	2	5	267	69	74
	4	**551**	1	0	67	452	31
	5	**666**	9	28	112	13	504

$$Accuracy = \frac{0 + 2 + 267 + 452 + 504}{1639} \times 100\% = 74.7\%$$

In addition to confusion matrix, the receiver operating characteristic (ROC) curve is also a frequently employed tool to illustrate the diagnostic ability of a classifier by plotting the true positive rate (TPR) against the false positive rate (FPR). As a threshold is used to divide continuous output value into dichotomous classification outcomes, each point on ROC curves is determined by the TPR and FPR corresponding to a threshold. ROC curves are insensitive to classification distribution, classification prior probability, and misclassification cost, and therefore are widely used for evaluating the performance of machine learning-based classifiers (*42*).

On a ROC curve plot, the top left corner, which suggests a false positive rate of zero, and a true positive rate of one, is the ideal point for a classifier. In addition, larger area under the curve (AUC) is usually desired.

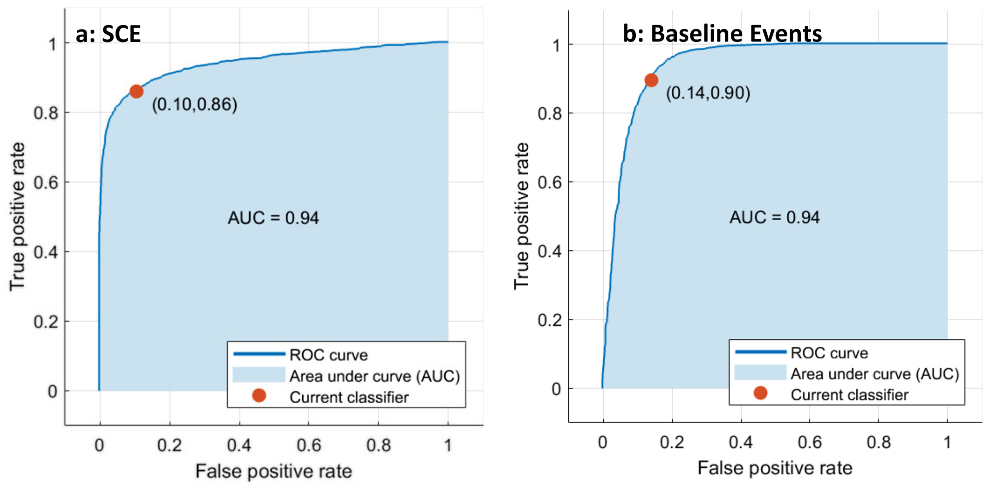


Figure 14. ROC Curves for Binary SVM Classifiers with Quadratic Kernel

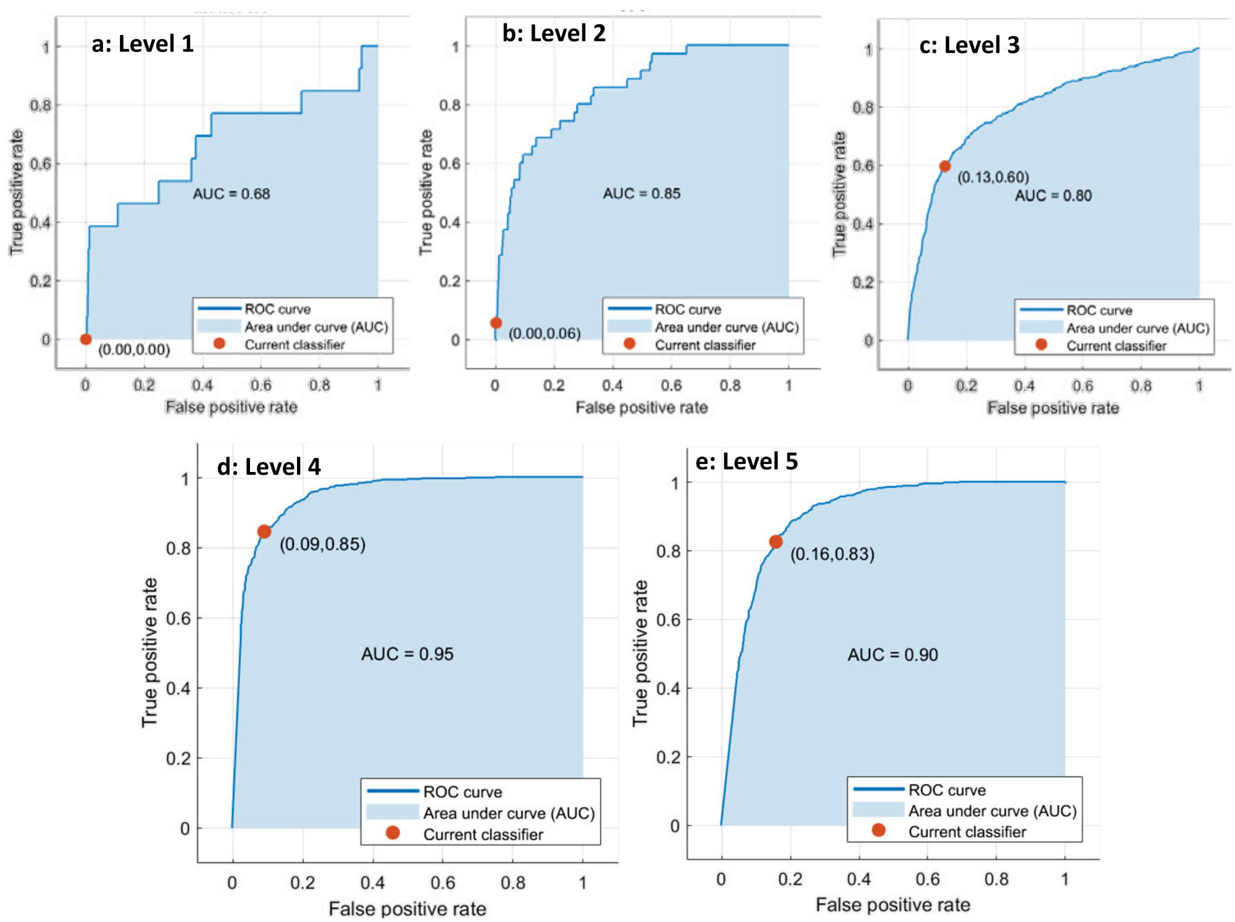


Figure 15. ROC Curves for Multi-Class SVM Classifier with Linear Kernel

Figure 14 and Figure 15 presented the ROC curves of two classifiers with different responses. The figures in general showed that the binary SVM classifier performed fairly well for the fixed-object event data. For the multi-class SVM classifier, the plots indicated that the classifier performed better for level 4 crash events, level 5 near crash events, and, to a lesser degree, level 3 events. The poor performance for level 1 and level 2 crash events were possibly due to the limited sample sizes.

6.1.2 Sensitivity Analysis Results

Figure 16 and Figure 17 showed the sensitivity analysis results of two SVM classifiers. Note that here the author only displayed variables of which the removal could result in a prediction accuracy reduction more than 0.5%. As the figures illustrated, driver behavior and critical speed were the variables identified to be significant to the binary classifier, which removal caused 4.22% and 2.74% reduction of classification rate. Driver behavior, struck object type, and driver reaction time were identified to be significant to the multi-class classifier, the removal of which resulted in 7.10%, 6.69% and 2.01% reduction of classification rate. Other variables held roughly equal sensitivity results or were less significant to the classifier.

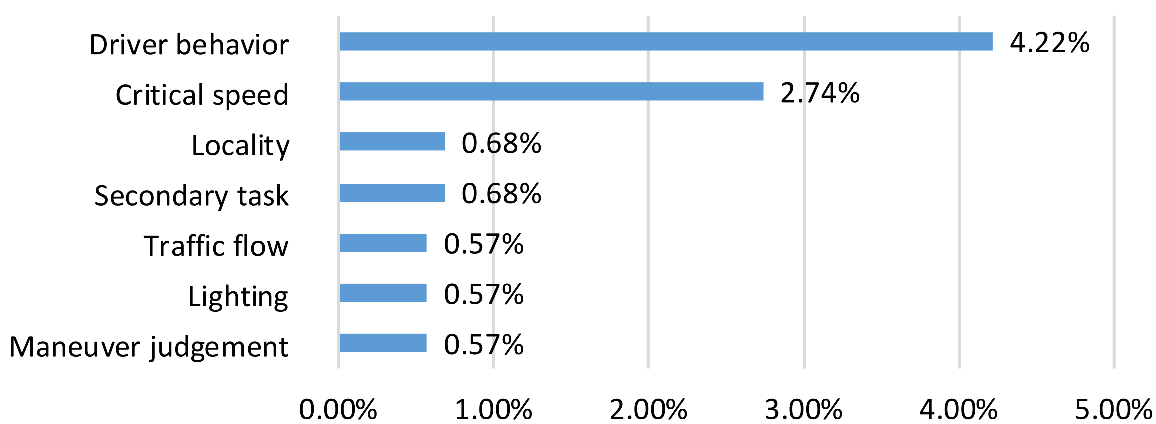


Figure 16. Sensitivity Analysis Results of Binary SVM Classifier

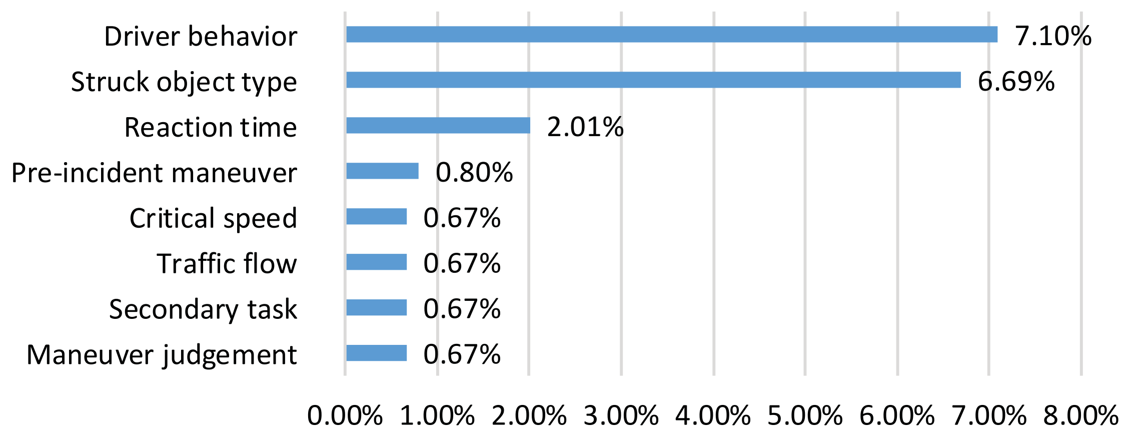


Figure 17. Sensitivity Analysis Results of Multi-class SVM Classifier

6.2 Results of ANN Classification

Similarly, the author also constructed two ANN models for fixed-object SCE occurrence and severity level respectively. The datasets and feature vectors were the same as SVM classifiers, while the responses were all dealt as vectors. The datasets were partitioned into 70% training set, 15% testing set, and 15% validation set. The author tested ANN models with different hidden layer numbers and hidden node numbers to identify the combination gave the best prediction performance. However, it was found that the prediction performance varied very significantly for different training trials. Table 17 simply listed the results of one trial. All the procedures were completed with MATLAB® Neural Network Toolbox ™ (*33*). Figure 18 showed a schematic illustration of a neural network structure for SCE occurrence prediction, with 85 input nodes as the feature vector representing an event, two hidden layers with 10 and 5 hidden nodes in each layer, and two output nodes exporting the probability of the event being a SCE and a baseline event.

Table 17. Classification Accuracy of ANN Classifiers with Different Structures

Structure	Accuracy	Accuracy
Single layer, 5	86.8%	75.0%
Single layer, 10	88.6%	75.2%
Single layer, 20	87.6%	73.9%
Double layer, [10, 5]	85.8%	75.5%
Double layer, [20, 5]	87.2%	74.3%
Double layer, [20, 10]	87.4%	75.0%

Note. The numbers in bracket [10, 5] meant 10 nodes in first hidden layer and 5 nodes in second hidden layer.

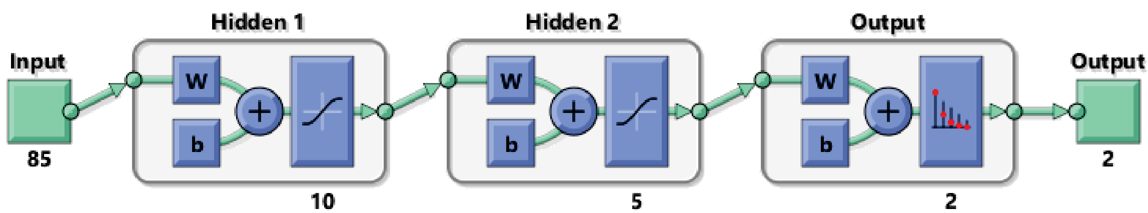


Figure 18. An Example of Neural Network Structure with Two Hidden Layer

As aforementioned, the prediction accuracy varied significantly (1% - 5%) for different training trials. Hence a higher value in Table 17 did not necessarily suggest that the associated network structure was superior to other structure combinations. They only provided a general insight of the ANN prediction results, that was approximate 87% for SCE occurrence prediction and 75% for SCE severity level prediction.

Table 18 and Table 19 showed the results of two ANN models with a single hidden layer and 10 hidden nodes. Moreover, Figure 19 presented the associated ROC curves. Note that although only 15% of the events were used for testing, the author integrated the prediction results of all participated events.

Table 18. Confusion Matrix of ANN Classifier for SCE, Baseline Classification

Total Events = 2,689			Actual	
			SCE	**Baseline**
			1,639	**1,050**
Predicted	**SCE**	**1,520**	1,443	110
	Baseline	**1,169**	196	940

$$Accuracy = \frac{1443 + 940}{2689} \times 100\% = 88.6\%$$

Table 19. Confusion Matrix of ANN Classifier for SCE Severity Classification

Total SCEs = 1,639			Counts by Actual Severity				
			1	**2**	**3**	**4**	**5**
			13	**35**	**447**	**534**	**610**
Counts by Predicted Severity	**1**	**0**	0	0	0	0	0
	2	**10**	2	5	3	0	0
	3	**414**	4	4	265	55	86
	4	**575**	1	0	76	468	30
	5	**640**	6	26	103	11	494

$$Accuracy = \frac{0 + 5 + 265 + 468 + 494}{1639} \times 100\% = 75.2\%$$

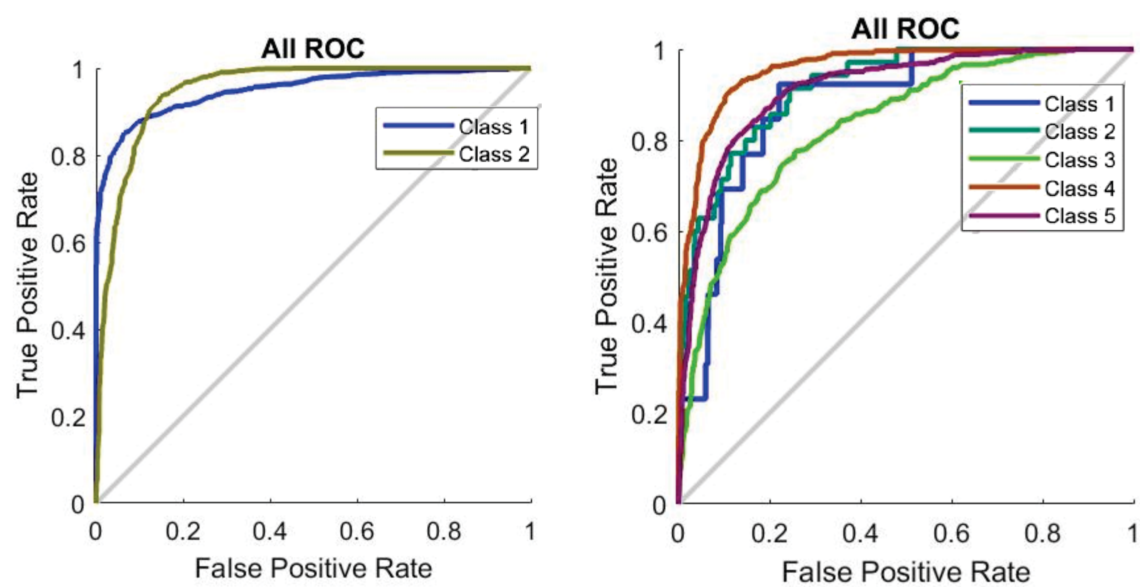


Figure 19. ROC Curves for ANN models

The author also attempted to conduct sensitivity analyses for ANN models. However, as the prediction accuracy varied significantly each training time, it was difficult to justify whether the output variance was due to the effect of removed variables, or simply was the variance between different training trials. Hence the author did not perform sensitivity analyses for two ANN models.

6.3 Conclusion

In this chapter, two machine learning methods: SVM and ANN, were employed for fixed-object crash analyses. SVM analyses tested six different kernel functions and used 5-fold cross validation. The results showed that binary SVM classifier with quadratic kernel and multi-class SVM classifier with linear kernel achieved the highest accuracy of 87.4% and 74.3% for two responses. In addition, sensitivity analyses were conducted to infer the cause and effect relationship between predictors and responses, which identified two significant variables: driver behavior and critical speed, for binary SVM classifier, and three significant variables: driver behavior, struck object type and driver reaction time for multi-class SVM classifier. In ANN analyses, six network structures with different hidden layer numbers and hidden node numbers were tested. The results showed roughly equal prediction performance for different combinations, that was about 87% for SCE occurrence prediction and 75% for SCE severity level prediction. However, the author found that the prediction accuracy varied significantly with different training trials. Hence no sensitivity analysis was performed for ANN models.

Chapter 7. Fixed-object Crash Case Studies

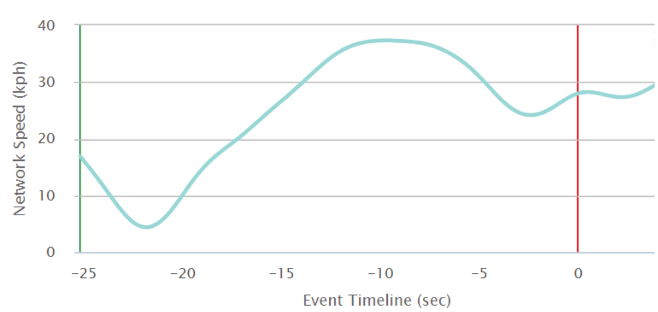
The before logistic regression and machine learning analyses revealed a variety of factors with potential impacts on the risks of fixed-object crash occurrence and outcome. In this chapter, the author further provides case studies representing a number of typical fixed-object crash scenarios based on SHRP2 event videos. These case studies provide additional information on the sequence of events and exact causes for the SHRP2 fixed-object crashes and near crashes.

For each case study, the author included a number of key data elements in an effort to unveil its causes, contributing factors, environmental conditions, and insights to potential countermeasures.

7.1 Fixed-object Crash Case Studies

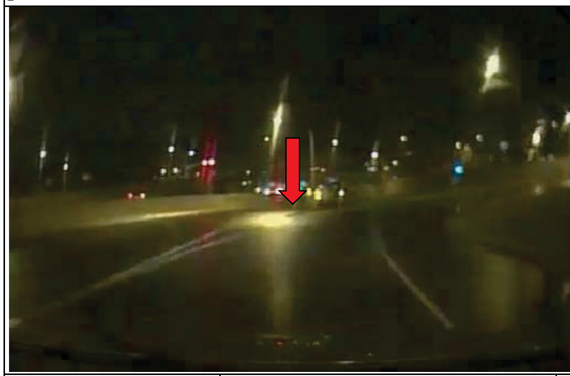
Example 1-3 show three sample fixed-object crashes that involve medians or channelization devices. These events were contributed by a combination of weather conditions, roadway design issues, lighting conditions, and/or driver errors.

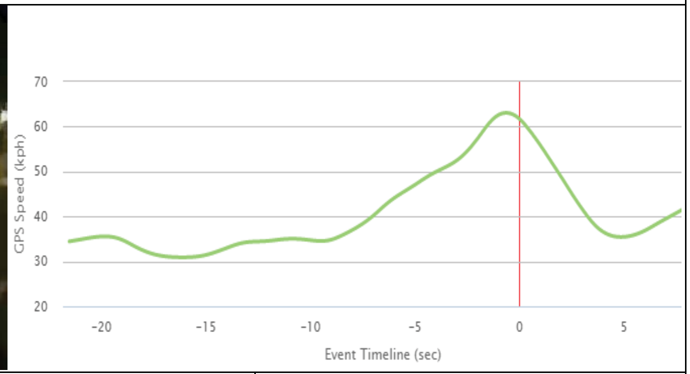
Example 1: Vehicle struck a raised median after making a left turn



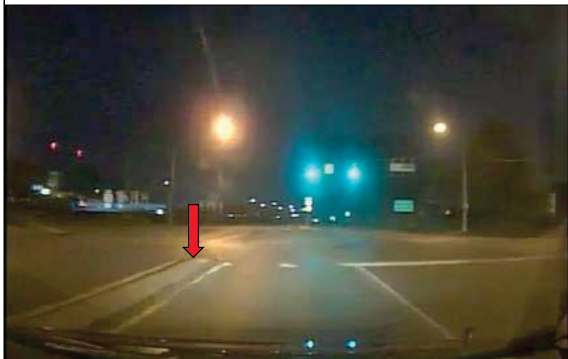
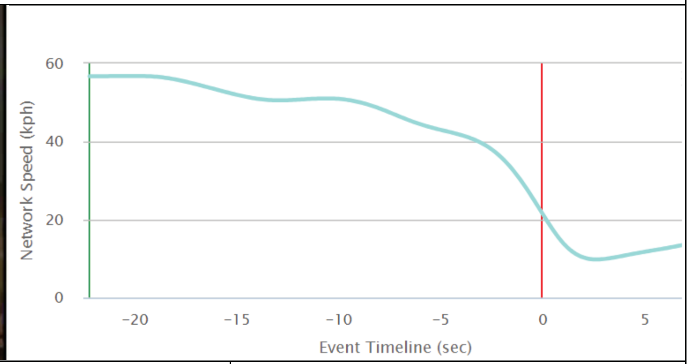
Roadway	Divided multilane	**Lighting**	Daylight
Weather	Light rain	**Location Type**	Signalized Intersection
Maneuver	Turning left	**Crash severity**	Level 3
Travel Speed	Medium (~20 mph)		
Description	The subject vehicle proceeded into a 4-leg signalized intersection from the dedicated left-turn lane. The subject driver turned too early and sharply, struck the raised median on the exit approach.		
Observation	Based on the vehicle turning trajectory, the driver did not notice the raised median possibly because the pillar on the driver side blocked his view. Pavement markings clearly delineate the left turn path can potentially prevent such events.		

Example 2: Vehicle running over raised gore at a lighted entrance ramp junction



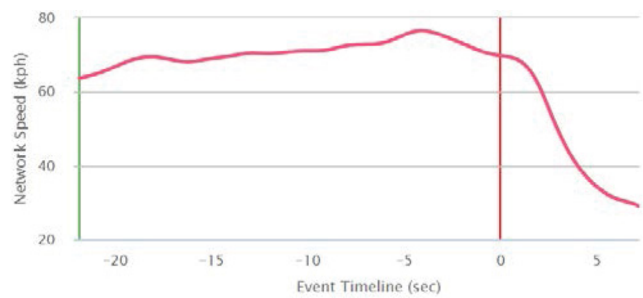


Roadway	C-D road	**Lighting**	Darkness with light
Weather	Light rain	**Location Type**	Entrance ramp
Maneuver	Going straight	**Crash severity**	Level 3
Travel Speed	Medium (~35 mph)		
Description	Subject vehicle was attempting to enter a freeway via a lighted two-lane C-D road. At the end of the C-D road, the lanes separated into two single-lane ramps with the right ramp curve to the right with a relatively sharp radius. The driver was supposed to merge to the left lane in order to enter the freeway main lanes. However, the driver appeared to be confused with the ramp configuration and went straight onto the raised gore area. Both ramp lanes were illuminated, but neither luminaire provided sufficient light to the gore point. The driver was a senior driver and appeared to be looking intently throughout the entire event. Notice that the roadway was wet, which might have contributed to the event as well.		
Observation	At this location, the right lane on the C-D road diverts from the entrance ramp via a relatively sharp curve towards the right. The trajectory of the subject vehicle appeared to be aiming straight to the correct entrance ramp. The senior driver appeared to have thought both lanes were part of the entrance ramp. While lighting was provided upstream of the ramp junction on both lanes, the gore was not sufficiently illuminated.		

Example 3: a crash involving a vehicle running over raised median when turning left			
			
Roadway	Divided multilane	**Lighting**	Darkness with light
Weather	No adverse weather	**Location Type**	Signalized Intersection
Maneuver	Turning left	**Crash severity**	Level 3
Travel Speed	Medium (~30 mph)		
Description	Subject driver approached a 4-leg signalized intersection in the dedicated left-turn lane. There was a raised median dividing the road. As the traffic light was green, the driver steered left to complete a turning maneuver. The driver apparently misjudged the turn and ran over the median. The intersection was lighted, but lighting seemed to be not sufficient to allow clear identification of the appropriate turning path.		
Observation	The subject driver failed to identify the proper turning path and made the left turn too early. Pavement markings clearly delineate the left turn path can potentially prevent such events.		

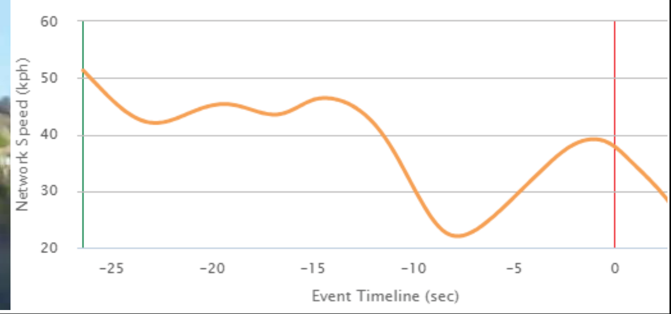
Example 4-6 are three representative cases of fixed-object crashes involving curbs. These cases illustrate how driver errors, lighting, and roadway design/traffic control issues contribute to fixed-object crashes.

Example 4: Vehicle hitting curb on the right when driver texted



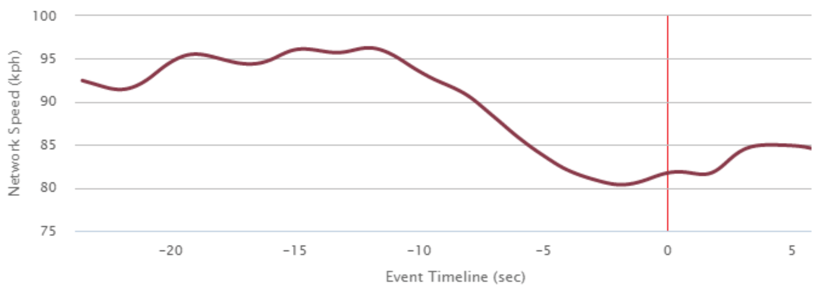
Roadway	Undivided local road	**Lighting**	Daylight
Weather	No adverse weather	**Location Type**	Bridge
Maneuver	Going straight	**Crash severity**	Level 2
Travel Speed	Medium (~45 mph)		
Description	Subject vehicle drifted unintentionally to the right, striking the curb on the bridge at a speed of about 44 mph. Subject driver was clearly distracted while texting on a phone. After the impact, subject driver steered left immediately to get vehicle back on the road, traveled shortly and pulled over the car to check vehicle damage.		
Observation	Driver was distracted.		

Example 5: Vehicle running over curb on the right



Roadway	Undivided two-lane road	**Lighting**	Daylight
Weather	No adverse weather	**Location Type**	Straight local road
Maneuver	Going straight	**Crash severity**	Level 3
Travel Speed	Low(~25 mph)		
Description	The subject driver traveled on an undivided local two-lane road. The road reduces width ahead, but the curb was not conspicuous due to the shadow of the adjacent building. The subject driver was on the phone intermittently.		
Observation	In this crash, there was an unexpected curb narrowing the street, which was covered by shadows during bright daylight, making it difficult for drivers to identify. Delineating the curb using retroreflective materials will increase its visibility.		

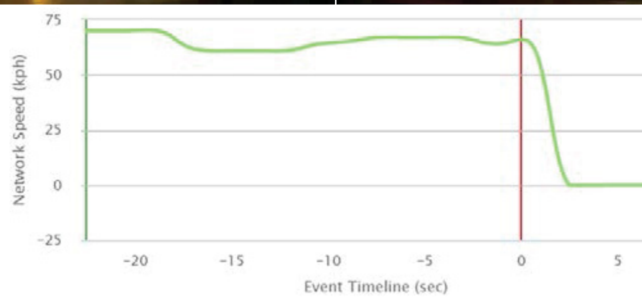
Example 6: Vehicle hitting curbs and left wall in an over-lighted underground ramp



Roadway	One-lane tunnel	**Lighting**	Darkness with light
Weather	Light rain	**Location Type**	Freeway underground ramp
Maneuver	Turning left	**Crash Severity**	Level 3
Travel Speed	High (~60 mph)		
Description	Shortly after entering a lighted tunnel at a freeway interchange, the subject vehicle failed to correctly follow the curved roadway and crashed into the left curb and then the tunnel wall on the left, causing the vehicle to lose control and then crash onto the curve on the right side. The tunnel had a much higher lighting level compared to the roadway prior to it and the driver appeared to be adapting to the sudden brightness while negotiating the curve. The driver was not distracted and there were no other vehicles in the close vicinity.		
Observation	In this scenario, the lighting level in the tunnel appeared to be too bright. The subject driver could not adapt to the condition quickly enough while negotiating the curved roadway. A transition should be provided at the tunnel entrance to improve the lighting condition.		

Example 7 is an example of a fixed-object crash involving a light pole at a curved intersection. The unconventional alignment of the intersection combined with low visibility condition during nighttime contributed to this event.

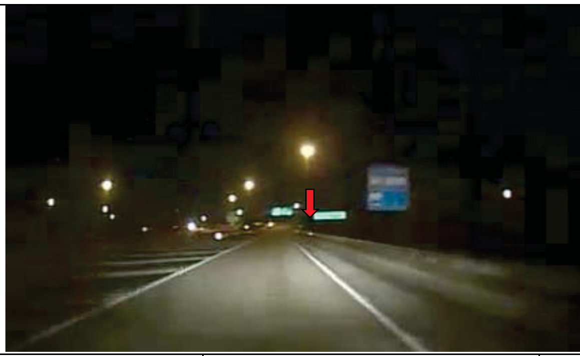
Example 7: Vehicle striking a utility pole at a Y intersection

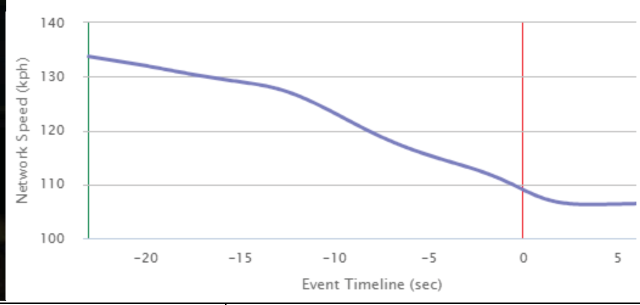


Roadway	Undivided two-lane	**Lighting**	Darkness with light
Weather	Light rain	**Location Type**	Signalized intersection
Maneuver	Going straight	**Crash severity**	Level 2
Travel Speed	Medium(~35 mph)		
Description	Subject driver approached a signalized intersection with yellow light. There was streetlight, but it appeared to be very dim, maybe partially due to the raining weather. The subject driver followed the pavement marking, crossed the intersection at a relatively high speed. The driver soon realized that the intersection curved sharply to the left when he approached to the mid of intersection. He braked hard to avoid colliding with the utility pole ahead, but it was unsuccessful.		
Observation	In this scenario, the intersection alignment was unconventional, and there was no sign warning drivers about roadway alignment ahead. The dim streetlight, raining weather, wet road surface and darkness all added to the difficulty for the driver to distinguish the road alignment earlier. Potential countermeasures include clearly delineating the travel lanes, providing advanced warning sign of unconventional intersection layout, and redesign lighting to illuminate the entire intersection.		

Examples 8-10 show cases of fixed-object crashes involving guardrails. For the first two cases, driver errors appeared to be the major contributing factor. The third case illustrates potential roadway and traffic design issues that may have contributed to the event.

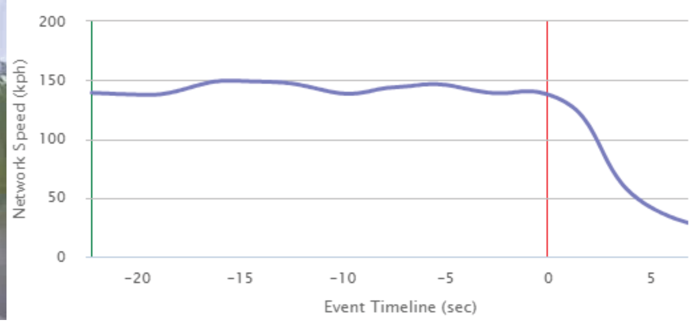
Example 8: Vehicle struck guardrail when the driver was impaired by drowsiness.





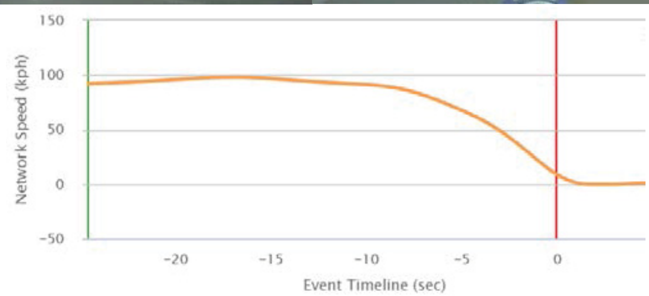
Roadway	Divided freeway	**Lighting**	Darkness with light
Weather	No adverse weather	**Location Type**	Freeway
Maneuver	Going straight	**Crash severity**	Level 1
Travel Speed	High(~80 mph)		
Description	Subject vehicle was about to exit a freeway while traveling in a deceleration lane. Subject driver appeared to be tired and drowsy. While decelerating, the vehicle drifted towards the right, hitting the guardrail on the right side.		
Observation	The subject driver was impaired by drowsiness.		

Example 9: Vehicle striking guardrail at high speed after running off road.



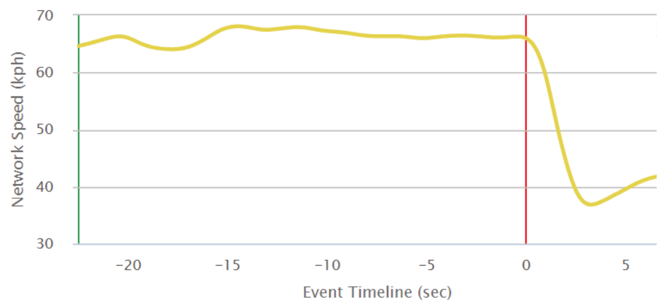
Event ID	29730844	**Lighting**	Daylight
Weather	No adverse weather	**Location Type**	Curvy local road
Maneuver	Going straight with unintentional drifting	**Crash severity**	Level 1
Roadway	Undivided two-lane road	**Travel Speed**	High(~85 mph)
Description	Subject driver was driving down a curvy, undivided, 2-lane road with no leading traffic. The speed limit was 40 mph, but the subject driver drove at 80 mph. After disregarded two yellow caution signs that warned drivers of the curvy road, the driver lost control on a curve, striking the guardrail on the right, spinning 720 degrees counterclockwise into the opposing lane.		
Observation	Speeding was the major cause of the event.		

Example 10: Vehicle striking guardrail terminal at the entrance of a split ramp.



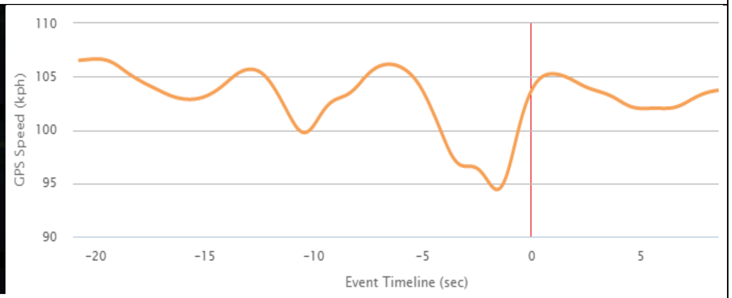
Roadway	Freeway exit ramp	**Lighting**	Daylight
Weather	No adverse weather	**Location Type**	Ramp gore
Maneuver	Going straight	**Crash severity**	Level 2
Travel Speed	High(~60 mph)		
Description	The subject driver decelerated in the exit lane of an interstate highway. The exit ramp split into two exits to different destinations identified by overhead signs. The subject driver drove on the way towards left exit initially. When the signs came into view, the subject driver tried to switch to the right lane. However, due to a conflicting vehicle occupying the lane, the subject vehicle could not successfully make the lane change and ran into the guardrail terminal dividing the two ramps while braking.		
Observation	The driver was not familiar with the roadway condition. The guardrail was provided to shield the sign post, resulting in additional risks for fixed-object crashes. In this scenario, the sign may be better located on the side, eliminating the need for the guardrail.		

Example 11 is an example of a fixed-object crash involving a crossing deer during nighttime. The roadway was not lighted and it was difficult to identify the deer well in advance.

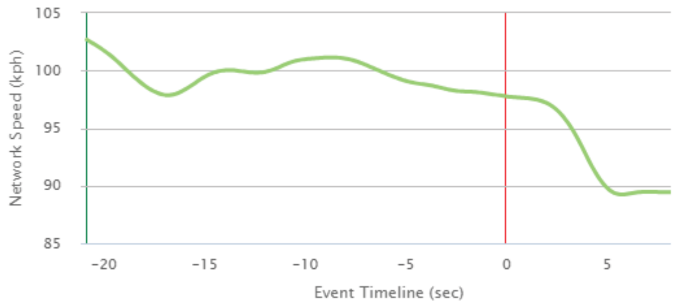
Example 11: Vehicle striking a deer during nighttime.			
			
Roadway	Undivided two-lane road	**Lighting**	Darkness without light
Weather	Mist/Light Rain	**Location Type**	Straight local road
Maneuver	Going straight	**Crash severity**	Level 2
Travel Speed	Medium (~40 mph)		
Description	The subject driver traveled on a straight two lane road during nighttime, hitting a deer walking across the road from the left. There was light rain and the road surface was wet. The road was unlighted and no leading traffic was present. Prior to the event, the driver was interacting with a passenger as well.		
Observation	It was dark and the vehicle had its headlamps on low beam. After identifying the dear, there was not enough time to react and avoid the crash.		

Example 12 and 13 illustrate two cases of crashes involving on-road debris. As the cases illustrate, large pieces of debris may be difficult to be identified particularly during nighttime and on high-speed roadways.

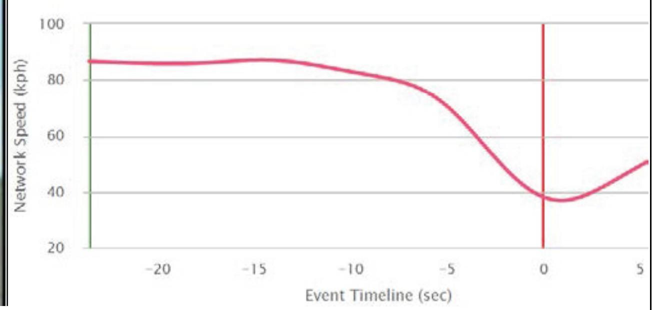
Example 12: A vehicle running over a large piece of debris on a unlighted freeway.



Roadway	Multilane freeway	**Lighting**	Darkness without light
Weather	None	**Location Type**	Freeway interchange
Maneuver	Going straight	**Crash severity**	Level 3
Travel Speed	High (~65 mph)		
Description	Subject vehicle was traveling straight on freeway main lanes at an interchange area. The driver failed to identify a large piece of debris (appeared to be a blown truck tire) in the travel lane and ran over it, almost losing control of the vehicle. The lane changing activities associated with the exiting traffic seemingly caused some speed fluctuation of the subject vehicle prior to the event.		
Observation	The roadway was not lighted. The subject vehicle was traveling at a relatively high speed and did not identify the debris on road. The driver was a senior person.		

Example 13: A vehicle running over a large wood piece when following other vehicle.			
 			
Roadway	Multilane freeway	**Lighting**	Daylight
Weather	None	**Location Type**	Freeway
Maneuver	Going straight	**Crash severity**	Level 3
Travel Speed	High (~65 mph)		
Description	The subject driver was following a lead vehicle on a freeway. There was a large piece of wood in the travel lane of the subject vehicle that was not visible due to the lead vehicle. Both the lead vehicle and the subject vehicle ran over the wood, causing a noticeable vibration to the subject vehicle.		
Observation	The subject driver followed closely with the leading vehicle, there was not enough time for the subject driver to take action to avoid hitting the wood after it became visible to the driver.		

Example 14 shows a case when a vehicle ran off the roadway at an intersection on a curved roadway segment due to not steer sufficiently.

Example 14: Vehicle ran off the road when making left turn			
			
Roadway	Undivided two-lane road	**Lighting**	Daylight
Weather	No adverse weather	**Location Type**	Unsignalized Y-intersection
Maneuver	Turning left	**Crash severity**	Level 3
Travel Speed	High (~50 mph)		
Description	The subject driver negotiated a curve to the left and approached to a Y-intersection. The destination road of the subject driver was the left leg, however the driver did not turn left sufficiently following the appropriate path, running off the road between the two legs and narrowly missing a sign.		
Observation	In this scenario, the intersection layout was uncommon. The centerline markings were not present at the intersection, leaving a wide paved area unmarked. The median area separating the two approaches was lower than the pavement, leaving the roadway edge of the left leg undelineated. Potential countermeasures would be better delineating the turning paths using centerline/edgeline markings.		

7.2 Conclusion

The scenarios illustrated by the case studies and other SHRP2 fixed-object events pointed at a number of contributing factors and potential countermeasures:

- Driver errors. Driver errors were a major contributing factor for many SHRP2 fixed-object crashes. Among the various driver errors, the following were examples of common contributors:
    - Improper turns were found present for many fixed-object crashes at intersections or curved roadways involving raised medians and curbs. About one fourth (26.05%) events relevant to fixed objects involved improper turn. Previous logistic regression

analysis results indicated that drivers were about 10 times more likely to involve in fixed-object SCEs during or after making a turn than going straight.

  - Driver distraction contributed to many fixed-object crashes. Distractions involving secondary tasks were found in a number of SHRP2 fixed-object crashes.
  - Driver impairments such as drowsy driving were found in a number of events. Drowsy driving tends to particularly occur during nighttime. Combining with limited visibility, drowsy driving can frequently lead to fixed-object crashes.
  - Speeding or driving too fast for conditions could contribute to roadway departures and therefore led to fixed-object crashes involving roadside objects. Speeding is particularly risky for fixed-object crashes at intersections or on curved roadway sections.

- Roadway and traffic control issues. The case studies showed that many fixed-object crashes involved raised medians at intersections. In many cases, these medians were not marked with retroreflective materials or not well lighted during nighttime. Many fixed-object crashes occurring due to improper turns may be prevented by properly delineating the turning paths at intersections (especially intersections with unconventional layouts) with pavement markings. In some cases, the use of traffic control devices introduced additional objects that created risks for fixed-object crashes.
- Lighting and visibility. Many fixed-object crashes occurred during nighttime. Some cases were on roadways without lighting while others occurred on roadways with lighting. However, a study of fixed object cases on roadways with lighting also unveiled that improperly designed lighting could not prevent and in some cases contributed to fixed-object crashes.
- Adverse weather conditions. Adverse weather conditions can be a contributing factor for fixed-object crashes, since such weather conditions frequently result in limited visibility and slippery roadways.

Chapter 8: Conclusions and Recommendations

8.1 Effects of Influential Factors on Fixed-object Crashes

The logistic regression analyses lead to a series of findings relevant to effects of influential factors on fixed-object crash occurrence and severity. The following summarizes the major findings of logistic regression analyses:

8.1.1 Effects of Driver Behavior Factors on Fixed-object Crashes

Pre-incident Maneuver

Pre-incident maneuver played a more significant role in the occurrence of a fixed-object event than the severity outcome. Drifting when going straight and making a turn were identified as significant contributors that increased the odds of fixed-object event occurrence by 4 and 10 times, respectively. However, these two factors only increased the risk of a more severe outcome by 1.6 and 1.3 times. Drifting when going straight is often due to driver inattention, frequently resulting in roadway departures and collisions with roadside infrastructures. Making a turn is frequently occurred at intersections and therefore has a higher likelihood of crashing with objects at intersections such as curb corners, raised medians, and other channelization and delineation devices.

Driver Behavior

Driver behavior or errors played a critical role in both fixed-object occurrence and severity level, but the identified values and significance levels were different among models. In particular, the event occurrence analysis identified behaviors such as failed or improper signal, sign/signal violation, and the improper turn to be significant contributors to the probability of the occurrence of a fixed-object crash while the severity analysis showed that these behaviors had little contribution to crash severity outcomes. On the other hand, distraction, drowsiness, fatigue, and speeding correlated with increased severity outcomes when a fixed-object crash took place. Interestingly, avoiding animals/vehicles was found to be associated with a higher probability of SCE occurrence, but the severity analysis indicated that this action tended to correspond to less severe events.

The sensitivity analyses of both SVM classifiers confirmed that driver behavior was one of the most significant factors affecting the event occurrence and severity outcomes.

Critical Speed

The event occurrence analysis showed that fixed-object crashes and near crashes were more likely to occur on low-speed traveling condition. This finding was consistent with the analysis results of locality, which suggested that fixed-object events were more likely to take place on local roadways than on freeways and continuous divided highway segments. The results of event severity analysis, however, indicated that increased speeds led to increased risk of a more severe crash. The sensitivity analysis of the binary SVM classifier also showed that critical speed was a significant variable distinguishing SCEs and baseline events.

Reaction Time

It was not surprising to conclude that sufficient driver reaction time was beneficial for successful evasion. The results showed that each extra second reduced the likelihood of an event being a severe fixed-object crash by a factor of 0.55. Driver distraction, engagement of secondary tasks, and impairments (drowsiness, fatigue, sleepiness, etc.) could all potentially reduce driver reaction time when a conflict occurs. Reaction time was also identified as a significant variable affecting event severity outcomes in the sensitivity analysis of multi-class SVM classifier.

Secondary Tasks

This variable was only identified to be significant in the event occurrence analysis. Involvement of secondary tasks such as reaching/moving objects in vehicle, personal hygiene related activities, and adjusting/monitoring vehicle devices increased the risks of fixed-object crashes, with odds ratios of 28.4, 2.6, and 2.2, respectively.

8.1.2 Effects of Roadway and Traffic-Related Factors on Fixed-object Crashes

Traffic Density

The results of both logistic regression models showed that LOS A1 (free flow, with no leading traffic) was the most hazardous traffic conditions associated with significantly higher risks of fixed-object event occurrence and more severe outcomes. This finding is consistent with the "cooperative safety" concept. Under this concept, vehicles following other vehicles tend to be safer since the leading vehicles can help following vehicles by showing the safe paths. This phenomenon is particularly evident during nighttime or at locations with complex roadway features potentially conflicting with traffic.

Locality

The results showed that more fixed-object events and more severe fixed-object crashes occurred on local roads such as business/industrial, residential, urban, and open country compared to freeways and continuous divided highway segments. In particular, roadways characterized as open country were found to have increased the probability of a fixed-object event by 3 times, and caused a more severe outcome by 16.6 times. Open country was assigned to roadways where only vegetation was visible on the roadside. Events falling into this category were commonly characterized as no leading traffic, rural two-lane road, and conflicts with animals such as deer. Note that only included 23 SCEs and 19 baseline events were included in the "open country" category.

Struck Object Type

The results showed that struck objects such as ditches, utility/light poles, tree/shrubs, and roadway debris generally increased the probability of severe fixed-object crash outcomes compared to curb strikes, but events involving animals and vehicles tended to be associated with less severe outcomes. A potential explanation was that drivers were more vigilant when encountering those objects than curbs, hence were more likely to make successful evasive maneuvers and resulted in near-crash cases instead of crashes. The significantly low odds ratios and negative parameters associated with vehicle and animal conflicts suggested that these two objects were the ones that drivers want to strike the least. The sensitivity analysis of the multi-class SVM classifier suggested that struck object type was a significant factor for fixed-object event severity outcomes.

AADT and IRI

These two variables only appeared significantly in the model comparing SCEs and baselines. Note that AADT in that analysis did not reflect actual traffic condition during the events (see the traffic density variable). It was more considered as a surrogate for certain roadway types. The results seemingly indicated a roughly direct correlation between AADT and fixed-object crash probability. While the results seemed to suggest that roadways with an IRI between 20-150 had lower odds for fixed-object crashes, either smoother or rougher pavement increased the likelihood of fixed-object crashes. Both AADT and IRI were obtained from RID for only a portion of the events. In addition, a number of AADT and IRI levels where not significant in the model.

8.1.3 Effects of Environment-related Factors on Fixed-object Crashes

Lighting Conditions

During the modelling of fixed-object SCEs versus baseline events, both darkness lighted and darkness not lighted conditions correlated with risk levels of fixed-object crash occurrence that was almost two times higher than that for daytime. Note that the odds ratio for darkness lighted was actually higher than that for darkness not lighted, which could be due to some of the following reasons:

- A large proportion of the events analyzed in this study occurred at or near intersections in urban areas. Street lighting is more likely to be present at such locations and therefore lighting in this context was merely a surrogate for the location conditions where lighting was present.
- Roadway lighting, when designed/installed improperly, results in poor uniformity across the roadway surface. Brighter areas in some cases may attract driver attention from darker areas on the roadway surface and therefore help hide fixed objects if they happen to be located in those darker areas.
- When lighting levels are inadequate to light on-road/roadside objects, the presence of lighting becomes irrelevant to the risk levels of fixed-object crashes.
- Lighting infrastructure (i.e., light poles) can also create additional risks for fixed-object crashes if not properly placed.

Roadway surface

According to the results, slippery roadway surface significantly contributed to the severity outcomes of the fixed-object crashes based on the ordinal event severity modelling, but did not significantly affect the occurrence probability of fixed-object crashes.

Adverse weather

The results showed that the presence of adverse weather contributed to fixed-object crashes with higher severity, but this factor did not influence the event occurrence significantly.

8.2 Comparison between Logistic Regression and Machine Learning Methods

Logit regression methods are widely adopted by transportation safety analyses. The inference is generated based on assumptions of population distribution, relationship between response and independent variables, independence of independent variables. Machine learning methods do not require beforehand knowledge of data distribution, and hence can be more robust in some cases. This thesis used traditional logit regression and two machine learning methods, SVM and ANN. For all analyses, the author applied the constructed models to original data to determine the model prediction accuracy, which illustrated how well the model could explain the original datasets. It was found that these three methods gave comparable prediction performance, while machine learning methods may perform better for SCE severity level (Table 20). It should be noted that machine learning models did not include RID data, and logit regression models only involved variables with statistical significance.

Table 20. Prediction Accuracy for Different Methods

Method	SCE occurrence	SCE severity level
Logit Regression	88.9%	69.4%
SVM	87.4%	74.7%
ANN*	~87%	~75%

*. The prediction accuracy of ANN models varied significantly with different trials, the values in the table were rough estimations.

Previous studies found that the correct prediction accuracy of machine learning methods for crash severity ranging from 48.8% to 90% (*26, 28, 43, 44*). The prediction performance determined in this study was generally satisfactory. An important factor leading to the satisfactory prediction performance was the usage of NDS data, which could capture a detailed crash scenario, and resulted in many variables depicting real-time driving conditions, such as driver behaviors, reactions, secondary tasks, traffic density. On the other hand, this high prediction accuracy also suggested the potential success of crash prediction system that combined real-time driver behavior, real-time traffic condition and roadway GIS data.

It was not the goal of this thesis to develop a machine learning model with strong prediction powers, that could identify fixed-object crashes when given additional datasets. The purpose of SVM and ANN analyses was to explore the viability and effectiveness of employing machine learning methods on transportation safety analyses. It can be found that machine learning methods gave comparative or even better prediction performance compared to logit regression analyses. However, logistic regression methods were advantageous in giving interpretable parameter estimates, that could explain the effects of independent variables on responses. Although in this study, the author used an indirect method, sensitivity analysis, to infer the contribution of independent variables on models. It only concluded that there was a strong association between the identified significant variables and responses, no inference was given on whether the presence of that variable was beneficial or hazardous. In addition, it was found that sensitivity analyses only identified a few variables (2 and 3 for binary and multi-class SVM classifiers), and would fail to work if the variance between each training trials was large. These indicated that even with the help of sensitivity analyses, machine learning methods are still weak in explaining the cause and effect relationship between input variables and responses.

8.3 Recommendations

The results based on the statistical, machine learning, and case study analyses pointed at the following recommendations to effectively mitigate fixed-object crash risks:

- Roadway improvements. Several findings suggested the importance of improving roadways to mitigate fixed-object crash risks. Transportation agencies should avoid sharp curves whenever possible to reduce roadway departures. At curved segments where pavement edges may not be clearly visible due to certain terrains or roadside ditches, retroreflective edge lines should be added to aid driver navigation. At rural intersections, including particularly intersections with unconventional layouts, retroreflective edge and lane markings should also be considered to clearly delineate the travel paths.

- Traffic control and signage. The findings of this study demonstrated the risks caused by insufficient and/or improperly designed traffic control devices and methods for fixed-object crashes. Based on the results, the author recommends:

    - Delineate raised channelizing islands at intersections. Raised medians and curbs (including curbs for roadway diets) were frequently involved in the analyzed fixed-object crashes, particularly during low visibility conditions. Running over such devices may cause vehicles to temporarily lose control and therefore increase the risks of secondary, and sometimes severe crashes. Channelization devices and curbs with a high risk of being struck should be sufficiently delineated with retroreflective materials to improve their visibility and conspicuity particularly during low light conditions.

    - Remove or relocate unnecessarily on-road or roadside structures. The location of signs and other traffic control devices should be designed appropriately, possibly consider factors such as alignments (e.g., direction and characteristics of horizontal and vertical curves) and sight distance. Case study 10, for example, illustrates a scenario where the sign pole may be better installed on the side instead of the middle of the diverging ramp lanes. In addition to adding the sign post in the middle of the two ramp lanes, the sign placement also required the addition of the longitudinal guardrail that further added to the risks of fixed-object crashes.

    - Provide pavement markings at intersections to guide turning vehicles. A number of fixed-object crashes involved improper turning movements as illustrated by the case studies. To reduce improper turning, a potentially effective measure is to add adequate pavement markings to clearly guide left-turning vehicles through the intersection. This measure is particularly beneficial for intersections with unconventional layouts, large paved areas, and on major roadways with unlevel pavement surface due to terrain or drainage requirements.

- Provide sufficient, proper lighting when necessary. Lighting, when used properly, is an effective method to improve visibility during nighttime. The case studies in particular demonstrated a number of issues related to the use and design of lighting. The current lighting design standards use minimum lighting levels as the primary design control. Designers frequently design lighting to meet this requirement without carefully examining the risk locations of a design area. At intersections, for example, the focus areas of the luminaires should be on high-risk features, such as the raised medians and/or curbs at intersections, and gore points where lanes divide. In addition, at intersections with unconventional layouts, lighting should be provided to clearly illuminate the paths and layout of the intersection approaches.

- The case studies also illustrated the needs for better designed lighting transition at locations where lighting levels change. Sufficient lighting transition allows drivers to visually adapt to the new lighting level and therefore avoid temporary blindness during driving.

- Consider driver education and enforcement whenever possible. Several findings showed the significant role of driver errors in the occurrence and severity outcome of fixed-object crashes. Engineering solutions alone can hardly eliminate crashes. It is important to conduct driver education and enforcement activities when it is possible to raise safety awareness and ensure safety driver behavior. Risky driver behaviors identified during this study included, for example, speeding, distraction, and drowsiness/fatigue.

As previously noted, the SHRP2 crashes and near crashes included a large proportion of minor crashes. To completely understand fixed-object crash risks, it is necessary to further compare the findings of this study with those of similar studies that used police-reported crashes.

REFERENCE

National Highway Traffic Safety Administration. *Traffic Safety Facts 2015*. Publication DOT HS 812 384. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812384. Accessed Dec. 7, 2017.

Hillary, F. G., Schatz, P., Moelter, S. T., Lowry, J. B., Ricker, J. H., & Chute, D. L. Motor vehicle collision factors influence severity and type of TBI. Brain injury, Vol. 16, No. 8, 2002, pp. 729-741.

Bambach, M. R., R. J. Mitchell, and G. A. Mattos. Mean Injury Costs of Run-Off-Road Collisions with Fixed Objects: Passenger Vehicles and Motorcycles. Journal of Transportation Safety and Security, Vol. 7, No. 3, 2015, pp. 228-242.

National Highway Traffic Safety Administration. Fatality Analysis Reporting System. http://www-fars.nhtsa.dot.gov. Accessed Dec. 7, 2017.

Insurance Institute for Highway Safety Highway Loss Data Institute, Roadway and Environment: Collisions with Fixed Objects and Animals. http://www.iihs.org/iihs/topics/t/roadway-and-environment/fatalityfacts/fixed-object-crashes/2016. Accessed Feb. 7, 2018

Cejun Liu, and R. Subramanian. *Factors Related to Fatal Single-Vehicle Run-Off-Road Crashes*. Publication DOT HS 811 232. NHTSA, U.S. Department of Transportation. 2009.

Neyens, D. M., & Boyle, L. N. The effect of distractions on the crash types of teenage drivers. *Accident Analysis & Prevention*, Vol. 39, No.1, 2007, pp. 206-212.

TAKEMOTO, A., HIRASAWA, M., & KASAI, S. Causes of and Countermeasures for Fixed Object Crashes. *Proceedings of the Eastern Asia Society for Transportation Studies, The 9th International Conference of Eastern Asia Society for Transportation Studies,* 2011.

Dissanayake, S., & Lu, J. J. Factors influential in making an injury severity difference to older drivers involved in fixed object–passenger car crashes. *Accident Analysis & Prevention*, Vol. 34, No. 5, 2002, pp. 609-618.

Souleyrette, R., Kamyab, A., Hans, Z., Knapp, K. K., Khattak, A., Basavaraju, R., & Storm, B. *Systematic identification of high crash locations*. Publication Iowa DOT TR-442, CTRE 0059, Final Report. 2001.

Holdridge, J. M., Shankar, V. N., & Ulfarsson, G. F. The crash severity impacts of fixed roadside objects. *Journal of Safety Research*, Vol. 36, No. 2, 2005, pp. 139-147.

Dissanayake, S., & Roy, U. Crash severity analysis of single vehicle run-off-road crashes. *Journal of Transportation Technologies*, Vol. 4, No. 1, 2014, p. 1-10.

Roadside Design Guide. American Association of State Highway and Transportation Officials, Washington, DC. 2002.

Transportation Research Board of the National Academies; Virginia Tech Transportation Institute, 2017, "The 2nd Strategic Highway Research Program Naturalistic Driving Study InSight Dataset (v3.0)", doi:10.15787/VTT1/IEKRD3, VTTI Root Dataverse, V1

Campbell, K. The SHRP2 Naturalistic Driving Study: Addressing driver performance and behavior in traffic safety. *TR News*, Vol. 282, 2012, pp. 30-35.

SHRP2 - Roadway Information Database. http://www.ctre.iastate.edu/shrp2-rid/rid.cfm. Accessed May. 21, 2018.

McLaughlin, S., & Hankey, J. *Naturalistic Driving Study: Linking the study data to the Roadway Information Database.* Publication SHRP2 Report S2-S31-RW-3. Virginia Tech Transportation Institute. 2015.

Hankey, J. M., M. A. Perez, and J. A. McClafferty. "Description of the SHRP2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets." Virginia Tech Transportation Institute. 2016.

Hallmark, S. L., Tyner, S., Oneyear, N., Carney, C., & McGehee, D. Evaluation of driving behavior on rural 2-lane curves using the SHRP2 naturalistic driving study data. *Journal of safety research*, Vol. 54, 2015, pp. 17-27.

Yamamoto, T., & Shankar, V. N. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects. *Accident Analysis & Prevention*, Vol. 36, No. 5, 2004, pp. 869-876.

Szumilas, M. Explaining odds ratios. *Journal of the Canadian academy of child and adolescent psychiatry*, Vol. 19, No. 3, 2010, pp. 227.

Hallmark, S., Hsu, Y. Y., Boyle, L., Carriquiry, A., Tian, Y., & Mudgal, A. *Evaluation of data needs, crash surrogates, and analysis methods to address lane departure research questions using naturalistic driving study data*. Publication SHRP2 Report S2-S01E-RW-1. Virginia Tech Transportation Institute. 2011.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, Vol. 113, No. 10, 2016, pp. 2636-2641.

Chang, L. Y., & Wang, H. W. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention*, Vol. 38, No. 5, 2006, pp. 1019-1027.

Das, S., Brimley, B. K., Lindheimer, T., & Pant, A. *Safety Impacts of Reduced Visibility in Inclement Weather*. Publication ATLAS-2017-19. 2017.

Li, Z., Liu, P., Wang, W., & Xu, C. Using support vector machine models for crash injury severity analysis. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 478-486.

Li, X., Lord, D., Zhang, Y., & Xie, Y. Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention*, Vol. 40, No. 4, 2008, pp. 1611-1618.

Delen, D., Tomak, L., Topuz, K., & Eryarsoy, E. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. *Journal of Transport & Health*, Vol. 4, 2017, pp. 118-131.

Tefft, B. C. The Prevalence of Motor Vehicle Crashes Involving Road Debris, United States, 2011-2014. AAA Foundation for Traffic Safety. 2016, pp. 1-10 http://mail.thenewspaper.com/rlc/docs/2016/aaadebris.pdf. Accessed June. 1, 2018.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E. Sudweeks, J. D., ... & Bucher, C. *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment*. Publication No. HS-810 593. 2006.

SAS Institute Inc. SAS/STAT 9.2 User's Guide, Second Edition. https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_reg_sect038.htm. Accessed May. 15, 2018.

Mathwork. Classification Learner. https://www.mathworks.com/help/stats/classificationlearner-app.html. Accessed May. 20, 2018.

Mathwork. Neural Network Toolbox. https://www.mathworks.com/products/neural-network.html. Accessed May. 20, 2018.

Hosmer, D.W. and S. Lemeshow. *Applied Logistic Regression* (3nd Edition). John Wiley & Sons, New York. 2013.

Cortes, C., & Vapnik, V. Support-vector networks. *Machine learning*, Vol. 20, No. 3, 1995, pp. 273-297.

Hsu, C. W., & Lin, C. J. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, Vol. 13, No. 2, 2002, pp. 415-425.

Wikipedia contributor. Support vector machine. Wikipedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/Support_vector_machine. Accessed May. 15, 2018.

Nielsen, M. A. *Neural networks and deep learning*. 2015. http://neuralnetworksanddeeplearning.com. Accessed Jun. 1, 2018

Wikipedia contributor. Cross-validation (statistics). Wikipedia, Wikimedia Foundation, https://en.wikipedia.org/wiki/Cross-validation_(statistics). Accessed May. 20, 2018.

Fish, K.E., Blodgett, J.G. A visual method for determining variable importance in an artificial neural network model: an empirical benchmark study. *Journal of Targeting Measurement and Analysis for Marketing*, Vol. 11, No. 3, 2013, pp. 244-254.

Delen, D., Sharda, R., & Bessonov, M. 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.

Wang, X., P. Shu, L. Cao, and Y. Wang. A ROC Curve Method for Performance Evaluation of Support Vector Machine with Optimization Strategy. *Proceedings of the 2009 International Forum on Computer Science-Technology and Applications*, December 25-27, 2009.

Alkheder, S., Taamneh, M., & Taamneh, S. Severity prediction of traffic accident using an artificial neural network. *Journal of Forecasting*, Vol. 36, No. 1, 2016, pp. 100-108.

Iranitalab, A., & Khattak, A. Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 27-36.

APPENDIX A. Detail Table of Explanatory Variables

Variable	Definition	Values	SVM vector composition	Baseline	SCE
Age Group	The age group participant driver falling in.	16-19	1	192	343
		20-29	2	360	511
		30-39	3	75	112
		40-49	4	71	112
		50-59	5	86	122
		60-69	6	91	151
		70-79	7	97	146
		80-89	8	62	121
		90-99	9	2	3
Annual Miles	The estimated average annual mileage over the past five years of participant driver.	0-5000 miles	1	100	192
		5000-10000	2	211	330
		10000-15000	3	344	503
		15000-20000	4	159	262
		20000-25000	5	94	131
		25000-30000	6	52	86
		30000 and more	7	64	96
Number of Violation	The number of violations that the participant driver has in last three years.	0	0	658	1058
		1	1	241	344
		2 or 2+	2	141	218
State	The state where the event happens.	FL	[1 0 0 0 0 0]	259	444
		IN	[0 1 0 0 0 0]	89	156
		NC	[0 0 1 0 0 0]	169	259
		NY	[0 0 0 1 0 0]	251	366
		PA	[0 0 0 0 1 0]	63	95
		WA	[0 0 0 0 0 1]	219	319
Years of Driving	The number of years since when the driver gets his/her license.	<=1	1	68	142
		(1,2]	2	86	130
		(2,5]	3	182	275
		(5,10]	4	185	258
		(10,20]	5	105	138
		>=20	6	415	687
Pre-incident Maneuver	The last action of driving maneuver that the driver engaged in prior to or at the time of event.	Changing lanes	[1 0 0 0 0 0]	30	63
		Going straight	[0 1 0 0 0 0]	863	607
		Going straight but with unintentional drifting	[0 0 1 0 0 0]	8	71
		Making a turn	[0 0 0 1 0 0]	8	621
		Negotiating a curve	[0 0 0 0 1 0]	131	240
		Other	[0 0 0 0 0 1]	10	37
Maneuver judgment	judgment of the safety and legality of pre-incident maneuver.	Safe and legal	[1 0 0 0]	993	1118
		Safe but illegal	[0 1 0 0]	5	65
		Unsafe and illegal	[0 0 1 0]	44	239
		Unsafe but legal	[0 0 0 0]	8	217
Driver Behavior	Driver behaviors that occurred within seconds prior to the event, or those contribute to the event occurrence.	Apparent unfamiliar with roadway	[1 0 0 0 0 0 0 0 0]	0	36
		Avoiding animal, or other vehicle	[0 1 0 0 0 0 0 0 0]	2	21
		Distracted	[0 0 1 0 0 0 0 0 0]	0	313
		Drowsy, sleepy, asleep, fatigued	[0 0 0 1 0 0 0 0 0]	24	30
		Exceeded safe speed, or speed limit	[0 0 0 0 1 0 0 0 0]	43	201
		Failed to signal, improper signal	[0 0 0 0 0 1 0 0 0]	4	34

		Improper turn	[0 0 0 0 0 0 1 0 0]	1	430
		None	[0 0 0 0 0 0 0 0 0]	958	481
		Sign, signal violation	[0 0 0 0 0 0 0 1 0]	2	37
		Other	[0 0 0 0 0 0 0 0 1]	16	56
Driver Impairments	Physical impairments (drowsiness, fatigue) or police-reported impairments (alcohol, drug) that possibly result in the event.	None apparent	0	1024	1562
		With apparent impairments	1	26	77
Passenger Existence	Whether other passengers exist in the vehicle.	No	0	689	1219
		Yes	1	361	420
Secondary Task	Any observable driver engagement in tasks irrelevant to driving prior, or at the time of event.	Adjusting/monitoring vehicle devices	[1 0 0 0 0 0 0 0 0 0]	21	56
		Cell phone usage	[0 1 0 0 0 0 0 0 0 0]	89	167
		Drinking/Eating/Smoking	[0 0 1 0 0 0 0 0 0 0]	24	44
		External distraction (objects, animal, pedestrian, and etc.)	[0 0 0 1 0 0 0 0 0 0]	59	141
		Interaction with passenger	[0 0 0 0 1 0 0 0 0 0]	162	209
		Internal distraction (object, pet, etc.)	[0 0 0 0 0 1 0 0 0 0]	33	51
		Other	[0 0 0 0 0 0 1 0 0 0]	27	38
		Personal hygiene	[0 0 0 0 0 0 0 1 0 0]	29	52
		Reaching, moving object in vehicle	[0 0 0 0 0 0 0 0 1 0]	6	106
		Talking/singing, audience unknown	[0 0 0 0 0 0 0 0 0 1]	75	121
		No secondary tasks	[0 0 0 0 0 0 0 0 0 0]	525	654
Hands on Wheel	How many hands the driver had on the steering wheels at the start of the event.	Both hands	[1 0 0]	420	737
		None or at least one hand off	[0 1 0]	26	36
		Only one hand or at least one hand on	[0 0 1]	604	866
Driver Seatbelt Use	The use of the seatbelt of driver at the time of the event.	Lap/shoulder belt not properly worn or not used	0	44	135
		Lap/shoulder belt properly worn	1	1006	1504
Lighting	Lighting condition at the time of event.	Daylight	[1 0 0 0]	790	1072
		Dawn, dusk	[0 1 0 0]	50	81
		Darkness, not lighted	[0 0 1 0]	95	143
		Darkness, lighted	[0 0 0 1]	115	343
Weather	Weather condition at the time of event.	Adverse weather	1	82	209
		No adverse weather	0	968	1430
Road Surface Condition	The type of roadway surface condition that would influence vehicle coefficient of friction at the time of event.	Dry	0	910	1258
		Icy/snowy/wet	1	140	381
Traffic Flow	Roadway design regarding to the presence of median or barrier. If the event occurs in intersection, it indicates the roadway design prior to that intersection.	Divided (median strip or barrier)	[1 0 0 0]	598	398
		No lanes	[0 1 0 0]	0	41
		Not divided	[0 0 1 0]	424	1119
		One-way traffic	[0 0 0 1]	28	81
Traffic Density	The level of service at the time of the event. Based on number of vehicles present in subject's travel direction and the ability of the subject driver	LOS A1: Free flow, no lead traffic	[1 0 0 0 0]	293	910
		LOS A2: Free flow, leading traffic present	[0 1 0 0 0]	366	366
		LOS B: Flow with some restrictions	[0 0 1 0 0]	305	312
		LOS C: Stable flow, maneuverability and speed are more restricted	[0 0 0 1 0]	56	44

	to maneuver between lanes and select driving speed.	LOS D/E/F	[0 0 0 0 1]	30	7
Contiguous Travel Lane	The total number of contiguous travel lanes at the time of event, that vehicles could easily maneuver into, including any turn lanes.	0	0	0	41
		1	1	22	96
		2	2	458	974
		3	3	253	228
		4	4	164	151
		5 and 5+	5	153	149
Traffic Control	Type of traffic control applicable to the subject's travel direction at the time of event.	No traffic control	[0 0 0 0]	931	1150
		Sign control	[1 0 0 0]	63	212
		Traffic lanes marked	[0 1 0 0]	13	40
		Traffic signal	[0 0 1 0]	40	175
		Other	[0 0 0 1]	3	62
Relation to Junction	The spatial relation of the roadway that the vehicle was traveling to the junction.	Driveway, alley access, etc.	[1 0 0 0 0 0]	59	156
		Entrance/Exit ramp	[0 1 0 0 0 0]	26	33
		Interchange area	[0 0 1 0 0 0]	144	67
		Intersection, intersection-related	[0 0 0 1 0 0]	135	500
		Parking lot entrance/exit	[0 0 0 0 1 0]	49	269
		Other	[0 0 0 0 0 1]	0	11
		Non-junction	[0 0 0 0 0 0]	637	603
Alignment	The roadway curvature in the subject vehicle's travel direction at the time of event.	Curve	1	155	327
		Straight	0	895	1312
Grade	The roadway profile in the subject vehicle's travel direction at the time of event.	Dip/Hillcrest	[1 0 0]	6	20
		Grade down	[0 1 0]	64	116
		Grade up	[0 0 1]	99	184
		Level	[0 0 0]	881	1319
Locality	Description of the surroundings that influence or might influence the flow of traffic at the time of event.	Business/Industrial	[1 0 0 0 0 0 0]	252	612
		Bypass/Divided highway with traffic signal	[0 1 0 0 0 0 0]	52	27
		Interstate/Bypass/Divided Highway with no traffic signal	[0 0 1 0 0 0 0]	451	173
		Open country	[0 0 0 1 0 0 0]	19	23
		Residential area	[0 0 0 0 1 0 0]	199	591
		School/Church/Playground	[0 0 0 0 0 1 0]	60	159
		Urban	[0 0 0 0 0 0 1]	15	54
		Other	[0 0 0 0 0 0 0]	2	0
Construction Zone	Description of whether the event occurs in or in relation to construction zone.	Construction zone, or related	1	49	74
		No construction zone influence	0	1001	1565
AADT	Annual average daily traffic, an indicator of traffic volume.	<=5000	1	92	99
		5000-10000	2	109	102
		10000-15000	3	94	78
		15000-20000	4	76	56
		20000-30000	5	97	93
		30000-50000	6	106	79
		50000-100000	7	79	34
		>100000	8	58	39
Speed Limit	The posted speed limit of the subject's traveling roadway at the time of event.	20		1	1
		25		0	2
		30		5	11
		35		12	22

		40		11	10
		45		40	16
		50		15	6
		55		63	22
		60		10	10
		65		31	11
		70		24	6
IRI	Internal roughness index, an indicator of road surface roughness.	<=20		20	57
		20-50		41	19
		50-100		307	103
		100-150		109	69
		150-200		28	43
		>200		10	41
Radius	The radius of roadway curvature.	0-500		11	25
		500-1000		18	36
		1000-2000		26	27
		2000-5000		60	31
		>5000		58	32
		Straight		397	434
Curve Direction	The curve direction of roadway in vehicle's travel direction.	-1		85	69
		0		397	434
		1		88	82
Stuck Object Type	The type of object that involved in the event.	Others	[1 0 0 0 0 0 0 0 0 0 0 0 0]		94
		Raised median	[0 1 0 0 0 0 0 0 0 0 0 0 0]		121
		Animal	[0 0 1 0 0 0 0 0 0 0 0 0 0]		366
		Concrete barrier	[0 0 0 1 0 0 0 0 0 0 0 0 0]		7
		Curb	[0 0 0 0 1 0 0 0 0 0 0 0 0]		640
		Ditch	[0 0 0 0 0 1 0 0 0 0 0 0 0]		15
		Pavement edge/edge line	[0 0 0 0 0 0 1 0 0 0 0 0 0]		201
		Roadway debris	[0 0 0 0 0 0 0 1 0 0 0 0 0]		96
		Sign post	[0 0 0 0 0 0 0 0 1 0 0 0 0]		10
		Stopped, backing, pulling car	[0 0 0 0 0 0 0 0 0 1 0 0 0]		58
		Tree/shrub	[0 0 0 0 0 0 0 0 0 0 1 0 0]		7
		Utility/light pole	[0 0 0 0 0 0 0 0 0 0 0 1 0]		6
		W-beam barrier	[0 0 0 0 0 0 0 0 0 0 0 0 1]		18
Critical speed	The speed representing general traveling speed.	Continuous			
Reaction time	The time that allowing driver to make an evasive maneuver.	Continuous			

APPENDIX B: SAS Output

SAS output for binomial logit regression model

Model Information		
Data Set	WORK.IMPORT	
Response Variable	eventSeverity	eventSeverity
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	2689
Number of Observations Used	2689

Response Profile		
Ordered Value	**eventSeverity**	**Total Frequency**
1	0	1050
2	1	1639

Probability modeled is eventSeverity='1'.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	3599.679	1588.039
SC	3605.576	2018.514
-2 Log L	3597.679	1442.039

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	2155.6398	72	<.0001
Score	1523.5849	72	<.0001
Wald	512.1695	72	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
54.0876	67	0.8724

Summary of Forward Selection						
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	driverBehavior	9	1	1051.5532	<.0001	driverBehavior
2	criticalSpeed	1	2	253.9842	<.0001	criticalSpeed
3	IRI	6	3	139.5487	<.0001	IRI
4	trafficDensity	4	4	95.6389	<.0001	trafficDensity
5	secondaryTask	10	5	66.7628	<.0001	secondaryTask
6	preIncidentManeuver	5	6	43.5485	<.0001	preIncidentManeuver
7	AADT	8	7	47.9849	<.0001	AADT
8	locality	7	8	57.6869	<.0001	locality
9	maneuverJudgment	3	9	24.3960	<.0001	maneuverJudgment
10	contigTravelanes	5	10	18.5644	0.0023	contigTravelanes
11	handsOnTheWheel	2	11	10.7570	0.0046	handsOnTheWheel
12	lighting	3	12	14.2809	0.0025	lighting
13	trafficControl	4	13	9.7811	0.0443	trafficControl
14	surfaceCondition	1	14	3.6566	0.0558	surfaceCondition
15	passengerExistence	1	15	3.2588	0.0710	passengerExistence
16	grade	3	16	6.6282	0.0847	grade

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
preIncidentManeuver	5	34.3704	<.0001
maneuverJudgment	3	16.0199	0.0011
driverBehavior	9	39.7489	<.0001
passengerExistence	1	2.9601	0.0853
secondaryTask	10	53.0622	<.0001
handsOnTheWheel	2	10.1661	0.0062
lighting	3	13.6342	0.0034
surfaceCondition	1	3.4882	0.0618
trafficDensity	4	21.9477	0.0002
contigTravelanes	5	13.7778	0.0171
trafficControl	4	9.8510	0.0430
grade	3	6.5556	0.0875
locality	7	36.7197	<.0001
criticalSpeed	1	60.5405	<.0001
AADT	8	76.1996	<.0001
IRI	6	46.0758	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	13.9110	722.9	0.0004	0.9846
preIncidentManeuver	Changing lanes	1	0.9423	0.3514	7.1918	0.0073
preIncidentManeuver	Going straight, but with unintentional "drifting" within lane or across lanes	1	1.3933	0.5978	5.4315	0.0198
preIncidentManeuver	Making a turn	1	2.2989	0.5041	20.7979	<.0001
preIncidentManeuver	Negotiating a curve	1	0.0560	0.1899	0.0871	0.7679
preIncidentManeuver	Other	1	-0.9005	0.5978	2.2691	0.1320
maneuverJudgment	Safe but illegal	1	-0.5142	1.0402	0.2443	0.6211
maneuverJudgment	Unsafe and illegal	1	1.1902	0.5671	4.4044	0.0358
maneuverJudgment	Unsafe but legal	1	2.4294	0.6582	13.6215	0.0002
driverBehavior	Apparent unfamiliarity with roadway	1	17.3350	808.9	0.0005	0.9829
driverBehavior	Avoiding animal, or other vehicle	1	3.6471	0.8443	18.6589	<.0001
driverBehavior	Distracted	1	18.3936	250.4	0.0054	0.9415
driverBehavior	Drowsy, sleepy, asleep, fatigued	1	0.2975	0.3859	0.5940	0.4409

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
driverBehavior	Exceeded safe speed, or speed limit	1	0.7757	0.5919	1.7179	0.1900
driverBehavior	Failed to signal, improper signal	1	1.8436	1.1162	2.7283	0.0986
driverBehavior	Improper turn	1	4.4753	1.0569	17.9293	<.0001
driverBehavior	Other	1	0.9841	0.5248	3.5164	0.0608
driverBehavior	Sign, signal violation	1	1.6710	0.9844	2.8814	0.0896
passengerExistence	Yes	1	-0.3237	0.1882	2.9601	0.0853
secondaryTask	Adjusting/monitoring vehicle devices	1	0.8780	0.4324	4.1228	0.0423
secondaryTask	Cell phone usage	1	-0.3334	0.2940	1.2863	0.2567
secondaryTask	Drinking/Eating/Smoking	1	0.4598	0.4413	1.0858	0.2974
secondaryTask	External distraction (objects, animal, pedestrian, and etc.)	1	0.0841	0.2904	0.0840	0.7720
secondaryTask	Interaction with passenger	1	0.1970	0.2486	0.6277	0.4282
secondaryTask	Internal distraction (object, pet, non-specific eye glance)	1	-0.4643	0.4586	1.0250	0.3113
secondaryTask	Other	1	-0.1547	0.4928	0.0985	0.7536
secondaryTask	Personal hygiene	1	0.8268	0.4065	4.1371	0.0420
secondaryTask	Reaching, moving object in vehicle	1	3.3542	0.5194	41.7081	<.0001
secondaryTask	Talking/singing, audience unknown	1	0.2038	0.2591	0.6184	0.4317
handsOnTheWheel	None or at least one hand off	1	-1.9070	0.6371	8.9612	0.0028
handsOnTheWheel	Only one hand or at least one hand on	1	-0.2209	0.1439	2.3571	0.1247
lighting	Darkness, lighted	1	0.6621	0.1994	11.0242	0.0009
lighting	Darkness, not lighted	1	0.4590	0.2161	4.5099	0.0337
lighting	Dawn, Dusk	1	0.3223	0.3078	1.0964	0.2951
surfaceCondition	Icy/snowy/wet	1	0.3492	0.1870	3.4882	0.0618
trafficDensity	Level-of-service A2: Free flow, leading traffic present	1	-0.6545	0.1804	13.1572	0.0003
trafficDensity	Level-of-service B: Flow with some restrictions	1	-0.3569	0.1927	3.4302	0.0640
trafficDensity	Level-of-service C: Stable flow, maneuverability and speed are more restricted	1	-0.6304	0.4018	2.4620	0.1166
trafficDensity	Level-of-service D/E/F	1	-3.8469	1.2064	10.1675	0.0014
contigTravelanes	1	1	-13.2558	722.9	0.0003	0.9854
contigTravelanes	2	1	0.1449	0.3971	0.1332	0.7151
contigTravelanes	3	1	-0.6208	0.2108	8.6733	0.0032
contigTravelanes	4	1	0.0858	0.2646	0.1052	0.7456
contigTravelanes	5 and 5+	1	-0.2246	0.2990	0.5641	0.4526
trafficControl	Other	1	1.7570	0.7486	5.5092	0.0189
trafficControl	Sign control	1	0.5333	0.2557	4.3504	0.0370
trafficControl	Traffic lanes marked	1	-0.2542	0.5659	0.2017	0.6533
trafficControl	Traffic signal	1	0.1250	0.3712	0.1135	0.7362
grade	Dip, hillcrest	1	1.3454	0.7117	3.5741	0.0587
grade	Grade Down	1	0.1670	0.2594	0.4141	0.5199
grade	Grade Up	1	0.3754	0.2142	3.0712	0.0797
locality	Business/Industrial	1	0.8187	0.2239	13.3666	0.0003
locality	Bypass/Divided Highway with traffic signals	1	0.4167	0.3869	1.1596	0.2816
locality	Open Country	1	1.0840	0.4637	5.4653	0.0194
locality	Other	1	-13.0047	3881.8	0.0000	0.9973
locality	Residential area	1	1.3483	0.2249	35.9477	<.0001
locality	School/church/playground	1	0.7997	0.3146	6.4612	0.0110
locality	Urban	1	1.0099	0.6166	2.6824	0.1015
criticalSpeed		1	-0.0213	0.00274	60.5405	<.0001
AADT	1	1	-1.2394	0.2663	21.6600	<.0001
AADT	2	1	0.0306	0.3267	0.0088	0.9254

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
AADT	3	1	-0.1253	0.3520	0.1268	0.7218
AADT	4	1	0.2026	0.4127	0.2411	0.6234
AADT	5	1	0.2493	0.4326	0.3321	0.5644
AADT	6	1	0.2295	0.4050	0.3212	0.5709
AADT	7	1	0.6368	0.4299	2.1943	0.1385
AADT	8	1	1.2827	0.4294	8.9248	0.0028
IRI	1	1	1.4429	0.4369	10.9066	0.0010
IRI	2	1	-2.6842	0.6409	17.5390	<.0001
IRI	3	1	0.4296	0.5270	0.6645	0.4150
IRI	4	1	-0.00318	0.3255	0.0001	0.9922
IRI	5	1	0.9646	0.5015	3.6997	0.0544
IRI	6	1	1.2233	0.6559	3.4779	0.0622

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
preIncidentManeuver Changing lanes vs Going straight (accelerate, decelerate, constant speed)	2.566	1.289	5.109
preIncidentManeuver Going straight, but with unintentional "drifting" within lane or across lanes vs Going straight (accelerate, decelerate, constant speed)	4.028	1.248	13.002
preIncidentManeuver Making a turn vs Going straight (accelerate, decelerate, constant speed)	9.963	3.709	26.759
preIncidentManeuver Negotiating a curve vs Going straight (accelerate, decelerate, constant speed)	1.058	0.729	1.534
preIncidentManeuver Other vs Going straight (accelerate, decelerate, constant speed)	0.406	0.126	1.312
maneuverJudgment Safe but illegal vs Safe and legal	0.598	0.078	4.593
maneuverJudgment Unsafe and illegal vs Safe and legal	3.288	1.082	9.991
maneuverJudgment Unsafe but legal vs Safe and legal	11.352	3.124	41.243
driverBehavior Apparent unfamiliarity with roadway vs None	>999.999	<0.001	>999.999
driverBehavior Avoiding animal, or other vehicle vs None	38.362	7.332	200.715
driverBehavior Distracted vs None	>999.999	<0.001	>999.999
driverBehavior Drowsy, sleepy, asleep, fatigued vs None	1.346	0.632	2.869
driverBehavior Exceeded safe speed, or speed limit vs None	2.172	0.681	6.929
driverBehavior Failed to signal, improper signal vs None	6.319	0.709	56.330
driverBehavior Improper turn vs None	87.820	11.065	697.012
driverBehavior Other vs None	2.675	0.956	7.483
driverBehavior Sign, signal violation vs None	5.318	0.772	36.615
passengerExistence Yes vs No	0.723	0.500	1.046
secondaryTask Adjusting/monitoring vehicle devices vs No Secondary Tasks	2.406	1.031	5.615
secondaryTask Cell phone usage vs No Secondary Tasks	0.716	0.403	1.275
secondaryTask Drinking/Eating/Smoking vs No Secondary Tasks	1.584	0.667	3.761
secondaryTask External distraction (objects, animal, pedestrian, and etc.) vs No Secondary Tasks	1.088	0.616	1.922
secondaryTask Interaction with passenger vs No Secondary Tasks	1.218	0.748	1.982
secondaryTask Internal distraction (object, pet, non-specific eye glance) vs No Secondary Tasks	0.629	0.256	1.544
secondaryTask Other vs No Secondary Tasks	0.857	0.326	2.251
secondaryTask Personal hygiene vs No Secondary Tasks	2.286	1.031	5.071
secondaryTask Reaching, moving object in vehicle vs No Secondary Tasks	28.624	10.343	79.218

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
secondaryTask Talking/singing, audience unknown vs No Secondary Tasks	1.226	0.738	2.037
handsOnTheWheel None or at least one hand off vs Both hands	0.149	0.043	0.518
handsOnTheWheel Only one hand or at least one hand on vs Both hands	0.802	0.605	1.063
lighting Darkness, lighted vs Daylight	1.939	1.312	2.866
lighting Darkness, not lighted vs Daylight	1.583	1.036	2.417
lighting Dawn, Dusk vs Daylight	1.380	0.755	2.524
surfaceCondition Icy/snowy/wet vs Dry	1.418	0.983	2.046
trafficDensity Level-of-service A2: Free flow, leading traffic present vs Level-of-service A1: Free flow, no lead traffic	0.520	0.365	0.740
trafficDensity Level-of-service B: Flow with some restrictions vs Level-of-service A1: Free flow, no lead traffic	0.700	0.480	1.021
trafficDensity Level-of-service C: Stable flow, maneuverability and speed are more restricted vs Level-of-service A1: Free flow, no lead traffic	0.532	0.242	1.170
trafficDensity Level-of-service D/E/F vs Level-of-service A1: Free flow, no lead traffic	0.021	0.002	0.227
trafficControl Other vs No traffic control	5.795	1.336	25.132
trafficControl Sign control vs No traffic control	1.705	1.033	2.814
trafficControl Traffic lanes marked vs No traffic control	0.776	0.256	2.351
trafficControl Traffic signal vs No traffic control	1.133	0.547	2.345
grade Dip, hillcrest vs Level	3.840	0.952	15.492
grade Grade Down vs Level	1.182	0.711	1.965
grade Grade Up vs Level	1.456	0.957	2.215
locality Business/Industrial vs Interstate/Bypass/Divided Highway with no traffic signals	2.268	1.462	3.517
locality Bypass/Divided Highway with traffic signals vs Interstate/Bypass/Divided Highway with no traffic signals	1.517	0.711	3.238
locality Open Country vs Interstate/Bypass/Divided Highway with no traffic signals	2.956	1.191	7.336
locality Other vs Interstate/Bypass/Divided Highway with no traffic signals	<0.001	<0.001	>999.999
locality Residential area vs Interstate/Bypass/Divided Highway with no traffic signals	3.851	2.478	5.984
locality School/church/playground vs Interstate/Bypass/Divided Highway with no traffic signals	2.225	1.201	4.122
locality Urban vs Interstate/Bypass/Divided Highway with no traffic signals	2.745	0.820	9.192
criticalSpeed	0.979	0.974	0.984

SAS output for ordinal logit regression model

Model Information		
Data Set	WORK.IMPORT	
Response Variable	crashSeverity	crashSeverity
Number of Response Levels	5	
Model	cumulative logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	1639
Number of Observations Used	1639

Response Profile		
Ordered Value	**crashSeverity**	**Total Frequency**
1	1	13
2	2	35
3	3	447
4	4	534
5	5	610

Probabilities modeled are cumulated over the lower Ordered Values.

Score Test for the Proportional Odds Assumption		
Chi-Square	**DF**	**Pr > ChiSq**
11268.0364	150	<.0001

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	3968.104	3293.890
SC	3989.711	3585.590
-2 Log L	3960.104	3185.890

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	774.2134	50	<.0001
Score	558.8899	50	<.0001
Wald	574.0044	50	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
69.3808	91	0.9554

Summary of Forward Selection						
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	structObjectType	12	1	366.8387	<.0001	struckObjectType
2	driverBehavior	9	2	102.7041	<.0001	driverBehavior
3	reactionTime	1	3	56.2504	<.0001	reactionTime
4	locality	6	4	27.1837	0.0001	locality
5	trafficDensity	4	5	16.2946	0.0026	trafficDensity
6	Radius	6	6	15.7592	0.0151	Radius
7	preIncidentManeuver	5	7	11.2375	0.0469	preIncidentManeuver
8	numViol	3	8	7.7316	0.0519	numViol
9	criticalSpeed	1	10	2.9397	0.0864	criticalSpeed
10	weather	1	11	2.7651	0.0963	weather

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
numViol	3	6.0748	0.1080
preIncidentManeuver	5	13.0414	0.0230
driverBehavior	9	107.2417	<.0001
weather	1	2.9861	0.0840
trafficDensity	4	12.3309	0.0151
locality	6	36.3792	<.0001
criticalSpeed	1	3.7883	0.0516
Radius	6	21.7426	0.0013
struckObjectType	12	266.8916	<.0001
reactionTime	1	84.5230	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	-5.4665	0.6525	70.1777	<.0001
Intercept	2	1	-4.0579	0.6050	44.9919	<.0001
Intercept	3	1	-0.8791	0.5845	2.2621	0.1326
Intercept	4	1	1.0690	0.5854	3.3350	0.0678
numViol	0	1	-0.9412	0.4561	4.2593	0.0390
numViol	1	1	0.0429	0.1286	0.1112	0.7388
numViol	2 or More	1	0.1834	0.1784	1.0579	0.3037
preIncidentManeuver	Changing lanes	1	-0.4470	0.2960	2.2802	0.1310
preIncidentManeuver	Going straight, but with unintentional "drifting" within lane or across lanes	1	0.4981	0.2766	3.2443	0.0717
preIncidentManeuver	Making a turn	1	0.3062	0.1892	2.6188	0.1056
preIncidentManeuver	Negotiating a curve	1	-0.1813	0.1856	0.9542	0.3287
preIncidentManeuver	Other	1	-0.1434	0.3579	0.1606	0.6886
driverBehavior	Apparent unfamiliarity with roadway	1	1.3754	0.3595	14.6353	0.0001
driverBehavior	Avoiding animal, or other vehicle	1	-1.1367	0.6053	3.5268	0.0604
driverBehavior	Distracted	1	1.0890	0.1739	39.1993	<.0001
driverBehavior	Drowsy, sleepy, asleep, fatigued	1	1.2856	0.3930	10.7030	0.0011
driverBehavior	Exceeded safe speed, or speed limit	1	1.3257	0.2085	40.4204	<.0001
driverBehavior	Failed to signal, improper signal	1	-0.1678	0.3694	0.2064	0.6496
driverBehavior	Improper turn	1	-0.0592	0.1984	0.0892	0.7652
driverBehavior	Other	1	0.0198	0.3199	0.0038	0.9507
driverBehavior	Sign, signal violation	1	0.5230	0.3553	2.1673	0.1410
weather	Adverse weather	1	0.2642	0.1529	2.9861	0.0840
trafficDensity	Level-of-service A2: Free flow, leading traffic present	1	-0.3368	0.1356	6.1728	0.0130
trafficDensity	Level-of-service B: Flow with some restrictions	1	-0.3219	0.1504	4.5792	0.0324

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
trafficDensity	**Level-of-service C: Stable flow, maneuverability and speed are more restricted**	1	-0.8883	0.3428	6.7163	0.0096
trafficDensity	**Level-of-service D/E/F**	1	0.2565	0.7188	0.1274	0.7212
locality	**Business/Industrial**	1	0.8038	0.2700	8.8657	0.0029
locality	**Bypass/Divided Highway with traffic signals**	1	0.8137	0.4518	3.2428	0.0717
locality	**Church/school/playground**	1	1.0214	0.3007	11.5393	0.0007
locality	**Open Country**	1	2.8109	0.4835	33.7909	<.0001
locality	**Residential area**	1	0.8056	0.2666	9.1334	0.0025
locality	**Urban**	1	0.6420	0.3873	2.7473	0.0974
criticalSpeed		1	0.00622	0.00319	3.7883	0.0516
Radius	**1**	1	0.3594	0.4284	0.7040	0.4014
Radius	**2**	1	-0.1376	0.5309	0.0672	0.7955
Radius	**3**	1	0.0884	0.5146	0.0295	0.8636
Radius	**4**	1	-1.5421	0.5810	7.0451	0.0079
Radius	**5**	1	2.1959	0.5620	15.2680	<.0001
Radius	**6**	1	-0.5585	0.3852	2.1021	0.1471
struckObjectType	**Others**	1	-0.2770	0.2358	1.3802	0.2401
struckObjectType	**Raised median**	1	0.0850	0.1959	0.1884	0.6642
struckObjectType	**animal**	1	-2.2632	0.2236	102.4069	<.0001
struckObjectType	**concrete barrier**	1	-0.7520	0.8428	0.7961	0.3723
struckObjectType	**ditch**	1	3.0768	0.5387	32.6201	<.0001
struckObjectType	**pavement edge/ edge line**	1	-0.0992	0.1717	0.3338	0.5634
struckObjectType	**roadway debris**	1	0.9190	0.2888	10.1226	0.0015
struckObjectType	**sign post**	1	0.9879	0.6423	2.3660	0.1240
struckObjectType	**stopped, backing, pulling out car**	1	-3.2329	0.4771	45.9126	<.0001
struckObjectType	**tree/shrub**	1	1.4932	0.7851	3.6178	0.0572
struckObjectType	**utility/light pole**	1	2.1344	0.8247	6.6982	0.0097
struckObjectType	**w-beam barrier**	1	3.1188	0.5156	36.5858	<.0001
reactionTime		1	-0.5962	0.0648	84.5230	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
preIncidentManeuver Changing lanes vs Going straight (accelerate, decelerate, constant speed)	0.640	0.358	1.142
preIncidentManeuver Going straight, but with unintentional "drifting" within lane or across lanes vs Going straight (accelerate, decelerate, constant speed)	1.646	0.957	2.830
preIncidentManeuver Making a turn vs Going straight (accelerate, decelerate, constant speed)	1.358	0.937	1.968
preIncidentManeuver Negotiating a curve vs Going straight (accelerate, decelerate, constant speed)	0.834	0.580	1.200
preIncidentManeuver Other vs Going straight (accelerate, decelerate, constant speed)	0.866	0.430	1.747
driverBehavior Apparent unfamiliarity with roadway vs None	3.957	1.956	8.005
driverBehavior Avoiding animal, or other vehicle vs None	0.321	0.098	1.051
driverBehavior Distracted vs None	2.971	2.113	4.178
driverBehavior Drowsy, sleepy, asleep, fatigued vs None	3.617	1.674	7.813
driverBehavior Exceeded safe speed, or speed limit vs None	3.765	2.502	5.665
driverBehavior Failed to signal, improper signal vs None	0.846	0.410	1.744
driverBehavior Improper turn vs None	0.942	0.639	1.390
driverBehavior Other vs None	1.020	0.545	1.909
driverBehavior Sign, signal violation vs None	1.687	0.841	3.385
weather Adverse weather vs No adverse weather	1.302	0.965	1.757

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
trafficDensity Level-of-service A2: Free flow, leading traffic present vs Level-of-service A1: Free flow, no lead traffic	0.714	0.547	0.931
trafficDensity Level-of-service B: Flow with some restrictions vs Level-of-service A1: Free flow, no lead traffic	0.725	0.540	0.973
trafficDensity Level-of-service C: Stable flow, maneuverability and speed are more restricted vs Level-of-service A1: Free flow, no lead traffic	0.411	0.210	0.805
trafficDensity Level-of-service D/E/F vs Level-of-service A1: Free flow, no lead traffic	1.292	0.316	5.288
locality Business/Industrial vs Interstate/Bypass/Divided Highway with no traffic signals	2.234	1.316	3.792
locality Bypass/Divided Highway with traffic signals vs Interstate/Bypass/Divided Highway with no traffic signals	2.256	0.931	5.470
locality Church/school/playground vs Interstate/Bypass/Divided Highway with no traffic signals	2.777	1.540	5.007
locality Open Country vs Interstate/Bypass/Divided Highway with no traffic signals	16.624	6.444	42.888
locality Residential area vs Interstate/Bypass/Divided Highway with no traffic signals	2.238	1.327	3.774
locality Urban vs Interstate/Bypass/Divided Highway with no traffic signals	1.900	0.889	4.060
criticalSpeed	1.006	1.000	1.013
struckObjectType Others vs curb	0.758	0.478	1.203
struckObjectType Raised median vs curb	1.089	0.742	1.598
struckObjectType animal vs curb	0.104	0.067	0.161
struckObjectType concrete barrier vs curb	0.471	0.090	2.459
struckObjectType ditch vs curb	21.688	7.545	62.341
struckObjectType pavement edge/ edge line vs curb	0.906	0.647	1.268
struckObjectType roadway debris vs curb	2.507	1.423	4.415
struckObjectType sign post vs curb	2.686	0.763	9.457
struckObjectType stopped, backing, pulling out car vs curb	0.039	0.015	0.100
struckObjectType tree/shrub vs curb	4.451	0.956	20.737
struckObjectType utility/light pole vs curb	8.452	1.679	42.551
struckObjectType w-beam barrier vs curb	22.620	8.234	62.143
reactionTime	0.551	0.485	0.626