Final Report

FDOT Contract No: BDV29-977-48

Strategies to Identify and Mitigate Secondary Crashes Using Real-time Traffic Data on Florida's Turnpike System

FDOT Contract No: BDV29-977-48

Prepared for:

Research Center
Florida Department of Transportation
605 Suwannee Street, MS 30
Tallahassee, FL 32399

Prepared by:

Florida International University
Dept. of Civil & Environmental Engineering
10555 West Flagler Street, EC 3628
Miami, FL 33174

University of North Florida
School of Engineering
1 UNF Drive
Jacksonville, FL 32224

Project Manager: Eric Gordin, P.E.
Co-Project Manager: Raj Ponnaluri, Ph.D., P.E., PTOE, PMP

May 2021

DISCLAIMER

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

METRIC CONVERSION TABLE

U.S. UNITS TO SI* (MODERN METRIC) UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
LENGTH				
in	inches	25.400	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.610	kilometers	km
mm	millimeters	0.039	inches	in
m	meters	3.280	feet	ft
m	meters	1.090	yards	yd
km	kilometers	0.621	miles	mi

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
AREA				
in^2	square inches	645.200	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.590	square kilometers	km^2
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.470	acres	ac
km^2	square kilometers	0.386	square miles	mi^2

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
VOLUME				
fl oz	fluid ounces	29.570	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
NOTE: volumes greater than 1,000 L shall be shown in m^3.				

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Strategies to Identify and Mitigate Secondary Crashes Using Real-time Traffic Data on Florida's Turnpike System		5. Report Date May 2021 6. Performing Organization Code
7. Author(s) Priyanka Alluri, Thobias Sando, Angela Kitali, Denis Monyo, and Haifeng Wang		8. Performing Organization Report No.
9. Performing Organization Name and Address Lehman Center for Transportation Research Florida International University 10555 West Flagler Street, EC 3628 Miami, FL 33174; School of Engineering, University of North Florida 1 UNF Drive, Jacksonville, FL 32224		10. Work Unit No. (TRAIS) 11. Contract or Grant No. BDV29-977-48
12. Sponsoring Agency Name and Address Florida Department of Transportation 605 Suwannee Street, MS 30 Tallahassee, FL 32399-0450		13. Type of Report and Period Covered Final Report November 2019 – May 2021 14. Sponsoring Agency Code

15. Supplementary Notes

Mr. Eric Gordin, P.E., of the Florida's Turnpike Enterprise, served as the Project Manager for this project; and Dr. Raj Ponnaluri, P.E., PTOE, PMP, of the FDOT Traffic Engineering and Operations Office, served as the Co-Project Manager.

16. Abstract

The primary goal of this research was to develop a comprehensive approach to identify and mitigate secondary crashes on the Florida Turnpike System in real time. The specific research objectives were to: (1) identify secondary crashes; (2) identify significant factors contributing to the occurrence of secondary crashes; (3) develop an algorithm that predicts the likelihood of secondary crashes in real time; and (4) explore the potential of connected vehicle (CV) applications in mitigating secondary crashes.

This research used four types of data: (1) incident data; (2) high-resolution traffic data; (3) roadway geometric data; and (4) high-resolution rainfall data. These data were collected from January 2014 to June 2019. The study corridors were primarily selected from the Florida's Turnpike Mainline and Turnpike Extension.

A data-driven approach was developed to accurately estimate the incident impact area and identify secondary crashes within the impacted area. The Least Absolute Shrinkage and Selection Operator (LASSO) penalized logistic regression model, fitted using the bootstrap resampling approach, was used to identify factors influencing the risk of secondary crashes. Traffic flow, incident, temporal, weather, and roadway geometric attributes were considered to influence the likelihood of secondary crashes. An algorithm was developed as a *proof-of-concept* to predict the likelihood of secondary crashes in real time. This algorithm uses incident data from SunGuide®, traffic data from Regional Integrated Transportation Information System (RITIS), rainfall data from National Oceanic and Atmospheric Administration (NOAA), and the collected roadway geometric characteristics data. The algorithm estimates the impact area of an incident and predicts the likelihood of secondary crashes in real time.

Finally, the microsimulation approach was used to explore the potential of CV applications in mitigating secondary crashes. The following simulation scenarios were considered: blockage of the inner lane, one outer lane, and two outer lanes. A sensitivity analysis was performed by considering varying market penetration rates (MPRs) of CVs. The significance of CV applications in mitigating secondary crashes was assessed using traffic conflicts derived from microsimulation models.

17. Key Word	18. Distribution Statement		
Real-time Secondary Crash Prediction, Freeways, Incident Data, Connected Vehicle (CV) Technology, Traffic Incident Management.			
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 141	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

This research was funded by the Research Center of the Florida Department of Transportation (FDOT). The authors are grateful to the Project Managers, Eric Gordin, P.E., of the Florida's Turnpike Enterprise, and Raj Ponnaluri, Ph.D., P.E., PTOE, PMP, of the FDOT Traffic Engineering and Operations Office, for their guidance and support throughout the project. Special thanks to the following individuals for providing the support needed to complete this project:

- Mr. Tony Abid, ITS Server Operations Manager, Florida's Turnpike Enterprise
- Mr. Junias Aldajuste, E.I., Traffic Engineering Specialist, Florida's Turnpike Enterprise
- Dr. Enock Mtoi, P.E., Transportation Engineer, Florida's Turnpike Enterprise
- Mr. Dennis Ramos, ITS Network Infrastructure Manager, Florida's Turnpike Enterprise
- Ms. Karla Smith, TMC Operations Manager, Florida's Turnpike Enterprise
- Mr. Andrew Velasquez, P.E., PTOE, Program Manager, Florida's Turnpike Enterprise

EXECUTIVE SUMMARY

Traffic incidents are the primary source of non-recurring congestion. In addition to affecting roadway operations, traffic congestion resulting from an incident exposes other vehicles to the risk of being involved in additional incidents, typically referred to as secondary crashes. Secondary crashes adversely affect traffic operations and impose risks to the safety of both road users and incident responders. Accurate identification of secondary crashes is the first and the most crucial step in devising strategies to mitigate their occurrence. Nonetheless, the identification of secondary crashes is not a straightforward process. The definition itself is subjective, and identifying secondary crashes depends on the accuracy of the approach used to estimate the incident impact area. Since identifying potential secondary crashes is difficult, investigating factors that may influence these crashes becomes even more challenging.

The primary goal of this research was to develop a comprehensive approach to identify and mitigate secondary crashes on Florida's Turnpike System in real time. The specific research objectives were to:

- identify secondary crashes;
- identify significant factors contributing to the occurrence of secondary crashes;
- develop an algorithm that predicts the likelihood of secondary crashes in real time; and
- explore the potential of connected vehicle (CV) applications in mitigating secondary crashes.

Secondary Crash Identification

A data-driven approach was developed to better estimate the incident impact area and identify secondary crashes within the impacted area. Traffic incidents from the SunGuide® database and high-resolution speed data from HERE Technologies were used to estimate the incident impact area. These data were collected from January 2014 to June 2019. The study area, which is located in Florida, included a 97-mile section of the Florida's Turnpike Mainline and a 48-mile Turnpike Extension. Overall, 4,549 secondary crashes in the upstream direction of the primary incident were identified from 3,977 primary incidents. The identified secondary crashes on the upstream direction of the primary incident accounted for 1.4% of the 322,259 incidents. This is equivalent to 5.7 secondary crashes per mile per year.

Factors Influencing the Occurrence of Secondary Crashes

The Least Absolute Shrinkage and Selection Operator (LASSO) penalized logistic regression model, fitted using the bootstrap resampling approach, was used to identify potential factors influencing the risk of secondary crashes. The model is considered to improve the prediction accuracy because it accounts for the asymmetric nature of secondary crashes, performs variable selection, and removes correlated variables.

The following factors were associated with an increase in the likelihood of secondary crashes:

- Hazard-related incidents and crashes compared to the incidents related to vehicle problems.
- Incidents attended to by more than one responding agency.
- Moderate and severe incidents.
- Incidents that occurred on wet road surface conditions.

- Incidents that occurred during morning peak hours.
- The presence of diverge influence area, horizontal curve, or vertical curve within the incident impact area.

The following factors were associated with a decrease in the likelihood of secondary crashes:

- Higher average prevailing speed or speed variation before the occurrence of the incident.
- The presence of merge influence area within the incident impact area.
- Wider shoulders within the incident impact area.

Real-time Secondary Crash Likelihood Prediction: *Proof-of-concept*

An algorithm was developed as a *proof–of-concept* to predict the likelihood of secondary crashes in real time. The algorithm consists of three main parts, which includes:

- *Internal Storage Database*, which stores incident, speed, and rainfall data collected in real time. It also archives the real time secondary crash prediction results, historical databases, and the statistical model equation with the secondary crash likelihood parameters.

- *Backend Applications* for collecting, parsing, and saving incident, traffic, and rainfall data in real time. One of the applications continuously accesses Florida's Turnpike SunGuide® database every two minutes and ping new incidents. Speed data are retrieved from the HERE real-time flow Extensible Markup Language (XML) feed every minute. Rainfall data are obtained from the Next-Generation Radar (NEXRAD) Level-II network at an interval of 4-6 minutes.

- *Secondary Crash Prediction Application*: The information from the *Internal Storage* database and *Real-time Data Backend Programs* are then combined to predict the likelihood of secondary crashes every 15 minutes until the incident is cleared.

Potential of CV Applications in Mitigating Secondary Crashes

The microsimulation approach was used to explore the potential of CV applications in mitigating secondary crashes. The following simulation scenarios were considered: blockage of the inner lane, one outer lane, and two outer lanes. The scenarios also considered the effect of varying traffic volumes by analyzing traffic data during morning and evening peak periods. A sensitivity analysis was performed by considering varying market penetration rates (MPRs) of CVs. The significance of CV applications in mitigating secondary crashes was assessed using traffic conflicts derived from microsimulation models.

Deployment of CV applications was found to result in up to 98% reduction in traffic conflicts. A more significant reduction was observed in less congested traffic, even with low MPRs of CVs. A higher MPR was required to achieve significant conflict reduction during congested periods. Additionally, more conflicts were observed when two lanes were blocked, compared to single lane closures. The detour advisory was found to be significant for incidents that block multiple lanes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AADT	Annual Average Daily Traffic
API	Application Programming Interface
ATDM	Active Transportation and Demand Management
ATIS	Advanced Traveler Information System
ATMS	Advanced Traffic Management System
AWS	Amazon Web Service
BSM	Basic Safety Message
CCTV	Closed-circuit Television
CDBA	Continuous Driving Behavior Adjustment
CLI	Command Line Interface
Cloglog	Complementary Log-log
COM	Component Object Model
NTM	Northern Turnpike Mainline
CV	Connected Vehicle
dBZ	decibel relative to reflectivity
DMS	Dynamic Message Sign
DSRC	Dedicated Short Range Communication
EMS	Emergency Medical Service
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
FNR	False Negative Rate
FPR	False Positive Rate
HAR	Highway Advisory Radio
HCM	Highway Capacity Manual
TE	Turnpike Extension
HTTP	Hypertext Transfer Protocol
IIA	Incident Impact Area
ILB	Inner Lane Blocked
LASSO	Least Absolute Shrinkage and Selection Operator
MM	Mile Marker
MPR	Market Penetration Rate
STM	Southern Turnpike Mainline
NCHRP	National Cooperative Highway Research Program
NEXRAD	Next Generation Weather Radar
NOAA	National Oceanic and Atmospheric Administration
OBU	On-board Unit
OLB	Outer Lane Blocked
PET	Post-encroachment time
RBC	Ring Barrier Controller
RCI	Roadway Characteristic Inventory
RITIS	Regional Integrated Transportation Information System
ROC	Receiver Operating Curve
RSU	Roadside Unit
SCDOT	South Carolina Department of Transportation
SD	Standard Deviation
SMOTE-NC	Synthetic Minority Oversampling TEchnique –Nominal Continuous
SR	State Road
SSAM	Surrogate Safety Assessment Model

STM	Southern Turnpike Mainline
TIM	Traffic Incident Management
TMC	Transportation Management Center
TPR	True Positive Rate
TRB	Transportation Research Board
TSM&O	Transportation Systems Management and Operations
TTI	Time to Collision
USDOT	United States Department of Transportation
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
VANET	Vehicle Ad-hoc Network
VBS	Visual Basic Scripting
VPN	Virtual Private Network
WCT	Weather Climatic Toolkit
WSDOT	Washington State Department of Transportation
XML	Extensible Markup Language

CHAPTER 1
INTRODUCTION

1.1 Background

Transportation agencies strive for an efficient transportation system that is safe and has minimal delays. Nevertheless, congestion and traffic incidents have continuously been deterring the performance of the transportation network. The cost of traffic congestion to Americans in 2019 was estimated to be approximately $88 billion, an average of $1,377 per driver (INRIX, 2019). This congestion is partly caused by an increased traffic volume, particularly during peak hours, and is commonly termed as recurrent congestion. Traffic incidents, including traffic crashes, disabled vehicles, and debris on roadways, are also a significant cause of congestion, generally referred to as non-recurrent congestion. Traffic incidents often lead to capacity reduction and deterioration of the level of service. They account for more than half of all urban traffic delays and almost all rural traffic delays (Baykal-Gürsoy et al., 2009).

In addition to affecting the operational quality of roadways, these events affect the safety of road users and incident responders. In the United States, traffic-related incidents are a leading cause of death for Emergency Medical Service (EMS) providers, law enforcement, and towing services. On average, one law enforcement officer is killed every month in the United States, and one towing professional is killed every six days (Hagen, 2017). Traffic incidents also expose other vehicles to the risk of being involved in additional crashes called secondary crashes (Owens et al., 2010). These incidents tend to occur within the prior incident queue as vehicles encounter unexpected congestion and are unable to brake in time. They can also occur near a traffic incident as drivers become distracted by the incident scene (Goodall, 2017).

Secondary crashes have progressively been perceived as a significant issue, particularly on freeways (Hirunyanitiwattana and Mattingly, 2006). As such, there has been a growing interest in mitigating secondary crashes. Secondary crashes are non-recurring, leading to reduced capacity, additional traffic delays, and increased fuel consumption and emissions. These crashes also affect the safety of both road users and incident responders. The United States Department of Transportation (USDOT) estimated that secondary crashes alone are responsible for approximately 18 percent of all freeway traffic fatalities and 20 percent of all freeway crashes (Owens et al., 2010). Further, compared to primary incidents, secondary crashes have a significant impact on traffic management resource allocation (Vlahogianni et al., 2012; Karlaftis et al., 1999).

Prevention of secondary crashes has, therefore, been highlighted as a high-priority for traffic incident managers (O'Laughlin and Smith, 2002) and Transportation Management Centers (TMCs) (Owens et al., 2010). The Federal Highway Administration (FHWA) uses the reduction of secondary crashes as one of the performance measures for state incident management systems (National Cooperative Highway Research Program [NCHRP], 2014). The Florida Department of Transportation (FDOT) included secondary crashes as a *Safety* performance measure in its Transportation Systems Management and Operations (TSM&O) Strategic Plan (FDOT, 2017). Specifically, to reduce the risk to responders, secondary crashes, and delays associated with incidents, FDOT has an Open Roads Policy of clearing all travel lanes within 90 minutes (FDOT, 2018). Several states also consider secondary crash mitigation strategies in allocating funding for

the development of Traffic Incident Management (TIM) programs and on-road help services, such as FDOT's Road Ranger freeway service patrol (Lou et al., 2011).

1.2 Secondary Crash Mitigation Challenges

Agencies have been looking for ways to mitigate secondary crashes to reduce non-recurrent delays and the adverse safety impacts associated with these crashes. However, some hurdles limit the implementation of approaches to reduce secondary crashes. First and foremost, the process of identifying secondary crashes is itself a challenge since there is no universal definition of a secondary crash. The inconsistency in defining secondary crashes limits the possibility of exploring the underlying relationship between secondary crash occurrences and influential factors. This limitation, in turn, hinders the mitigation efforts.

It is difficult to identify factors associated with the occurrence of secondary crashes. Previous studies have considered several incident-related, traffic-related, geometric-related, and weather-related factors when developing secondary crash risk models. However, simply incorporating all variables in the model may lead to biased results, considering the possible significant correlation among the variables.

Not all incidents lead to secondary crashes. The proportion of incidents that results in secondary crashes (primary incidents) is much less than the proportion of incidents that do not cause secondary crashes (normal incidents). Secondary crash risk models use the two incident categories, i.e., normal and primary incidents, as the response variable, and thus making the modeling of the secondary crash likelihood an imbalanced classification problem. Neglecting this imbalance characteristic can lead to serious consequences, both in the model's coefficient estimates and prediction accuracy (Kitali et al., 2019b).

The likelihood of secondary crashes depends on several factors, including traffic flow characteristics, incident characteristics, weather conditions, roadway geometric conditions, etc. An in-depth understanding of these factors will help agencies on several fronts. In general, three major challenges are encountered when modeling the risk of secondary crashes: (1) infrequent nature of secondary crashes, (2) selection of the most important variables, and (3) identification of variable correlation. Therefore, any candidate model needs to account for these issues.

Research on approaches to mitigate secondary crashes is limited (Park et al., 2018; Park and Haghani, 2016b; Yang et al., 2017; Karlaftis et al., 1999; Kopitch and Saphores 2011). Numerous studies have indicated incident duration as one of the most important factors influencing the occurrence of secondary crashes (Kitali et al., 2018; Goodall, 2017; Wang et al., 2016; Zhan et al., 2009). That is, the longer it takes to clear an incident, the higher the likelihood of a secondary crash. The impact of incident duration on the risk of secondary crashes was found to increase even further when traffic transitioned from a free-flow state to a congested state (Park and Haghani, 2016b). Considering the interdependency between incident duration, prevailing traffic conditions, and the probability of secondary crashes, it is crucial to devise a proactive approach to mitigate the risk of secondary crashes (Park et al., 2018).

1.3 Research Goal and Objectives

The main goal of this research project was to develop a comprehensive approach to identify and mitigate secondary crashes on the Florida's Turnpike System in real time. The specific objectives were to:

- Investigate ways to accurately identify secondary crashes on the Florida's Turnpike System using real-time traffic and incident data.
- Identify factors that influence the occurrence of secondary crashes on the Florida's Turnpike System.
- Develop an algorithm that predicts the likelihood of secondary crashes in real time.
- Explore the potential of connected vehicle (CV) applications in mitigating secondary crashes.

1.4 Report Organization

The rest of this report is organized as follows:

- Chapter 2 entails a comprehensive synthesis of the literature on the main approaches used to identify and mitigate secondary crashes. The chapter also focuses on exploring factors associated with the occurrence of secondary crashes.

- Chapter 3 discusses the data used to achieve the research goal and objectives. Specifically, the chapter describes, in detail, the types of data used, data sources, data collection strategy, and data processing steps.

- Chapter 4 focuses on identifying secondary crashes. It first presents the methodology used to identify secondary crashes. It further discusses the characteristics of the identified secondary crashes.

- Chapter 5 presents the methodology used to predict the likelihood of secondary crashes based on the characteristics of primary incidents, traffic flow parameters, weather conditions, and roadway geometric characteristics.

- Chapter 6 discusses the algorithm developed to predict the likelihood of secondary crashes in real time.

- Chapter 7 quantifies the potential benefits of CV applications in mitigating secondary crashes. Different microsimulation scenarios based on lane blockage, times of day, and CV market penetration rates (MPRs) were explored.

- Chapter 8 summarizes this research effort.

CHAPTER 2
LITERATURE REVIEW

This chapter presents a synthesis of previous studies that focused on identifying secondary crashes, analyzing the risk factors influencing the occurrence of these crashes, and the potential approaches to mitigate their occurrence. The first section of this chapter provides a summary of existing methods used to identify secondary crashes. A summary of studies that explored factors associated with the occurrence of secondary crashes is presented next. This section also discusses major issues accompanying the modeling of secondary crash likelihood. The third section presents previous literature that explored strategies to mitigate secondary crashes. The research areas that need further investigation associated with the identification, prediction, and prevention of secondary crashes are discussed in the last section.

2.1 Existing Methods to Identify Secondary Crashes

Secondary crashes are traffic incidents that occur within the spatial and temporal impact area of the primary incidents (Zhang and Khattak, 2010a; Moore et al., 2004; Yang et al., 2014a; Karlaftis et al., 1999). Unlike other traffic incidents that incident responders easily identify, detecting secondary crashes is not straightforward since the definition itself is subjective. It is difficult to determine visually, either directly at the crash site or through closed-circuit television (CCTV) cameras, if the crash is a result of the backup caused by another incident, especially since the backup may also be due to recurrent congestion. As summarized in Table 2-1, three major approaches are often used to identify secondary crashes: (1) manual method, where personnel visually identify secondary crashes within the queue of the prior incident; (2) static method that uses predefined spatiotemporal thresholds; and (3) dynamic approach that estimates the primary incident influence area as a function of its impact on traffic flow characteristics, e.g., speed, volume, and/or density, and identify crashes that occurred within the impact area as secondary crashes.

Table 2-1: Methods Used to Identify Secondary Crashes

Method	Approach	Advantages	Challenges
Manual	Personnel visually identify secondary crashes: • On-site approach using incident responders, e.g., Highway Patrol, etc. • Off-site approach using CCTV, etc.	• Simple • Does not require any data processing	• Subjective • Unreliable • Inconsistent • Random
Static	Identify secondary crashes based on predefined distance and time thresholds for each primary incident (e.g., 2 miles upstream and 2 hours after the primary incident)	• More reliable than the manual method • Relatively easy to implement	• Less reliable compared to the dynamic method
Dynamic*	Identify secondary crashes based on the actual queue length of the primary incident	• Most reliable • Accurate	• Resource intensive • Limited by data availability

Note: *Can be reliably implemented in real time; CCTV = Closed Circuit Television.

An extensive literature review revealed that several studies have focused on identifying secondary crashes. For brevity, only a summary of these studies is presented in this report. Readers are referred to Sando et al. (2019) for a detailed discussion on this topic.

Figure 2-1 graphically summarizes previous studies that identified secondary crashes based on fixed spatial and temporal thresholds (Chang and Rochon, 2011; Green et al., 2012; Hirunyanitiwattana and Mattingly, 2006; Jalayer et al., 2015; Karlaftis et al., 1999; Kopitch and Saphores, 2011; Latoski et al., 1999; Moore et al., 2004; Raub, 1997; Tian et al., 2016; Zhan et al., 2008).

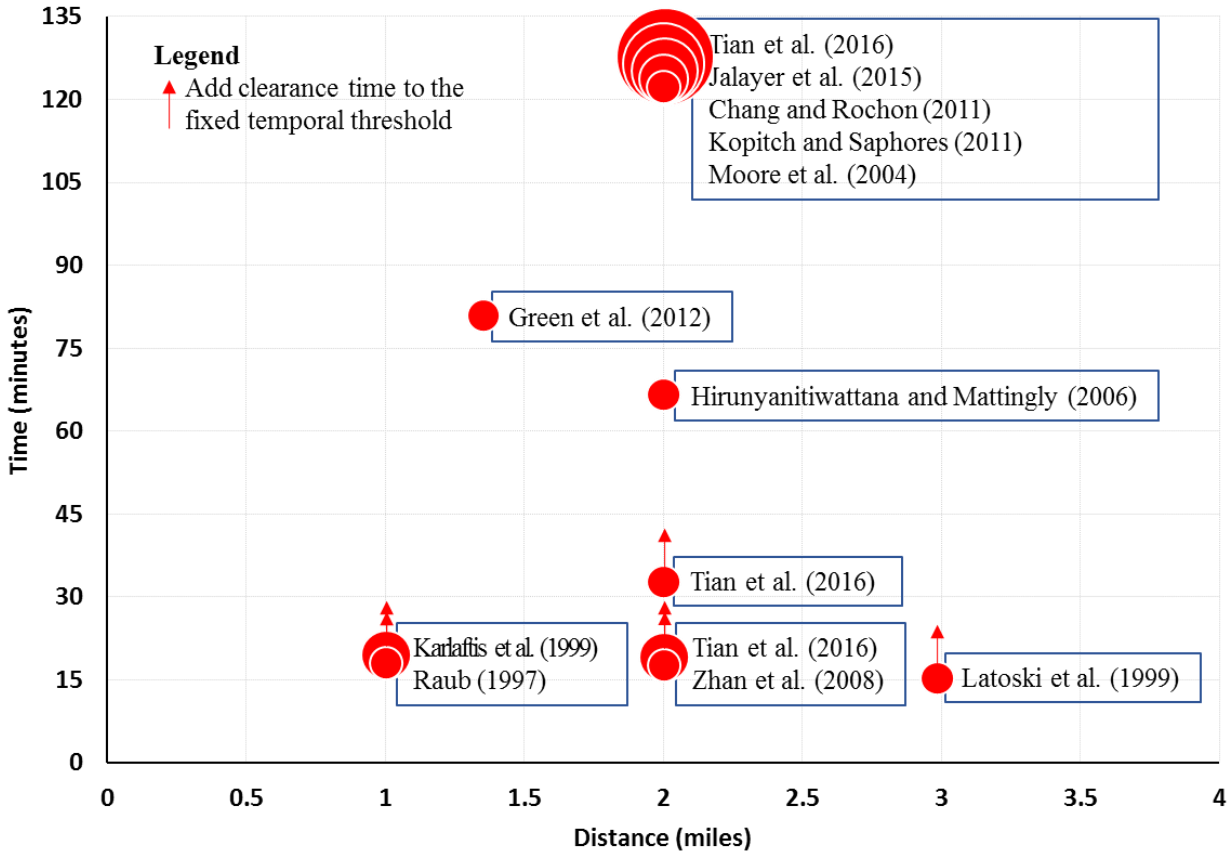


Figure 2-1: Studies That Used Static Method to Identify Secondary Crashes in the Upstream Direction

Similarly, secondary crashes that occurred in the opposite direction of the primary incident – because of the onlooker effect – were also commonly identified using different predefined thresholds. Figure 2-2 summarizes the studies that used the static method to identify secondary crashes in the opposite direction of the primary incident (Chang and Rochon, 2011; Green et al., 2012; Kopitch and Saphores, 2011; Moore et al., 2004).

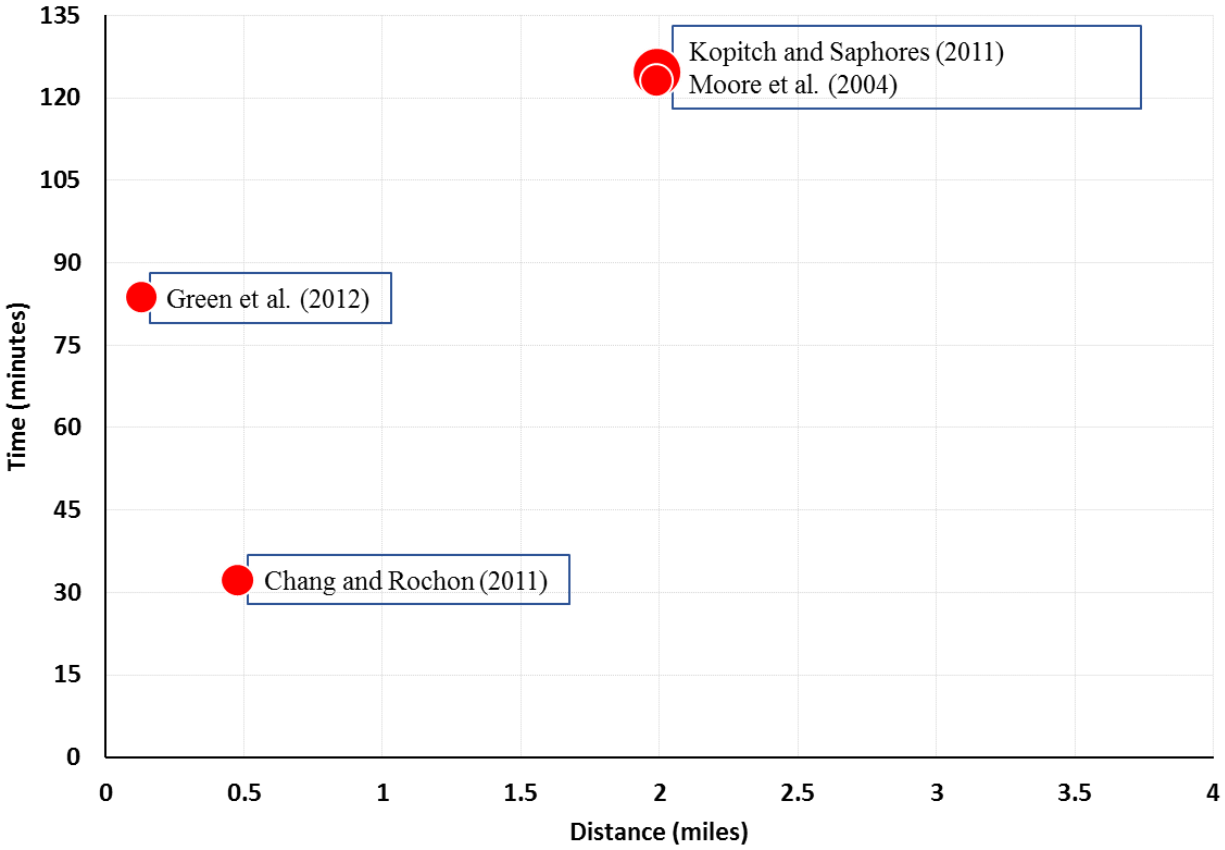


Figure 2-2: Studies That Used Static Method to Identify Secondary Crashes in the Opposite Direction

Unlike the manual method, the static method is more reliable because it is a function of predefined spatiotemporal parameters and not solely based on human judgment. However, the static method's one-size-fits-all approach of using fixed spatiotemporal thresholds does not yield reliable results (Kitali et al., 2019a). In other words, the fixed spatiotemporal thresholds do not effectively reflect the dynamic impact of incidents with varying characteristics and, therefore, may under- or overestimate the impact area (Ou et al., 2020).

To overcome the limitations associated with the static approach, some studies have focused on identifying secondary crashes based on prevailing traffic flow conditions at the time of the primary incident. In this case, spatiotemporal thresholds vary depending on the impact of the primary incident on traffic flow parameters, hence the term *dynamic*. The dynamic methods used in previous studies to identify secondary crashes can generally be grouped into three categories, i.e., queuing model-based, shockwave-based, and traffic data- based. Figure 2-3 visually summarizes some of the previous studies that explored the use of these models to capture the impact area of the primary incident (Chung 2013; Dougald et al., 2016; Goodall 2017; Imprialou et al., 2014; Kitali et al., 2018, 2019a, 2019b, 2021; Li et al., 2020; Mishra et al., 2016; Park and Haghani, 2016a, 2016b; Sando et al., 2019; Sarker et al., 2017; Sun and Chilukuri, 2006, 2010; Vlahogianni et al., 2010; 2012; Wang et al., 2016; 2019; Xu et al., 2016; Yang et al., 2014a, 2014b, 2014c; Zhan et al., 2009; Zhang and Khattak, 2010a; Zheng et al., 2014).

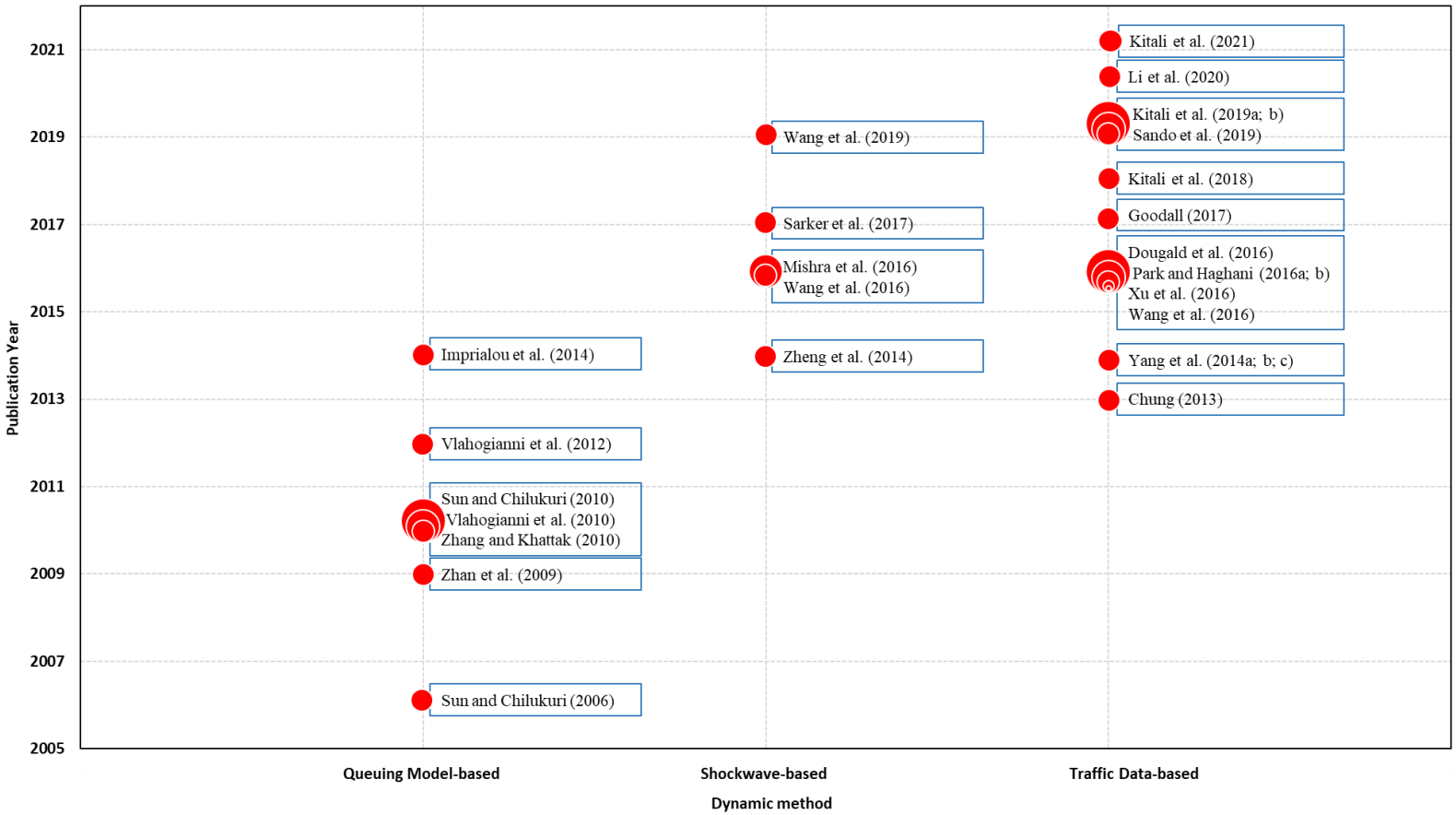


Figure 2-3: Studies That Used Dynamic Method to Identify Secondary Crashes

Although queuing methods provide a more realistic representation of incident impact areas – compared to the static approach – they generally rely upon the number and nature of the accessible variables, such as the assumed roadway capacities, arrival rate, etc. Different roadway segments are subject to different queuing formation processes because of their unique traffic, geometry, incident characteristics, and prevailing weather conditions.

Besides the queuing model-based approaches, other studies have used shockwave principles to dynamically identify secondary crashes, as shown in Figure 2-3 (Mishra et al., 2016; Sarker et al., 2017; Wang et al., 2019). In this case, the incident impact area is triangular. The spatiotemporal thresholds comprise the backward forming and discharging shockwaves linked with the occurrence and clearance of the incident (Yang et al., 2018). The backward-forming shockwave impacts the growth rate of the queue generated by the incident. When the incident is cleared, a forward-recovery shockwave initiates and ultimately reaches the backward-forming shockwave resulting in queue dissipation.

Several issues limit the application of the shockwave-based approach for identifying secondary crashes. The simplified assumption on the prevailing traffic conditions and modeling of the shockwave propagation remains to be an issue since they cannot accurately depict the dynamic progression of traffic states (Yang et al., 2018). Non-constant discharge and arrival rates make it challenging to model the complicated shockwaves with the assumption of a constant speed. Overall, both the queuing and shockwave dynamic methods use some prior assumptions to simplify the traffic conditions' complex characteristics, resulting in an incorrect estimation of the incident impact areas. Further, both methods cannot accurately distinguish the recurrent congestion from the non-recurrent congestion caused by the incident (Ou et al., 2020).

To overcome the limitations of the queuing model-based and shockwave-based methods, recent efforts in estimating the impact area of a primary incident have been shifted towards the use of data-driven approaches. Empowered by the advancements in traffic data collection technologies, several studies have explored data-driven approaches to identify secondary crashes (Dougald et al., 2016; Goodall, 2017; Kitali et al., 2018; Sando et al., 2019; Xu et al., 2016). Some traffic data sources include infrastructure-based traffic sensors, probe-vehicles, crowdsourced traffic data from third-party vendors, and CV technologies. The data-driven approaches use readily available traffic data to estimate incident impact areas while accounting for recurrent congestion (Ou et al., 2020). Further consideration must be made to mimic how congestion builds up and dissipates along the segments impacted by the primary incident. This is an essential step towards the accurate identification of secondary crashes. Failure to correctly estimate the incident impact area may lead to over- or under-estimation of the impact area and hence the number of secondary crashes.

2.2 Factors Influencing the Likelihood of Secondary Crashes

Following the identification of secondary crashes, the next step towards developing strategies to mitigate secondary crashes is to identify potential factors associated with the occurrence of secondary crashes. Identifying risk factors that influence the likelihood of secondary crashes is critical to the development and implementation of efficient and resilient traffic management strategies.

2.2.1 Secondary Crash Risk Prediction Models

The literature review indicated that only a few studies have explored factors associated with the occurrence of secondary crashes. Both parametric and non-parametric models have been used to analyze the likelihood of secondary crashes. Most of these studies adopted the respective models following the binary nature of secondary crash occurrence, i.e., given a primary incident or normal incident. Notably, primary incidents refer to incidents that resulted in a secondary crash(es), while normal incidents refer to incidents that did not result in secondary crashes. In these studies, geometric, weather, traffic conditions, and incident characteristics associated with primary incidents were compared with those of normal incidents.

Most of the studies that developed parametric models used binary regression models such as logit, probit, or complementary log-log (cloglog) models to analyze the likelihood of secondary crashes (Goodall, 2017; Wang et al., 2016; Karlaftis et al., 1999; Kopitch and Saphores, 2011; Zhan et al., 2009, 2008). In this case, the response variable is dichotomous with a "*yes*" category representing incidents that resulted in a secondary crash and a "*no*" category being an incident not resulting in a secondary crash. As mentioned earlier, these two categories of incidents are generally referred to as primary incidents and normal incidents, respectively.

In secondary crash risk models, the independent variables include a list of potential factors that may contribute to the occurrence of secondary crashes. The coefficients obtained by estimating the relationship between the probability of a secondary crash following a primary incident – based on a set of explanatory variables – can hence be used to quantify the impact of each contributing factor on the secondary crash risk. Table 2-2 presents a summary of studies that used parametric modeling approaches to explore risk factors that influence the likelihood of secondary crashes. The table includes secondary crash identification methodologies, secondary crash risk prediction models, and significant variables in each study. In addition to logistic regression models, other studies have examined the likelihood of secondary crashes using other approaches, including the proportional test, probit models, and cloglog models (Hirunyanitiwattana and Mattingly, 2006; Khattak et al., 2009; Kitali et al., 2018).

As indicated in Table 2-3, non-parametric models such as Bayesian neural networks and decision trees have also been used to model secondary crash risk (Vlahogianni et al., 2010, 2012). A fundamental difference between non-parametric models and parametric models is that the non-parametric models lack an inherent mechanism for explicitly describing the significance of input variables and hence considered black-box (Yang et al., 2018). The need for developing non-parametric models with explanatory power is related to the decision-making process in transportation. Instinctively, any decision in transportation and traffic operations ought to be founded on a strong comprehension of the mechanism by which various variables interface with and impact transportation phenomena (Vlahogianni et al., 2012).

Table 2-2: Summary of Literature on Parametric Secondary Crash Risk Models

Reference	Secondary Crash Identification Method	Method	No. of Var.	Variable Selection Method	Significant Variables
Karlaftis et al. (1999)	Static (1 mile and 15 min)	LR	18	Not Applicable	Season, clearance time, type of vehicle involved, and lateral location
Zhan et al. (2008)	Static (2 miles and 15 min + clearance)	LR	18	Not Applicable	Number of lanes, primary incident duration, time-of-day, number of vehicles, and vehicle rollover.
Zhan et al. (2009)	Cumulative arrival and departure	LR	19	Forward conditional criterion	Primary incidents type and lane-blockage duration, time of day, and direction where the incident occurred
Kopitch and Saphores (2011)	Static (2 miles and 2 h)	LR	9	Not Applicable	Number of vehicles, number of trucks, changeable message sign, and road work project
Khattak et al. (2012)	Static (1 mile and duration of primary incident (+15 min if lane blocked))	LR	13	Not Applicable	Incident duration, crashes, peak hours, number of vehicles, and AADT
Yang et al. (2014b)	Data-driven approach	LR (rare events)	10	Statistical significance level (0.1)	Daytime off-peak hours, daytime peak hours, duration, rear-end crashes, lane closure, and winter season
Wang et al. (2016)	Shockwave principle	LR	12	Not Applicable	Shockwave originating in the wake of a primary incident, duration, unsafe speed, and weather
Mishra et al. (2016)	Shockwave principle	Linear probability model, LR, and MNL	16	VIF correlation factor and significance level	Average speed of upstream traffic, upstream flow, AADT, incident type, number of vehicles, weather condition, and functional class
Wang et al. (2019)	Shockwave principle	LR	13	Not Applicable	Shockwave speed that occurred at the time of the primary incident, shockwave speed generated when incident responders arrive at the scene to control traffic, shockwave speed during dissipation, incident processing duration, unsafe speed, and rain.

Note: AADT = Annual Average Daily Traffic; Cloglog = complementary log-log, LASSO = Least Absolute Shrinkage and Selection Operator, LR = Logistic Regression, MNL = Multinomial Logistic Regression, No. of var. = Number of variables, VIF = Variance Inflation Factor.

Table 2-2 (Cont'd): Summary of Literature on Parametric Secondary Crash Risk Models

Reference	Secondary Crash Identification Method	Method	No. of Var.	Variable Selection Method	Significant Variables
Xu et al. (2016)	Data-driven approach	Bayesian LR	24	Pearson correlation and stepwise logit	Average speed, traffic volume, standard deviation of detector occupancy, volume difference between adjacent lanes, crash severity, crash type, day of the week, road surface condition, and number of lanes
Goodall (2017)	Data-driven approach	LR	3	Not Applicable	Congestion and incident duration
Sarker et al. (2017)	Shockwave principle	Generalized ordered response probit	15	Not Applicable	AADT, traffic composition, land use, number of lanes, right side shoulder width, posted speed limits, and ramp indicator
Kitali et al. (2018)	Data-driven approach	Bayesian cloglog	21	Random Forest	Average occupancy, incident severity, percent of lanes closed, incident type, incident clearance duration, incident impact duration, and incident occurrence time.
Kitali et al. (2019b)	Data-driven approach	Penalized LR (with resampling)	23	LASSO	Mean of detector occupancy, coefficient of variation of equivalent hourly volume, mean of speed, incident type, percent lane closed, incident occurrence time, shoulder blocked, number of responding agencies, incident impact duration, incident clearance duration, and roadway alignment

Note: AADT = Annual Average Daily Traffic; Cloglog = complementary log-log, LASSO = Least Absolute Shrinkage and Selection Operator, LR = Logistic Regression, MNL = Multinomial Logistic Regression, No. of var. = Number of variables, VIF = Variance Inflation Factor.

Table 2-3: Summary of Literature on Non-parametric Secondary Crash Risk Models

Reference	Secondary Crash Identification Method	Method	Explanatory Function	Number of Variables	Significant Variables
Vlahogianni et al. (2010)	Method based on spatiotemporal impact area of primary crash	Bayesian Neural Network	Mutual information	8	Maximum queue length, queue duration, and primary crash duration
Vlahogianni et al. (2012)	Automatic tracking of moving traffic jams	Bayesian Neural Network	Mutual information and partial derivatives	11	Traffic speed, changes in traffic speed and volume, duration of the primary crash, hourly volume, rainfall intensity, number of vehicles involved, blocked lanes, percentage of trucks, and upstream geometry
Park and Haghani (2016b)	Data-driven approach based on Gaussian Mixture Model and Bayesian structure equation model	A principled Bayesian learning approach to Neural Network and Logit model	Multilayer perceptron	13	Location area, incident type, and time of day
Park et al. (2017); (2018)	Data-driven approach based on Gaussian Mixture Model and Bayesian structure equation model	A principled Bayesian learning approach to Neural Network and Stochastic Gradient Boosted Decision Trees	A pedagogical rule extraction	13	Unexpected traffic congestion caused by a primary incident and onlooker factors

While most primary incidents result in one secondary crash, some primary incidents result in multiple crashes, in this case, referred to as cascading crashes. Events consisting of multiple secondary crashes are expected to have a longer impact duration and severe impacts on traffic. This situation presents additional impedance and increases interference among vehicles, particularly in upstream traffic (Zhang and Khattak, 2010b). Although generally uncommon, such events present a significant challenge to transportation agencies. They are expected to be attended by multiple responding agencies at different time stamps and locations. Moreover, incidents attended to by numerous incident responders may require lane closures, which further reduces the roadway's capacity, resulting in more congestion.

Identifying factors associated with the likelihood of cascading crashes and their possible interdependency is the first step towards devising strategies to mitigate them. Only a few studies have explored factors associated with the occurrence of cascading crashes (Mishra et al., 2016; Xu et al., 2019; Zhang and Khattak, 2010b). In addition to understanding independent factors contributing to the risk of cascading crashes, it is also crucial to understand the possible interdependencies between the influential factors. For example, a cascading crash that occurred during peak hours is expected to have a different impact from a crash that occurred during peak hours while it is raining. Thus, understanding a combination of factors significantly associated

with the risk of cascading crashes is a crucial step towards devising strategies to mitigate cascading crashes.

As indicated in Table 2-4, few studies have modeled the risk of cascading crashes (Mishra et al., 2016; Xu et al., 2019; Zhang and Khattak, 2010b). Zhang and Khattak (2010b) used an ordered logit model to investigate factors contributing to cascading crashes. Multiple vehicles involved in the primary incident and lane blockage increased the likelihood of cascading crashes. In addition to multiple vehicles, Mishra et al. (2016) found the following variables to be associated with a higher risk of cascading crashes: high upstream traffic flow, rear-end crash, inclement weather, and crashes that occurred on arterial corridors rather than freeways.

Table 2-4: Summary of Studies that Explored Factors Associated with the Occurrence of Cascading Crashes

Study	Objective	Study Location	Real-time Data Use	Incident Types used in the Analysis	Methodology	Factors that Significantly Influenced the Likelihood of Cascading Crashes
Mishra et al. 2016)	Likelihood of cascading crashes	Shelby County, Tennessee	• Freeway (15-min speed, flow, and occupancy) • Arterials (MPO travel demand model)	• Crashes only	Multinomial logit (MNL) model	• AADT (+) • Freeway vs. arterials (-) • Upstream flow (+) • Rear-end or single-vehicle crashes vs. other crash types (+) • Good weather condition (-)
Zhang and Khattak (2010b)	Contributing factors to cascading crashes	Hampton Roads, Virginia	• Did not use real-time data	• Crashes • Disabled vehicle • Abandoned vehicle • Others	Partial proportional odds (PPO) model	• Crashes (+) • Longer incident durations (+) • Multiple vehicle involvement (+) • High percentage of lane blockage (+) • Shorter segment length (+) • High AADT (+)
Xu et al. (2019)	Effect of real-time traffic conditions on the occurrence of cascading crashes	I-5 freeway, California	• 30-sec raw loop detector data, i.e., count, speed, and occupancy	• Crashes only	Zero inflated ordered probit model	• Hit-and-run primary crashes (+) • Average detector occupancy (+) • Rainy weather (+) • Severe primary crashes (+)

Note: AADT = Annual Average Daily Traffic; MPO = Metropolitan Planning Organization, (+) = variable increases the cascading crash risk; (-) = variable reduces the cascading crash risk.

Xu et al. (2019) used a zero-inflated ordered probit regression model to study the effects of prevailing traffic characteristics on the likelihood of cascading crashes. Two states were considered in modeling the frequency of secondary crashes caused by one primary incident. The

first state, the *secondary-crash-free* state, predicted whether the initial incident would lead to secondary crashes. The second state, referred to as the *secondary-crash-prone* state, determined the frequency of secondary crashes caused by one primary incident. The following factors were influential in the *secondary-crash-free* state: average traffic volume, average speed, and the difference between the number of on-ramps and off-ramps in a segment. In the *secondary-crash-prone* state, the significant factors that were found to influence the likelihood of multiple secondary crashes included hit-and-run primary crashes, average detector occupancy, rainy weather, and primary crash severity. The three studies presented in Table 2-4 identified primary incident characteristics, traffic flow conditions, inclement weather conditions as attributes that significantly influence the likelihood of cascading crashes.

2.2.2 Issues Accompanying Modeling of Secondary Crash Risk

Modeling the risk of secondary crashes is accompanied by several challenges. The infrequent nature of secondary crashes is one of the significant issues that need to be addressed. The selection of the most important variables, the detection of variable correlation, the use of more representative traffic variables, and the issue with missing information are among the issues encountered with explanatory variables used in secondary crash risk models. The following subsections discuss these issues in detail.

Imbalanced Data

As indicated earlier, secondary crashes are infrequent in nature. A majority of secondary crash risk models developed using either logit or probit link functions are symmetrical, i.e., the likelihood of secondary crash occurrence is presumed to rise to a probability of 0.5, then decrease toward the asymptote at one (1) (Kitali et al., 2017). In other words, in secondary crash likelihood prediction, symmetric models, such as logit or probit models, yield more reliable results when the proportion of normal incidents (~50%) is equal to the proportion of primary incidents (~50%). However, secondary crashes account for less than 20% (Owens et al., 2010; Sando et al., 2019) of total incidents, meaning that the proportion of primary incidents is much less than the proportion of normal incidents (i.e., primary incidents and normal incidents are asymmetrically distributed).

To account for the imbalanced nature of the response variable in a secondary crash risk model, Yang et al. (2014b) introduced the rare-event logistic regression model, and Kitali et al. (2019b) used a Synthetic Minority Over-sampling TEchnique-Nominal Continuous (SMOTE-NC) technique. Kitali et al. (2018) used a cloglog as an alternative prediction model over the conventional logit and probit models. Unlike the logit and probit models, the cloglog model is asymmetrical with a fat tail as it departs from zero (0) and sharply approaches one (1) (Kitali et al., 2017; Martin and Wu, 2017). Accounting for the infrequent nature of the secondary crash events has proven to improve the prediction accuracy of the secondary crash risk models (Kitali et al., 2019b).

Variables Selection

As indicated in Figure 2-4, researchers have considered several incident-related, traffic-related, temporal-related, geometric-related, and weather-related factors when developing secondary crash

risk models. However, it may not be possible to include all variables in the model due to the possible significant correlation among the factors. Moreover, the use of less important variables will introduce noise in the model and hence, reduce its accuracy.

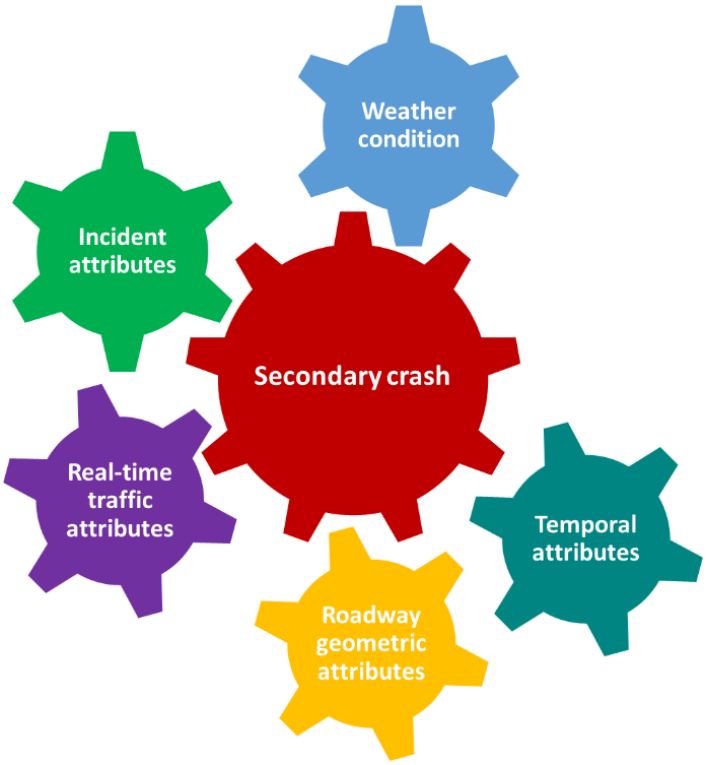


Figure 2-4: Factors Contributing to Secondary Crash Occurrence

One way to address this issue is to select and include only the most important variables. Variable subset selection methods, such as a stepwise technique, were used in several studies to add one best-fit variable at a time during model fitting (Mishra et al., 2016; Xu et al., 2016; Zhan et al., 2009). Nevertheless, this criterion has several drawbacks, including the result that each addition of a new variable may render one or more of the already included variables non-significant. Also, because the stepwise variable selection process is discrete, it often exhibits high variance and may not reduce the full model's prediction error. In other words, small changes in the data can result in different variables being selected, and this can potentially reduce the model's prediction accuracy (Menard and Torelli, 2014; Tibshirani, 1996).

As an alternative to stepwise variable selection, Kitali et al. (2018) used random forests, a non-parametric approach, to select the most important variables for inclusion in the secondary crash risk prediction model. In a later study, Kitali et al. (2019b) applied the Least Absolute Shrinkage and Selection Operator (LASSO) penalized likelihood, a regression analysis method that performs both variable selection and regularization. The LASSO method enhances the prediction accuracy and interpretability of the statistical model (Tibshirani, 1996). LASSO shrinks some coefficients of a regression model, in this case, logistic regression, and sets others to zero (0) to obtain variables with a substantial effect on the outcome (Tibshirani, 1996). LASSO also performs variable selection and variable correlation simultaneously. That is, between a pair of highly correlated

variables, LASSO tends to pick the most important variable and discard the other by shrinking them towards zero.

Because the LASSO method performs variable selection through a continuous process, it does not suffer as much from high variability, i.e., it simultaneously does both continuous shrinkage and automatic variable selection (Kitali et al., 2019b). The penalty term introduced by LASSO during the variable selection process ensures better estimation of the prediction error while avoiding overfitting. Selecting an optimal subset of explanatory variables is expected to improve the classification accuracy and make the model's interpretation easier. Since some of the variables will be minimized to zero, model thriftiness is achieved as well (Kitali et al., 2019b). Similar to the imbalanced data issue, fitting secondary crash risk models with important variables only improves the model's prediction accuracy (Kitali et al., 2019b).

Use of Aggregated Traffic Flow and Weather Characteristics

Traditional traffic data, such as annual average daily traffic (AADT) and speed limit, have often been included as explanatory variables in secondary crash risk models (Chimba and Kutela, 2014; Khattak et al., 2012; Mishra et al., 2016; Zhang and Khattak, 2010a). These data limit the reliability of results simply because they are aggregated values and do not reflect the prevailing traffic conditions at the time of an incident. Following the advancement in traffic data collection, high-resolution traffic data, instead of AADT and speed limit, have been increasingly used in developing secondary crash risk prediction models (Kitali et al., 2018, 2019b; Park and Haghani, 2016a, 2016b; Sando et al., 2019; Vlahogianni et al., 2012; Xu et al., 2016). The high-resolution traffic flow data provides a more accurate measurement of traffic flow conditions than traditional aggregated static traffic data.

Xu et al. (2016) used the random-effect logistic regression to develop a secondary crash risk prediction model using the high-resolution traffic flow data before the occurrence of primary incidents. The results suggested that the inclusion of high-resolution traffic variables significantly increases the model's predictive performance. Traffic volume, average speed, occupancy variation, and volume difference between adjacent lanes are the main traffic variables contributing to the increased risk of secondary crashes.

Inclement weather conditions, particularly rainfall, is one factor that could potentially exacerbate the occurrence of secondary crashes. Rainfall decreases the driver's sight distances and increases the vehicle's stopping distance (Haule et al., 2020; Kidando et al., 2019). During rainy conditions, approaching vehicles may not have an adequate opportunity to make emergency maneuvers, leading to an increased possibility of secondary crashes (Li et al., 2014). It is imperative to incorporate weather conditions as one of the potential variables in the secondary crash likelihood model.

Previous research that included rainfall as one of the secondary crash influential factors obtained the data either from an incident database (Wang et al., 2016; Khattak et al., 2012, 2009; Xu et al., 2016; Zhan et al., 2008) or rain gauges (Kopitch and Saphores, 2011; Vlahogianni et al., 2012). Incident report-based rainfall data is qualitatively recorded by incident responders only once and mostly at the incident notification time. As such, this value of rainfall information may not reflect

the prevailing rainfall intensity, especially in locations that experience short duration rainfall, when the incident impact duration is relatively long (Andrew, 2019). Gauge-based rainfall data are retrieved from weather stations that are usually sparsely distributed (Andrew, 2019). Similar to traffic flow characteristics, rainfall intensity varies both spatially and temporally. However, both incident-based and gauge-based rainfall data do not account for the spatiotemporal nature of rainfall.

Missing Potential Variables

While previous studies have considered numerous variables in secondary crash likelihood models, some variables have rarely been considered. Some of these variables include the presence of vertical curves, merging, and diverging ramps within the incident impact area. Unlike other roadway sections, merge and diverge influence areas are accompanied by more lane changes and high speed differentials by drivers attempting to enter or exit the freeway. This situation may increase the risk of secondary crashes. Thus, it is essential to incorporate merge and diverge influence areas in secondary crash risk models. Few studies have considered ramps as a potential variable that may influence the likelihood of secondary crashes (Karlaftis et al., 1999; Khattak et al., 2012, 2009; Park and Haghani, 2016b). Of those studies, the influence of ramps on secondary crash occurrence was not found significant.

In summary, researchers have used both parametric and non-parametric models to link secondary crash risks with geometric, incident, weather, and traffic characteristics. Understanding factors associated with secondary crash occurrence will help devise effective strategies to alleviate the effects of primary incidents, thus reducing the likelihood of secondary crash occurrence.

2.3 Strategies to Mitigate Secondary Crashes

Mitigating the risk of secondary crashes is critical for effective traffic incident management. Nonetheless, only a few studies have focused on drafting and deploying specific countermeasures to mitigate secondary crashes (Park et al., 2018; Park and Haghani, 2016b; Yang et al., 2017; Karlaftis et al., 1999; Kopitch and Saphores, 2011). Incident responding agencies could be better prepared to respond to potential secondary crashes when conditions associated with a high likelihood of occurrence exist. The success of this approach will depend on the availability of incident responders and the time they arrive at the incident scene. An incident responder may be hindered by a long queue, thus delaying the process of incident clearance (Yang et al., 2018).

Numerous studies have indicated incident duration as one of the most important factors influencing the occurrence of secondary crashes (Kitali et al., 2018; Goodall, 2017; Wang et al., 2016; Zhan et al., 2009). Khattak et al. (2012) observed a significant correlation between incident duration, the likelihood of a secondary crash, and the primary incident characteristics. A 10-minute increase in the primary incident duration was found to be associated with a 0.2 percent increase in the likelihood of secondary crashes (Khattak et al., 2009). Similarly, Goodall (2017) found the probability of a secondary crash occurrence to increase by approximately one percentage point for each additional two to three minutes spent on the scene under congested traffic. Compared with other traffic incidents whose occurrences are quite stochastic, the occurrence of secondary crashes is more deterministic as they are mostly caused by either turbulent traffic conditions initiated by

the primary incident or the onlooker effect (Xu et al., 2019). The impact of incident duration on the risk of secondary crashes was found to increase even further when traffic transitioned from a free-flow state to a congested state (Park and Haghani, 2016b).

Considering the interdependency between incident duration, prevailing traffic conditions, and the probability of secondary crashes, it is crucial to devise a proactive approach to mitigate the risk of secondary crashes (Park et al., 2018). A proactive secondary crash mitigation strategy can be implemented by disseminating advanced warning messages to inform upstream drivers of the potential secondary crash risk. Figure 2-5 illustrates a concept that focuses on disseminating advanced warning messages to drivers upstream a potential primary incident. This concept plan provides a framework for the processes required to effectively communicate information to upstream drivers in real time.

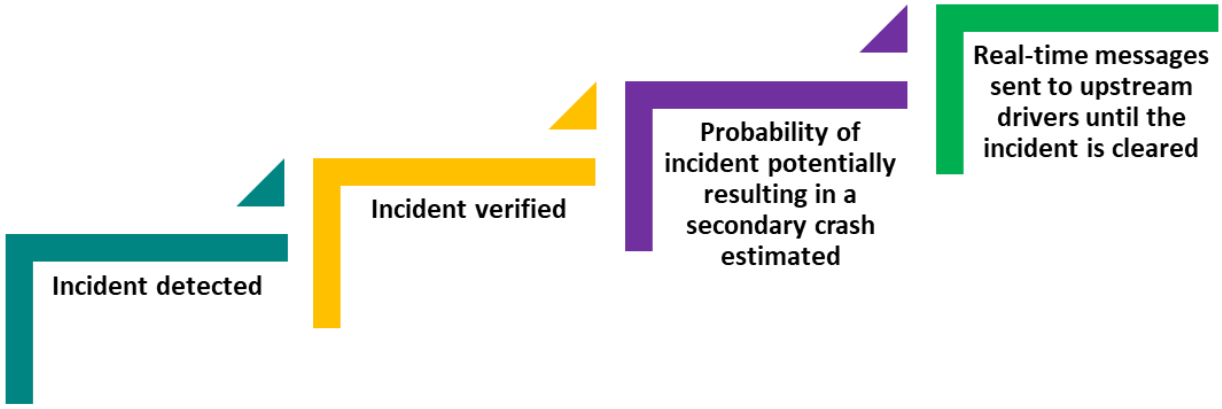


Figure 2-5: A Conceptual Plan to Disseminate Secondary Crash Risk Warning Messages to Upstream Drivers in Real Time

2.3.1 Potential Warning Messages

As can be inferred from Figure 2-5, following the detection and verification of a traffic incident, the risk of occurrence of a secondary crash can be estimated, in real time, as a function of prevailing conditions. If the detected traffic incident is deemed to have a probability of resulting in a secondary, information about this incident could be disseminated to motorists upstream of the primary incident location in real time. The disseminated information will provide motorists with an opportunity to take necessary precautions to avoid being involved in a secondary crash, such as slowing down, changing lanes in advance, and/or diverting to alternate routes.

Speed Advisory

Speed advisory information may be provided in addition to other warning messages to increase the likelihood of driver compliance and prevent secondary crashes. Driver compliance and improved network performance have been reported when advisory speeds are only slightly lowered (Riggins et al., 2016). Li et al. (2014) proposed using variable speed limits to reduce the risks of secondary crashes during inclement weather conditions. By analyzing the risk of secondary crashes, the speed limits can be adjusted (i.e., lowered) depending on prevailing traffic and weather conditions. Based on safety surrogate measures, the proposed variable speed limit system was found to reduce the risk of secondary crashes by 40-50 percent (Li et al., 2014). Introducing a

variable speed limit when the risk of a secondary crash is high can help achieve the desired speed reduction to minimize hard-braking and high deceleration conditions that could potentially lead to secondary crashes.

Lane Change Advisory

Lane change advisory messages inform drivers of lane blockages resulting from traffic incidents downstream. Figure 2-6 shows two examples of lane change messages used for Active Transportation and Demand Management (ATDM) by the Washington State Department of Transportation (WSDOT). The distance between the downstream incident and the upstream lane change message varies, depending on the method of dissemination.

Figure 2-6: Examples of Lane Change Advisory Messages (WSDOT, 2019)

Since a majority of the DMS locations are permanent, advisory messages may be displayed well upstream of an incident. CV messages, however, can be delivered to vehicles at variable distances within the range of the vehicle's signal. Therefore, the algorithm should vary the advisory information to be disseminated based on incident characteristics and traffic flow parameters, such as queue formation, traffic flow, and density.

Detour/Alternate Route Advisory

The FHWA has identified the alternate route plan as a key traffic management strategy for reducing the effects of non-recurring congestion-causing events on highways (Dunn Engineering Associates, 2006). Alternate routes serve to reduce traffic demand upstream of an unplanned event, which helps minimize traffic congestion. Figure 2-7 shows an example of a detour advisory issued by the South Carolina Department of Transportation (SCDOT) using the Dynamic Message Signs (DMSs).

Previous studies show that diversion of only 10-15% of the traffic flow can produce optimal benefits, and higher diversion rates may further decrease travel time on the main route and overall network (Liu et al., 2012; Park and Smith 2012; Zhou 2008). However, high diversion rates may also reduce performance on alternate routes. Therefore, diversion of traffic should be advised so long as motorists do not experience longer travel times while using the alternate route. This can be accomplished by tracking the diverting traffic to advise only a proportion to be directed to the alternate route.

Figure 2-7: An Example of Detour/Alternate Route Advisory (Cabbagestalk, 2017)

The greatest challenge associated with the dissemination of advance warning messages (i.e., speed advisory, lane change advisory, and detour/alternate route advisory) to inform upstream drivers of potential secondary crash risks is determining the location, time, and delivery method to alert motorists of the potential primary incident.

2.3.2 Approaches to Disseminate Warning Messages to Upstream Drivers

The following methods could be used to broadcast warning messages to upstream motorists:

- DMSs (Kopitch and Saphores, 2011), and
- Advanced Traveler Information Systems (ATIS), such as Florida's FL511 service; navigation applications, such as *Waze*; and CV applications (Soloka, 2019; Yang et al., 2017).

The following subsections discuss these communication avenues to inform drivers upstream of a primary incident that may help to mitigate potential secondary crashes.

Dynamic Message Signs

DMSs are programmable devices that can display any combination of letters and/or symbols/graphics to deliver messages to motorists. They can provide real-time information and are used for traffic warnings, regulations, routing, and traffic management (Montes et al., 2008). Some messages provided by DMSs suggest a course of action to motorists, such as change travel speed, change lanes, or divert to a different route. Other messages may serve to inform motorists of changes in current or future traffic conditions (e.g., *Congestion Ahead*) or state regulations (e.g., *Buckle Up It's the Law*, *Click It or Ticket*, etc.).

DMS messages may reduce potential secondary crashes and downstream speed differentials by informing motorists of downstream traffic conditions (e.g., congestion caused by an incident) and encouraging safer driving (Chatterjee et al., 2002; Mounce et al., 2007). Kopitch and Saphores (2011) used the distance from the primary incident location to the nearest upstream DMS as a proxy to quantify the impacts of DMS messaging on secondary crash prevention. The DMS

location was assumed to be at least two miles away from the primary incident for it to be effective (Kopitch and Saphores, 2011). The effectiveness of DMSs in reducing secondary crashes increased between 2 and 11.15 miles and decreased between 11.15 miles and 22.3 miles. Although DMSs were found to influence the probability of secondary crashes, this finding was not statistically significant (Kopitch and Saphores, 2011).

In addition to safety benefits, DMS messages have also been observed to improve mobility. DMS messages may be used to divert traffic to an alternate route, thereby reducing delays (Mounce et al., 2007; Chiu et al., 2001). The willingness of drivers to adhere to DMS messages and divert to alternate routes depends on several factors. A primary concern for drivers is the characteristics of the alternate route. Motorists are more willing to divert if they deem the alternate route to be suitable. If motorists believe the alternate route will have a similar amount of congestion as their current route, they may not be willing to divert. Additionally, motorists may not divert routes if they suspect the alternate route will not provide shorter travel times.

Advanced Traveler Information Systems

In addition to DMSs, other platforms that could be used to disseminate proactive safety messages to upstream drivers are ATIS, including:

- Florida's FL511 service, and
- navigation applications, especially those that leverage crowdsourced user reports for providing services such as *Waze*.

Navigation applications such as Waze can allow users to create and send highway advisory messages from their smartphone at the incident scene (Figure 2-8).

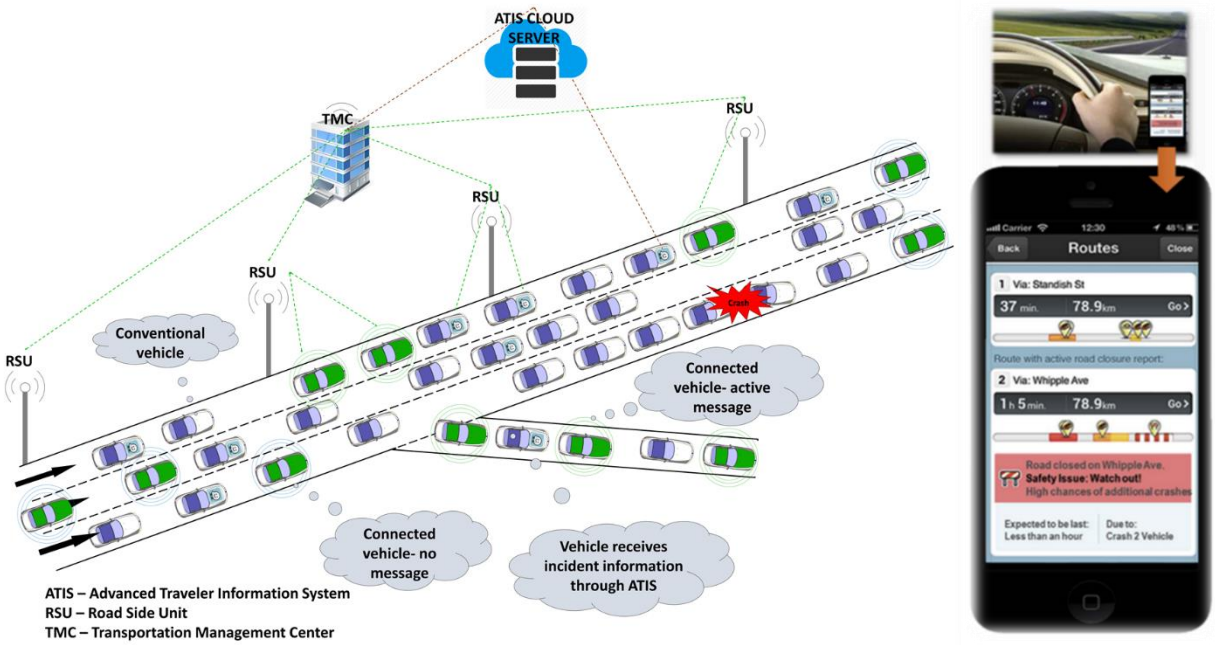


Figure 2-8: Potential of Advanced Traveler Information System in Mitigating Secondary Crashes

The utilization of this correspondence innovation enables drivers to know what is happening on the road and alert them about traffic conditions, incidents, police presence, construction, and even route change suggestions to save time (Imani, 2019). The *Waze* platform has already been integrated into the SunGuide® software used by many TMCs in Florida for traffic management (Glotzbach, 2014). The incidents reported on *Waze* are linked directly to SunGuide® in real time. Likewise, the *Waze* database collects the incidents reported in the SunGuide® system (Glotzbach, 2014).

Amin-Naseri et al. (2017) evaluated the reliability, coverage, and added value of crowdsourced traffic incident reports from *Waze* in Iowa. The study concluded that the crowdsourced data stream from *Waze* is an invaluable source of information for broad coverage of traffic monitoring systems, covering 43.2% of Iowa's Advanced Traffic Management System (ATMS) crash and congestion reports. The *Waze* application also provided timely reporting, 9.8 minutes earlier than the probe-based alternative, on average, and with reasonable geographic accuracy. *Waze* reports currently make significant contributions to incident detection and further complement the ATMS coverage of traffic conditions.

CV Applications

Given the emerging CV technologies, it is likely that many vehicles will soon connect with the surrounding infrastructure. CVs are equipped with certain technologies that help them communicate with their environment. This connected environment allows the CVs to communicate (i.e., send and receive messages) with other vehicles, known as V2V communication, as well as communicate with the surrounding infrastructure, known as vehicle-to-infrastructure (V2I) communication (Harding et al., 2014). This potential, coupled with roadside equipment, can subsequently provide the ability to alert drivers of downstream incidents, which can lead to enhanced safety and mobility. A notable benefit of such technology could be the prevention of secondary crashes.

V2V communication uses on-board Dedicated Short Range Communication (DSRC) devices to convey information about a vehicle's status, such as speed, direction, acceleration, braking status, and other information, to other vehicles and receive similar information from other CVs. Figure 2-9 shows some of the CV technologies that utilize V2V communication. CVs also use the vehicle ad hoc network (VANET) to convey and process signals to and from roadside units (RSUs), as well as other vehicles in the stream (Ghori et al., 2018). The messages exchanged between vehicles have the range and line-of-sight capabilities that exceed current stand-alone vehicle sensing technologies (Harding et al., 2014).

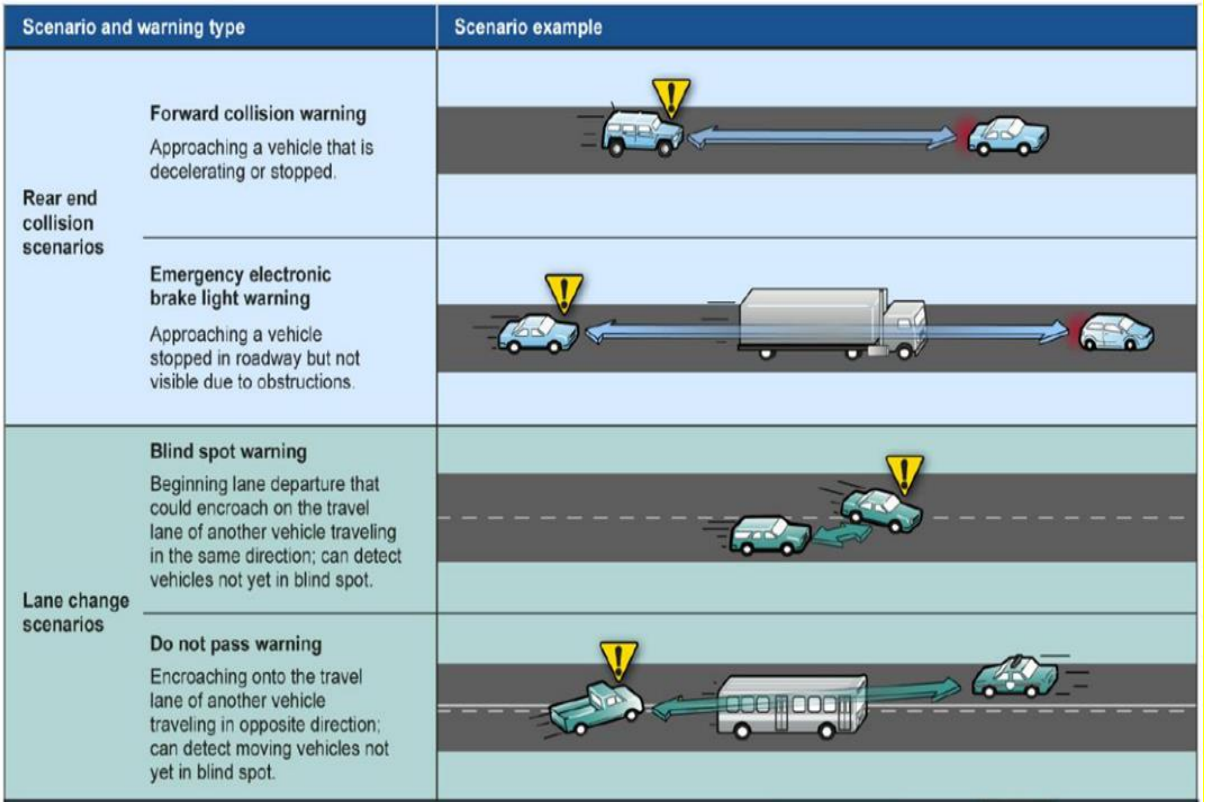


Figure 2-9: An Example of V2V Technology (Yue et al., 2018)

The benefits of CV applications in preventing secondary crashes have, so far, been evaluated using microsimulation studies. Yang et al. (2017) examined the impact of V2V communication on improving the situational awareness of drivers to mitigate secondary crashes. The risk of secondary crashes, measured by the number of simulated conflicts, was found to be significantly reduced if the MPR of CVs on a highway was at least 15% in dense traffic conditions (Yang et al., 2017).

Unlike V2V communication, V2I communication involves communication between vehicles and infrastructure equipment. Currently, the majority of V2I deployments are along arterial roadways, where the communication is primarily between vehicles and traffic signals and the surrounding environment. Through this communication system, CVs share information with RSUs linked to ground servers and traffic control centers (Yang et al., 2017). As illustrated in Figure 2-10, V2I communication could convey various real-time traffic information, including downstream congestion and advisory speed limits on sharp curves. Such real-time information promotes driver awareness, which may improve safety for motorists (Marshall et al., 2017).

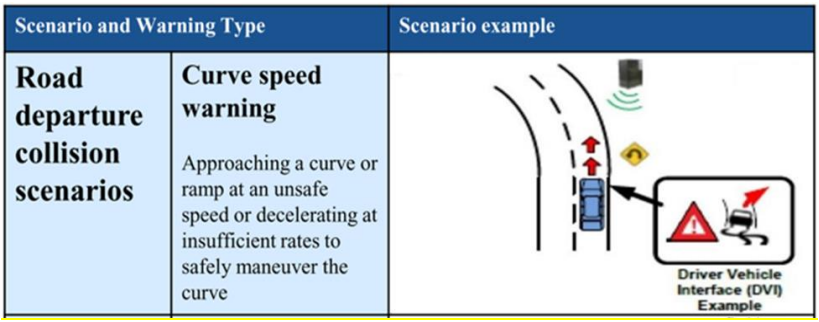


Figure 2-10: An Example of V2I Technology (Yue et al., 2018)

2.4 Summary

The FHWA has established the reduction of secondary crashes as one of the performance measures for incident management programs. Proper identification of secondary crashes is pivotal to accurate reporting of the effectiveness of the programs deployed to mitigate secondary crashes. However, the limited knowledge of the nature and characteristics of secondary crashes has largely impeded their mitigation efforts. The following subsections discuss the research gap pertaining to the identification of secondary crashes, understanding factors influencing the likelihood of secondary crashes, and the prediction of secondary crashes.

2.4.1 Challenges in Identifying Secondary Crashes

Primarily three methods have been used to identify secondary crashes: (1) manual method; (2) static method; and (3) dynamic method. In the "manual" method, secondary crashes are manually identified by either TMC personnel or incident responders. In this case, the impact area of primary incidents is estimated visually based on the observer's judgment. However, despite being the most commonly used method, it is subjective, unreliable, inconsistent, and random.

Instead of relying on the manual method to identify secondary crashes, some studies defined the impact area of the primary incident based on fixed spatiotemporal thresholds and detected secondary crashes within the predefined area. Although the static method is better than the manual method, the one-size-fits-all approach of using fixed spatiotemporal thresholds does not yield reliable results. This is because the impact area of the primary incident depends on the prevailing traffic conditions, i.e., uncongested or congested conditions. To overcome the limitations of the manual and static methods, recent studies have adopted a data-driven dynamic method. In this case, spatiotemporal thresholds vary depending on the impact of the primary incident on traffic flow parameters, hence the term *dynamic*. Although the dynamic method is proven to yield accurate and reliable results, applying it requires high-resolution traffic data, which are only available at limited locations. To accurately identify secondary crashes, this approach needs to be able to distinguish non-congestion patterns from congestion patterns. Further consideration must be made to emulate how congestion conditions develop and disseminate.

Prevailing traffic data were used to automatically estimate the impact area of individual incidents and identify secondary crashes within the affected area. The developed approach considered how the queue caused by the primary incident grows and dissipates upstream of the incident. This

approach is considered to more accurately identify secondary crashes since it better reflects the changes in traffic characteristics caused by the primary incident.

2.4.2 Challenges in Identifying Factors Influencing Secondary Crashes

After identifying secondary crashes, understanding the contributing factors is crucial to developing strategies to mitigate them. Both parametric and non-parametric models have been used to estimate the secondary crash likelihood. The response variable, which is the probability of a secondary crash, is modeled as a binary variable, i.e., primary incident as the "yes" category and normal incident as the "no" category. Primary incident characteristics and traffic flow characteristics, followed by weather conditions, and geometric characteristics, have been observed to have a significant impact on the likelihood of secondary crashes.

Modeling the risk of secondary crashes has the following challenges: (1) accounting for the infrequent nature of secondary crashes; (2) selecting the most important variables with minimal correlation; (3) considering prevailing traffic conditions; and (4) including other potential variables that are rarely considered in the literature.

A penalized logistic regression, fitted using the bootstrap resampling approach, was used to identify the risk factors that influence secondary crashes. Traffic flow, incident, temporal, weather, and roadway geometric attributes were considered as potential factors that may influence the likelihood of secondary crashes. Potential secondary crashes influential variables that were rarely considered in previous studies, i.e., presence of vertical curve, merge influence area, and diverge influence area within the incident impact area, were explored.

2.4.3 Challenges with Deploying Secondary Crash Mitigation Strategies

It is important to devise proactive strategies to promptly reduce the risk of secondary crashes because their occurrence is largely influenced by the severity of the primary incident and how quickly the incident is cleared. Previous research that explored strategies to mitigate secondary crashes used traffic data to identify and predict the likelihood of secondary crashes in real time (Kitali et al., 2018; Xu et al., 2016). However, these studies neglected the influence of prevailing traffic conditions on the likelihood of a secondary crash following the occurrence of the initial incident.

An algorithm that predicts the likelihood of secondary crashes in real time was developed as a *proof-of-concept* in this research effort. This algorithm could be used to develop an ATMS to proactively prevent secondary crashes. The potential of CV applications in mitigating secondary crashes was also explored. The explored applications include speed advisory, lane-change advisory, and detour advisory.

CHAPTER 3
DATA AND STUDY CORRIDORS

This chapter discusses the data and study corridors used to achieve the research goal and objectives. Four main types of data, i.e., incident data, real-time speed data, real-time rainfall data, and roadway geometric characteristics data, were used. The study sections constituted the roadway corridors within the Florida's Turnpike Mainline and Turnpike Extension.

3.1 Data Requirements

The following types of data were required to achieve the research goal: (1) incident data; (2) high-resolution traffic data; (3) roadway geometric data, including work zone information; and (4) high-resolution rainfall data. Incident data were obtained from the SunGuide® database. High-resolution traffic data were retrieved from HERE Technologies, and work zone data were obtained from the FDOT Open Data Hub (FDOT, n.d.). Other roadway geometric characteristics were extracted from the Roadway Characteristic Inventory (RCI) database, Google Earth Pro, and Google Maps. High-resolution rainfall data were retrieved from the National Oceanic and Atmospheric Administration (NOAA) Next Generation Weather Radar (NEXRAD) Level-II network. These data were collected for 5.5 years, from January 2014 to June 2019. The following subsections discuss each of these data sources.

3.1.1 SunGuide®

SunGuide® is an ATMS software used by the FDOT to process and archive incident data on freeways. The database stores several incident attributes, including:

- incident identification (Event ID),
- roadway name,
- latitude and longitude of the incident location,
- incident notification time,
- incident type,
- number and categories of responding agencies (e.g., EMS, towing, Road Ranger, etc.),
- lane closure information,
- incident severity,
- weather condition, and
- road surface condition.

The categories of incident events included in the SunGuide® database are crash, disabled vehicles, debris on roadway, emergency vehicles, police activity, vehicle fire, flooding, pedestrian, abandoned vehicles, construction, wrong-way driver, etc. These categories were further summarized into four groups: crashes, vehicle problems, hazards, and other events. Accordingly, the *crashes* group contained crash events. *Vehicle problems* included all events that were not crashes, but were vehicle-related, e.g., disabled vehicles, abandoned vehicles, etc. *Hazards* included all objects on the roadway with the potential of causing crashes, e.g., debris on roadway, wildlife, etc. *Other events* encompassed all events that do not fit in the three aforementioned event

categories, e.g., other, bridge work, amber alert, wrong-way driver, etc. These event types were excluded from the analysis.

Within the SunGuide® database, incident severity is categorized into three groups: minor, moderate, and severe. In the context of incident management, incident severity represents the impact of an incident on traffic conditions and the extent of lane blockage at the incident location. Illustratively, incidents that led to all lanes blocked were the most severe, while incidents without any lane blockages were the least severe (i.e., minor). In this research, incident severity was divided into two categories: minor and moderate/severe.

3.1.2 HERE Technologies

HERE Technologies record the space mean speed for roadways by dividing them into segments referred to as Traffic Message Channels. HERE Technologies speed data were collected from the Regional Integrated Transportation Information System (RITIS) platform. There are 406 Traffic Message Channels along the study corridor, 284 along the Mainline (State Road (SR)-91), and 122 along Turnpike Extension (TE) (Figure 3-1).

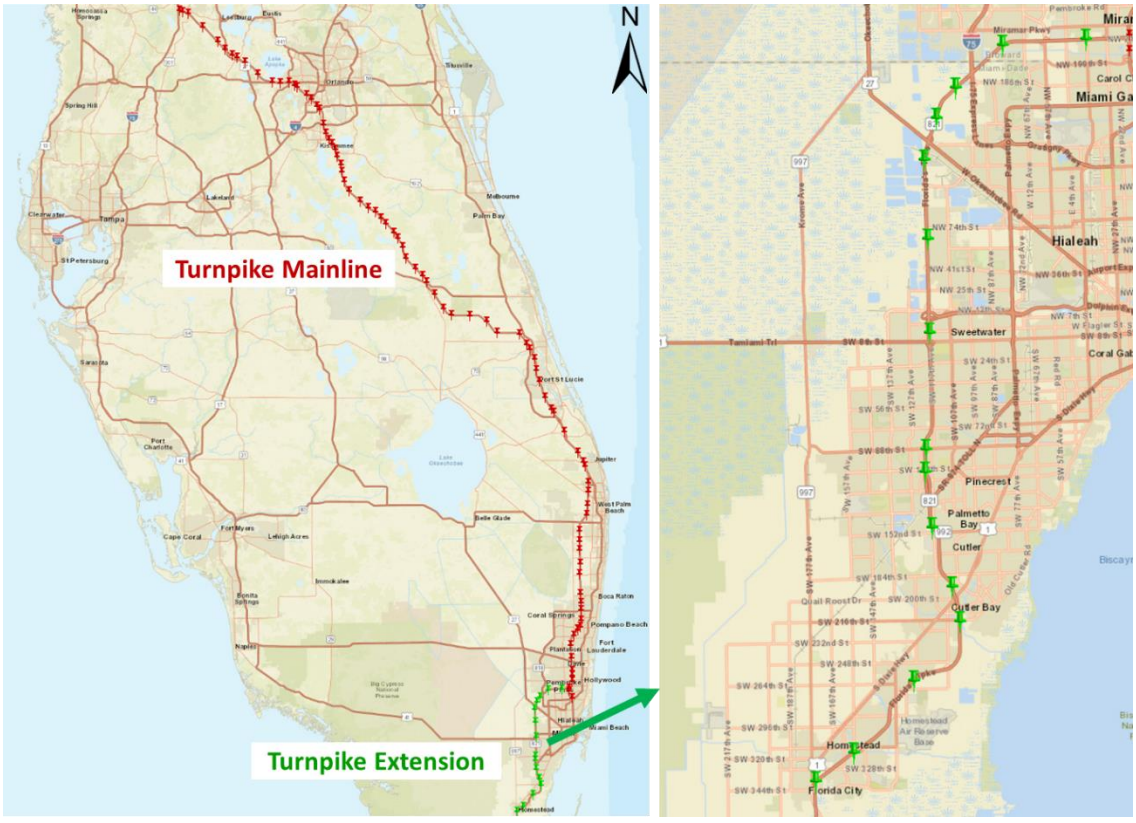
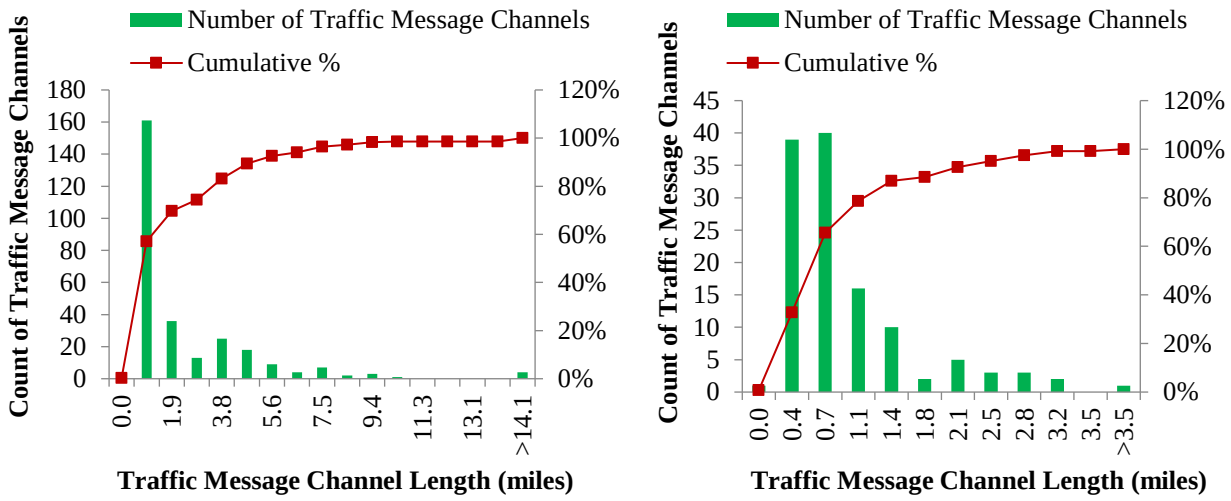


Figure 3-1: Network of HERE Traffic Message Channels within the Study Corridor

On average, Traffic Message Channels along the study corridors are 1.9 miles and 0.7 miles on Mainline and TE, respectively. As depicted in Figure 3-2(a), 65% of the Traffic Message Channels along the Mainline are shorter than 1.4 miles. On the other hand, 88% of Traffic Message Channels

along TE are shorter than 1.5 miles (Figure 3-2(b)). Only 7% of the Traffic Message Channels along TE are longer than 2 miles.

The Traffic Message Channel length affects the estimation of the impact caused by the incident on traffic flow, including queue formation and dissipation. That is, the use of traffic data from overly long Traffic Message Channels may result in an inaccurate estimation of the traffic flow characteristics changes caused by the incident. Since the longest Traffic Message Channel along TE is approximately 4 miles, while the longest Traffic Message Channel along the Mainline is 15 miles, the cut-off of 4 miles was considered as a criterion to include a Traffic Message Channel in the analysis. Notably, only 15% of the Mainline Traffic Message Channels are longer than 4 miles. Thus, as depicted in Figure 3-3, the final study corridor has three main segments: 48-mile long TE, 69-mile long Northern Turnpike Mainline (NTM), and 28-mile long Southern Turnpike Mainline (STM).



(a) Turnpike Mainline (b) Turnpike Extension

Figure 3-2: Spatial Distribution of Traffic Message Channels

The 5-minute speed data from HERE Technologies were first used to identify secondary crashes. Next, speed data (i.e., mean and standard deviation (SD)) in the Traffic Message Channel where the incident occurred and within 10 minutes before the occurrence of the incident were collected to capture the traffic conditions before the occurrence of the incident. To determine the prevailing traffic conditions, speed data within the Traffic Message Channels impacted by the incident, from the time the incident was detected to the time when the traffic flow returned to normal, were used. Since the incident impact duration along different Traffic Message Channels may differ, the incident impact area was individually defined for each Traffic Message Channel.

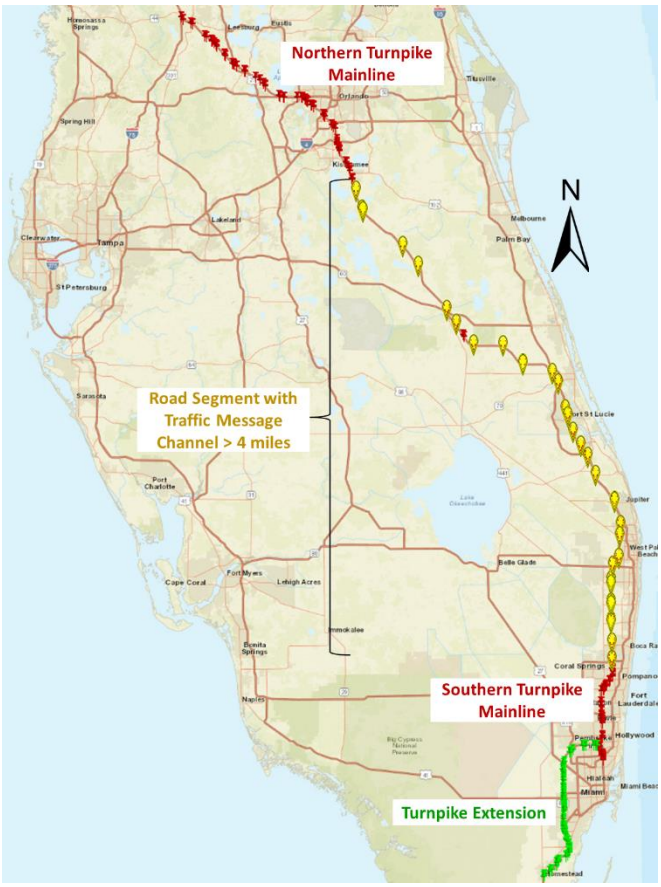


Figure 3-3: Selected Roadway Sections within the Study Area

3.1.3 Roadway Geometric Characteristics and Work Zone Data Sources

Roadway geometric characteristics that may significantly impact traffic flow characteristics and hence the likelihood of secondary crashes were considered. The following geometric variables were considered: shoulder width, horizontal curves, vertical curves, merging segment, and diverging segment. Other potential geometric variables that were considered include service plazas and toll plazas. Since there are very few service plazas and toll plazas within the study area, these variables were excluded from the analysis.

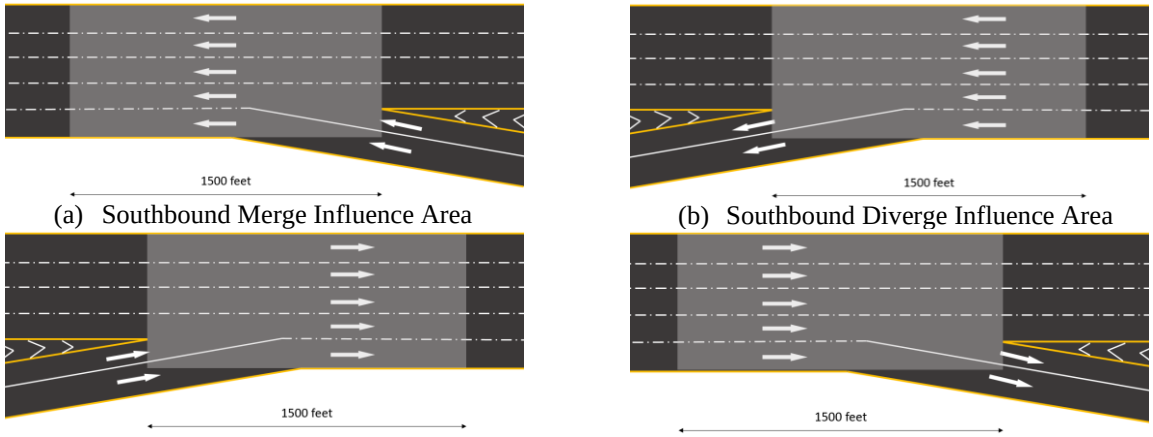
Shoulder width, horizontal curves, and vertical curves variables were collected from the RCI database for the years 2014 through 2019. The shoulder width variable was derived for the outside shoulder located adjacent to the outside travel lane. Outside shoulders provide for the accommodation of stopped vehicles, emergency use, and lateral support of the roadbed (FDOT, 2016). Since the entire roadway section has a median, the shoulder width variable was collected from both directions. The final shoulder width corresponding with each incident was calculated as a weighted average of all the shoulder widths within the incident impact area:

$$Weighted\ shoulder\ width = \frac{\sum_i^n Shoulder\ width_i \times Incident\ impact\ area_i}{Total\ incident\ impact\ area} \quad (3\text{-}1)$$

where $Shoulder\ width_i$ is the shoulder width within the incident impact area and $Incident\ impact\ area_i$ is the portion of the incident impact area with $Shoulder\ width_i$. The subscript i represents the different shoulder width values within the incident impact area, where n is the total number of segments within the impact area.

The horizontal curve variable was aggregated into two categories: incidents with a horizontal curve within their impact area and those without a horizontal curve within their impact area. The vertical curve variable was also aggregated in the same manner as the horizontal curve.

The merge and diverge influence areas were derived from Google Earth Pro and Google Maps using the Historical Imagery and the Street View tools. The Historical Imagery tool was used to verify the location of the identified ramps during the study period. The merge and diverge influence areas were defined based on the Highway Capacity Manual (HCM) (Transportation Research Board [TRB], 2016). A *merge influence area* spans from the point where the edges of the travel lanes of the merging roadways meet to a point 1,500 feet downstream of that point. Similarly, a *diverge influence area* spans from the point where the edges of the travel lanes of the diverging roadways meet to a point 1,500 feet upstream of that point. While the HCM defines the ramp influence area as one that includes only lanes 1 and 2, both merge and diverge influence areas cover the entire roadway section (i.e., all travel lanes) since they are measured within the impact area of an incident (see Figure 3-4).



(a) Southbound Merge Influence Area
(b) Southbound Diverge Influence Area
(c) Northbound Merge Influence Area
(d) Northbound Diverge Influence Area

Figure 3-4: Definition of Merge and Diverge Influence Areas

The final merge/diverge influence area considered was also a dichotomous variable, similar to the horizontal/vertical curve variables. That is, incidents with a merge/diverge influence area within their impact area were grouped into the "*yes*" category and those without merge/diverge influence area within their impact area were categorized as "*no*". Note that the "presence of merge influence area" and the "presence of diverge influence area" were treated as two separate variables.

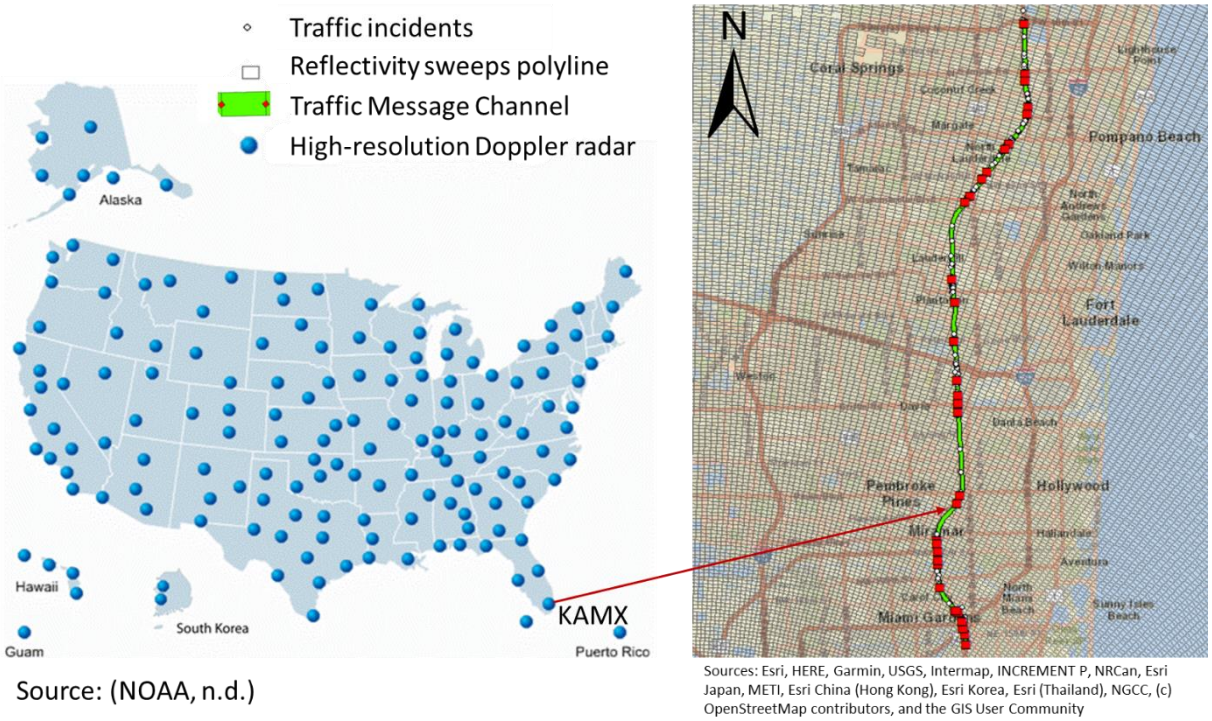
The work zone activities data were retrieved from the Active Construction Project database service that is updated nightly in the FDOT Open Data Hub. The database provides the work zone construction location and duration. The direction of the construction activities is not mentioned in the Open Data Hub and is therefore verified manually using the Google Map Historical Tool.

3.1.4 NOAA

NOAA preserves, monitors, and assesses climate and historical weather data. One of the systems maintained by NOAA is NEXRAD. NEXRAD is a network of 160 high-resolution Doppler radar sites that detect precipitation and atmospheric movement and disseminate near real-time data in approximately 5-minute intervals from each site (Barr, 2018). With these high-resolution data, it is possible to obtain the actual rainfall intensity over the road network in short time intervals.

Original data from the NEXRAD network, referred to as NEXRAD Level-II data, were used. These data included reflectivity, one of the meteorological base data quantities. Radar measures rainfall intensity using radiations reflected on a target surface, in this case, a roadway network. The proportion of a target's productivity in capturing and returning radiofrequency energy is alluded to as reflectivity. Reflectivity can simply be defined as a measure of fractions of radiations reflected by a given surface. It is expressed as the ratio of the radiant energy reflected and the total amount of energy incident upon that surface (Andrew, 2019).

As indicated in Figure 3-5, reflectivity data were downloaded from the radar located in Miami, Florida (KAMX – Miami, FL). This radar is positioned at latitude: 25.61056, longitude: -80.41306, and has been operational since April 20, 1995. Specifically, the NEXRAD Level-II data were accessed from Amazon S3 through the following link https://noaa-nexrad-level2.s3.amazonaws.com. Similar to other high-resolution Doppler radars under NEXRAD, the KAMX radar covers a 248.5-mile radius.



Source: (NOAA, n.d.)

Figure 3-5: Collection of High-resolution Rainfall Data from Radar (NOAA, n.d.)

Figure 3-6 describes the approach used to retrieve rainfall data from NEXRAD. Reflectivity data were obtained for incidents that occurred during inclement weather conditions, as indicated in the

incident database. The data were retrieved at 5-minute intervals, from the time when the incident began impacting traffic to the time when either a secondary crash occurred for primary incidents, or when the traffic flow returned to normal for normal incidents. The downloaded radar data from Amazon S3 are in a unique digital binary format. Thus, as indicated in Figure 3-6 (Step 2), the NOAA Weather Climatic Toolkit (WCT) was used to visualize and convert data into a conventional scientific format, a shapefile in this case. ArcGIS software was then used to merge the downloaded radar data with the Traffic Message Channels impacted at time interval (*t*).

The recorded reflectivity values were converted to rainfall intensity using the following reflectivity-rainfall intensity relationship (Andrew, 2019):

$$R = \frac{10^{\frac{dBZ^{\frac{1}{2}}}{10}}}{250} \tag{3-2}$$

where, R is the rainfall intensity expressed in millimeters per hour (mm/hr), and dBZ is an abbreviation for decibel relative to reflectivity (Z). The *dBZ* is used to compare the reflectivity of a target surface in mm^6 per m^3 to the return of a droplet of rain with a diameter of 1 mm. In other words, it measures the strength of the energy reflected to the radar by the target surface, in this case, the roadway segment.

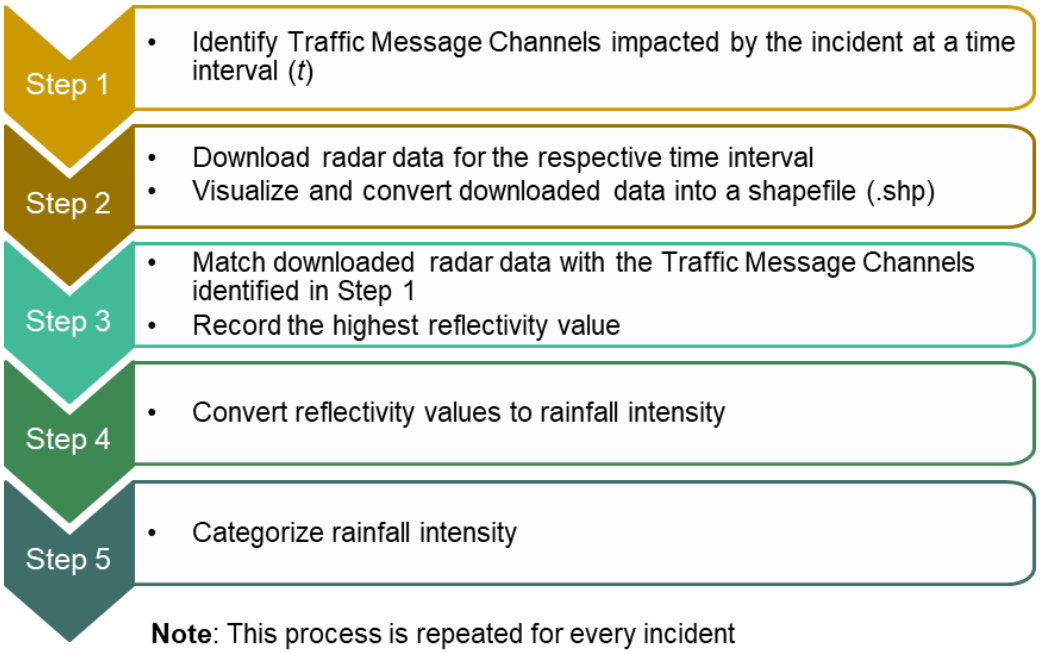


Figure 3-6: Workflow for Collecting and Processing Reflectivity Data

Finally, the rainfall intensity data were grouped into three categories according to the American Meteorological Society (AMS) rainfall intensity classification (American Meteorological Society [AMS], n.d.). The three groups include light rainfall (Trace – 0.10 in/hr), moderate rainfall (0.10 – 0.30 in/hr), and heavy rainfall (> 0.30 in/hr). Table 3-1 shows a sample of rainfall data retrieved from KAMX radar on June 30, 2019, 11:34:55 AM, on Florida's Turnpike System Mainline [mile

marker (MM) 50.1 – 54.8. The high-resolution rainfall data were obtained from the NOAA database.

Table 3-1: Sample Rainfall Data from NEXRAD on June 30, 2019

Sweep Time	Begin MM	End MM	Rainfall (mm/hr)	Rainfall (in/hr)	Rain Category
11:34:55 AM	50.1	50.5	0.0051	0.0002	Light
11:34:55 AM	50.5	54.1	5.1292	0.2019	Moderate
11:34:55 AM	54.1	54.4	42.3367	1.6668	Heavy
11:34:55 AM	54.4	54.5	42.3367	1.6668	Heavy
11:34:55 AM	54.5	54.8	9.1212	0.3591	Heavy

3.2 Study Area

The study corridors were selected from the Florida's Turnpike Mainline and Turnpike Extension. As shown in Figure 3-7, the Turnpike Mainline is a 312-mile corridor consisting of two main roadways: the Florida Turnpike Mainline (or SR-91) and TE (or SR 821). The two roadways are 265 miles and 48 miles, respectively. The following subsections discuss the criteria considered while selecting the study corridors.

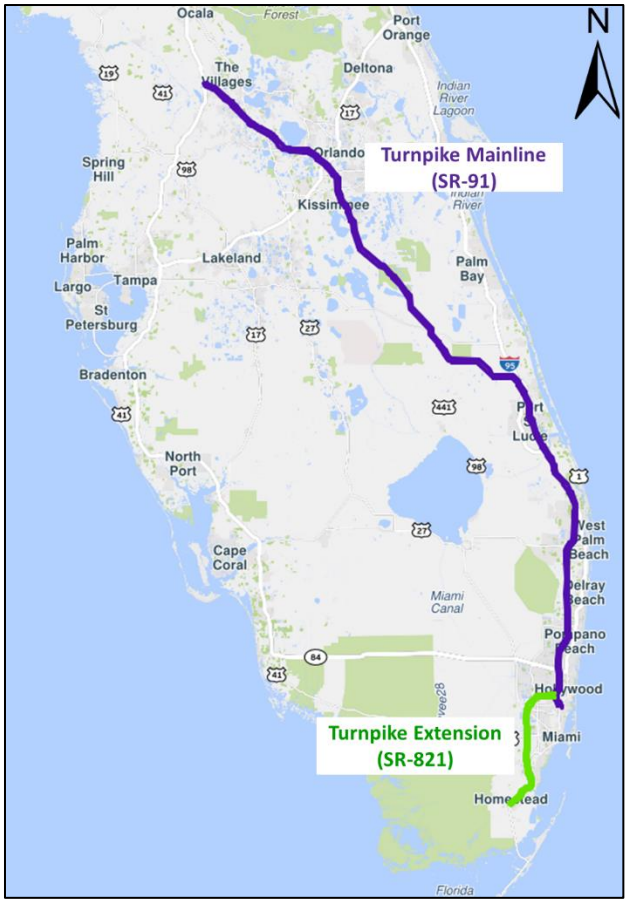
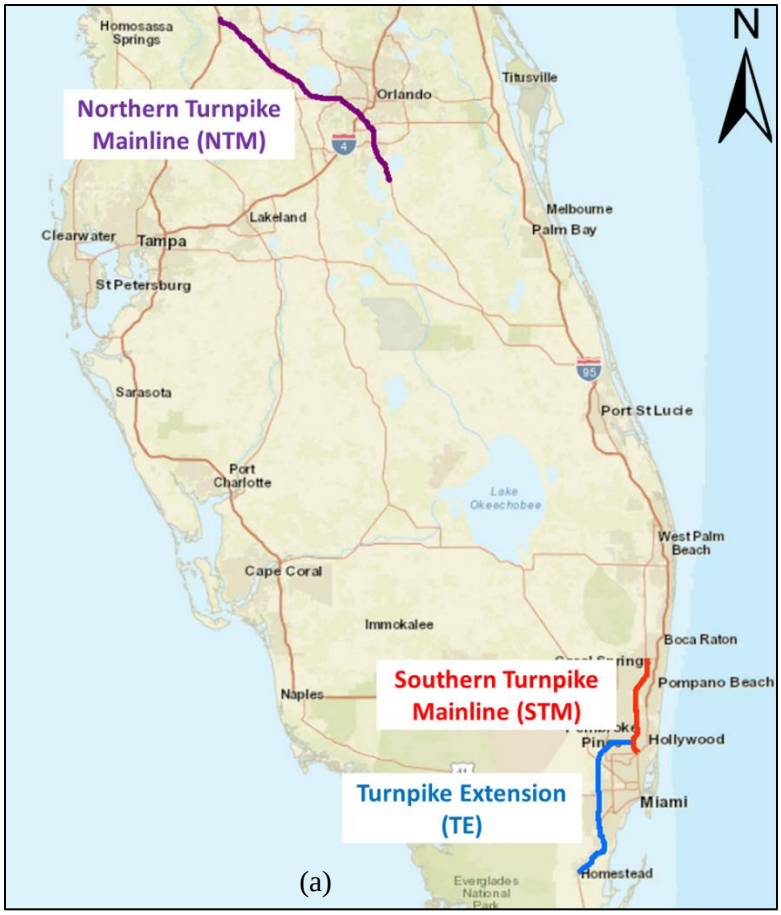


Figure 3-7: Florida's Turnpike Mainline and Turnpike Extension

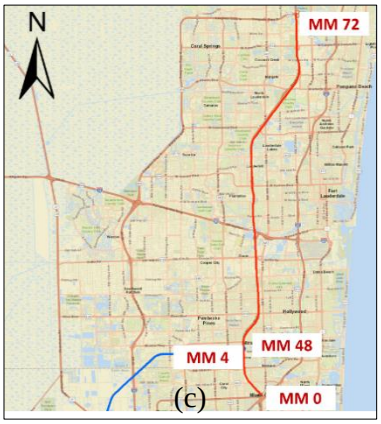
3.2.1 Study Corridors for Identifying Secondary Crashes

Two main data sources are required to estimate an incident impact area: (1) traffic incidents; and (2) high-resolution traffic data. The HERE Technologies record the speed for roadways by dividing them into Traffic Message Channels. The study corridors used to identify secondary crashes were selected based on the availability of high-resolution traffic data. As discussed in Section 3.1.1, segments with overly long Traffic Message Channels were excluded. This is because the traffic information from these segments may result in an inaccurate estimation of traffic flow changes caused by the incident.

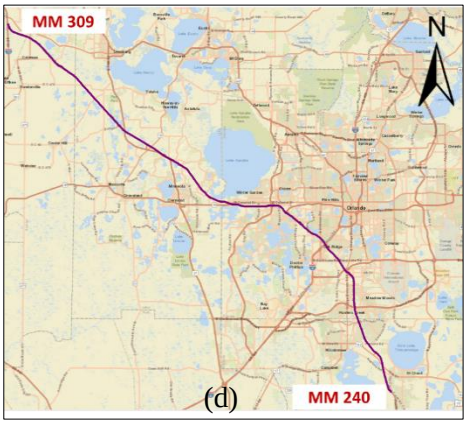
Figure 3-8(a) shows the location of the study area. The TE section is from MM 0 to MM 48 (see Figure 3-8(b)). The STM is located from MM 0 through MM 4, which is the Turnpike Mainline Spur, and from MM 48 through MM 72, which is the junction between SR-91 and SR-869 (Sawgrass Expressway) (see Figure 3-8(c)). The NTM is located from MM 240 through MM 309 (see Figure 3-8(d)). Table 3-2 summarizes the HERE Traffic Message Channels along the selected study corridors.



Turnpike Extension (TE)



Southern Turnpike Mainline (STM)



Northern Turnpike Mainline (NTM)

Sources: Esri, HERE, Garmin, USGS, Intermap, INCREMENT P, NRCan, Esri Japan, METI, Esri China (Hong Kong), Esri Korea, Esri (Thailand), NGCC, OpenStreetMap contributors, and the GIS User Community.

Figure 3-8: Selected Roadway Sections along Florida's Turnpike Mainline and Turnpike Extension

Table 3-2: Distribution of HERE Traffic Message Channels along the Study Corridors

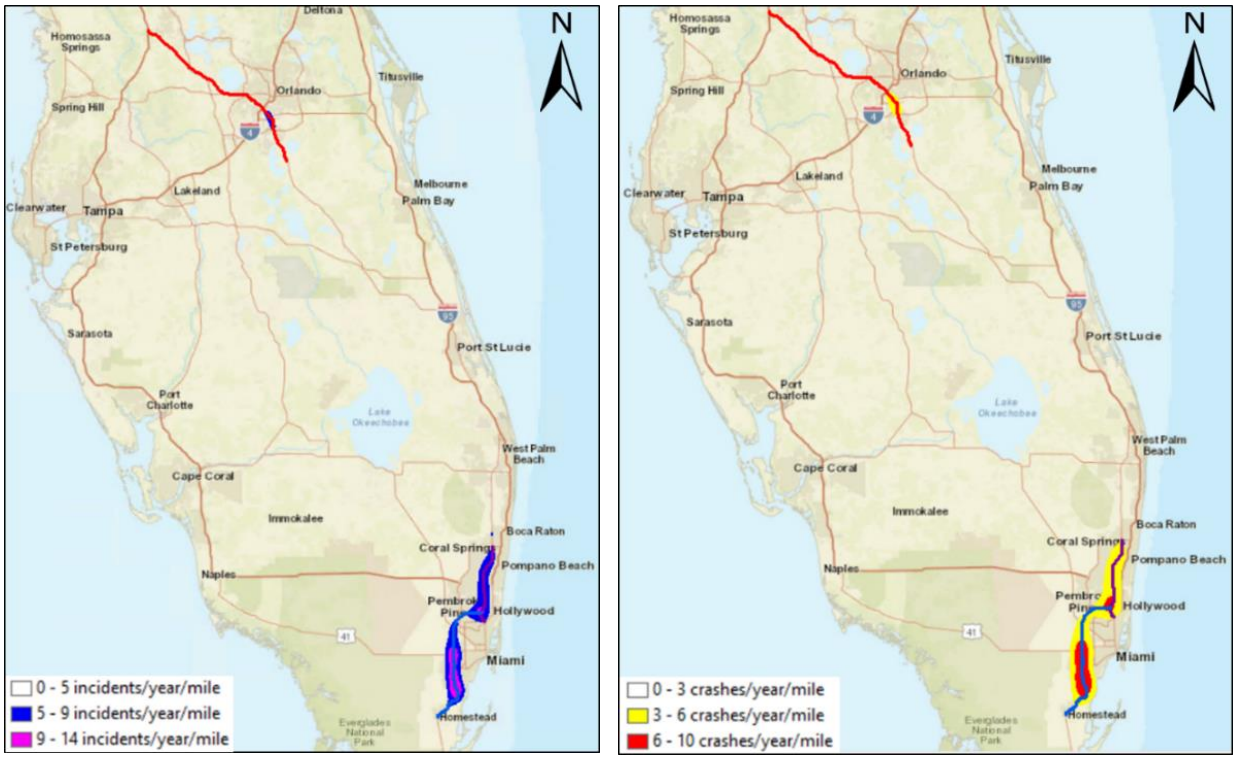
Roadway	Number of Traffic Message Channels			Length of Corridor (miles)
	Northbound	Southbound	Total	
Northern Turnpike Mainline (NTM)	46	47	93	69
Southern Turnpike Mainline (STM)	34	35	69	28
Turnpike Extension (TE)	61	61	122	48

3.2.2 Study Corridor for Developing the Secondary Crash Likelihood Model

Of the three corridors used to identify secondary crashes, only STM was used to develop the secondary crash likelihood models. Incident hotspots were considered one of the criteria for selecting the study corridors for developing the secondary crash likelihood model. The Kernel Density function in ArcGIS was used to identify high incident segments within the Florida's Turnpike System. The hotspot analysis was conducted based on traffic incidents that occurred along the study corridors, i.e., TE, STM, and NTM, during the study period.

The research also ensured there were no major construction activities, i.e., lane widening, bridge maintenance, interchange improvements, etc., taking place along the selected study corridor. Previous research indicated that factors that significantly affect the traffic flow characteristics are more likely to increase the risk of secondary crashes (Kitali et al., 2019b, 2018; Xu et al., 2016). The constrained driving environment in work zones tends to disturb the normal traffic flow, leads to speed reductions, and reduces road capacity. As such, the presence of the work zone, particularly construction activities, could exacerbate the occurrence of secondary crashes compared to other corridors without major construction activities. The Active Construction Projects shapefile from the FDOT Open Hub Data website was used to identify major construction activities along the study corridors during the study period.

As indicated in Figure 3-9, both the incident hotpot analysis and the crash hotspot analysis identified the TE and STM as corridors that experienced the highest number of traffic incidents and crashes. Nonetheless, the exploratory analysis of the data in the Active Construction Projects shapefile indicated that lane widening construction activities were taking place within the TE section during the study period. Meanwhile, in STM, there were no such activities during the study period. Thus, this section was used to develop the secondary crash likelihood model.



(a) Incident hotspot (b) Crash hotspot

Sources: Esri, HERE, Garmin, USGS, Intermap, INCREMENT P, NRCan, Esri Japan, METI, Esri China (Hong Kong), Esri Korea, Esri (Thailand), NGCC, OpenStreetMap contributors, and the GIS User Community.

Figure 3-9: Corridors with High Incidents and Crashes along Florida's Turnpike Mainline and Turnpike Extension

3.2.3 Study Corridor for Developing the Secondary Crash Risk Prediction Model

The STM corridor used in the secondary crash likelihood model was also used to predict the occurrence of secondary crashes in real time. As mentioned earlier, the STM corridor is a 28-mile section of the Florida's Turnpike Mainline (SR-91) from MM 0 through MM 4, which is the Turnpike Mainline Spur, and from MM 48 through MM 72, which is the junction between SR-91 and SR-869 (Sawgrass Expressway).

3.2.4 Study Corridor for Exploring the Potential of Mitigating Secondary Crashes Using CV Applications

The potential of CV applications to mitigate secondary crashes was explored using a microsimulation approach. The selected study corridor, shown in Figure 3-10, is a 7.8-mile, 6-lane (3-lanes in each direction) road segment on the Florida's Turnpike Mainline (SR-91). The freeway segment is in Broward County and crosses four roadways: Sample Road, Copans Road, Coconut Creek Road, and Atlantic Boulevard. Note that the interchanges are 1 to 2 miles apart, with access to the Turnpike at each interchange except the Copans Road crossing. This site was chosen over other segments along the Florida's Turnpike due to its relatively high number of crashes in 2016-2019.

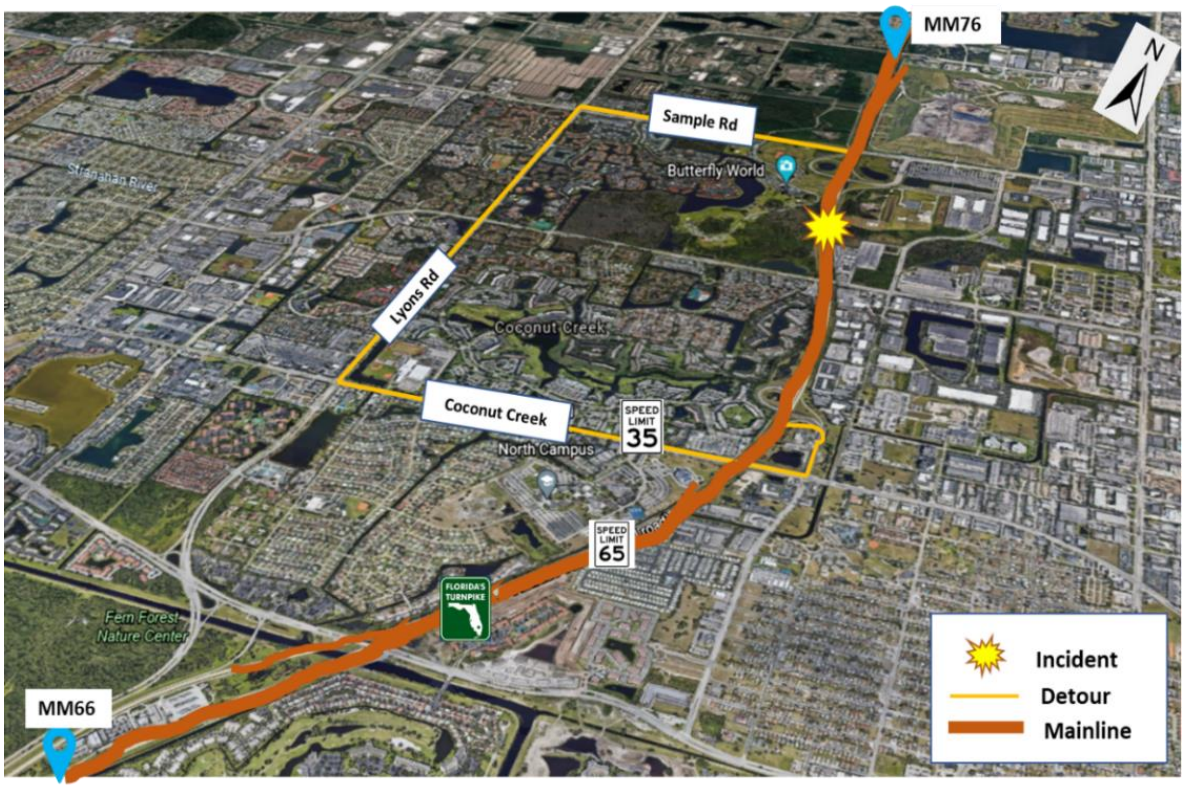


Figure 3-10: Selected Study Corridors for the Microsimulation Model: Turnpike Mainline (SR-91) and Detour Route (Lyons Road)

Furthermore, a 4.2-mile section of Lyons Road, with two lanes in each direction, was considered for detouring purposes, as shown in Figure 3-10. The detour route is expected to divert a portion of the Turnpike's mainline traffic, particularly while the incident is being cleared. The diverted traffic exits on Coconut Creek Road before traveling along Lyons Road and returns to the Turnpike mainline via W Sample Road. Compared to other possible detour routes along the Turnpike corridor, the selected detour is one of the sections with the closest consecutive interchanges (~ two miles).

3.3 Summary

The following four data types were used in this research:

- Traffic incidents from the SunGuide® database,
- High-resolution speed data from HERE Technologies,
- Rainfall data from NOAA NEXRAD Level-II network, and
- Roadway geometric attributes data from FDOT RCI, Google Maps, and Google Earth Pro.

Table 3-3 summarizes the data requirements for different steps accomplished in this research effort. These data were collected both historically and in real time (except for roadway geometric variables).

Table 3-3: Data Needs for Predicting Secondary Crashes in Real Time

Data Source	Data Type	Identify SC	SC Likelihood Model	SC Prediction Model
SunGuide®	Incident			
HERE Technologies	Speed			
RCI	Shoulder width, horizontal curve, and vertical curve			
Google Maps and Google Earth Pro	Merge ramps and diverge ramps			
NOAA	Rainfall intensity			

Note: FDOT = Florida Department of Transportation; RCI = Roadway Characteristic Inventory; NOAA = National Oceanic and Atmospheric Administration; SC = Secondary Crash.

Table 3-4 summarizes the study corridors selected from the Florida's Turnpike Mainline to achieve the research goal and objectives.

Table 3-4: Summary of Study Corridors Considered in this Research Effort

Research Objective	Study Corridor
Identifying SCs	48-Mile TE; 69-mile NTM; 28-mile STM
Developing the SC Likelihood Model	28-mile STM
Developing the SC Risk Prediction Model	28-mile STM
Exploring the Potential of Mitigating SC Using CV Applications	7.8-mile section of STM; 4.2-mile section of Lyons Road

Note: SC = Secondary Crash; TE = Turnpike Extension; Northern Turnpike Mainline (NTM); STM = Southern Turnpike Mainline.

CHAPTER 4
IDENTIFY SECONDARY CRASHES

This chapter discusses the methodology used to identify secondary crashes along the Florida's Turnpike Mainline and Turnpike Extension. A data-driven approach that accurately estimates the primary incident's impact area and accurately identifies secondary crashes was developed. A critical component of this approach is that it considers that the queue built by the primary incident will grow and dissipate at different rates. The impact of the incident on the location where it occurred is expected to be different from the farther upstream segments. Additionally, using the developed approach, the incident's non-recurrent congestion caused by the incident can be separated from any recurrent congestion within the incident's spatiotemporal impact area. Note that this research focused on secondary crashes that occurred in the upstream direction of the primary incident.

4.1 Background

The impact area of individual incidents was first estimated as the first step towards identifying secondary crashes. The extent of the impact area was characterized by the primary incident duration and the length of the queue initiated by the incident. The fundamental questions addressed while developing an algorithm to identify secondary crashes included:

- How to accurately determine the spatial and temporal boundaries of the impact area when an incident occurs?
- How to determine whether the change in traffic states at the location of the secondary crash is due to the prior incident or other factors such as recurrent congestion?

Two main data sources were required to estimate the impact area of a primary incident: (1) traffic incidents; and (2) real-time traffic data. These data were collected within the period of study, which ranges between January 2014 to June 2019. High-resolution speed data were retrieved from HERE Technologies, and traffic incident data were retrieved from the SunGuide® database.

4.2 Methodology

A data-driven approach was used to identify secondary crashes. This method accurately estimates the impact area of the primary incident using speed data from HERE Technologies and identifying secondary crashes occurring within the impact area of the primary incident. The developed approach aims to better capture traffic flow characteristics, such as speed, that change over space and time and affect the queue formation caused by the primary incident. As discussed in Section 3.2.1, the study area included the TE corridor, a 48-mile extension of the Florida's Turnpike, and a 97-mile section on the Florida's Turnpike Mainline, i.e., a 69-mile NTM and a 28-mile STM. As indicated in Figure 4-1, four major steps were used to identify secondary crashes using the developed data-driven approach. The following subsections discuss each of these four steps in detail.

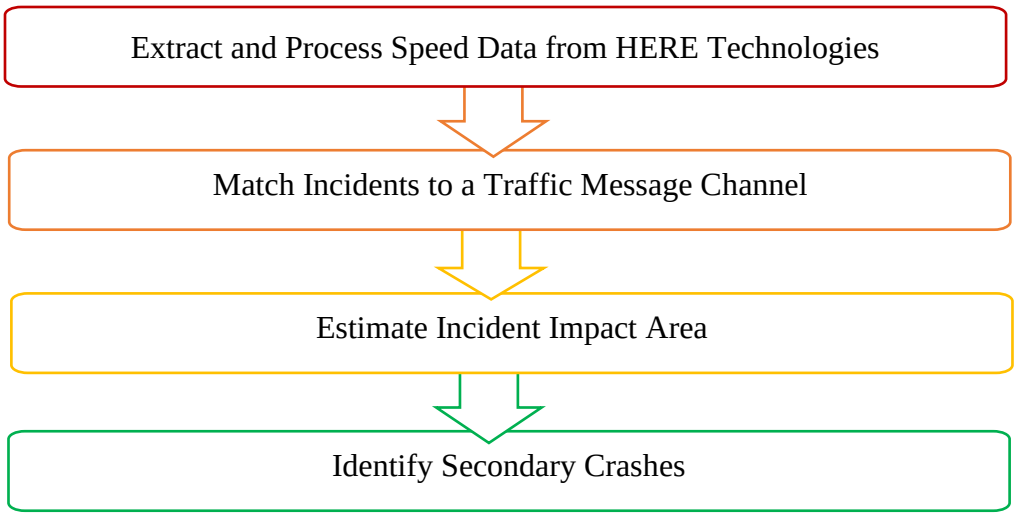


Figure 4-1: Data-driven Approach to Identify Secondary Crashes

4.2.1 Extract and Process Speed Data from HERE Technologies

The 5-minute speed data from HERE Technologies were retrieved from the 284 Traffic Message Channels along the study corridor from January 2014 through June 2019. These data were used to establish the recurrent speed profile of the section under normal traffic conditions. The average speed in each 5-minute interval was used to establish the recurrent speed profile. Additionally, a confidence interval of two standard deviations was established to define the lower and upper bounds (i.e., speed bandwidth) of the speed profile to account for the variation in speeds on a roadway segment. For each Traffic Message Channel, seven speed profiles were generated, one for each day of the week. Independent speed profiles for different days of the week and times of the day were established to account for the recurrent traffic congestion. Figure 4-2 shows a typical speed profile for a 24-hour period on a weekday. As expected, there is a significant drop in speed during the morning peak hours, while the average speeds were the highest between midnight and 5:00 AM.

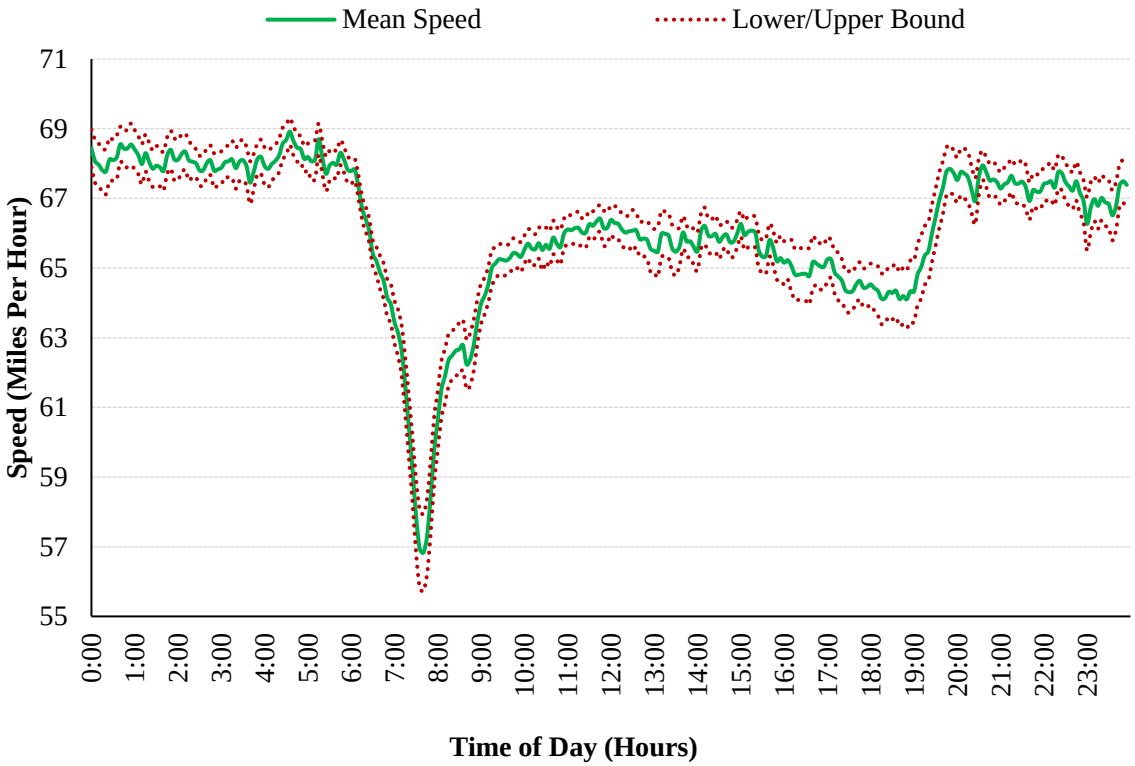


Figure 4-2: Sample Speed Profile for Estimating Normal Traffic Conditions

4.2.2 Match Incidents to a Traffic Message Channel

The geographic location of both the incidents and the Traffic Message Channels is the most critical information required for matching an incident to the Traffic Message Channel. MMs of incidents and Traffic Message Channels (start and end) were used instead of the geographic coordinates, i.e., longitudes and latitudes. Through the ArcGIS tool, the *Toll Roads Polyline shapefiles* extracted from the FDOT Transportation Data and Analytics Office website were used to assign MMs to the incidents and the start and end of the Traffic Message Channels. This approach ensures that roadway alignment characteristics, especially on curved segments, do not affect the accurate computation of the spatial relationship between incidents and Traffic Message Channels.

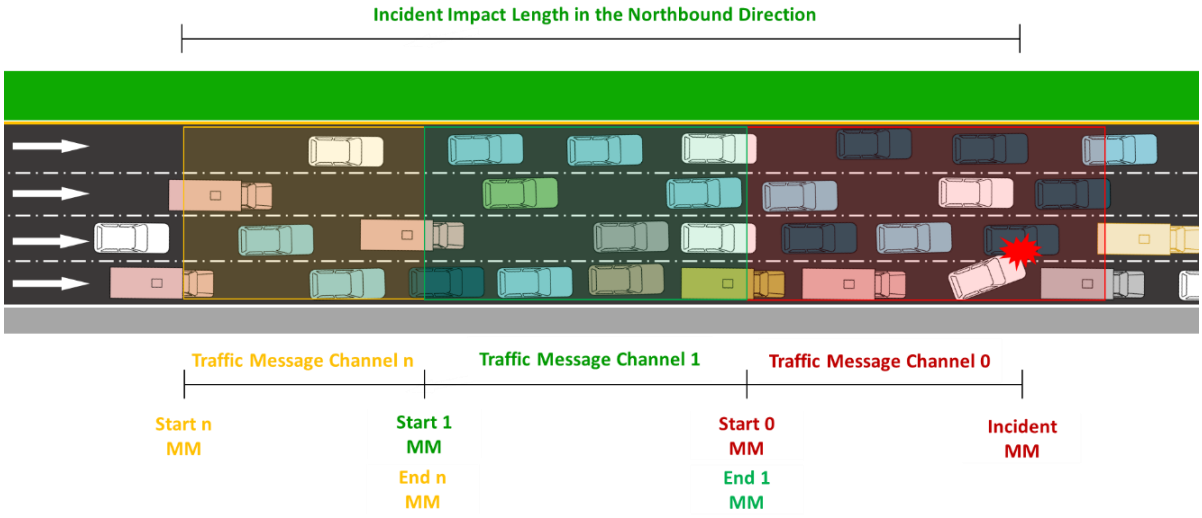
Using the assigned MMs, each incident was matched to a Traffic Message Channel at the incident location. For northbound incidents, since MMs increase in the northbound direction, the MM of the northbound incident must be greater than or equal to the MM of the start of the Traffic Message Channel and less than or equal to the MM of the end of the Traffic Message Channel. Similarly, for southbound incidents, since MMs decrease in the southbound direction, the MM of the incident must be less than or equal to the MM of the start of the Traffic Message Channel and greater than or equal to the end of the Traffic Message Channel. In other words, the start and end of each Traffic Message Channel is direction dependent. Figure 4-3 provides an example of a 0.97-mile-long Traffic Message Channel in the TE section. The date, day, and reported time of incidents that were successfully matched with the Traffic Message Channels were extracted and used in the next steps.

Figure 4-3: An Example of Assigning Incidents to Traffic Message Channels

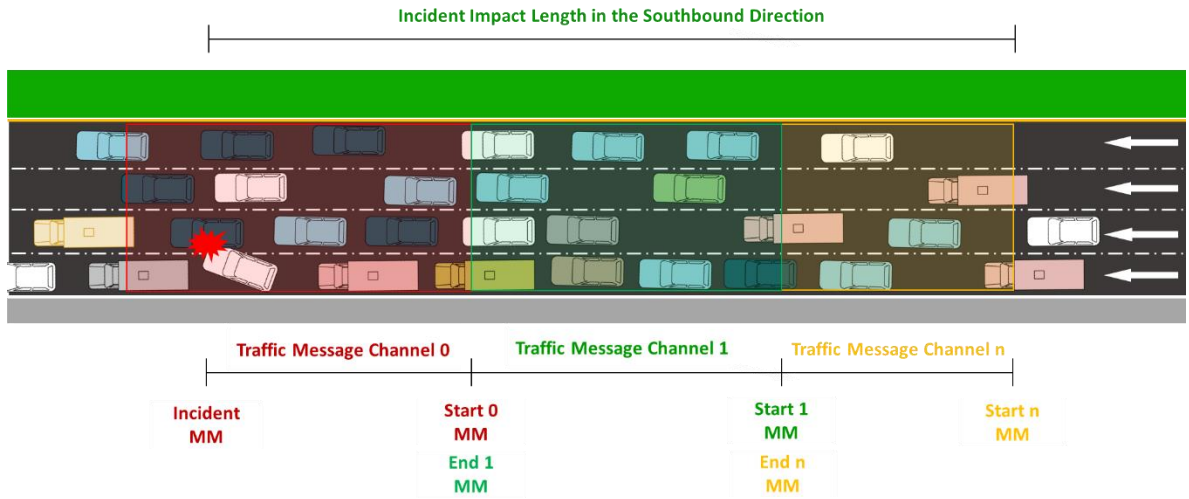
4.2.3 Estimate Incident Impact Area

Traffic incidents and real-time traffic data were required to estimate the incident impact area. The impact area was computed for incidents that were successfully matched to the Traffic Message Channels, as discussed in an earlier section. This process was achieved by tracking the reported speeds at the segment where the incident occurred, from the time the incident was detected to the time when the traffic flow returned to normal. An incident was considered to have affected the traffic flow characteristics of the segment when the average speed along the segment was less than the lower boundary of the speed profile. The same procedure was repeated for all the upstream Traffic Message Channels affected by the incident. Next, the time taken for the traffic to return to normal, following the occurrence of an incident, was recorded for each affected Traffic Message Channel. Since the incident impact duration along different Traffic Message Channels may differ, the incident impact area was defined for each Traffic Message Channel individually.

In summary, this process enabled the accurate estimation of the spatiotemporal impact area of the incident. That is, for each impacted Traffic Message Channel, the temporal thresholds were defined by the incident impact duration, i.e., from the time the incident was first detected to the time traffic returned to normal. As indicated in Figure 4-4 (a), the incident impact length in the northbound direction is defined by the difference in distance between the location of the incident and the start of the last impacted Traffic Message Channel (n). Similarly, the incident impact length in the southbound direction refers to the difference in distance between the location of the incident and the end of the last impacted Traffic Message Channel (n).



(a) Definition of Incident Impact Length in the Northbound Direction



(b) Definition of Incident Impact Length in the Southbound Direction

Figure 4-4: Definition of Incident Impact Length

Figure 4-5 shows an example of the impact area caused by an incident **I-1**, where the x- and y-axes represent the time and length of the affected roadway segments, respectively. Note that each cell in Figure 4-5 represents a speed measurement by the Traffic Message Channel at the t^{th} time interval, i.e., 5 minutes in this case. As indicated in Figure 4-5, the impact duration and impact length vary across the different Traffic Message Channels impacted by the incident. While the segment where the incident occurred, i.e., Traffic Message Channel 0, has the most extended impact duration, the farthest segment impacted by incident **I-1**, i.e., Traffic Message Channel 6, has the shortest impact duration.

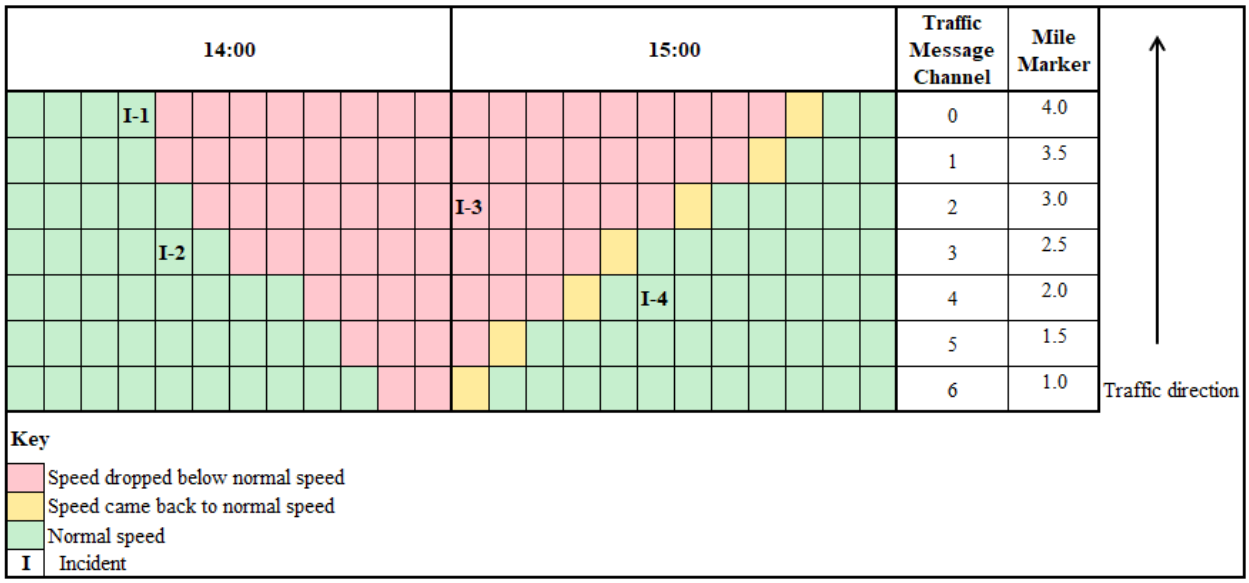


Figure 4-5: Illustration of the Approach to Estimate Incident Impact Area

4.2.4 Identify Secondary Crashes

Following the establishment of the area impacted by each incident, the last step was to identify secondary crashes. A traffic incident was considered a secondary crash if it occurred within the prior incident's spatiotemporal impact area. Referring to Figure 4-5, since incident **I-1** occurred earlier than incidents **I-2, I-3**, and **I-4**, the main task was to determine whether these three incidents occurred because of incident **I-1**. Considering the impact area in Figure 4-5, incident **I-3** was considered a secondary crash to incident **I-1** since it occurred within the impact area of incident **I-1**.

4.3 Results and Discussion

Along the study corridor, the SunGuide® database included a total of 622,264 incidents from January 2014 – June 2019. After excluding incidents with missing information, the remaining data consisted of a total of 322,259 incidents. Table 4-1 provides more information about incidents used to identify secondary crashes. Of these 322,259 incidents, 116,521 incidents occurred along TE, 95,583 incidents occurred on the South Section of the Mainline (STM corridor), and the remaining 110,155 incidents occurred on the Northern Section of the Mainline (NTM corridor).

Incidents that occurred on ramps were also not included in the analysis. Compared to the mainline segments, ramps have a complex geometry that significantly affect the traffic transition states, i.e., from free-flow to breakdown, congested, recovery, and eventually back to free-flow. For this reason, incidents that occur on ramps cannot be combined with the incidents that occurred on the freeway mainline (Sando et al., 2019).

Table 4-1: Incidents Used to Identify Secondary Crashes

Criteria	Count
Total incidents from Jan 2014-June 2019	622,264
Missing Information from SunGuide®	
Incidents with missing coordinates	7,016
Incidents that were plotted outside the study corridor	144
Incidents that occurred on ramps	2,522
Incidents with "other" event types	42,240
Missing Information from HERE Technologies	
Incidents with no matched Traffic Message Channels data	5,276
Incidents along sections with overly long Traffic Message Channels	221,652
Incidents with extreme incident impact duration and incident impact length	20,990
Incidents without HERE speed data	165
Total incidents excluded from the analysis	300,005
Total incidents included in the identification of secondary crashes	**322,259**

As indicated in Table 4-1, to identify secondary crashes, 322,259 incidents from the SunGuide® database and high-resolution speed data from HERE Technologies were evaluated. Table 4-2 provides a summary of the secondary crashes identified along the study corridors. As indicated in the table, a total of 4,549 secondary crashes were identified from 3,977 primary incidents. This is an equivalent of 5.7 secondary crashes per mile per year along the 148-mile study corridor. In other words, about six secondary crashes per mile occurred annually along the study corridors.

The identified secondary crashes account for 1.4% of all traffic incidents. While the proportion of secondary crashes, when compared to all incidents, may not seem alarming at first glance, secondary crashes account for 12.2% of all crashes included in the analysis. As indicated in Table 4-2, the highest proportion of secondary crashes was identified along the TE corridor, followed by the STM, and finally, the NTM.

Table 4-2: Secondary Crashes Identified along the Study Corridors

Seg.	Seg. Len. (miles)	NI	PI	SC	All Inc.	All Crash	SC/ mile/year	Prop. of SC/Inc. (%)	Prop. of SCs/Crash (%)
TE	48	111,274	2,516	2,964	116,521	19,369	11.2	2.5	15.3
STM	28	93,709	932	1,008	95,583	9,020	6.5	1.1	11.2
NTM	69	109,090	529	577	110,155	8,818	1.5	0.5	6.5
Overall	145	314,073	3,977	4,549	322,259	37,207	5.7	1.4	12.2

Note: TE = Turnpike Extension; STM = Southern Turnpike Mainline; Northern Turnpike Mainline (NTM); Seg. Len. = Segment Length; NI = Normal Incident; PI = Primary Incident; SC = Secondary Crash; Inc. = Incident.

4.3.1 Spatiotemporal Distribution of Secondary Crashes

Figures 4-6 and 4-7 show the spatial and temporal characteristics of secondary crashes in relation to primary incidents. The median distance between primary incidents and secondary crashes was found to be 2.5 miles. About half of all secondary crashes occurred within 40 minutes after the

primary incident. Almost half of all secondary crashes (47%) occurred within 2 miles upstream of the primary incident. Meanwhile, more than three-quarters of secondary crashes (93%) occurred within two hours. Overall, 40% of secondary crashes occurred within two hours of the onset of a primary incident and two miles upstream of the primary incident, the most commonly considered static spatiotemporal threshold.

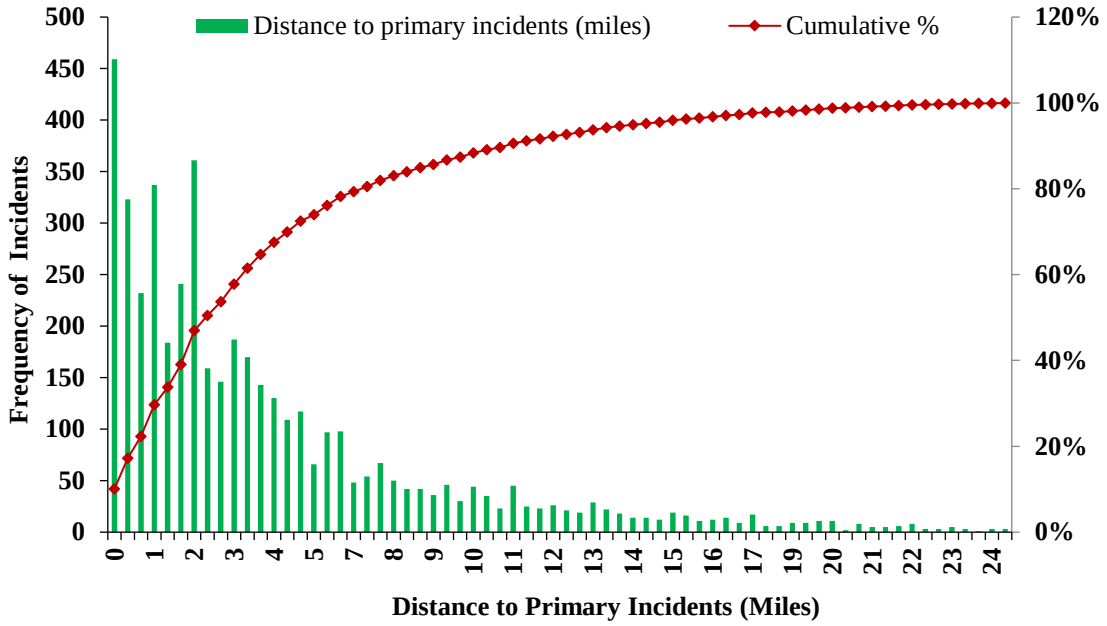


Figure 4-6: Spatial Distribution of Secondary Crashes in Relation to Primary Incidents

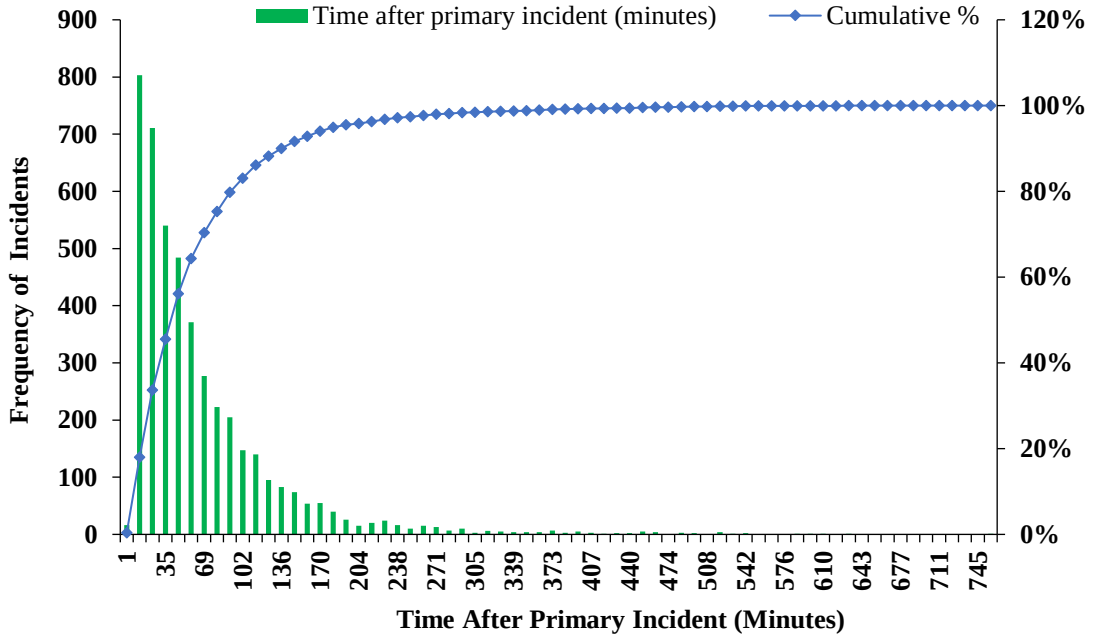


Figure 4-7: Temporal Distribution of Secondary Crashes in Relation to Primary Incidents

4.3.2 Time of Day and Day of Week Distribution

Figure 4-8 shows the distribution of the 4,549 secondary crashes, 3,977 primary incidents, and 314,073 normal incidents by different periods. More than three-quarters of secondary crashes (85%) occurred during peak hours, i.e., morning peak, 6:00 AM to 10:00 AM, and evening peak, 3:00 PM to 8:00 PM. Specifically, 33% of secondary crashes occurred during the morning peak, while the remaining 52% occurred during the evening peak. The highest proportion of secondary crashes during morning peak hours occurred from 8:00 AM to 9:00 AM (11%), while the highest proportion of secondary crashes during the evening peak period (13%) occurred between 5:00 PM and 6:00 PM, summing to a total of 24% of all secondary crashes that occurred along the study corridors.

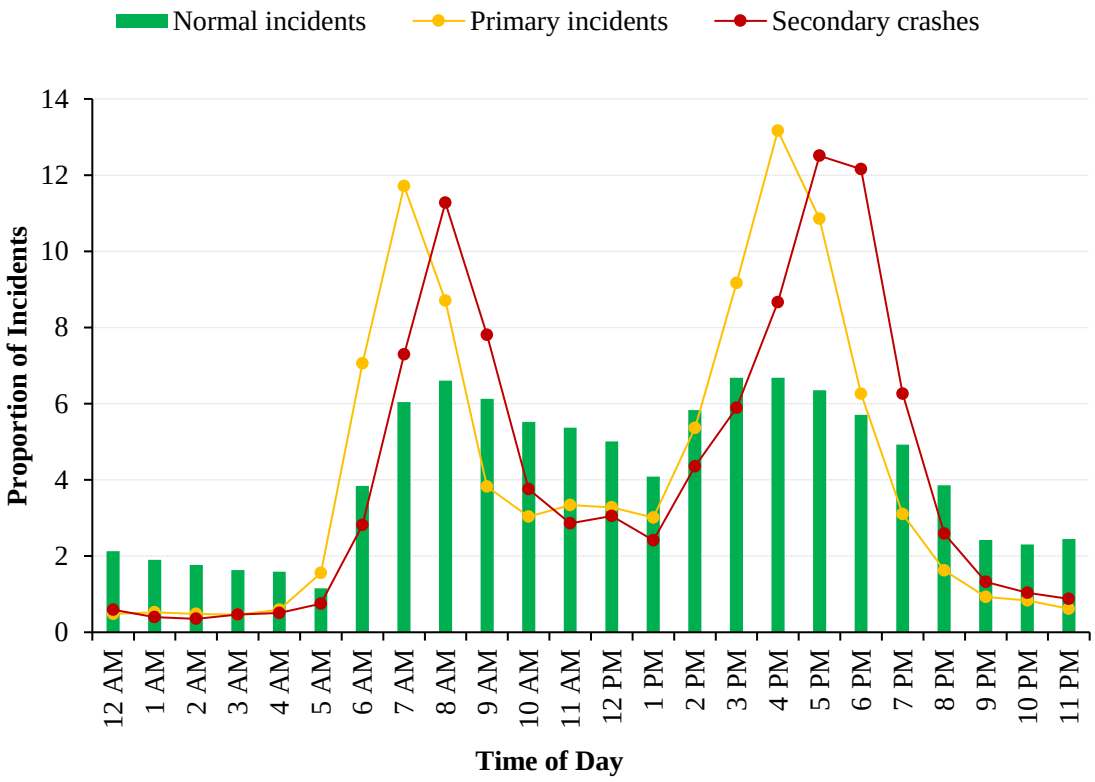


Figure 4-8: Distribution of Traffic Incidents by Time of Day

The highest proportion of primary incidents was observed during the evening peak period between the hours of 2:00 PM and 8:00 PM, accounting for 50% of all primary incidents. As can be inferred from Figure 4-8, the peaks of primary incidents and secondary crashes are one hour apart. Unlike primary incidents and secondary crashes, there is no significant distinction in the distribution of normal incidents during peak hours. More than three-quarters of normal incidents (94%) occurred between the hours of 6:00 AM and 8:00 PM. As can be observed from Table 4-3, approximately half of normal incidents occurred during peak hours (53%), while the remaining half occurred during off-peak hours.

Table 4-3: Distribution of Traffic Incidents by Time of Day

Temporal Characteristic	Category	Incident Category (%)		
		Normal Incidents	Primary Incidents	Secondary Crashes
Time of Day	Peak Hours	68	84	85
	Off-peak Hours	32	16	15

More than three-quarters of both primary incidents (84%) and secondary crashes (85%) occurred during peak hours. Compared to off-peak hours, peak-hour traffic flow characteristics were found to contribute more to the occurrence of secondary crashes. This is expected because smaller gaps between vehicles characterize congested traffic, providing less maneuvering room for drivers to avoid a crash (Mishra et al., 2016; Kitali et al., 2019b).

Figure 4-9 presents the distribution of incidents by day of the week. It can be inferred from the figure that the proportion of normal incidents, primary incidents, and secondary crashes is much higher on weekdays than on weekends. Compared to other days of the week, Friday was found to experience the highest proportion of secondary crashes (20%). Only 13% of secondary crashes occurred on weekends.

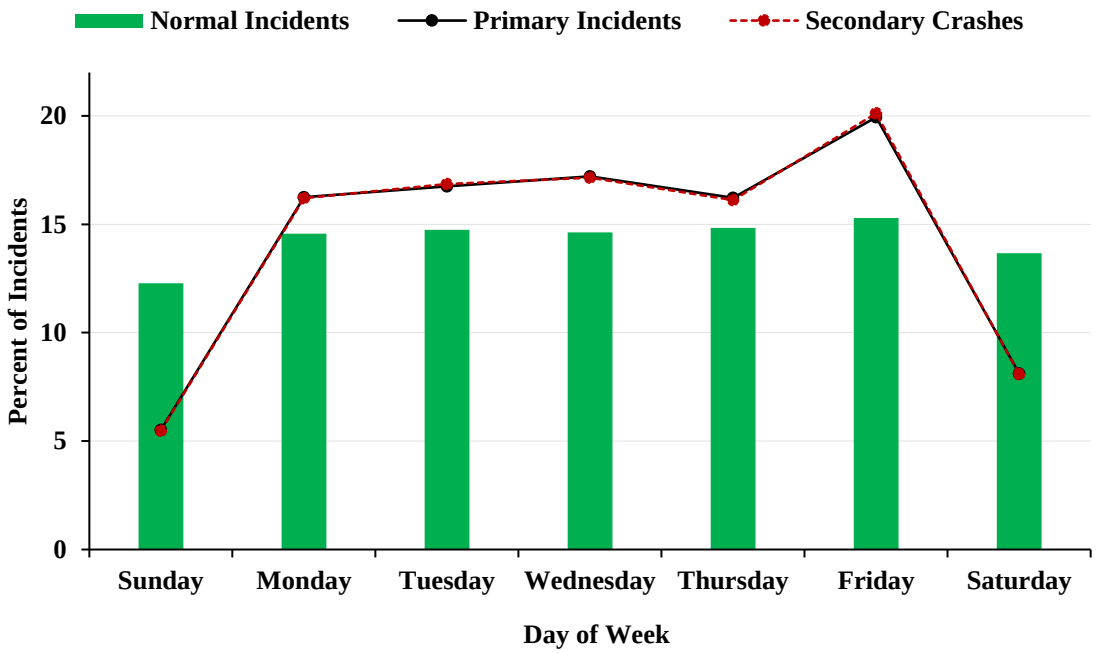


Figure 4-9: Distribution of Traffic Incidents by Day of Week

4.3.3 Incident Characteristics

Figure 4-10 provides the distribution of the incident clearance duration for towing-involved and no-towing-involved incidents. From Figure 4-10, it can be inferred that 94% of traffic incidents that did not involve towing were cleared within 90 minutes, while only 64% of traffic incidents that involved towing were cleared within 90 minutes, a value adopted from the FDOT's Open Road Policy.

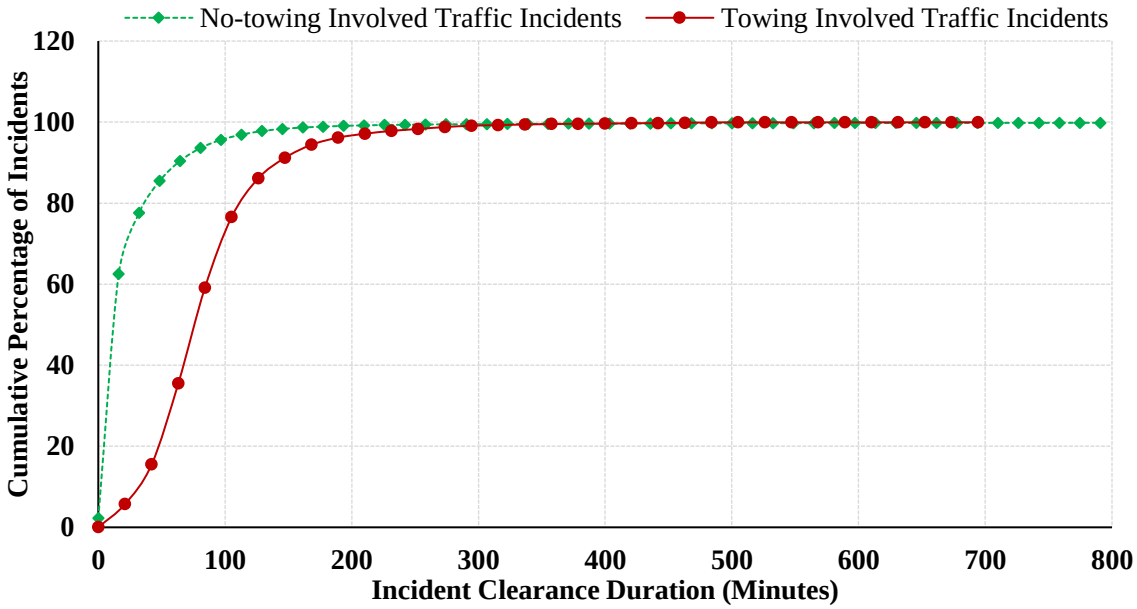


Figure 4-10: Distribution of Incident Clearance Duration for Towing Involved and No Towing Involved Incidents

In addition to towing, the EMS presence at the incident scene was also identified as one of the factors that increase the incident clearance duration. This observation is evident in Figure 4-11, where 95% of traffic incidents that did not involve EMS were cleared within 90 minutes, while only 64% of traffic incidents that involved EMS were cleared within 90 minutes. As expected, traffic incidents involving towing and EMS resulted in longer incident clearance durations as they tend to require more time to be cleared.

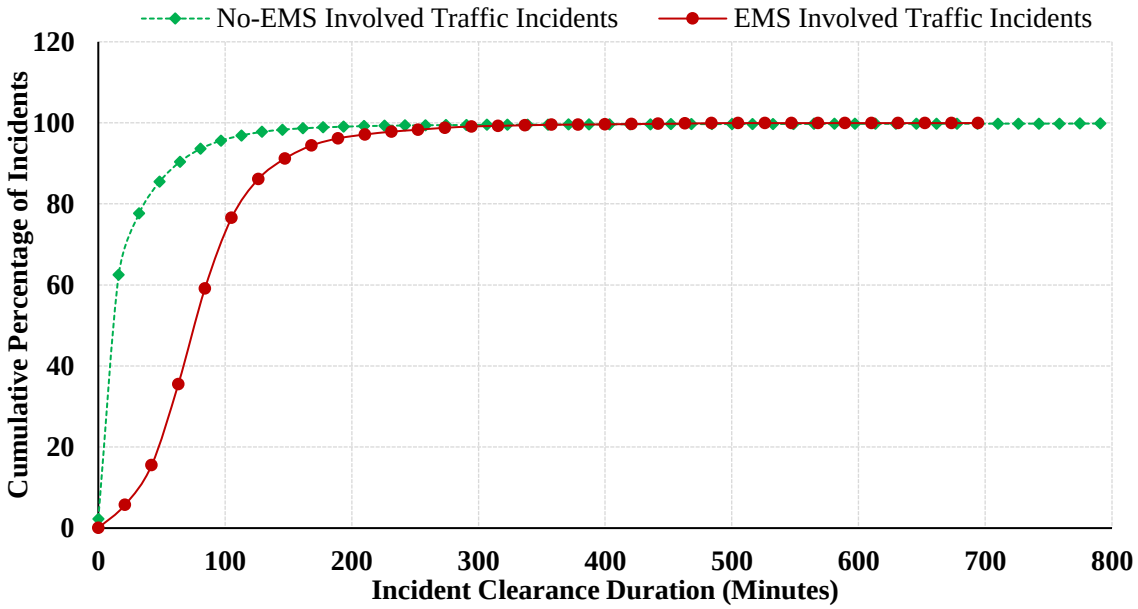


Figure 4-11: Distribution of Incident Clearance Duration for EMS Involved and No EMS Involved Incidents

As indicated in previous studies, the likelihood of secondary crashes increases with an increase in incident clearance duration (Xu et al., 2016; Kitali et al., 2018). This is evident from the data, as 13% of primary incidents required towing, while only 3% of normal incidents required towing (see Table 4-4). Similarly, a higher percentage of incidents involving EMS resulted in secondary crashes (11%). Furthermore, while only 28% of normal incidents involved more than one responding agency, 51% of primary incidents and 55% of secondary crashes involved multiple responding agencies. These statistics suggest that incidents involving a greater number of responding agencies increase the likelihood of secondary crashes.

Table 4-4: Incident Distribution Based on Responders' Characteristics

Incident Characteristics	Category	Incident Category (%)		
		Normal Incidents	Primary Incidents	Secondary Crashes
Towing Involved	No	97.0	86.6	85.2
	Yes	3.0	13.4	14.8
Emergency Involved	No	98.2	89.4	89.1
	Yes	1.8	10.6	10.9
Number of Responding Agencies	1	71.9	49.0	45.2
	2	24.3	31.7	33.8
	3	1.8	7.0	9.0
	4	0.9	4.7	5.3
	5	0.8	4.9	5.0
	6+	0.3	2.6	1.7

As can be observed from Table 4-5, 97% of normal incidents did not result in a lane closure, while 21% of primary incidents resulted in a lane closure. The percentage of lanes closed is an indicator of the severity of the primary incident, as severe incidents tend to result in an increased number of lanes closed. About 9% of primary incidents resulted in moderate to severe impacts on traffic, while only 1% of normal incidents were moderate to severe.

Table 4-5: Incident Characteristics

Incident Characteristics	Category	Incident Category (%)		
		Normal Incidents	Primary Incidents	Secondary Crashes
Percentage of Lanes Closed	0	97.0	79.3	99.7
	0-50	0.4	2.7	0.2
	50-100	2.6	18.0	0.1
Incident Severity*	Minor	98.9	90.6	93.4
	Moderate	0.7	5.9	4.9
	Severe	0.4	3.5	1.7

Note: *Incident severity refers to the extent to which the incident impacted the traffic.

As indicated in Figure 4-12, only 10% of normal incidents were crashes, a proportion similar to all incidents (12%), while approximately half of the primary incidents were crashes (47%). In other words, the probability of secondary crashes was found to be higher when primary incidents were crashes. Note that the category "Other" in Figure 4-12 includes emergency vehicles, vehicle fire, and police activity. All incidents include normal incidents, primary incidents, and secondary crashes.

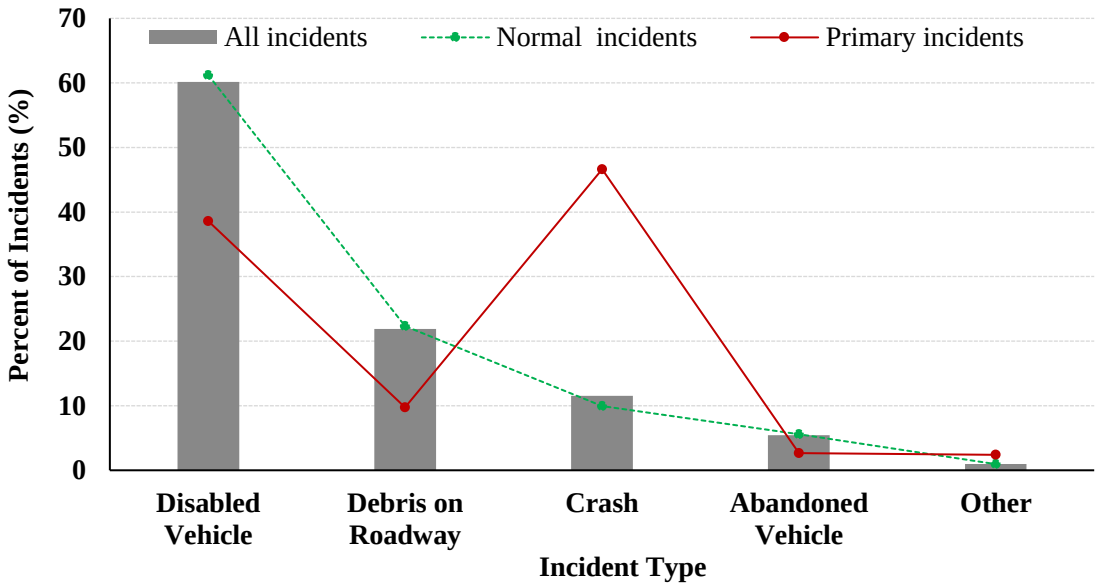


Figure 4-12: Distribution of Incidents by Incident Type

Figure 4-13 shows the distribution of the incident clearance duration for normal incidents and primary incidents. Overall, normal incidents were cleared more quickly than primary incidents; approximately 94% of the normal incidents were cleared within 90 minutes, while only 82% of the primary incidents were cleared within 90 minutes. The longer clearance time of the primary incidents could be considered one of the factors that may have contributed to the occurrence of secondary crashes.

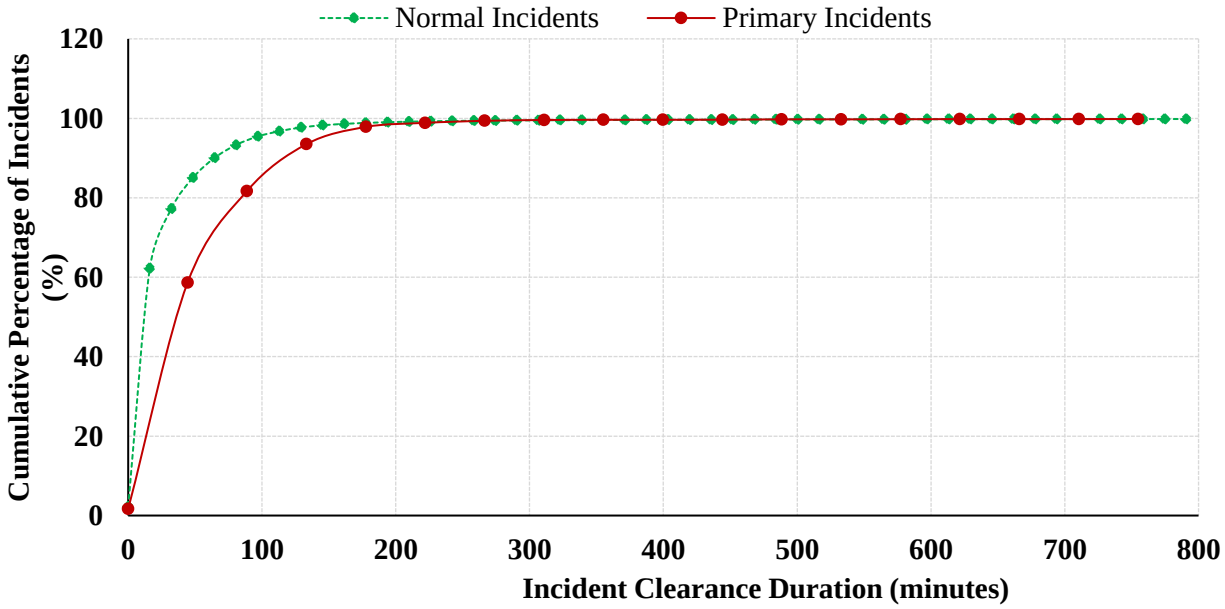


Figure 4-13: Distribution of Incident Clearance Duration for Normal and Primary Incidents

Figure 4-14 presents the distribution of the incident clearance duration for the identified primary incidents and secondary crashes. Approximately 77% of the secondary crashes were cleared within 90 minutes, while 82% of the primary incidents were cleared within 90 minutes. This observation implies that primary incidents were cleared slightly faster than secondary crashes.

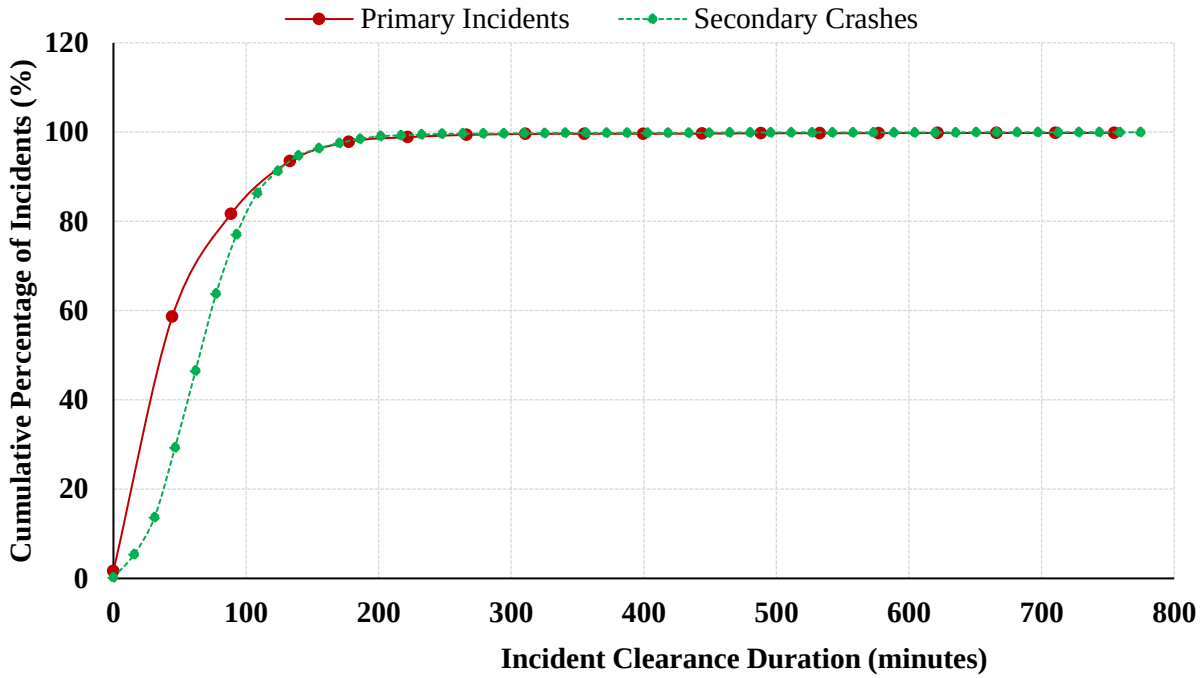


Figure 4-14: Distribution of Incident Clearance Duration for Primary Incidents and Secondary Crashes

4.3.4 Environmental Conditions

Environmental conditions (i.e., weather, roadway surface, and lighting) have been identified as some of the factors that influence the likelihood of secondary crashes (Vlahogianni et al., 2012). Table 4-6 summarizes the variation of weather condition, roadway surface condition, and lighting condition by incident category, i.e., normal incident, primary incident, and secondary crash.

Table 4-6: Environmental Conditions

Environmental Condition	Category	Incident Category (%)		
		Normal Incidents	Primary Incidents	Secondary Crashes
Weather	Clear	97.9	87.3	79.9
	Cloudy/Fog/Rain	2.1	12.7	20.1
Roadway Surface Condition	Dry	98.7	88.7	81.6
	Wet	1.3	11.3	18.4
Lighting Condition	Daylight	71.3	80.2	77.5
	Dark/Dusk/Down	28.7	19.8	22.5

Regarding weather condition, as indicated in Table 4-6, more than three-quarters of all the three incident categories occurred under clear weather condition. However, compared to normal

incidents (2%), a higher proportion of primary incidents (13%) occurred during cloudy/fog/rainy conditions. Similarly, a higher percentage of primary incidents (11%) and secondary crashes (18%) occurred on wet surface conditions. These statistics imply that inclement weather conditions and adverse road surface conditions are among the factors that increase the probability of secondary crashes. Compared to normal incidents, a higher proportion of primary incidents and secondary crashes occurred during the daytime.

4.4. Summary

Prevailing traffic data were used to estimate the impact area of individual incidents and identify secondary crashes within the affected area. The developed approach considered how the queue caused by the primary incident grows and dissipates upstream of the incident. Traffic incidents from the SunGuide® database and high-resolution speed data from HERE Technologies were used to estimate the impact area of a primary incident. These data were collected from January 2014 to June 2019. The study area, which is located in Florida, included a 97-mile section of the Florida's Turnpike Mainline, and TE, a 48-mile adjoining corridor. The Mainline study corridor consisted of a 69-mile NTM and a 28-mile STM.

The analysis was based on 322,259 traffic incidents that occurred along the study corridors between January 2014 and June 2019. Overall, 4,549 secondary crashes in the upstream direction of the primary incident were identified from 3,977 primary incidents. The identified secondary crashes on the upstream direction of the primary incident accounted for 1.4% of the 322,259 incidents. This is equivalent to 5.7 secondary crashes per mile per year.

Descriptive statistics of the secondary crashes indicated that 93% of the secondary crashes occurred within two hours after the occurrence of the primary incidents. Spatially, 47% of the secondary crashes occurred within two miles from the primary incident. Overall, 40% of secondary crashes occurred within two hours of the onset of a primary incident and within two miles upstream of the primary incident, the most commonly considered spatiotemporal threshold. The following are some of the key characteristics of the primary incidents and secondary crashes:

- Only 3% of secondary crashes occurred between midnight and 5:00 AM, whereas 85% occurred during peak hours. Specifically, 33% of secondary crashes occurred during the morning peak (i.e., 6:00 AM - 10:00 AM), while the remaining 52% occurred during the evening peak (i.e., 2:00 PM - 8:00 PM). The highest proportion of primary incidents (13%) occurred between 4:00 PM and 5:00 PM, while the highest proportion of secondary crashes (13%) occurred an hour after the primary incident, i.e., between 5:00 PM and 6:00 PM.

- The proportion of normal incidents and secondary crashes was much higher on weekdays than on weekends. Compared to other days of the week, Friday experienced the highest proportion of secondary crashes (20%).

- While secondary crashes were found to occur on Mondays and Fridays, normal incidents were found to occur primarily on weekdays (i.e., Monday through Friday). Only 13% of secondary crashes were found to occur on weekends.

- As expected, traffic incidents involving towing and/or EMS resulted in longer incident clearance durations, as they tend to require more time to be cleared. Approximately 94% of normal incidents were cleared within 90 minutes, while 82% of primary incidents were cleared within 90 minutes. Likewise, 94% of traffic incidents that did not involve EMS were cleared within 90 minutes, while only 64% of traffic incidents that involved EMS were cleared within 90 minutes. The longer clearance time of the primary incidents could be considered to have contributed to the occurrence of secondary crashes.

- The severity of primary incidents was found to be one of the factors that influence the occurrence of secondary crashes. About 9% of primary incidents were moderate/severe, while only 1% of normal incidents were moderate/severe. Besides the severity of primary incidents, the number of responding agencies, percentage of lanes closed, incident type, and incidents that required towing and/or EMS were also considered to be good indicators of incident severity. About 99% of normal incidents did not result in lane closure, while 21% of primary incidents resulted in a lane closure. Only 10% of normal incidents were identified as crashes, while 47% of primary incidents were crashes. About 13% of primary incidents required towing, while only 3% of normal incidents required towing. Similarly, a higher percentage of incidents involving EMS resulted in secondary crashes (11%). While only 28% of normal incidents involved more than one responding agency, 51% of primary incidents and 55% of secondary crashes involved more than one responding agency. These statistics indicate that the severity of primary incidents influences the occurrence of secondary crashes.

- Compared to normal incidents (2%), a higher proportion of primary incidents (13%) occurred during cloudy/foggy/rainy conditions. Similarly, a higher percentage of primary incidents (11%) and secondary crashes (18%) occurred on wet surface conditions. These statistics imply that inclement weather conditions and adverse road surface conditions are among the factors that increase the probability of secondary crashes.

CHAPTER 5
ANALYZE FACTORS INFLUENCING OCCURRENCE OF SECONDARY CRASHES

Not all incidents lead to secondary crashes. The likelihood of secondary crashes depends on several factors, including characteristics of incidents, weather conditions, geometric conditions, traffic flow characteristics, etc. An in-depth understanding of these factors will help agencies on several fronts. It will assist in proactively preventing secondary crashes. First responders will be more vigilant and prepared in case secondary crashes occur. And finally, motorists upstream of the primary incident could be warned about potential secondary crashes.

This chapter discusses the methodology used to identify factors influencing the occurrence of secondary crashes. Variables from four data types (incidents, traffic flow characteristics, weather conditions, and geometric characteristics) were considered as potential factors that influence the likelihood of secondary crashes. Incident data were obtained from the SunGuide® database. Real-time traffic data were extracted from the HERE Technologies. Weather conditions were obtained from NEXRAD Level-II network, and roadway geometric data were retrieved from the FDOT's RCI, Google Earth Pro, and Google Maps.

5.1 Study Area

As indicated in Section 3.2.2, the selected study corridor for identifying factors associated with the occurrence of secondary crashes is the 28-mile STM section. Figure 5-1 shows the location of this 28-mile study corridor. This section was selected based on the availability of HERE data which was earlier used to identify secondary crashes. Besides, incident hotspots and major construction activities, i.e., lane widening, bridge maintenance, interchange improvements, etc., were used as additional criteria to select the study corridor.

Both the incident hotpot analysis and crash hotspot analysis identified the TE and STM as corridors that experienced the highest number of traffic incidents and crashes. Nonetheless, the exploratory analysis of the data in the Active Construction Projects shapefile indicated that lane widening construction activities were taking place within the TE section during the study period. Meanwhile, there were no such activities during the study period in the STM corridor. Thus, this 28-mile section in the STM corridor was considered for predicting the likelihood of secondary crashes.

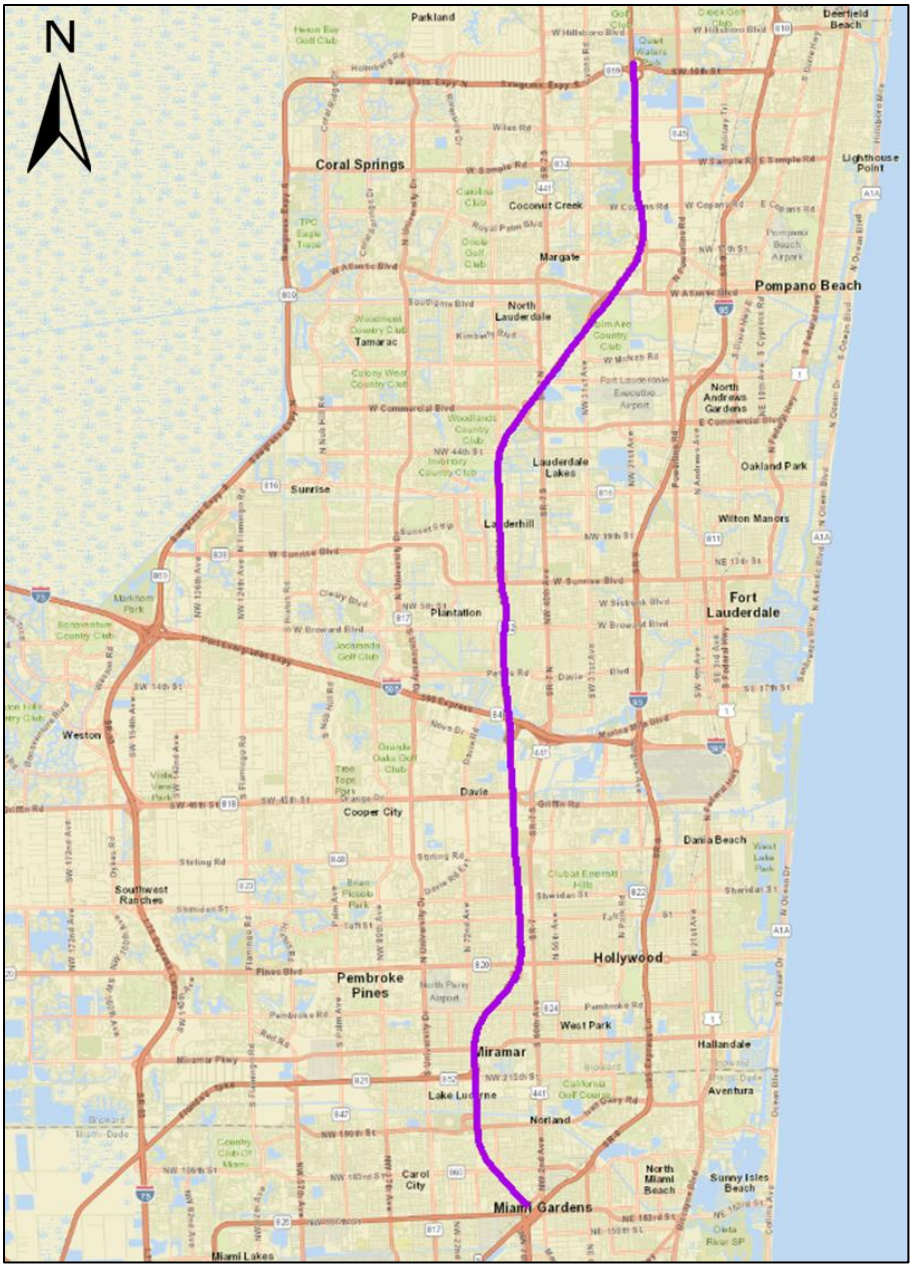


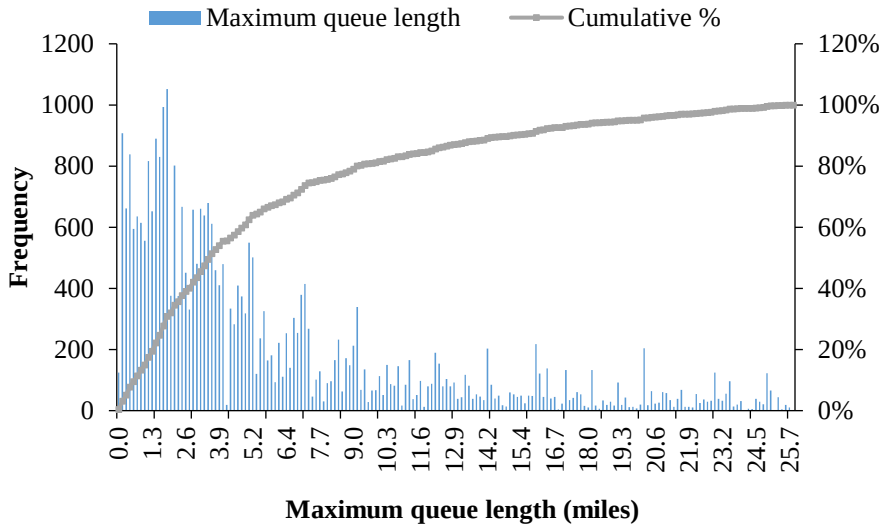
Figure 5-1: Study Corridor for Identifying Factors Associated with the Likelihood of Secondary Crashes

5.2 Data

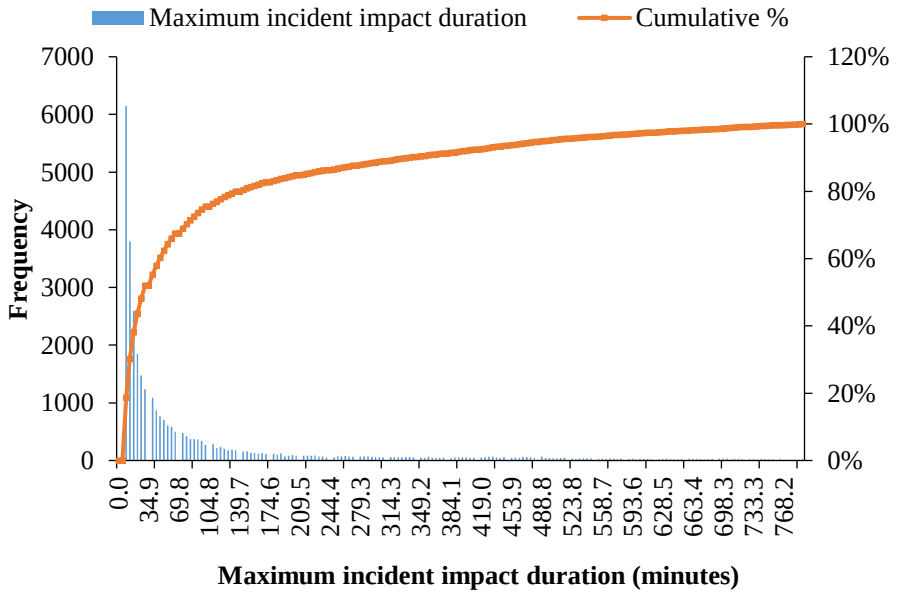
The following data were used to develop the secondary crash likelihood prediction model:

- Incident data from SunGuide®,
- Traffic flow attributes from HERE Technologies, and
- Roadway geometric attributes from RCI database, Google Earth Pro software, and Google Maps.

About 65% of the incidents that occurred along the study corridor during the study period (January 2014 – June 2019) did not impact traffic. Figures 5-2(a) and 5-2(b) present the distribution of the maximum queue length and the maximum incident impact duration, respectively, for the incidents that impacted traffic, which is about 35% of the total incidents included in the likelihood model.



(a) Maximum Queue Length



(b) Maximum Incident Impact Duration

Figure 5-2: Spatiotemporal Distribution of Incident Impact Area

From these figures, it can be inferred that 78% of incidents had a maximum impact duration of 2 hours. About 32% of incidents had a maximum queue length of 2 miles. Overall, 10% of incidents had a maximum impact area within 2 miles and 2 hours, the most commonly considered static spatiotemporal threshold.

Of the 95,583 incidents included in the likelihood model analysis, 7,319 incidents were further dropped from the analysis. As described in Table 5-1, of these 7,319, a total of 6,377 incidents were missing information on some of the attributes. The remaining 942 incidents were marked as non-tertiary crashes, meaning that these are secondary crashes that did not result in additional secondary crashes; in other words, they are neither normal incidents nor primary incidents. In total, 1,008 incidents (66 being tertiary secondary crashes) were identified as secondary crashes accounting for 1.1% of the 95,583 incidents included in the analysis. The 1,008 secondary crashes were caused by 932 primary incidents indicating that 76 primary incidents resulted in more than one secondary crash. Thus, a total of 88,340 incidents (calculated as: (95,583 - 7,319) + 76) were included in developing the likelihood model.

Table 5-1: Incidents used to Model the Likelihood of Secondary Crashes

Criteria	Count
Incidents used to identify secondary crashes	95,583
Unique secondary crashes	942
Duplicated primary incidents	76
Incidents to be used in the likelihood model	94,717
Missing information from HERE Technologies	
Incidents without information on speed before the incident	844
Missing information from SunGuide®	
Incidents with "NULL" responders	264
Incidents with "Unknown" severity	5,268
Incidents with "NULL/turn lane closure/blocked" worst lane blockage	5
Total incidents excluded from the likelihood model	6,381
Total incidents included in the secondary crash likelihood model	**88,340**

Speed data (i.e., mean and standard deviation) in the Traffic Message Channel where the incident occurred and within 10 minutes before the occurrence of the incident were collected to capture the traffic conditions before the occurrence of the incident. Besides, to determine the prevailing traffic conditions, speed data within the Traffic Message Channels impacted by the incident from the time the incident was detected to the time when the traffic flow returned to normal were used. Since the incident impact duration along different Traffic Message Channels may be different, the incident impact area was defined for each Traffic Message Channel individually. The data preparation efforts for the roadway geometric characteristics are covered in detail in Section 3.1.3 of this report.

5.3 Methodology

Secondary crash risk models determine the probability that a secondary crash will occur given a prior incident. From a statistical learning point of view, secondary crash risk modeling can be viewed as a binary classification problem. Suppose that the incident dataset has n observations $(\boldsymbol{X}^i, y_i), i\epsilon 1,2,\dots,n$ with p explanatory variables $\boldsymbol{X}^i = x_{i1}, x_{i2}, \dots, x_{ip} = x_i^T$. Let $y = (y_1, \dots, y_n)^T$ be the response variable, which is binary in nature, i.e., y_i represents the secondary crash indicator

(1 indicates a secondary crash is caused by a primary incident (i), and 0 indicates that no secondary crash occurred).

Researchers have used several methods to identify factors influencing the risk of a secondary crash. Of all the previously adopted methods, logistic regression has an exceptional advantage since it provides a direct estimate of class probability and does not require a tuning parameter. As shown in Equation 5-1, the logistic regression model presents the class-conditional probabilities through a linear function of the predictors.

$$log\frac{\Pr(y_i=1|x_i)}{\Pr(y_i=0|x_i)} = \beta_0 + x_i^T\beta \tag{5-1}$$

where $\beta = (\beta_1, \dots, \beta_p)^T$ is the vector of coefficients of the p predictors to be estimated excluding the intercept β_0 and $\Pr(y_i = 1|x_i)$, $\Pr(y_i = 0|x_i)$ denote the conditional probabilities of the class labels 1 and 0, respectively. A maximum likelihood approach is commonly used in calculating the coefficients and the log-likelihood can be written as shown in Equation 5-2.

$$\begin{aligned} l(\beta_0,\beta) &= \sum_{i=1}^{N}\{y_i logPr(Y=1;\beta) + (1-y_i)log(1 - logPr(Y=1;\beta))\} \\ &= \sum_{i=1}^{N}\left\{y_i(\beta_0 + x_i^T\beta) - log\left(1 + e^{(\beta_0 + x_i^T\beta)}\right)\right\} \end{aligned} \tag{5-2}$$

Instead of using a conventional logistic regression, a LASSO penalized logistic regression fitted using the bootstrap resampling approach was used to model the likelihood of secondary crashes. The proposed method aimed at addressing the three major challenges encountered while modeling the secondary crash risk, i.e., variable correlation, variable selection, and imbalanced nature of the response variable. Specifically, the present study used the LASSO penalized estimator to extract the most important explanatory variables, with minimal correlation, influencing the risk of secondary crashes. Since the proportion of primary incidents is smaller than the proportion of normal incidents, the bootstrap resampling method was used to fit the penalized logistic regression. The next subsections describe in detail the penalized logistic regression and the bootstrap resampling approach.

5.3.1 Penalized Logistic Regression

LASSO penalized logistic regression is a regression analysis method that performs both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model (Tibshirani, 1996). The LASSO penalized estimator shrinks some coefficients of a regression model and set others to zero to obtain variables with a substantial effect on the outcome (Tibshirani, 1996). LASSO performs important variable selection and variable correlation simultaneously. That is, between a pair of highly correlated variables, LASSO tends to pick the most important variable and discard the other by shrinking it toward zero.

Because the LASSO method performs variable selection through a continuous process, it does not suffer as much from high variability, i.e., it simultaneously does both continuous shrinkage and automatic variable selection. The penalty term introduced by LASSO during the variable selection process ensures better estimation of the prediction error while avoiding overfitting. Selecting an optimal subset of explanatory variables is expected to improve the classification accuracy and

make the model's interpretation easier. Since some of the variables will be shrunk to zero, the model parsimony is also achieved.

The logistic regression model in Equation 5-1 can further be extended into the LASSO logistic regression model by adding the L_1 constraint on β parameters (Equation 5-2). The L_1 constraint is added to minimize the negative log-likelihood function with the penalty term. The generated coefficients can be expressed as a sparse linear combination of p number of predictor variables when solving the following optimization problem (Equation 5-3).

$$\underset{(\beta_0,\beta)}{min}\left\{\sum_{i=1}^{N}-\frac{1}{n}\left[y_i(\beta_0+x_i^T\beta)-log\left(1+e^{(\beta_0+x_i^T\beta)}\right)\right]+P_\lambda(\beta)\right\} \quad (5\text{-}3)$$

where $P_\lambda(\beta)$ is the penalty term that depends on λ, a vector of non-negative regularization parameters, commonly referred to as a tuning parameter. The tuning parameter λ controls the strength of shrinkage in the explanatory variables, i.e., when λ takes larger values, more weight will be given to penalty term and vice versa (Tibshirani, 1996). In this way, both shrinkage and variable selection are done simultaneously and it is also this property that makes LASSO generally easier to interpret. Depending on the property of the LASSO penalty, some coefficients in β will be exactly equal to zero. Further, it is also because of the penalty term λ that a LASSO model can include any number of variables.

While there are numerous penalty terms, a good penalty produces an estimator that is not biased or over-penalize large parameters (Algamal and Lee, 2015a). The adaptive LASSO penalty was thus selected because it applies adaptive weights when penalizing parameters (Zou, 2006). The adaptive LASSO imposes higher weight to the small coefficients and lower weight to the large coefficients to reduce the selection bias and fit the model consistently (Algamal and Lee, 2015b). This approach is thus said to have an oracle property, and it is the main advantage of adaptive LASSO as compared to other penalty terms such as conventional LASSO, ridge penalty, and elastic net (Algamal and Lee, 2015a). The estimation of the vector β_j is thus obtained by minimizing Equation 5-4 where w_j is a vector of data-driven weights. Although various methods have been used to estimate the weights (e.g., LASSO estimates), a ridge regression was used in this case to estimate initial weights (SAS Institute Inc., 2019) because of the limitations of LASSO as pointed by Algamal and Lee (2015b).

$$\hat{\beta}=\arg\min\beta\left[-L(\beta|Y)+\lambda\sum_{j=1}^{p}w_j|\beta_j|\right] \quad (5\text{-}4)$$

5.3.2 Bootstrap Resampling

The bootstrap resampling method was used to estimate the logistic regression parameters to resolve the data imbalance caused by a disproportionally high proportion of normal incidents than primary incidents. The bootstrap resampling involves estimating parameters by repeatedly and randomly sampling subsets of data hence providing more accurate estimates (Hastie et al., 2009; Kassambara, 2017; Pei et al., 2016). The conventional bootstrapping approach involves drawing a sample randomly and evenly with replacement. The resampling focused on neutralizing the effect of a significantly low percentage of primary incidents.

A three-step resampling approach was thus applied to the dataset. First, the incident data were divided into two groups: normal incidents, and primary incidents. Then, *k* samples (where *k* equals the number of primary incidents) were randomly drawn from all groups in each bootstrap replication. The resulting subset of data contained an equal number of normal incidents and primary incidents. The new dataset was then used to fit the penalized logistic regression. Finally, the procedure of drawing samples of *k* observations and fitting the model was repeated 5,000 times (arbitrarily selected as a trade-off between prediction accuracy and computation time), and the standard errors and confidence intervals of the estimates were calculated based on these 5,000 estimates.

The model coefficients were obtained by calculating the mean of all the estimates of the bootstrap samples. The odds ratio (OR), which represents how the dependent variable varies with the predictor variable, is computed relative to the base category. The odds ratio was calculated as shown in Equation 5-5.

$$Odds\ Ratio = e^{coefficient} \tag{5-5}$$

5.4 Results and Discussion

The following subsections discuss the descriptive statistics of the explanatory variables (i.e., traffic flow, incident attributes, temporal, weather, and roadway geometric attributes) and present the results of the fitted secondary crash likelihood model.

5.4.1 Descriptive Statistics

Tables 5-2 and 5-3 summarize the data used in this analysis. The data variables were categorized into incident, temporal, weather, traffic flow, and roadway geometric characteristics. From Table 5-2, it can be inferred that a majority of normal incidents were vehicle related. The proportion of primary incidents that are crashes (37%) was about four times that of normal incidents that are crashes (8%). The majority of normal incidents were responded to by one agency (72%). In comparison, the proportion of primary incidents attended by one and more than one agency was almost equal, 52% and 48%, respectively. A higher proportion of primary incidents (13%) had EMS as one of the responding agencies compared to normal incidents, which had only 2% of incidents with EMS involvement. A similar observation can be made on incidents where towing was involved and with moderate or severe severity.

While an equal proportion of normal incidents occurred during peak and off-peak hours, more than two thirds (72%) of primary incidents occurred during peak hours. Compared to normal incidents, a higher proportion of primary incidents occurred during adverse weather conditions and on wet road surfaces.

Table 5-2: Descriptive Statistics of Categorical Variables

Attribute	Attribute Category	Secondary Crash Likelihood				Total
		No (Normal Incident)		Yes (Primary Incident)		
		Count	Percent	Count	Percent	
Incident Attributes						
Incident type	Vehicle problem	62,309	71	457	49	62,766
	Hazard	17,704	20	125	13	17,829
	Crash	7,392	8	349	37	7,741
Number of responding agencies	1	63,158	72	482	52	63,640
	2+	24,247	28	449	48	24,696
EMS involvement	No	85,944	98	813	87	86,757
	Yes	1,461	2	118	13	1,579
Towing involvement	No	85,257	98	805	86	86,062
	Yes	2,148	2	126	14	2,274
Lane closure	No	85,048	97	745	80	85,793
	Yes	2,357	3	186	20	2,543
Incident severity	Minor	86,574	99	831	89	87,405
	Moderate/severe	831	1	100	11	931
Temporal Attributes						
Day of week	Weekday	65,901	75	811	87	66,712
	Weekend	21,504	25	120	13	21,624
Time of day	Off-peak	46,505	53	263	28	46,768
	Morning peak	17,507	20	302	32	17,809
	Evening peak	23,393	27	366	39	23,759
Weather Attributes						
Weather condition	Clear	86,303	99	843	91	87,146
	Adverse	1,102	1	88	9	1,190
Road surface condition	Dry	86,575	99	851	91	87,426
	Wet	830	1	80	9	910
Roadway Geometric Attributes						
Presence of horizontal curve within IIA	No	35,628	41	239	26	35,867
	Yes	51,777	59	692	74	52,469
Presence of vertical curve within IIA	No	43,301	50	276	30	43,577
	Yes	44,104	50	655	70	44,759
Presence of diverge influence area within IIA	No	46,958	54	294	32	47,252
	Yes	40,447	46	637	68	41,084
Presence of merge influence area within IIA	No	26,351	30	264	28	26,615
	Yes	61,054	70	667	72	61,721
Response variable	Secondary crash likelihood	87,405	99	931	1	88,336

Note: EMS is Emergency Medical Service; IIA is Incident Impact Area.

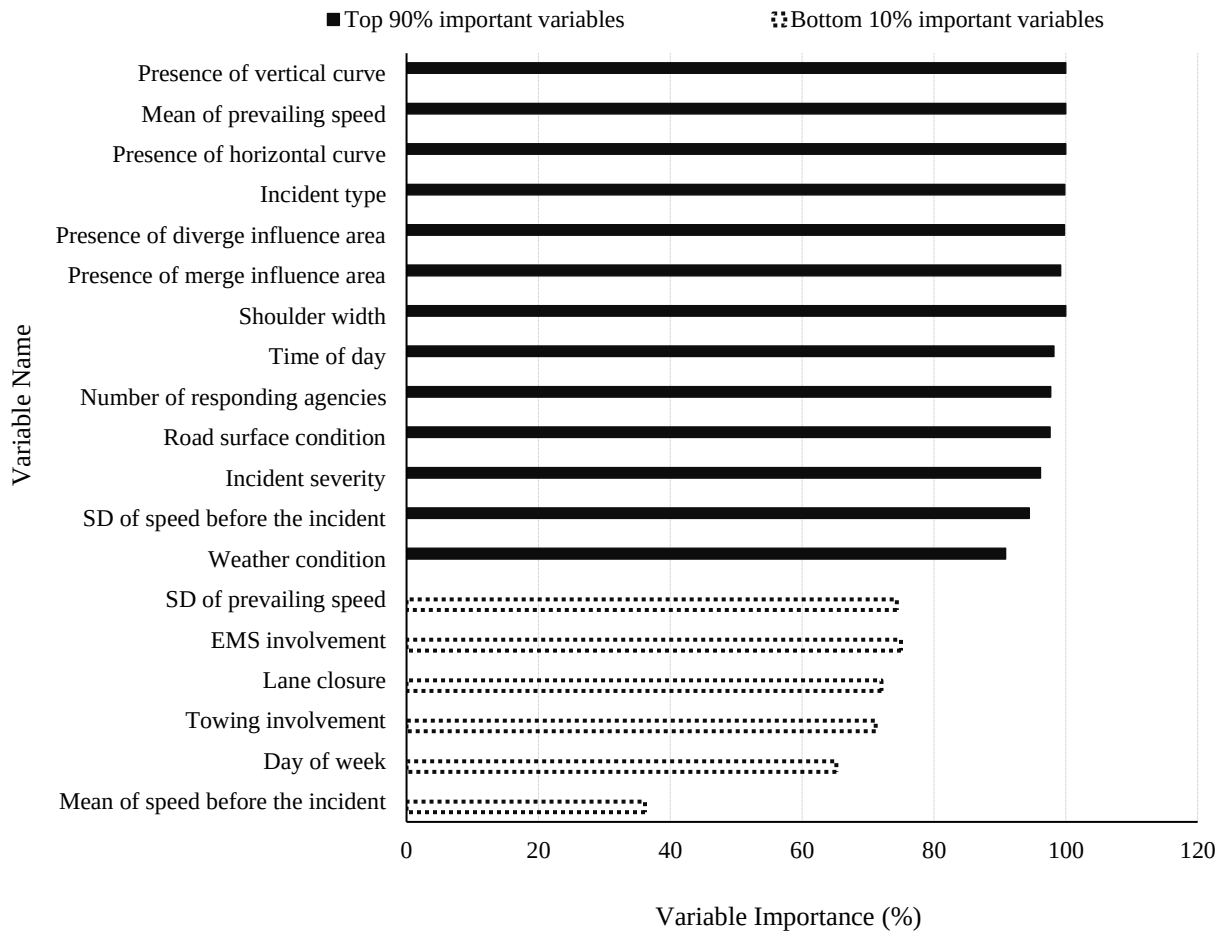
Table 5-3: Descriptive Statistics of Continuous Variables

Variable	Minimum	Mean	Median	SD	Maximum
Shoulder width (feet)[a]	4.0	10.5	10.5	1.6	22.0
Mean of speed before the incident (mph)[b]	1.0	64.1	66.3	9.5	80.7
SD of speed before the incident (mph) [b]	0.0	2.1	1.6	2.0	34.0
Mean of prevailing speed (mph) [b]	4.4	63.9	66.2	9.1	80.9
SD of prevailing speed (mph) [b]	0.0	3.1	1.9	3.7	30.9

Note: [a] roadway geometric attribute; [b] traffic flow attribute; SD is standard deviation.

5.4.2 Results of the Likelihood Model

The penalized logistic regression was used to identify factors that influence the risk of secondary crashes. Based on the percent of times selected, the variable importance is illustrated in Figure 5-3.



Note: SD = Standard Deviation.

Figure 5-3: Selection of the Important Variables

The top 90% of the selected variables, when fitting the penalized logistic regression on the bootstrapped samples, were considered the most important variables. Out of the 19 variables included in the analysis, the following 13 were selected as the most important variables:

- Traffic flow attributes
  - mean of prevailing speed
  - standard deviation of speed before the incident

- Incident-related variables
  - incident type

    - lane closure
    - number of responding agencies

- Weather-related variables
    - road surface condition
    - weather condition

- Temporal attribute
    - Time of day

- Geometric-related variables
    - vertical curve
    - horizontal curve
    - merge influence area
    - diverge influence area
    - shoulder width

Table 5-4 shows the penalized logistic regression model results and the number of times the variable was selected in the model as an important variable. The following subsections discuss the results in detail. Note that only the most important variables that are significant at the 95% credible interval are discussed.

Traffic Flow Attributes

Of the four traffic flow-related variables that were included in the model, the following two variables were identified as the most important and are significant at the 95% credible interval: mean of prevailing speed and SD of speed before the incident. As shown in Table 5-4, the negative parameter of the average of the prevailing speed indicates that the risk of secondary crashes decreases as the average prevailing speed increases. That is, for a unit increase in the average prevailing speed, the likelihood of secondary crashes decreased by 11%. The decreasing speed represents an increase in traffic density and queue formation. The disturbances caused by the primary incident are easier to propagate these queuing traffic formations, leading to an elevated risk of secondary crashes. This finding is consistent with the previous studies, which reported that the risk of secondary crashes increases with the decrease in average speed (Kitali et al., 2019b; Xu et al., 2016).

To prevent secondary crashes in this scenario, traffic control strategies could be developed to accelerate the dissipation of queues resulting from primary incidents. For example, variable speed limits could be used to gradually increase the speed downstream and decrease the speed upstream of the primary incident location at the same time. Further, information about the location and severity of the incident should be disseminated to upstream drivers, thereby allowing them to make decisions (e.g., change lanes, detour) in advance.

Interestingly, the standard deviation of speed before the incident was negatively associated with the risk of a secondary crash, meaning that the risk of a secondary crash increases with a decrease in variation of speed ten minutes before the incident. Specifically, a unit increase in the standard deviation of the speed before the incident reduced the likelihood of secondary crashes by 7%.

Table 5-4: Secondary Crash Likelihood Prediction Model Results

Variable	Category	Mean	OR	Med	SD	CI (%)		Count	%
						2.5	97.5		
Intercept	N/A	6.94	N/A	6.92	0.76	5.49	8.48	5,000	100
Traffic Flow Attributes									
Mean of speed before the incident (mph)	N/A	0.02	1.02	0.02	0.01	0.01	0.03	1,808	36
SD of speed before the incident (mph)	N/A	**-0.07**	**0.93**	**-0.07**	**0.02**	**-0.12**	**-0.04**	**4,723**	**94**
Mean of prevailing speed (mph)	N/A	**-0.12**	**0.89**	**-0.12**	**0.01**	**-0.15**	**-0.10**	**5,000**	**100**
SD of prevailing speed (mph)	N/A	0.02	1.02	0.02	0.02	-0.02	0.05	3,719	74
Incident Attributes									
Incident type	Vehicle problem								
	Hazard	**0.25**	**1.28**	**0.24**	**0.10**	**0.05**	**0.45**	**4,816**	**96**
	Crash	**0.53**	**1.71**	**0.53**	**0.19**	**0.17**	**0.92**	**4,992**	**100**
Lane closure	No								
	Yes	-0.06	0.94	-0.05	0.34	-0.80	0.59	3,601	72
Number of responding agencies	1								
	2+	**0.23**	**1.26**	**0.23**	**0.10**	**0.05**	**0.42**	**4,886**	**98**
EMS involvement	No								
	Yes	0.21	1.23	0.21	0.39	-0.62	1.00	3,751	75
Towing involvement	No								
	Yes	0.03	1.03	0.04	0.35	-0.71	0.72	3,558	71
Incident severity	Minor								
	Moderate/severe	**0.89**	**2.44**	**0.86**	**0.54**	**0.00**	**2.10**	**4,809**	**96**
Temporal Attributes									
Day of week	Weekday								
	Weekend	-0.04	0.96	-0.04	0.11	-0.24	0.19	3,260	65
Time of day	Off-peak								
	Morning peak	**0.23**	**1.25**	**0.23**	**0.12**	**0.01**	**0.46**	**4,911**	**98**
	Evening peak	-0.27	0.76	-0.28	0.13	-0.52	-0.03	4,153	83
Weather Attributes									
Weather condition	Clear								
	Adverse	**0.53**	**1.71**	**0.50**	**0.54**	**-0.46**	**1.67**	**4,544**	**91**
Road surface condition	Dry								
	Wet	**1.07**	**2.91**	**1.05**	**0.59**	**0.00**	**2.29**	**4,882**	**98**
Roadway Geometric Attributes									
Shoulder width (feet)	N/A	**-0.14**	**0.87**	**-0.14**	**0.02**	**-0.19**	**-0.10**	**5,000**	**100**
Presence of horizontal curve within IIA	No								
	Yes	**0.50**	**1.65**	**0.50**	**0.10**	**0.31**	**0.69**	**5,000**	**100**
Presence of vertical curve within IIA	No								
	Yes	**0.82**	**2.27**	**0.82**	**0.09**	**0.63**	**1.01**	**5,000**	**100**
Presence of diverge influence area within IIA	No								
	Yes	**0.27**	**1.32**	**0.27**	**0.10**	**0.08**	**0.47**	**4,989**	**100**
Presence of merge influence area within IIA	No								
	Yes	**-0.30**	**0.74**	**-0.30**	**0.09**	**-0.48**	**-0.11**	**4,962**	**99**

Note: Bolded numbers are significant at 95% credible interval; OR is odds ratio; Med is median; SD is standard deviation; IIA is incident impact area.

A high standard deviation indicates a higher variability, and vice versa. This metric is included to help capture the effect of rapid changes in traffic conditions (e.g., shockwaves and braking maneuvers) associated with pre-incident conditions. It is worth noting that high traffic speeds are associated with low standard deviation, whereas low traffic speeds usually have high speed variations. In other words, if the incident occurred when the traffic speed is higher, higher variability in speed is likely to occur as traffic is transitioning from the free-flow state to the congested state, a situation that increases the likelihood of secondary crashes. On the other hand, if the incident occurred when the variation of the traffic speed estimates is high (or the average speed is low), the likelihood of a secondary crash to occur is expected to be low since traffic is already in a congested state and a significant variation in speed is not expected.

Incident Attributes

The most important incident-related variables included incident type, number of responding agencies, and lane closure. Compared with vehicle problem-related incidents, hazard and crash-related incidents were more likely to result in a secondary crash. Specifically, hazard-related incidents were 28% more likely to result in a secondary crash than vehicle problem-related incidents. Meanwhile, crashes were 71% more likely to result in a secondary crash than vehicle problem-related incidents. From this finding, it can be inferred that the risk of crashes to result in secondary crashes is twice as much as that of hazard-related incidents. A possible explanation for this observation may be related to the extent of impact different incident types may have on traffic. In general, crashes are more likely to cause congestion than other incident types, such as hazards and vehicle problems.

As expected, the number of responding agencies was also identified as a significant predictor variable that influenced the risk of secondary crash occurrence. More specifically, incidents attended to by more than one responding agency were 26% more likely to result in a secondary crash compared to incidents attended to by only one responding agency. The number of responding agencies is an indicator of the severity of the incident as severe incidents tend to be attended by more responding agencies than less severe ones. Moreover, incidents attended to by multiple incident responders may require lane closures, a situation that further reduces the capacity of the roadway resulting in more congestion and hence increases the likelihood of a secondary crash. This fact is proven by the positive coefficient of the incident severity variable, which indicates that moderate/severe incidents were twice as likely to result in secondary crashes compared to incidents with minor severity.

Weather Attributes

Wet road surface condition was found to be positively associated with the risk of secondary crashes, indicating that incidents that occurred on wet road surface conditions were more likely to result in secondary crashes than those on dry surface conditions. The corresponding odds ratio of 2.91 suggested that wet road surface conditions were twice as likely to cause secondary crashes. A similar observation was found in a previous study (Xu et al., 2016).

Temporal Attributes

The results in Table 5-4 show that the time of day variable was among the most important variables. Compared to off-peak hours, incidents that occurred during morning peak hours were 25% more likely to result in secondary crashes. This finding indicates that secondary crashes are more likely to occur during congested periods.

Roadway Geometric Attributes

All the five geometric-related variables considered were identified as the most important variables that significantly influence the likelihood of a secondary crash. The considered roadway geometric-related variables were diverge influence area, merge influence area, shoulder width, horizontal curve, and vertical curve. Incidents with impact areas within the diverge influence area were 32% more likely to result in secondary crashes. Diverge influence areas increase the risk of secondary crashes since they are accompanied by more lane changes and high speed differentials because of drivers who are attempting to exit the freeway. On the contrary, the estimated parameter of merge influence area was negative, implying that incidents with a merge influence area within the impact area were 26% less likely to cause secondary crashes.

The estimated parameter of shoulder width was negative, implying that a unit increase in shoulder width was accompanied by a 13% decrease in the likelihood of a secondary crash. A possible explanation is that shoulders provide room for veering away from a possible crash. Further, when a platoon of vehicles is suddenly forced to slow down, some of the drivers in the middle of the platoon who are unaware of the downstream traffic condition tend to complete the deceleration process on the shoulders.

As indicated in Table 5-4, incidents with horizontal curves within the impact area, compared to tangents, were associated with the increased risk of secondary crashes. When the curved segment is within the impact area of the incident, the likelihood of a secondary crash increased by 65%. This is expected as the queue along a curved section may not be quickly visible to the upstream drivers. A similar observation was found on incidents with a vertical curve within the incident impact area. Incidents with elevated sections within their impact area were 127% more likely to cause secondary crashes than those on flat sections. The presence of vertical curves may limit the sight distance, making it difficult for upstream drivers to easily recognize the queue built by the initial incident.

5.5 Summary

The LASSO penalized logistic regression model, fitted using the bootstrap resampling approach, was used to identify factors influencing the risk of secondary crashes. Traffic flow, incident, temporal, weather, and roadway geometric attributes were considered as potential factors that may influence the likelihood of secondary crashes. The study area included a 28-mile section of the Florida's Turnpike Mainline SR-91. Data used were collected from the following sources: HERE Technologies (traffic flow attributes); SunGuide® database (incident); RCI database, Google Earth Pro software, and Google Maps (roadway geometric attributes). These data were collected for a period of five and a half years from January 2014 through June 2019.

Next, the LASSO penalized estimator was used to extract the most important explanatory variables, with minimal correlation, influencing the risk of secondary crashes. Because the proportion of primary incidents is smaller than the proportion of normal incidents, the bootstrap resampling method was used to fit the penalized logistic regression. The model is considered to improve the predictive accuracy of the secondary crash risk model because it accounts for the asymmetric nature of secondary crashes, performs variable selection, and removes correlated variables.

Out of the 19 variables included in the analysis, the following 13 were selected as the most important variables:

- Traffic flow attributes
    - mean of prevailing speed
    - SD of speed before the incident

- Incident-related variables
    - incident type
    - lane closure
    - number of responding agencies

- Weather-related variables
    - road surface condition
    - weather condition

- Temporal attribute
    - Time of day

- Geometric-related variables
    - vertical curve
    - horizontal curve
    - merge influence area
    - diverge influence area
    - shoulder width

All the aforementioned 13 most important variables except weather condition were found to be significant at the 95% credible interval. The following are some of the key findings on the influence of these factors on the likelihood of secondary crashes.

- A unit increase in the average prevailing speed reduces the likelihood of a secondary crash by 11%. In other words, a unit decrease in average prevailing speed was accompanied with an 89% increase in the risk of secondary crashes.

- A unit increase in the standard deviation of speed before the occurrence of the incident reduced the risk of a secondary crash by 7%.

- Hazard-related incidents were 28% more likely to result in secondary crashes compared to vehicle problem-related incidents.

- Crashes were 71% more likely to result in secondary crashes than vehicle problem-related incidents.

- Incidents attended to by more than one responding agency were 26% more likely to result in secondary crashes compared to incidents attended by only one responding agency.

- Moderate/severe incidents were twice as likely to result in secondary crashes compared to minor incidents.

- Incidents that occurred on wet road surface conditions were twice more likely to result in secondary crashes than those on dry surface conditions.

- Compared to off-peak hours, incidents that occurred during morning peak hours were 25% more likely to result in secondary crashes.

- Incidents with diverge influence area within their impact area were 32% more likely to result in secondary crashes.

- Incidents with merge influence area within their impact area were 26% less likely to cause secondary crashes.

- A unit increase in shoulder width was associated with a 13% decrease in the likelihood of a secondary crash.

- Incidents with a curved segment within their impact area increased the risk of secondary crashes by 65%.

- Incidents with a vertical curve segment within their impact area increased the risk of secondary crashes by 127%.

CHAPTER 6
REAL-TIME PREDICTION OF SECONDARY CRASHES: PROOF-OF-CONCEPT

The availability of real-time traffic and weather data offers an opportunity to proactively mitigate secondary crashes. This chapter discusses the algorithm developed to predict the likelihood of secondary crashes in real time. The study area included a 28-mile section of the Florida's Turnpike Mainline (SR-91).

6.1 Databases

As depicted in Table 6-1, four main types of data are required to predict the likelihood of secondary crashes in real time.

Table 6-1: Data Requirements for the Real-time Secondary Crash Prediction Algorithm

Data Type	Historical Data Source	Real-time Data Source	Ping Rate (Minutes)
Incident: • Incident type • Time of day	SunGuide®	SunGuide®	2
Traffic: • Standard deviation of speed before the incident • Mean prevailing speed	RITIS	HERE Technologies	1
Rainfall	NEXRAD Level-II	NEXRAD Level-II	4-6
Roadway geometric characteristics: • Shoulder width • Horizontal curves • Vertical curves • Merging segment • Diverging segment	• RCI • Google Earth Pro • Google Maps	N/A	N/A

Note: NEXRAD = Next Generation Weather Radar; RCI = Roadway Characteristic Inventory; RITIS = Regional Integrated Transportation Information System.

- *Incident Data*: In Chapter 5, the following incident-related variables were identified as the most important: incident type, number of responding agencies, and incident severity. Of all the incident-related variables, only incident type was included in the real-time secondary crash prediction model. Number of responding variables and incident severity were not included since it is not clear at what time these variables are reported after the incident occurred. Since these two variables could be considered a surrogate measure of congestion, time of day, the most important temporal attribute, was used instead. A program that uses a Virtual Private Network (VPN) was used to access incident data in real time from the SunGuide® database every two minutes, i.e., at a 2-minute interval (see Table 6-1).

- *Traffic Data*: As discussed in Chapter 5, the following two traffic-related attributes were identified as the most important: the standard deviation of speed before the incident and the mean prevailing speed. Note that historical speed data from HERE Technologies were retrieved from RITIS. The real-time traffic data were obtained from the HERE real-time flow Extensible Markup Language (XML) feed and accessed via Hypertext Transfer Protocol (HTTP).

- *Rainfall Data*: The radar-based historical and real-time rainfall data were retrieved from the NEXRAD Level-II network archived by the Amazon Web Service (AWS) Command Line Interface (CLI). The rainfall intensity values from NEXRAD are usually updated every 4-6 minutes. Note that the downloaded rainfall data are categorized into two groups, i.e., *no/light* rainfall and *medium/heavy* rainfall.

- *Roadway Geometric Characteristics*: In Chapter 5, all considered geometric variables were found important: shoulder width, horizontal curves, vertical curves, merging segment, and diverging segment. Shoulder width, horizontal curves, and vertical curves variables were collected from the RCI database for the years 2014 through 2019. The merge and diverge influence areas were derived from Google Earth Pro and Google Maps using the Historical Imagery and the Street View tools. Since the roadway geometric variables do not change as often, they were stored in an internal database and will be updated based on changes made along the study corridor.

6.2 Real-time Secondary Crash Prediction Algorithm

The real-time secondary crash prediction algorithm is divided into three main parts:

- Internal Storage Database,
- Real-time Data Backend Applications, and
- Secondary Crash Prediction Program.

6.2.1 Internal Storage Database

The first part of the algorithm is the *Internal Storage* database consisting of the following information: collected real-time incident, speed, and rainfall data, real-time secondary crash prediction results, historical databases, statistical model equation and parameters for the secondary crash likelihood, and other potential attributes required to predict the likelihood of secondary crashes but are not collected in real time, i.e., roadway geometric characteristics. Specifically, the *Internal Storage* database consists of:

- *Speed Profiles* database, which contains the Traffic Message Channel's specific historical average speed (*speed*) and the lower bound speed (*speed_LB*) for different days of the week at a 5-minute interval.

- *Upstream Traffic Message Channels* database, which includes the corresponding upstream Traffic Message Channels for each Traffic Message Channel in the study corridor.

- *Secondary Crash Likelihood Model Parameters* table, which includes the model coefficients (i.e., β values) for each explanatory variable in the model and the intercept.

- *Roadway Geometric Characteristics* database, which includes shoulder width (feet), presence of merge influence area, diverge influence area, horizontal curve, and vertical curve.

In addition to archiving historical data, the *Internal Storage* database also stores the real-time incident information from the SunGuide® database, real-time speed data from HERE Technologies, and real-time rainfall data from the NEXRAD Level-II AWS system. The database also stores the real-time secondary prediction likelihood results along with the data used to derive them.

6.2.2 Real-time Data Backend Applications

The second part of the algorithm, referred to as the *Real-time Data Backend Applications*, has programs that read, parse, and save speed and rainfall data in real time.

- *Real-time Speed Data*: One of the *Backend Applications*, referred to in the algorithm as *RealtimeSpeed*, retrieves the real-time speed data from the HERE real-time flow XML feed. It accesses the XML feed every minute via HTTP. Since the XML feed data are available in a .gp format, the application has a sub-step that converts the .gp data into a readable format, such as the .xml format. Within the *RealtimeSpeed* application, a sub-process continuously aggregates the data every ten minutes to obtain the standard deviation of speed before the incident. This application also computes the mean of the prevailing speed every 15 minutes from the time the incident is pinged to the time the incident is cleared/traffic comes back to normal.

- *Real-time Rainfall Data*: The *RealtimeRain* application retrieves real-time rainfall information from NEXRAD Level-II data hosted in the "noaa-nexrad-level2" Amazon S3 bucket. The rainfall intensity values from NEXRAD are usually updated every 4-6 minutes. The *RealtimeRain* application also has a function to convert the maximum value of the rainfall intensity into no/light rainfall and medium/heavy rainfall.

6.2.3 Secondary Crash Prediction Application

The information from the *Internal Storage* database and *Real-time Data Backend Programs* are then combined to predict the likelihood of secondary crashes. The secondary crash prediction process makes the third and final part of the algorithm, referred to as the *SecondaryCrashPrediction* application. The *SecondaryCrashPrediction* application has three main functionalities: (1) ping new incidents from the SunGuide® database, (2) estimate the queue of the pinged incident, and (3) estimate the probability of the pinged incident resulting in a secondary crash. The application uses VPN to identify and retrieve incidents from the SunGuide® database every two minutes. The application has a filter that retains only incidents that occurred within the study area. The main incident characteristics that are saved include the incident first notified date and time, incident first notified day, incident direction, incident location (MMs and/or latitudes and longitudes), and incident type. The *SecondaryCrashPrediction* application also categorizes the incident's first notified time into off-peak hours, morning peak hours, or evening peak hours.

Once the *SecondaryCrashPrediction* application pings a new incident, it then combines the information from the *Internal Storage* database, *RealtimeRain* application, and *RealtimeSpeed* application to predict the likelihood of secondary crashes. The application performs the following

two steps: (1) estimate the impact area of an incident; and (2) predict the likelihood of secondary crashes.

Estimate the Incident Impact Area

Incident occurrence time and incident location are required to estimate its impact area. This information is retrieved through the *SecondaryCrashPrediction* application. When an incident is reported in the *SunGuide® database*, its location based on MM (I_{MM}) and direction of travel are recorded. The time of day and day of the week when the incident is reported are also captured. The location and direction of the incident are then used to identify the Traffic Message Channel (TMC0) where it occurred. Once this Traffic Message Channel is determined, the corresponding Traffic Message Channel (TMC1) upstream TMC0 is extracted from the *Upstream Traffic Message Channel Database*.

The incident impact area is estimated to (1) determine whether the incident can result in secondary crashes and (2) identify the spatial and temporal extent for disseminating the messages to upstream motorists. Both historical and real-time traffic data are required for this purpose. The real-time traffic speed on TMC0 V_{TMC0_i} is retrieved from the HERE real-time flow XML feed using the *RealtimeSpeed* application. The value of V_{TMC0_i} is compared with the lower bound speed of TMC0 at the time and day, similar to the incident's first notified time and day. Note that the Traffic Message Channel's lower bound speed is obtained from the *Speed Profile Database*.

If the prevailing traffic speed is lower than the Traffic Message Channel's lower bound speed, the next step involves predicting the likelihood of secondary crashes at TMC0.

Predict the Likelihood of Secondary Crashes

A logistic regression model is used to predict the likelihood that a secondary crash will occur given a set of explanatory variables. More information about this approach is provided in Chapter 5. Specifically, the probability $P(i)$ of an incident (i) causing a secondary crash, with $P(i)$ taking values from 0 to 1, is obtained using Equation 6-1.

$$P(i) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_p}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_p}} \quad (6\text{-}1)$$

Due to the probability output of the logistic regression model, there is a dilemma in selecting the correct threshold because a high threshold typically fails to identify many potential primary incidents. In contrast, a low threshold falsely identifies normal incidents as primary incidents. This problem is more pronounced in imbalanced classification data. That is, the proportion of the primary incident is overly lower than the proportion of normal incidents. Thus, an optimal indicator that will maximize the prediction of actual primary incidents and minimize false normal incidents is necessary.

The selection of an optimal cutoff value involves balancing the false positive rate (FPR) and false negative rate (FNR). The receiver operating curve (ROC) could visualize and quantify the tradeoff between the two measures. This curve is created by plotting the true positive rate (TPR) on the y-

axis against the false positive rate (FPR) on the x-axis at various cutoff settings, i.e., between 0 and 1. A false negative (FN) means that an incident resulted in a secondary crash, but the model failed to detect it. In contrast, a false positive (FP) indicates that an incident did not result in a secondary crash, but the model predicted it as a primary incident. If $P(i)$ is greater than the optimal threshold, an incident (i) is considered to have a higher potential of resulting in one or multiple secondary crashes.

This process is repeated to the corresponding upstream Traffic Message Channels every *15 minutes* until the prevailing speed is higher than the TMC0's lower bound speed. Natural traffic flow at shorter time intervals will contain a large amount of noise (Guo et al., 2018). Previous literature has recommended using a minimum of 15-minute measurement intervals to obtain stable traffic flow rates (Smith and Ulmer, 2003). Figure 6-1 summarizes the data required to identify incidents with a higher likelihood of resulting in secondary crashes.

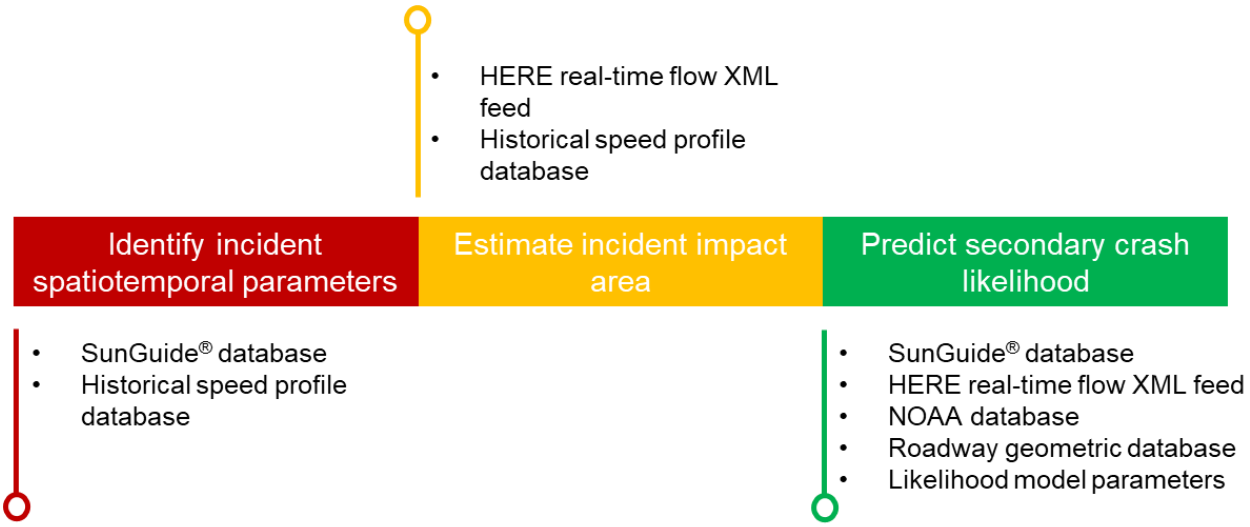


Figure 6-1: Data Used in the Real-time Secondary Crash Prediction Algorithm

6.3 Steps to Predict Secondary Crash Likelihood in Real Time

The main goal of the algorithm is to accurately determine if the traffic incident has a high likelihood of resulting in secondary crashes. Once the *SecondaryCrashPrediction* application detects an incident within the study corridor, it first determines whether an incident significantly impacted traffic. If yes, then the application predicts the likelihood of the respective incident resulting in a secondary crash. This process is repeated until traffic comes back to normal in the Traffic Message Channel where the incident occurred and the immediate upstream Traffic Message Channel.

Figure 6-2 shows the main steps used in the algorithm to estimate the incident impact area and predict the likelihood of the respective incident resulting in a secondary crash. To better illustrate the functionalities in the algorithm, the process is explained using the following example.

6.3.1 Incident Impact Area Estimation

Step 1: Consider an incident I pinged by the *SecondaryCrashPredictio*program at 8:00 AM. Following the detection of incident I, its location based on MM (I_{MM}) and direction of travel are recorded. The time of day and day of the week when the incident is reported are also captured.

Step 2: The location and direction of the incident are then used to identify the Traffic Message Channel (TMC0) where it occurred. Once this Traffic Message Channel is determined, the corresponding Traffic Message Channel (TMC1) upstream TMC0 is extracted from the *Upstream Traffic Message Channel Database*.

Step 3: As mentioned earlier, the *RealtimeSpeed* application continuously reads, parses, and saves speed data every minute for each Traffic Message Channel. The real-time traffic speed of the Traffic Message Channel where the incident occurred (TMC0) and the time when the incident was detected (8:00 AM), i.e., V_{TMC0_i}, is retrieved from the saved speed data. The value of V_{TMC0_i} is compared with the lower bound speed of TMC0 at the time and day, similar to the incident's first notified day and time. Note that the Traffic Message Channel's lower bound speed is obtained from the *Speed Profile Database*.

Step 4: The incident I is considered to impact the traffic significantly when the prevailing traffic speed is lower than the Traffic Message Channel's lower bound speed. In this case, the next step involves predicting the potential of incident I resulting in a secondary crash.

6.3.2 Secondary Crash Prediction

Step 5: A logistic regression equation is used to predict the likelihood that a secondary crash will occur given a set of explanatory variables. The model has the following ten explanatory variables:

a. *incident type* grouped into three categories, i.e., vehicle problems (e.g., disabled vehicles, emergency vehicles, abandoned vehicles, and vehicle fire), debris on roadway, and crashes.

b. *incident first notified time* categorized into three groups, i.e., off-peak hours, morning peak hours, and evening peak hours.

c. *Rainfall intensity* categorized into no/low rainfall and medium/heavy rainfall

d. *Shoulder width (feet)*

e. *Presence of merge influence area*

f. *Presence of diverge influence area*

g. *Presence of horizontal curve*

h. *Presence of vertical curve*

i. *Standard deviation of speed before the incident (miles per hour)*

j. *Mean of prevailing speed (miles per hour)*

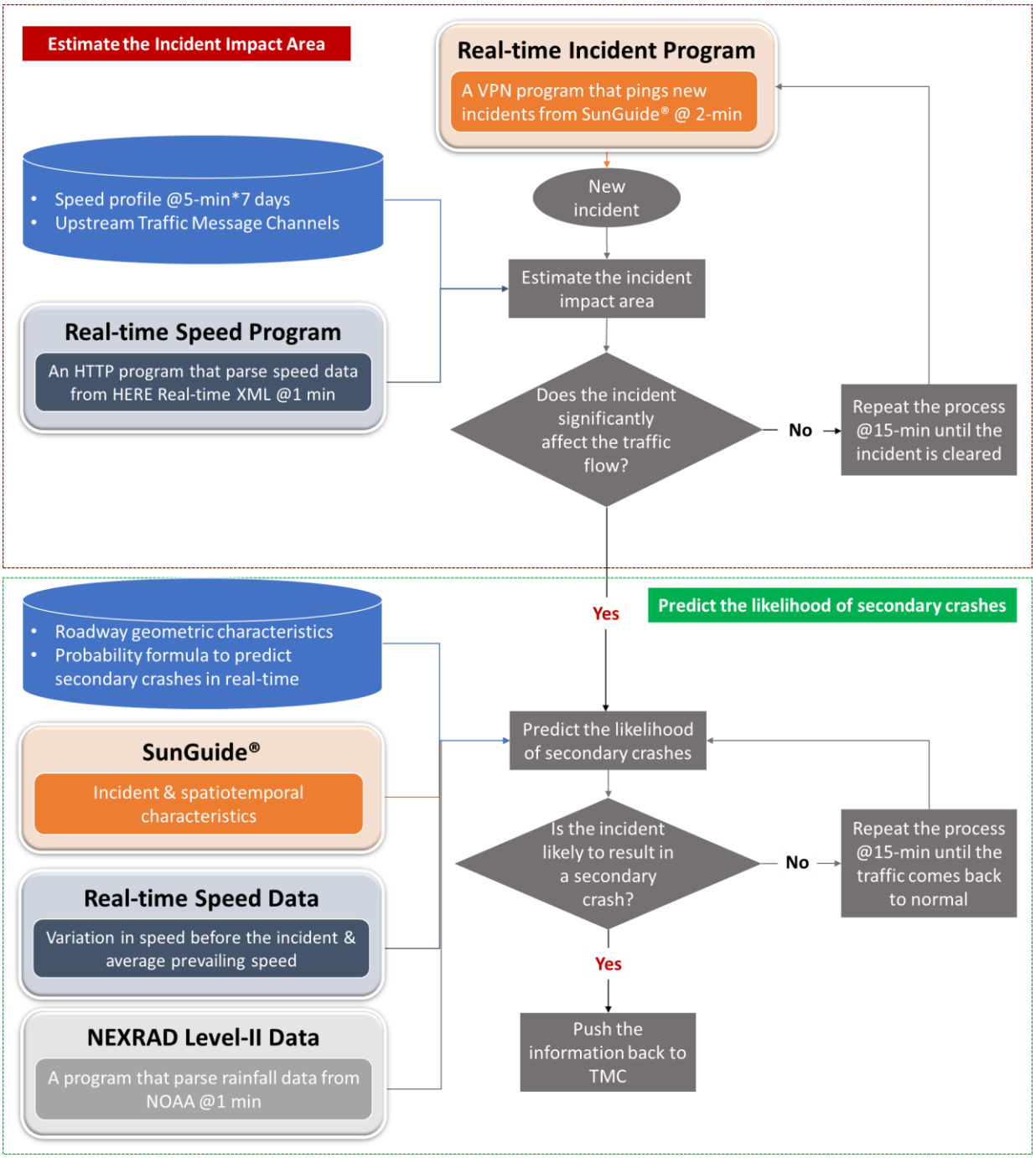


Figure 6-2: Real-time Secondary Crash Prediction Algorithm

Step 6: As mentioned earlier, the incident type and incident first notified time variables are saved internally using the *SecondaryCrashPrediction* application. For the analysis at the time when the incident occurred, real-time traffic and rainfall data were extracted 5 minutes before the incident reported time, i.e., 07:55 AM and 08:00 AM. The purpose was to account for potential inaccuracies in the incident reported time (Golob and Recker, 2003; Kitali et al., 2018; Xu et al., 2016).

Step 7: The rainfall intensity is obtained from the NEXRAD Level-II database through the R*ealtimeRainfall* application. Specifically, the maximum rainfall intensity between 07:55 AM and 08:00 AM is used as the rainfall intensity value in the model.

Step 8: The speed data are retrieved from the HERE real-time flow XML feed through the R*ealtimeSpeed* application. The standard deviation of the speed before the incident is obtained by computing the sample standard deviation of the speed values between 07:45 AM and 07:55 AM. The mean of the prevailing speed is obtained by averaging speed values between 07:55 AM and 08:00 AM.

Step 9: The roadway geometric attributes, i.e., shoulder width, presence of merge influence area, presence of diverge influence area, presence of horizontal curve, and presence of vertical curve are retrieved from the *Roadway Geometric Attributes Database*. Except for the shoulder width variable, which is used as a continuous variable, the rest of the roadway geometric variables are grouped into the "yes" category implying Traffic Message Channels with the merge/diverge/horizontal curve/vertical curve and the "*no*" category meaning Traffic Message Channels without merge/diverge/horizontal curve/vertical curve.

Step 10: The collected explanatory variables and the parameters from the *Likelihood Model Parameters Database* are used to predict the probability of secondary crashes. The equation for predicting the likelihood of secondary crashes is shown in Equation 6-2. If the estimated secondary crash probability is higher than 0.4, incident I is considered to have a higher likelihood of resulting in a secondary crash. Steps 1-10 are repeated to the corresponding upstream Traffic Message Channels (TMC1).

$$Secondary\ Crash\ Probability = 1/(1 + EXP(-1 * (7.36 + (-0.08 * SD\ of\ speed\ before\ the\ incident) + (-0.13 * Mean\ of\ prevailing\ speed) + (-0.25 * Debris\ on\ Roadway) + (0.44 * Crash) + (0.01 * Morning\ Peak) + (-0.63 * Evening\ peak) + (0.72 * Rainfall\ intensity) + (-0.1 * Shoulder\ width) + (0.67 * Horizontal\ curve) + (1.17 * Vertical\ curve) + (0.43 * Diverge\ influence\ area) + (-0.53 * Merge\ influence\ area)))) \quad (6\text{-}2)$$

Step 11: Steps 1-10 are repeated every 15 minutes until the prevailing speed is higher than the TMC0's lower bound speed. For the example incident I, the next analysis period will be 08:15 AM. Table 6-2 provides the secondary crash predicted probabilities for incident I. The unit of the speed values in columns "Current speed", "Low bound speed", "standard

deviation of speed before the incident", and "Mean of prevailing speed" are miles per hour (mph).

Table 6-2: Predicted Secondary Crash Probabilities for Incident *I*

Timestamp	Current speed (mph)	Lower bound speed (mph)	Standard deviation of speed before the incident (mph)	Mean of prevailing speed (mph)	Debris on roadway	Crash	Morning peak	Evening peak	Rainfall	Shoulder width (ft)	Horizontal curve	Vertical curve	Diverge influence area	Merge influence area	Secondary crash probability
3/5/2021 8:02	40.3	53.2	3.2	47.8	0	1	1	0	1	10.6	1	1	1	1	0.94
3/5/2021 8:15	33.4	51.9	3.2	40	0	1	1	0	1	10.6	1	1	1	1	0.98
3/5/2021 8:30	67.8	69.0	3.2	67.3	0	1	1	0	0	10.6	1	1	1	1	0.37
3/5/2021 8:45	74.3	70.4	-	-	-	-	-	-	-	-	-	-	-	-	-

The shoulder width values are in feet. Except for the speed-related and shoulder width variables, the rest of the variables are dummy variables. That is, if the cell value is 1, it implies incident *I* is associated with the characteristics shown in the column heading. For instance, incident *I* in this example, is a crash, occurred during morning peak hours, on a horizontal curve, etc. otherwise, if the cell value is 0, it implies the opposite, i.e., incident *I* does not have the characteristics mentioned in the column heading. As shown in this table, the traffic returned to normal at 8:45 AM, i.e., the current speed at 08:45 AM was higher than the lower bound speed of the speed profile for the Traffic Message Channel where the incident occurred. Since the predicted probabilities of a secondary crash at 8:02 AM and 8:15 AM are higher than 0.4 (optimal threshold), it implies that incident *I* has a higher likelihood of resulting in a secondary crash at these timestamps.

6.4 Summary

This chapter presented the algorithm, developed as a *proof-of-concept*, to predict the likelihood of secondary crashes in real time. The study area included a 28-mile section of SR-91. The algorithm consists of three main parts, the first one being the *Internal Storage* database, which includes:

- historical databases,
- statistical model equation and parameters for the secondary crash likelihood, and
- other potential attributes required to predict the likelihood of secondary crashes but are not collected in real time, i.e., roadway geometric characteristics.

The second part of the algorithm consists of *Backend Programs* for collecting, parsing, and saving incident, traffic, and rainfall data in real time. One algorithm application continuously accesses the Florida's Turnpike SunGuide® database through VPN every two minutes and pings new incidents. The R*ealtimeSpeed* application functions to retrieve, process, and save real-time speed data from the HERE real-time flow XML feed. This process is implemented for each Traffic Message

Channel within the study corridor every minute. The *RealtimeRainfall* application retrieves real-time rainfall information from the NEXRAD Level-II network hosted in the "noaa-nexrad-level2" Amazon S3 bucket. The rainfall intensity values are available every 4-6 minutes.

The information from the *Internal Storage* database and *Real-time Data Backend Programs* are then combined to predict the likelihood of secondary crashes. The secondary crash prediction process makes the third and final part, referred to as the *Secondary Crash Prediction Program*. This program is a two-fold process that focuses on (1) estimating the impact area of an incident and (2) predicting the likelihood of secondary crashes.

CHAPTER 7
EXPLORE THE POTENTIAL OF CONNECTED VEHICLE (CV) APPLICATIONS

This chapter quantifies the potential benefits of CV applications in mitigating secondary crashes. Since CVs are not yet fully operational, a sensitivity analysis was conducted at varying MPRs. The first section of this chapter provides a brief introduction to the CV applications and their potential to mitigate secondary crashes. It also discusses microsimulation and surrogate safety measures as a suitable approach to analyze CV applications. The second section focuses on the methodology used to quantify the benefits of CVs in mitigating secondary crashes. The results are presented and discussed in the third section. The final section includes the summary.

7.1 Background

To effectively mitigate the risk of secondary crashes, information about primary incidents must promptly be communicated to upstream drivers (Kitali et al., 2018). TIM agencies use advance warning messages to alert drivers about downstream incidents (Yang et al., 2018). Conventionally, the advance warning messages are posted along the roadside by incident response agencies. Responders can be hindered by incident-induced congestion (Yang et al., 2018, 2017), leading to potential delays in posting advance warning messages, and hence, increasing the likelihood of secondary crash occurrence. Other methods, such as DMSs, highway advisory radio (HAR), traveler information websites, and 511 systems, are used to disseminate incident information to travelers in real time (Carson, 2010; Motamed and Machemehl, 2014; Pearce and Subramaniam, 1998). Nonetheless, once the incident occurred, the longer it takes for the information to be disseminated to the public, the greater is the likelihood of secondary crashes (Kitali et al., 2019b; Yang et al., 2018).

CV applications have the potential to reduce the delay in relaying advance warning messages to upstream drivers by automating the incident detection process and instantly sending a message to upstream drivers after TMC has verified the incident. Because CVs are not yet fully deployed, traffic simulation models and surrogate safety measures are typically used to assess their safety performance (Paikari et al., 2014; Rahman et al., 2018; Yang et al., 2017). Among the surrogate measures proposed by the FHWA, time to collision (TTC) is the most well-known time-based safety indicator (Gettman and Head, 2003). TTC is defined as the time remaining before two vehicles collide, assuming both maintain their course and speed (Saffarzadeh et al., 2013; Shahdah et al., 2015). The TTC threshold (range 1.5 to 4.0 seconds) defines the potential traffic conflict and is effective for rear-end, head-on, and weaving conflicts (Mahmud et al., 2019). Post-encroachment time (PET) is another commonly used surrogate measure and is more efficient for intersection conflicts (Mahmud et al., 2019). PET refers to the time interval between two instances when the first vehicle leaves a conflict point and when the second vehicle enters it (Sen et al., 2007).

The following lane blockage scenarios were analyzed as a primary incident: one outer lane closed, one inner lane closed, and two outer lanes closed. Since full market penetration of CVs is not anticipated anytime soon, a sensitivity analysis was performed at varying CV MPRs. The analysis also considered the effect of the time period (AM peak and PM peak), which reflects situations with different traffic volumes. The VISSIM microscopic software was used to model a freeway

road segment on the Florida's Turnpike System and Lyons Road, an arterial that runs parallel to the freeway. The software was used to replicate CV applications to issue advisories, such as speed, lane-change, or detour advisory to drivers during an incident. The safety evaluation was performed using the Surrogate Safety Assessment Model (SSAM) software by importing trajectory files from VISSIM to analyze conflicts. The change in the number of simulated conflicts was used to evaluate the effectiveness of CV technology in mitigating secondary crashes.

7.2 Methodology

The VISSIM model was used as a simulation testbed. Incident modeling and V2I communication were modeled using the V2X module in the VISSIM software through the component object model (COM) application programming interface (API). This module simulates wireless communication and data exchange within the connected environment. Scripting was done using the Visual Basic Scripting (VBS) language. Another critical stage of this methodology was a safety evaluation, which employed the SSAM software. The details regarding SSAM are provided in the safety evaluation section. Finally, a statistical analysis was conducted to assess the significance of CVs in mitigating secondary crashes at different MPRs. Figure 7-1 summarizes steps for VISSIM model development and a safety evaluation process by SSAM software. The following subsections describe in detail the VISSIM modeling, simulation of the CV environment, and safety evaluation.

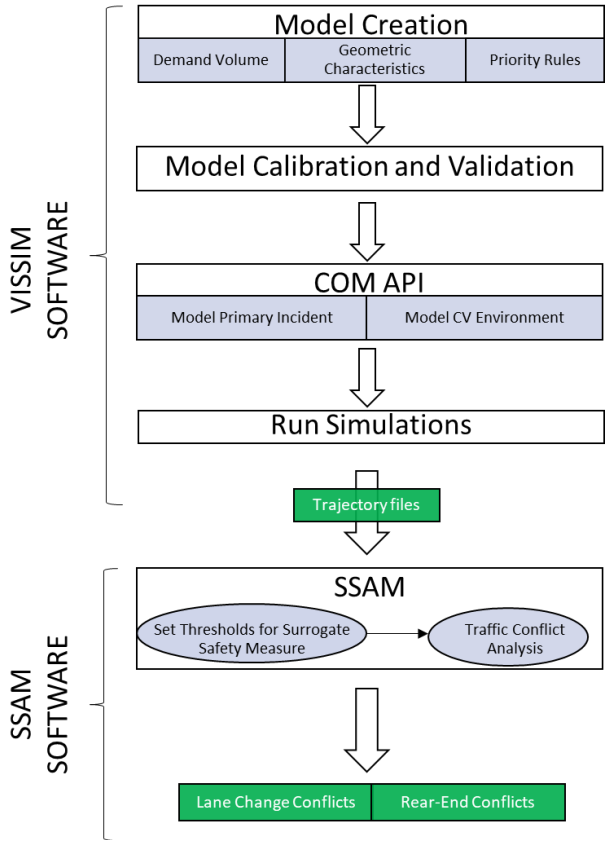


Figure 7-1: Methodology Framework for Exploring the Potential of CV Applications in Mitigating Secondary Crashes

7.2.1 Model Network Development

VISSIM is a microscopic simulation software that evaluates vehicle data based on car-following and lane-changing models. The software is capable of modeling complex traffic situations, such as CV communication, by virtual creation of sensors and wireless communication through the use of COM. The VISSIM model used was partly developed and calibrated by the FDOT by merging the previously developed VISSIM models for the Sawgrass Expressway and Interstate 95 (I-95). To include the real-world queues, the model limits of the arterials were extended to approximately 0.5 miles outside the main construction project limits. Another VISSIM model for a detour was created and merged with the previous model to analyze the potential safety benefits of using detour strategies in the CV environment.

Florida's Turnpike Mainline

As shown by the rectangle in Figure 7-2, the segment that experienced more crashes from 2016-2019 was clipped from the parent Turnpike model and used for further analysis.

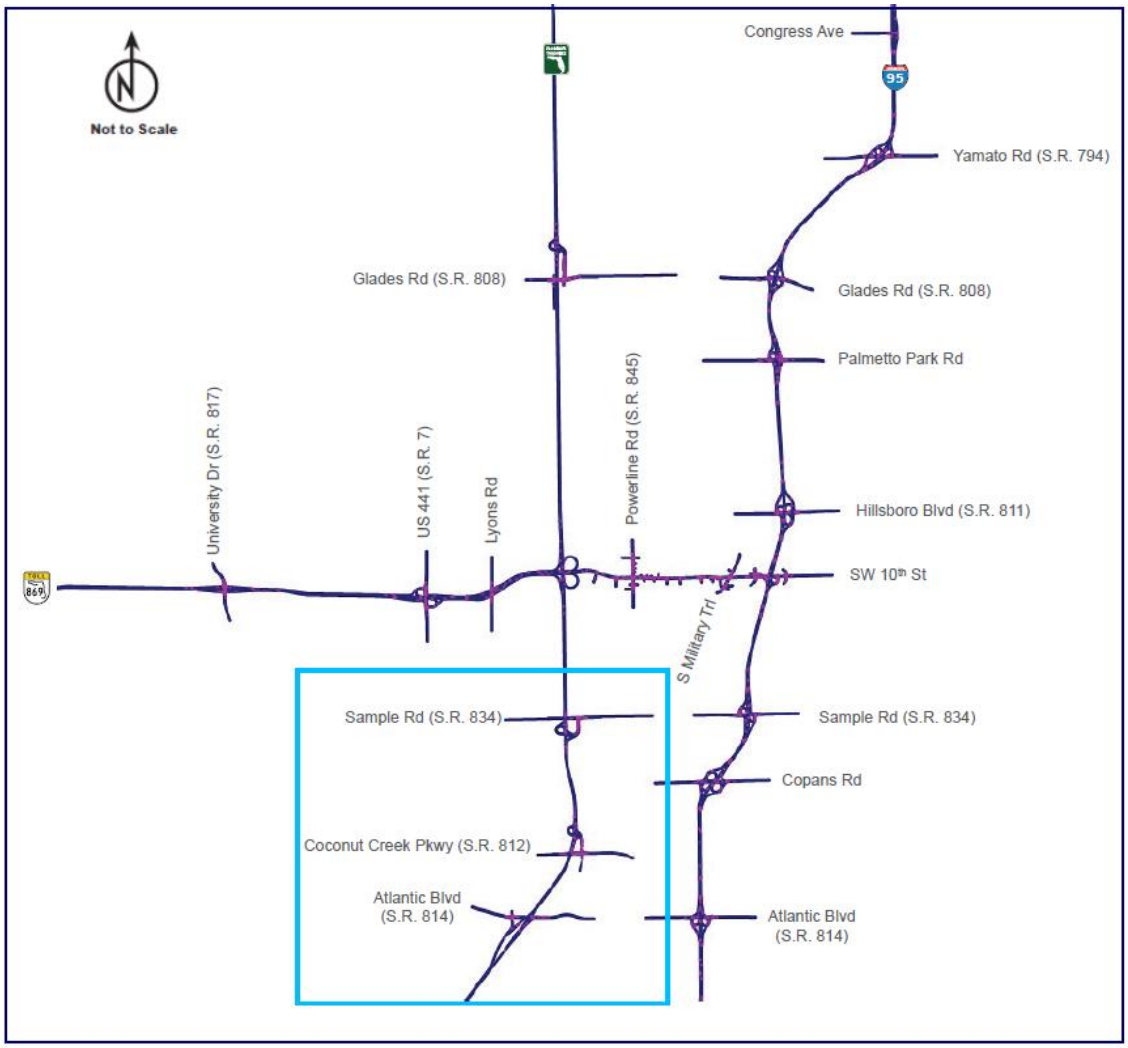


Figure 7-2: Turnpike VISSIM Model

The morning peak period was from 6:30 AM to 9:30 AM, and the evening peak period was from 4:00 PM to 7:00 PM. Within the peak period, the AM and PM peak hours were 7:30 AM to 8:30 AM and 5:00 PM to 6:00 PM, respectively. A 30-minute seeding time was added to load the network with traffic to attain equilibrium between the number of vehicles entering and exiting the traffic flow. The total simulation time was marked as 6:00 AM to 9:30 AM and 3:30 PM to 7:00 PM. Notably, traffic volumes represented traffic conditions of three different one-hour periods, including the pre-peak hour, peak hour, and the post-peak hour. The volumes were given in 15-minute intervals with the percentages of passenger cars and heavy vehicles.

Table 7-1 provides a breakdown of traffic volumes along the Turnpike Mainline and the ramp volumes on each on-ramp and off-ramp used in the clipped model. Additionally, Table 7-2 summarizes the loading factors used to convert the balanced 2016 peak hour volume into pre-peak hour and post-peak hour volumes. It shows the hourly conversion factors for each 15-minute interval in the total simulation period. These factors were generated based on time-slicing factors obtained from the hourly traffic volume distribution recorded in the field.

Table 7-1: Mainline and Ramp Traffic Volumes Used in the Microsimulation Model

(a) AM Period			
Location	*Demand Volume*[1]	*Location*	*Demand Volume*[1]
Florida's Turnpike Northbound		*Florida's Turnpike Southbound*	
Mainline before Atlantic Boulevard off-ramp	6,090	Mainline after on-ramp from Sawgrass Expressway	5,460
Mainline after Atlantic Boulevard off-ramp	4,860	Mainline after on-ramp from Sample Road	5,740
Mainline after on-ramp from Coconut Creek Road	4,810	Mainline after on-ramp from Coconut Creek Road	4,910
Mainline after on-ramp from Sample Road	4,160	Mainline after on-ramp from Atlantic Boulevard	5,840
Off-ramp to Atlantic Boulevard	1,230	On-ramp from Atlantic Boulevard	930
Off-ramp to Coconut Creek Parkway	710	Off-ramp to Coconut Creek Parkway	1,230
On-ramp from Coconut Creek Parkway	660	On-ramp from Coconut Creek Parkway	400
Off-ramp to Sample Road	1,200	Off-ramp to Sample Road	690
On-ramp from Sample Road	550	On-ramp from Sample Road	970
(b) PM Period			
Location	*Demand Volume*[1]	*Location*	*Demand Volume*[1]
Florida's Turnpike Northbound		*Florida's Turnpike Southbound*	
Mainline before Atlantic Boulevard off-ramp	5,720	Mainline after on-ramp from Sawgrass Expressway	3,980
Mainline after Atlantic Boulevard off-ramp	4,830	Mainline after on-ramp from Sample Road	4,610
Mainline after on-ramp from Coconut Creek Road	5,560	Mainline after on-ramp from Coconut Creek Road	4,660
Mainline after on-ramp from Sample Road	5,140	Mainline after on-ramp from Atlantic Boulevard	5,900
Off-ramp to Atlantic Boulevard	890	On-ramp from Atlantic Boulevard	1,240
Off-ramp to Coconut Creek Parkway	400	Off-ramp to Coconut Creek Parkway	620
On-ramp from Coconut Creek Parkway	1,130	On-ramp from Coconut Creek Parkway	670
Off-ramp to Sample Road	1,060	Off-ramp to Sample Road	400
On-ramp from Sample Road	640	On-ramp from Sample Road	1,030

Note: [1] Demand volume in vehicles per hour.

Table 7-2: Hourly Volume Conversion Factors

Time Interval	Simulation Time (Seconds)	AM Condition		PM Condition	
		15 minutes	Hourly	15 minutes	Hourly
Seed Time	0 - 900	9.38%	22.07%	22.08%	45.34%
	900 – 1,800	12.69%		23.26%	
Pre-Peak Hour AM: 6:30 - 7:30 AM PM: 4:00 - 5:00 PM	1,800 – 2,700	16.57%	81.55%	22.37%	92.31%
	2,700 – 3,600	19.38%		22.92%	
	3,600 – 4,500	21.29%		23.20%	
	4,500 – 5,400	24.31%		23.82%	
Peak Hour AM: 7:30 - 8:30 AM PM: 5:00 - 6:00 PM	5,400 – 6,300	25.50%	100.00%	24.25%	100.00%
	6,300 – 7,200	25.32%		25.20%	
	7,200 – 8,100	24.74%		25.39%	
	8,100 – 9,000	24.44%		25.17%	
Post-Peak Hour AM: 8:30 - 9:30 AM PM: 6:00 - 7:00 PM	9,000 – 9,900	23.60%	87.19%	24.44%	92.82%
	9,900 – 10,800	22.38%		24.07%	
	10,800 – 11,700	20.74%		22.83%	
	11,700 – 12,600	20.47%		21.48%	

Lyons Road

A parallel arterial, Lyons Road, was included in the analysis to evaluate the benefits of a detour strategy. Therefore, a VISSIM model was developed from the Turnpike exit on Coconut Creek Road and then connected to Lyons Road, and finally to W Sample Road. The model replicated the existing road geometric conditions, desired speed, and priority rules along the detour. The traffic volumes were retrieved from the Florida Traffic Online database for the AM and PM peak periods. The traffic volume balance was considered between the volume obtained from FTE and those entering and exiting the Turnpike network. The balanced volumes were then entered in 15-minute intervals, the same as in the Turnpike mainline model.

Among other traffic-related functions, PTV Vistro does optimization of traffic signal timings. Considering its compatibility with PTV VISSIM, the PTV Vistro 2020 software was used to optimize signal timings for major signalized intersections along the Lyons Road detour segment. The software exported Ring Barrier Controller (RBC) files with optimized signal timings for each intersection. The RBC files were imported into the VISSIM software to provide priority rules at signalized intersections along the considered detour.

7.2.2 Model Calibration and Validation Processes

The Florida's Turnpike Enterprise provided the VISSIM model containing the Turnpike network, and the initial calibration and validation process is well documented in their VISSIM model calibration report (Florida's Turnpike Enterprise, 2017). The report provides details of the model development and calibration processes for the existing 2016 AM and PM peak conditions for the SW 10th Street project in Broward County. Calibration targets, including capacity, traffic volume, travel time, speed, intersection delay, queue length, and visualization, were used as recommended by the FDOT *Traffic Analysis Handbook* (FDOT, 2014). These targets were established based on average speed, vehicle flows, and queues to ensure the developed model replicates the existing traffic conditions.

The calibration process resulted in parameters indicated in Tables 7-3 and 7-4. Table 7-3 presents the freeway calibration parameters, and Table 7-4 shows the calibration parameters for the arterials. As mentioned before, the model included an extended segment of arterials outside the main construction project limits to include the real-world queues. Therefore, the calibration parameters for arterials were adopted for the merged detour model.

Table 7-3: Calibration Parameters for the Freeway Microsimulation Model

Lane Change Parameters		Default	Freeway Calibration Parameters
Necessary Lane Change (Route)			
Maximum deceleration		-13.12 ft/s^2 (Own) -9.84 ft/s^2 (Trail)	-13.12 ft/s^2 -9.84 ft/s^2
-1 ft/s^2 per distance		200 ft (Freeway)	200 ft
Accepted deceleration		-3.28 ft/s^2 (Own) -1.64 ft/s^2 (Trail)	-3.28 ft/s^2 -1.64 ft/s^2
Waiting time before diffusion		60 sec	180 sec
Minimum headway (front/rear)		1.64 ft	0.98 ft and 1.51 ft
To Slower Lane if Collision Time Above (seconds)		0.00	0.00
Safety distance reduction factor		0.6	0.25 and 0.40
Maximum deceleration for cooperative braking		-9.84 ft/s^2	-29.99 and -31.99 ft/s^2
Overtake reduced speed areas		Uncheck	Checked
Advanced Merging		Checked	Checked
Cooperative lane change		Unchecked	Checked especially for freeway merge/diverge areas
If Checked	Maximum Speed Difference	6.71 mph	6.71 mph
	Maximum Collision Time	10 sec	10 sec

Table 7-4: Calibration Parameters for the Arterial Microsimulation Model

Lane Change Parameters		Default	Arterial Calibration Parameters
Necessary Lane Change (Route)			
Maximum deceleration		-13.12 ft/s^2 (Own) -9.84 ft/s^2 (Trail)	-13.12 ft/s^2 -9.84 ft/s^2
-1 ft/s^2 per distance		100 ft (Arterial)	100 ft
Accepted deceleration		-3.28 ft/s^2 (Own) -3.28 ft/s^2 (Trail)	-3.28 ft/s^2 -3.28 ft/s^2
Waiting time before diffusion		60 sec	180 sec
Minimum headway (front/rear)		1.64 ft	1.51 ft
To Slower Lane if Collision Time Above (seconds)		0.00	0.00
Safety distance reduction factor		0.6	0.25, 0.40, 0.50
Maximum deceleration for cooperative braking		-9.84 ft/s^2	-29.99 and -31.99 ft/s^2
Overtake reduced speed areas		Uncheck	Checked
Advanced Merging		Checked	Checked
Cooperative lane change		Unchecked	Checked
If Checked	Maximum Speed Difference	6.71 mph	6.71 mph
	Maximum Collision Time	10 sec	10 sec

7.2.3 Incident Modeling

A lane blockage was considered to represent a primary incident. The trajectory files for conflict analysis were filtered based on spatiotemporal relation with the modeled primary incident. In other words, the traffic conflicts were considered to be directly influenced by the lane blockage. The

vehicular conflicts were therefore used as a proxy for secondary crash risk upstream of the primary incident location.

The VISSIM 2020 simulation software does not have a specific built-in incident formation module. Therefore, the present study adopted a stopped vehicle approach to simulate incident blockages. This approach utilizes the *"AddVehicleAtLinkPosition"* function that exists within the COM interface. The function lets users add or remove a vehicle at a specific time at a chosen location (Chou and Miller-Hooks, 2011). To replicate an incident with this function, a stationary vehicle is added and stays at the incident location for the chosen incident duration. It is set to be removed when the incident is cleared. To simulate situations where two or more lanes are blocked, multiple vehicles with the same time of placement and removal can be added to adjacent lanes.

A total of 90 scenarios were considered in the analysis, as described in Table 7-5. The scenarios involved one outer lane closed, one inner lane closed, and two outer lanes closed, performed at varying CV MPRs. The present study adopted an incident duration of 30 minutes for all the lane blockage scenarios. The chosen span was between the minimum and maximum duration for one lane and two lanes closed recommended by HCM (HCM, 2016). The 90 scenarios described in Table 7-5 were simulated by incorporating Visual Basic scripts in the VISSIM model through the event-based script, enabling creating the CV environment and controlling traffic behaviors during the simulation time. In the simulation model, a resolution is defined as simulation time-steps per second. In other words, the resolution is the time steps in a second for which vehicle data are collected and can be accessed during simulation time. The resolution was set to 10-time steps per simulation second, which replicate a transmission frequency of 10 Hz for the Basic Safety Messages (BSMs). Per the recommendation provided in the VISSIM User Manual, five simulation runs were conducted for each scenario, with a random seed increment of 10 for each run (PTV, 2020).

Table 7-5: Lane Blockage Scenarios

Scenario	Period	CV MPR
One outer lane blocked (AM)	Pre-peak, Peak, Post-peak	0%, 25%, 50%, 75%, and 100%
One outer lane blocked (PM)	Pre-peak, Peak, Post-peak	0%, 25%, 50%, 75%, and 100%
One inner lane blocked (AM)	Pre-peak, Peak, Post-peak	0%, 25%, 50%, 75%, and 100%
One inner lane blocked (PM)	Pre-peak, Peak, Post-peak	0%, 25%, 50%, 75%, and 100%
Two outer lanes blocked *without* detour (AM)	Pre-peak, Peak, Post-peak	0%, 25%, 50%, 75%, and 100%
Two outer lanes blocked *with* detour (AM)	Pre-peak, Peak, Post-peak	0%, 25%, 50%, 75%, and 100%

7.2.4 CV Applications in Microsimulation

VISSIM provides add-on APIs, such as COM, which allow users to incorporate their applications and access objects data during the simulation. The COM steps are task-specific depends on the objective of the study. A COM API was used to access vehicle attributes and integrate CV applications during the run time. COM API was used in steps such as a real-time collection of vehicle data, incident modeling, processing, and disseminating advisory messages, which involves changing CV's driving behavior to maximize safety benefits during the incident, and termination of incident-related messages.

To the extent possible, the study used VB-scripting and the COM API interface in VISSIM to simulate CV applications to mitigate secondary crashes. This method enhances the CV environment and modeling of traffic behavior and travel conditions during the simulation process. This section describes the procedures and techniques adopted to simulate CV applications to mitigate secondary crashes on freeways.

Roadside Unit

In the connected environment, RSUs collect and transmit a wide range of vehicle data from CVs that are in range. Modeling of RSU capabilities was considered in the COM environment to enable real-time data collection. The function "*GetbyLocation*" in the COM environment was used to return vehicle collection data within a 2-mile communication range. The following vehicle attributes were collected: vehicle identification (Veh ID), speed, desired speed, link identification (Link ID), lane number, desired lane, location, coordinate at front of each vehicle (CordFront), and vehicle type (Veh Type). The "*GetbyLocation*" mimics RSU functions in the real-world, which collects and communicates CV information within a range of two miles under DSRC technologies. All CV data were available for other processes in the COM, enabling the generation of advisory messages to be broadcast to CVs within range.

Advisory Messages

Figure 7-3 shows the algorithm used to process advisory messages to be disseminated to motorists in the presence of an incident. The algorithm provides speed, lane change, and/or detour advisory to drivers upstream of the incident location. The dashed rectangle in Figure 7-3 shows the part of the detour advisory algorithm, which was used for "*with detour"* scenarios. The algorithm first determines whether or not the vehicle approaching the incident location is CV. If the vehicle is a CV, the algorithm checks whether it is within the DSRC range of the RSU close to the incident scene. When the CV is within range, the algorithm adjusts the CV driving behavior and retrieves the real-time vehicle data necessary to process the advisory messages.

Notably, conventional vehicles must passively interact with each other as well as with the CVs. As such, drivers of these conventional vehicles may adjust their car-following behavior and lane-changing behavior after perceiving and assessing other surrounding vehicles' status. These induced interactions and the adjusted behavior of CVs may change the overall safety performance of the traffic flow. Besides, as the conventional vehicles approach the incident scene, they usually slow down and/or change lanes.

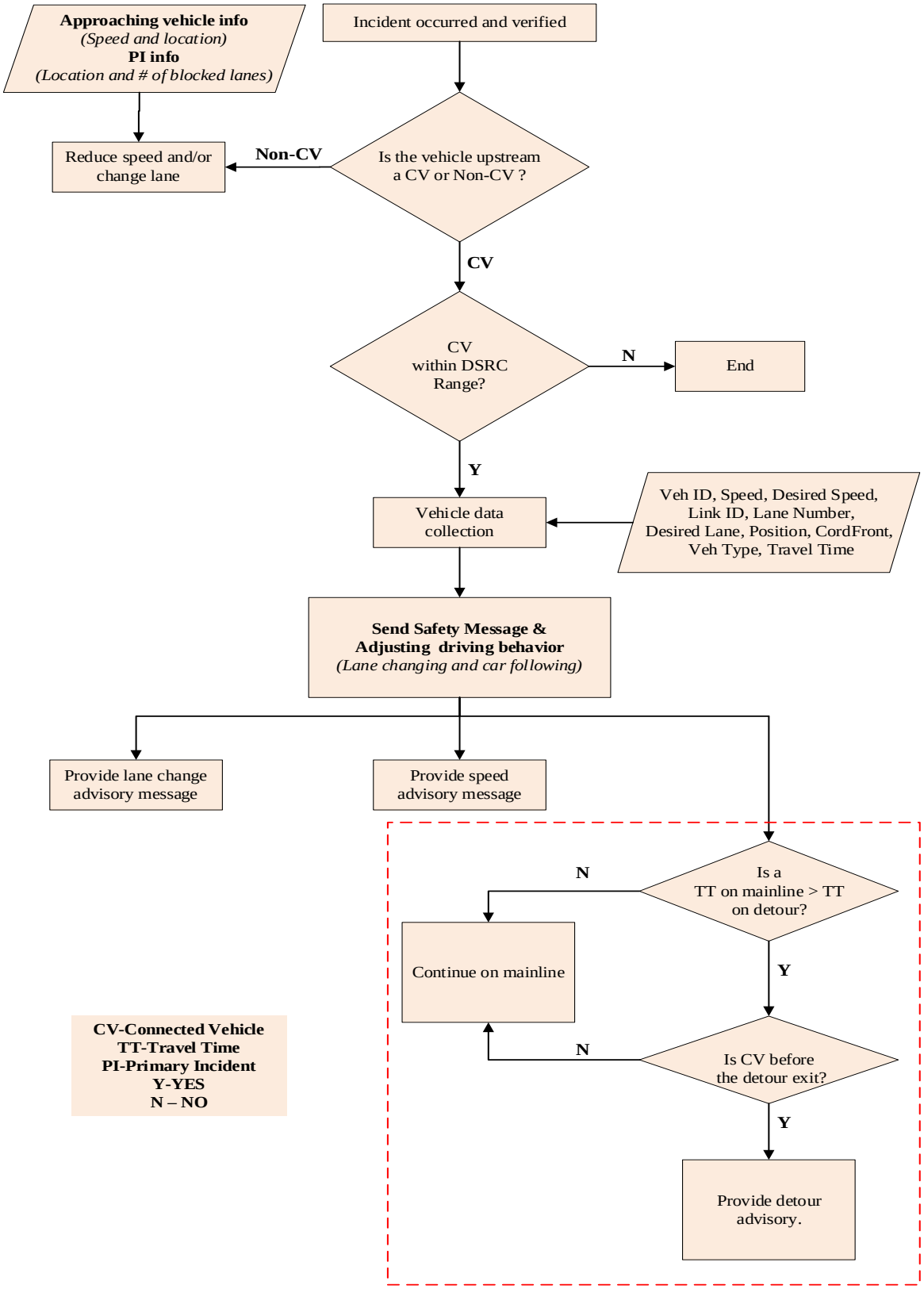


Figure 7-3: Algorithm to Process Advisory Messages

The kind of advisory message for an individual CV is a function of the position of that CV relative to an incident location. Moreover, the algorithm also considers conventional car behaviors as they approach close to the incident location. The following operational assumptions were made while modeling the V2I communication:

- An incident was modeled in the northbound direction, and the analysis considered only northbound traffic.
- The communication plan was set to operate at 10 Hz. i.e., the vehicle data collection, processing of an appropriate advisory message, and dissemination of advisory messages were done every 0.1 seconds, a time interval which has been used for BSMs (Kenney, 2011).
- The effective range of V2I communication was two miles, with the use of the DSRC mode.
- The study assumed no communication latency or information loss in V2I communication.
- Driver compliance with advisory messages was assumed to be 100%.

Driver Behavior in CV Environment

The continuous driving behavior adjustment (CDBA) method was used to model driving behavior in the CV environment in VISSIM. The method is used for applications that give drivers continuous instructions to adjust their driving behavior for a particular goal (i.e., to mitigate the risk of secondary crashes) during a specific period, as shown in Figure 7-4 (Songchitruksa et al., 2016). CDBA is applicable for CV applications, such as lane change assist system and variable speed advisory.

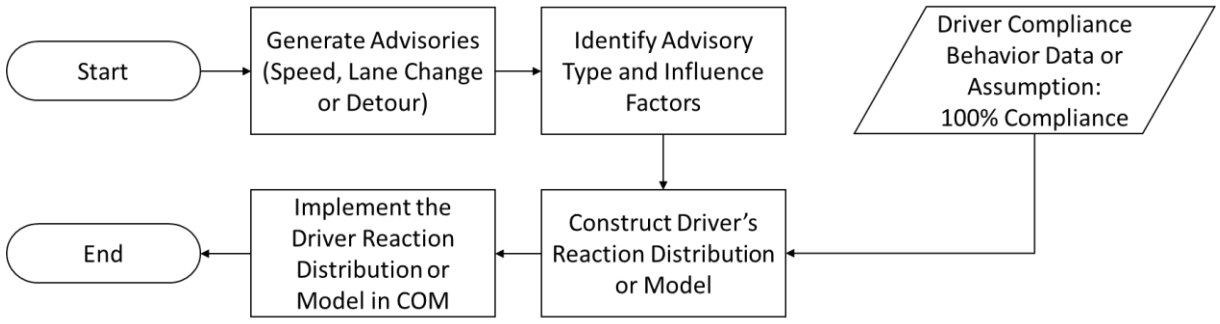


Figure 7-4: Process for Continuous Driving Behavior Adjustment

Different models, including car-following, lane-changing, and gap-acceptance models, control vehicle movements in microscopic simulation. While the car-following model controls the interaction between two vehicles in the same lane, the lane-changing model governs vehicles' lateral movements. On the other hand, the gap-acceptance model dictates the merging of vehicles to a destined lane. The findings from a previous study, which performed a sensitivity analysis of VISSIM's driver behavior parameters on vehicles' safety (Habtemichael and Santos, 2013), were used as a basis for adjusting CV driving behavior to enhance safety benefits.

In longitudinal driving behavior, the number of observed preceding vehicles (in the 'look ahead distance' parameter set) controls how well vehicles can predict and react to the movement of other vehicles in the link. For the car-following model, the simulated vehicle's safety distance is

governed by two parameters – CC0 (standstill distance) and CC1 (headway time), as shown in Equation 7-1.

$$Safety\ distance = CC0 + CC1 * speed \quad (7\text{-}1)$$

The safety distance is more affected by the multiplicative part (i.e., CC1) than the additive part (i.e., CC0), particularly for high-speed facilities such as freeways. A higher value of CC1 reflects a more cautious driver, enhancing safety by increasing the safety distance. The 'following' parameter (i.e., CC2) controls how much more distance than the desired safety distance a driver allows before intentionally moving closer to the leading vehicle. Thus, the CC1 and CC2 values were set slightly higher for CVs than for conventional vehicles. Another important car-following parameter that significantly impacts the simulated vehicle's safety is an empirical parameter called the threshold for entering 'following' (i.e., CC3). It controls the deceleration process, in particular, the number of seconds before reaching the safety distance. The driver recognizes a preceding slower vehicle at this stage, and the larger the CC# value, the more safety distance is assigned to the vehicle.

For the study corridor, both free and necessary lane-changing parameters were adjusted to enhance safe lane-changing maneuvers. In free lane-changing parameters, the safety distance reduction factor controls the gap which a vehicle accepts for a lane-changing maneuver. The larger the factor, the fewer the number of conflicts in the simulation. Regarding the necessary lane-changing parameters, factors such as lane-changing position, maximum deceleration of trailing vehicles, and acceleration reduction for trailing vehicles, are known to influence traffic safety (Habtemichael and Santos, 2013). In the present study, these three factors were adjusted to increase the safety of simulated CVs during an incident.

Driver Behavior Adjustments in CV Environment

A CV will change its behavior by adjusting car-following and lane-change behaviors when it receives the safety message regarding an incident (Yang et al., 2017). In the simulation model, this was done by creating a separate driving behavior for a vehicle class, *"CVs with the active message",* which contained adjusted driving behaviors discussed in the previous section and was added to the *main link behavior types* for links in the VISSIM model. This behavior was activated for all CVs within range by changing their vehicle type to one under the vehicle class "*CVs with the active message"*. The vehicle type was returned to its original type after passing the incident location, which restores the original driving behavior. This change of the vehicle class was performed within the event-based script. Figure 7-5 is a conceptual diagram that visualizes how lane change and detour advisories were disseminated in the simulation model. Speed advisory was also disseminated to CVs within range to facilitate a smooth reduction in speed as they pass the incident location. It was assumed that there is a connection between TMC and RSU, so messages from TMC reach CVs via RSU. In the VISSIM model, information such as advisory speed information, detours, and lane change were used as potential TIM messages from TMC and were disseminated during the incident.

Lane-change Advisory: Lane change messages were sent to all CVs within the communication range at a distance of 0.75 miles upstream of the incident location. Once received, the vehicles'

desired lanes were set to those not blocked by the incident, making them change lanes once they get sufficient gaps in the adjacent lanes.

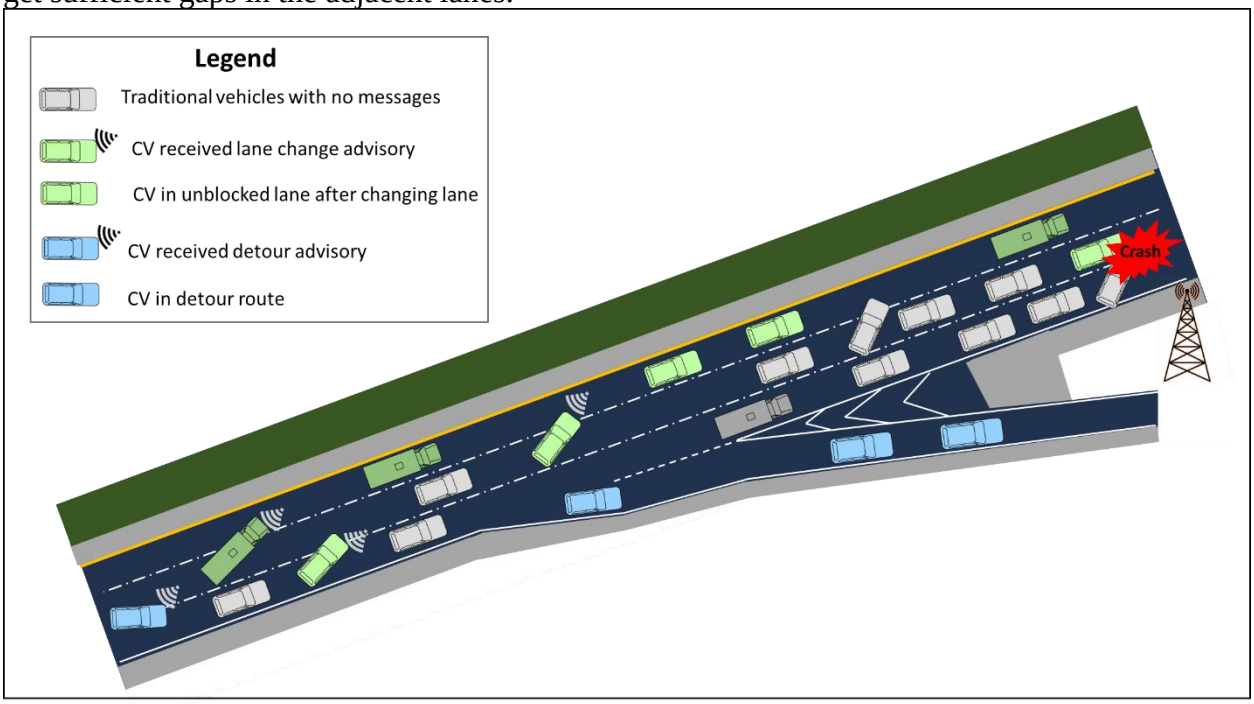


Figure 7-5: Flow of Advisory Information in Simulation Model

Speed Advisory: For speed advisory systems, speeds are sent directly to the On-Board Units (OBUs) of the individual vehicle through V2I communication (Grumert and Tapani, 2012). The suggested speed sent to a CV within a communication range depends on its distance from the incident location, its current speed, and the advised speed. The driver of the CV adjusts speed (100% driver compliance was assumed), based on the desired speed distribution generated, to attain a smooth deceleration rate as the driver approaches the new recommended speed point. Vehicles were only advised to reduce speeds when the average speed within 300 ft of the incident location dropped to 10% less than normal. Speed reductions were advised at 20 mph less than the speed limit within 0.5 miles upstream of the incident and 10 mph below the speed limit within 0.75 miles upstream of the incident. It should be noted that the speed advisory varied depending on whether the primary incident resulted in a single or a two-lane blockage.

Detour Advisory: Travel time was used as a performance measure to assess the impact of traffic diversion during the incident. Vehicles were advised to consider a detour whenever the travel time through the incident scene becomes longer than using a detour.

7.2.5 Safety Evaluation

The SSAM software uses developed algorithms to identify conflicts from vehicle trajectory files generated in traffic microscopic simulation software. It calculates surrogate measures of safety corresponding to each vehicle-to-vehicle interaction and determines whether or not each interaction satisfies the criteria to be deemed an official conflict. Thresholds for analyzed surrogate measures can be altered or changed in the software to match desired thresholds of analysis,

including TTC, PET, Maximum deceleration, Speed difference, and Deceleration Rate. Simulated conflicts analyzed in SSAM are categorized according to the degree of collision as either lane-change conflicts, rear-end conflicts, or crossing conflicts. Rear-end and lane-change conflicts, common in freeway operation, were considered to evaluate the potential of CVs in mitigating secondary crashes. A 2-mile section upstream of the incident location was created in the VISSIM model to collect vehicle trajectories. This section replicates the V2I communication range of the RSU at the incident location using DSRC technology. For each scenario, the vehicle trajectory files generated in each run were then imported into the SSAM software. The SSAM facilitates the identification of conflicts through statistical analysis of vehicle trajectory files generated from the microscopic simulation. TTC and PET parameters were used as measurable traffic indicators to obtain traffic conflict data in the SSAM. Specific thresholds were adopted to determine critical conflicts, i.e., 1.5 seconds for TTC and 5.0 seconds for PET, which are consistent with a previous study by Yang et al. (2017).

7.3 Results and Discussion

This section discusses the conflict analysis results from the SSAM software. The TTC and PET parameters were used as the surrogate safety measures; the adopted thresholds were 1.5 seconds and 5.0 seconds for TTC and PET, respectively. Notably, SSAM classified the conflicts as rear-end or lane-changing conflicts, which are likely to occur on the freeway. Thus, the obtained conflicts were used as a surrogate measure of secondary crashes.

7.3.1 Scenarios with One Lane Blocked

One of the three northbound lanes was closed to evaluate scenarios with one lane blocked. The study presents and compares conflicts resulting from scenarios with a blocked inner lane and a blocked outer lane. Additionally, the detour strategy was analyzed when the outer or inner lane was blocked, and the results are presented in this section.

Traffic Conflicts During AM Peak Period

Figure 7-6 presents the number of conflicts during the AM pre-peak hour, AM peak hour, and AM post-peak hour. It shows the number of rear-end conflicts, lane-change conflicts, and total conflicts found for each 25% increment of CV penetration (i.e., MPR of 0%, 25%, 50%, 75%, 100%) in each analysis period. Total conflicts are calculated as the sum of rear-end and lane-change conflicts. For the same analysis period and the same CV penetration rate, the graphs show results for the two one-lane blockage scenarios: *Inner Lane Blocked (ILB)* and *Outer Lane Blocked (OLB)*. Furthermore, Table 7-6 shows a matrix that summarizes the percentage change in the number of conflicts as the CV penetration rate increases.

Total Conflicts: Overall, for both the ILB and the OLB scenarios, as expected, CVs reduced total conflicts. In ILB scenarios, as the MPR of CVs was increased, a more or less similar trend in conflict reduction was observed during all the analysis periods (i.e., pre-peak, peak, and post-peak). On the other hand, a varying pattern was observed when the outermost lane was blocked. Notably, the number of total conflicts is higher in the OLB scenarios than ILB scenarios. This

finding could be due to a relatively high frequency of merge and diverge maneuvers when the outer lane was blocked.

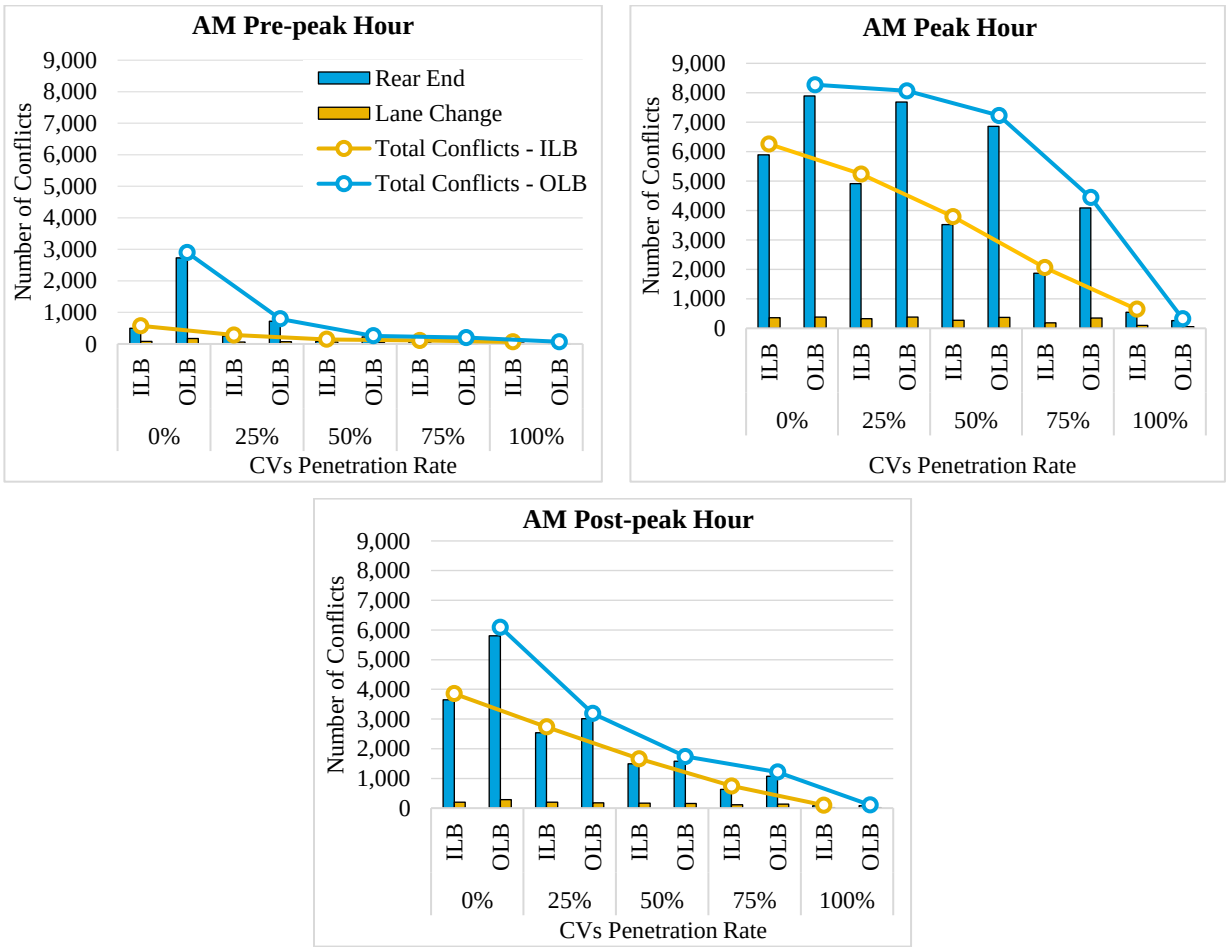


Figure 7-6: Traffic Conflicts for ILB and OLB Scenarios (AM Peak Period)

Regarding ILB scenarios, a reduced number of total conflicts, by up to 98%, with full CV market penetration was observed. A reduction in conflicts was observed even at 25% MPR of CVs during the pre-peak, peak, and post-peak periods, as shown in Table 7-6. A consistent trend in reduction in conflicts was observed for each 25% increment of CV market penetration. For the OLB scenarios, a pronounced reduction in conflicts was noticed at 25% CV MPR during the pre-peak and post-peak hours. In these periods, successive increments of CV composition beyond 25% resulted in a slight decline in conflicts.

Table 7-6: Percent Change in Conflicts for ILB and OLB Scenarios (AM Peak Period)

WHEN THE INNER LANE IS BLOCKED (ILB)

	AM Pre-peak hour				AM Peak hour				AM Post-peak hour			
Total Conflicts												
	Initial Composition (%)				Initial Composition (%)				Initial Composition (%)			
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-44.90%				-16.30%				-29.00%			
50	-75.00%	-65.90%			-39.40%	-27.60%			-56.90%	-39.30%		
75	-82.00%	-73.10%	-28.10%		-67.10%	-60.70%	-45.80%		-79.80%	-71.60%	-53.20%	
100	-89.60%	-86.20%	-58.20%	-41.90%	-94.90%	-94.00%	-91.60%	-84.60%	-97.50%	-96.50%	-94.20%	-87.70%
Rear-End Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-48.90%				-16.70%				-30.50%			
50	-82.60%	-65.90%			-40.20%	-28.30%			-59.00%	-41.00%		
75	-86.30%	-73.10%	-21.20%		-68.30%	-61.90%	-46.90%		-82.60%	-75.00%	-57.60%	
100	-92.90%	-86.20%	-59.50%	-48.50%	-95.60%	-94.80%	-92.70%	-86.20%	-97.90%	-96.90%	-94.80%	-87.70%
Lane-Change Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-19.80%				-10.70%				-2.20%			
50	-27.40%	-9.50%			-26.40%	-17.60%			-18.80%	-16.90%		
75	-55.30%	-44.20%	-38.40%		-48.60%	-42.40%	-30.10%		-28.90%	-27.30%	-12.50%	
100	-68.40%	-60.50%	-56.40%	-29.20%	-83.60%	-81.70%	-77.80%	-68.20%	-91.10%	-90.90%	-89.10%	-87.50%

WHEN THE OUTER LANE IS BLOCKED (OLB)

	AM Pre-peak hour				AM Peak hour				AM Post-peak hour			
Total Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-72.60%				-2.50%				-47.70%			
50	-91.10%	-69.30%			-12.60%	-10.40%			-71.50%	-45.50%		
75	-93.10%	-75.30%	-22.90%		-46.30%	-44.90%	-38.60%		-80.00%	-61.80%	-29.90%	
100	-97.60%	-92.30%	-72.90%	-64.90%	-96.20%	-96.10%	-95.60%	-92.90%	-98.30%	-96.70%	-93.90%	-91.40%
Rear-End Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-73.60%				-2.60%				-48.10%			
50	-91.90%	-69.30%			-13.10%	-10.80%			-72.90%	-47.70%		
75	-93.50%	-75.30%	-19.60%		-48.10%	-46.80%	-40.30%		-81.50%	-64.30%	-31.70%	
100	-98.00%	-92.30%	-74.90%	-68.70%	-96.70%	-96.70%	-96.20%	-93.70%	-98.50%	-97.20%	-94.60%	-92.10%
Lane-Change Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-57.60%				-0.80%				-38.60%			
50	-78.30%	-48.90%			-3.00%	-2.20%			-44.20%	-9.10%		
75	-87.50%	-70.60%	-42.40%		-8.80%	-8.10%	-6.00%		-51.10%	-20.30%	-12.40%	
100	-91.60%	-80.30%	-61.40%	-33.00%	-84.40%	-84.30%	-84.00%	-82.90%	-92.80%	-88.30%	-87.10%	-85.30%

Conflict reduction Conflict increase

However, during the peak hour, a greater reduction in conflicts was observed in the transition between 50% and 75% CV market penetration. It should be noted that the reduction in total conflicts was due to a reduction in either rear-end or lane-change conflicts or both. A detailed discussion on conflict reduction, per specific conflict type, is presented in the following subsections.

Rear-end Conflicts: In all scenarios, rear-end conflicts were more prominent than lane-change conflicts. This finding is expected since drivers on a freeway may not find enough time to react to the incident ahead of their path. This effect may propagate to upstream traffic in a short time and result in relatively more rear-end conflicts. These results are consistent with a previous study (Atamo, 2012), which also used VISSIM and SSAM tools and reported more rear-end conflicts than other conflicts on freeways. As shown in Figure 7-6, fewer rear-end conflicts were observed for ILB scenarios compared to OLB scenarios. This finding could be due to a relatively high frequency of merge and diverge maneuvers when the outer lane was blocked. Notably, for both OLB and ILB scenarios, the overall reduction in the number of rear-end conflicts follows a similar trend as the total conflicts. With a 25% composition of CVs, a vast conflict reduction was reported for ILB scenarios during pre- and post-peak hours, by about 45% and 29%, respectively. Similarly, for the OLB scenarios, 73% and 48% of conflicts were reduced during pre- and post-peak hours, respectively. However, with only 25% of CV penetration, a lower conflict reduction was reported (approximately 2% and 16% for the OLB and ILB, respectively) during the peak hour. Unlike in the peak hour, relatively lower traffic volumes for the pre- and post-peak hours might be the reason for the difference in conflict reduction among the peak periods. In general, rear-end conflicts decreased as the MPR of CVs increased. The high reductions in rear-end conflicts in the presence of CVs are associated with the decrease in vehicle speeds and change in driver behaviors, which were accounted for by early warnings of the incident and the advance advisory messages.

Lane-change Conflicts: Fewer lane-change conflicts were observed compared to rear-end conflicts. Only a slight difference in the number of lane-change conflicts was observed between the ILB and the OLB scenarios. As MPR of CVs increased, a relatively small reduction in lane-change conflicts resulted, despite which lane was blocked. The reduction was more pronounced in low traffic periods, pre- and post-peak hours in particular than in the peak hour. For the OLB scenarios, fewer conflict reductions were observed in the peak hour until 100% penetration of CVs. This is possibly due to a situation where either a leading or lagging vehicle is not a CV, resulting in not enough cooperation in creating a gap for a lane-changing maneuver. Thus, with 100% MPR of CVs, both leading and lagging vehicles are aware and behave cooperatively for lane-changing maneuvers, which results in a considerable reduction of the lane-change conflicts.

Traffic Conflicts During PM Peak Period

Figure 7-7 presents the number of conflicts during the PM peak period. It shows the conflicts and scenarios similar to those discussed in the earlier section (i.e., during the AM peak periods). The results are presented in a 25% increment of CV composition for pre-peak, peak, and post-peak hours. Again, the outermost (OLB) or innermost (ILB) lane blockage was considered in the analysis. Table 7-7 shows a matrix that summarizes the percentage change in the number of conflicts as the CV penetration rate increases.

Total Conflicts: The PM peak period analysis was found to experience more conflicts compared to the AM peak period. During the PM period, the OLB scenarios reported more conflicts compared to the ILB. As expected, the adoption of CVs was found to reduce the number of total conflicts. With a blocked inner lane, the increase in CV composition resulted in conflict reduction, with a similar trend for the pre-peak, peak, and post-peak hours. A large reduction was found in the transition from 75% to 100% composition of CVs. Another consistent pattern of conflict reduction was seen in the OLB scenarios. Up to 50% CV penetration, a 5% reduction was observed for the peak and post-peak hours, whereas it was approximately 16% for the pre-peak hour. The transition from 50% to 75% penetration of CVs resulted in a considerable reduction in total conflicts for the entire peak period. The following section gives a detailed discussion on conflict reduction per specific conflict type.

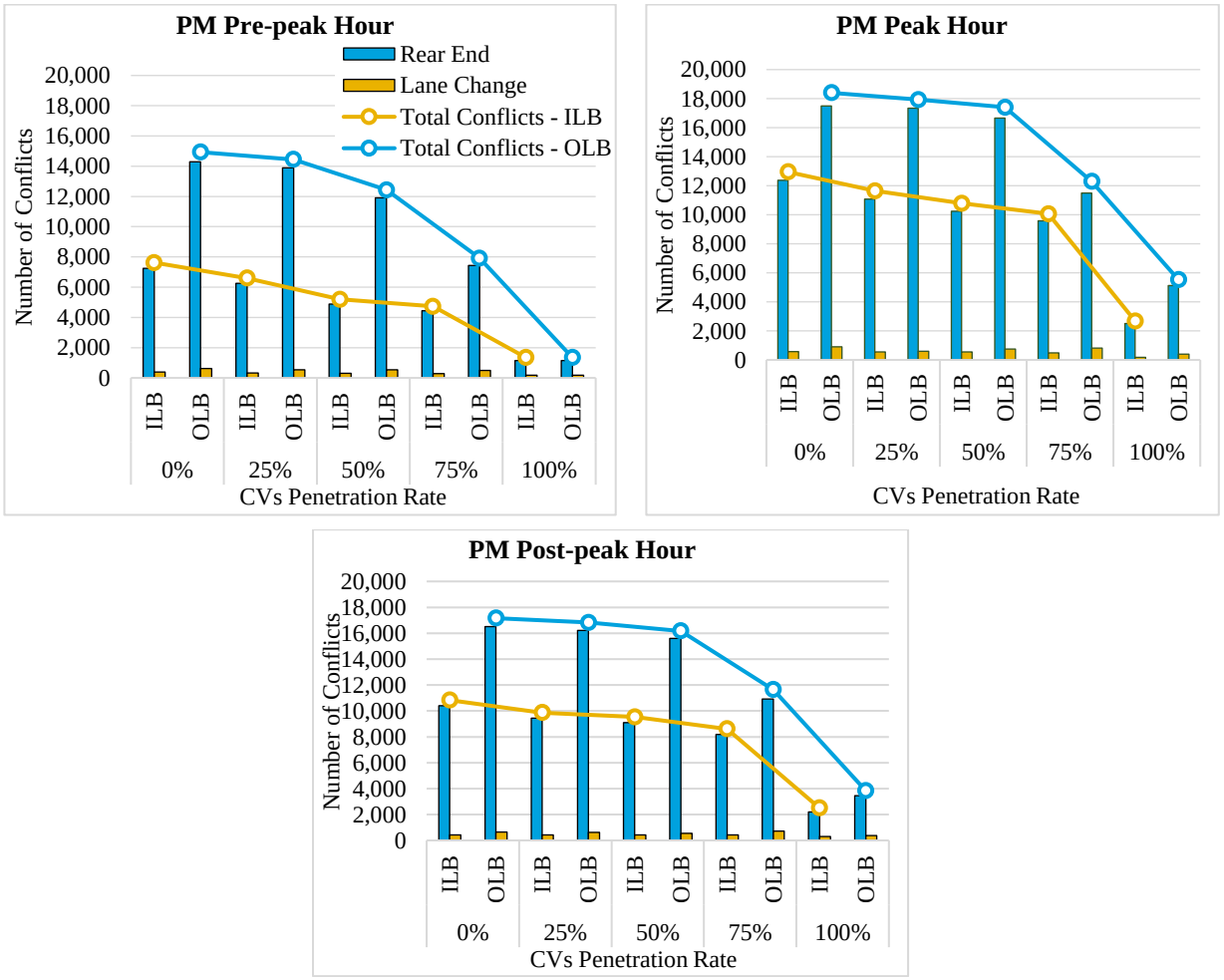


Figure 7-7: Traffic Conflicts for ILB and OLB Scenarios (PM Peak Period)

Rear-end Conflicts: There are many rear-end conflicts in the PM period than the corresponding AM peak period. These conflicts were more prominent than the lane-change conflicts. Unlike the AM peak period, the PM period experienced a lower reduction of rear-end conflicts with the increase in the CV MPR. For instance, with up to 50% of CV penetration for OLB scenarios, a 5% reduction was observed for the peak and post-peak hours, whereas it was about 16% for the pre-peak hour.

Table 7-7: Percent Change in Conflicts for ILB and OLB Scenarios (PM Peak Period)

WHEN THE INNER LANE IS BLOCKED (ILB)

PM Pre-peak Hour					PM Peak Hour				PM Post-peak Hour			
Total Conflicts												
	Initial Composition (%)				Initial Composition (%)				Initial Composition (%)			
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-13.50%				-10.90%				-9.20%			
50	-31.90%	-21.30%			-17.40%	-7.40%			-11.90%	-3.00%		
75	-38.60%	-29.10%	-9.80%		-22.90%	-13.60%	-6.70%		-20.50%	-12.50%	-9.70%	
100	-82.40%	-79.70%	-74.10%	-71.30%	-79.50%	-77.00%	-75.20%	-73.40%	-76.90%	-74.60%	-73.80%	-71.00%
Rear-End Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-13.40%				-11.20%				-9.20%			
50	-32.70%	-22.20%			-18.10%	-7.70%			-12.40%	-3.50%		
75	-38.70%	-29.20%	-9.00%		-23.30%	-13.60%	-6.40%		-21.20%	-13.30%	-10.10%	
100	-84.00%	-81.50%	-76.20%	-73.90%	-80.00%	-77.50%	-75.60%	-73.90%	-78.80%	-76.70%	-75.80%	-73.10%
Lane-Change Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-14.60%				-3.10%				-8.90%			
50	-18.50%	-4.50%			-3.90%	-0.80%			-1.70%	7.90%		
75	-37.40%	-26.70%	-23.20%		-15.00%	-12.20%	-11.50%		-3.10%	6.40%	-1.40%	
100	-53.10%	-45.10%	-42.50%	-25.10%	-69.50%	-68.50%	-68.20%	-64.10%	-31.10%	-24.40%	-30.00%	-29.00%

WHEN THE OUTER LANE IS BLOCKED (OLB)

PM Pre-peak Hour					PM Peak Hour				PM Post-peak Hour			
Total Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-3.30%				-2.50%				-1.90%			
50	-16.70%	-13.90%			-5.40%	-2.90%			-5.80%	-3.90%		
75	-47.00%	-45.20%	-36.30%		-33.20%	-31.40%	-29.40%		-32.20%	-30.90%	-28.00%	
100	-91.00%	-90.70%	-89.20%	-83.10%	-70.00%	-69.20%	-68.30%	-55.10%	-77.60%	-77.20%	-76.20%	-67.00%
Rear-End Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-2.80%				-0.90%				-1.70%			
50	-16.80%	-14.40%			-4.80%	-3.90%			-5.40%	-3.70%		
75	-48.00%	-46.50%	-37.50%		-34.30%	-33.70%	-31.00%		-33.90%	-32.70%	-30.10%	
100	-91.90%	-91.70%	-90.30%	-84.40%	-70.80%	-70.50%	-69.30%	-55.50%	-79.10%	-78.70%	-77.90%	-68.30%
Lane-Change Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-14.10%				-33.80%				-6.40%			
50	-15.70%	-1.80%			-16.40%	26.20%			-15.00%	-9.20%		
75	-22.90%	-10.20%	-8.50%		-10.40%	35.20%	7.20%		10.20%	17.60%	29.60%	
100	-71.10%	-66.40%	-65.70%	-62.60%	-55.00%	-32.00%	-46.10%	-49.80%	-41.10%	-37.10%	-30.70%	-46.50%

Conflict reduction Conflict increase

There was a greater reduction in total conflicts at low penetration rates (up to 50%) for the ILB than the OLB scenarios. Similar to the AM period results, merge and diverge maneuvers may have contributed to more rear-end conflicts in the OLB than in the ILB scenarios. Notably, in both scenarios, maximum benefits in conflict reduction (about 80%) were reported at a full CVs penetration. A lower reduction in conflicts at lower penetration of CVs during the PM peak hours was observed. This result could be because of the higher traffic volume along the study corridor. Higher traffic volumes cause vehicles to move close to each other, and when the incident happens, only CVs get information and advisory messages. Thus, at low CV penetration rates, there are more conventional vehicles with neither incident information nor advisory messages. Therefore, when a CV receives a speed reduction warning, a conventional vehicle(s) behind the CV may delay reducing speed. Consequently, the lack of cooperation among conventional vehicles and CVs to perform safe maneuvers may have resulted in more rear-end conflicts. Nevertheless, with the full penetration, all vehicles receive an early warning and speed advisory, resulting in a considerable reduction in rear-end conflicts.

Lane-change Conflicts: High traffic demand during the PM peak period is associated with fewer lane-change conflicts as CV composition increased. The maximum reduction (approximately 70%) was experienced with full penetration of CVs during the PM pre-peak hour under the OLB scenario. Surprisingly, in some scenarios, conflicts increased with the increase in CV composition, as shown in Tables 7-8, 7-9, and 7-10. For example, during the peak hour for the OLB scenarios, conflicts increased as CV penetration increased from 25% to 75% (Table 7-9). This result may be attributed to many CVs receiving a lane-change advisory, unlike conventional vehicles. Thus, the lack of cooperation for lane-changing maneuvers among vehicles results in an overall increase in lane-change conflicts.

Statistical Analysis

The surrogate safety indicators used for statistical analysis were the number of conflicts obtained from the SSAM safety evaluation. For a standard comparison among scenarios with varying CV compositions, a 2-mile segment (CV communication range) upstream of the primary incident location was used as an exposure variable. The objective was to check the significance of CV technologies at various penetration rates in reducing traffic conflicts, which are considered as a surrogate measure for secondary crashes. The statistical analysis of the average number of conflicts employed a one-tailed t-test with the null and alternative hypotheses. The student *t*-test was used to determine if the means of two sets of data are significantly different from each other.

The null hypothesis was that the mean difference between the average number of conflicts between 0% and a succeeding percentage of CV compositions is zero. The null hypothesis was tested against an alternate hypothesis that the mean difference between the average number of conflicts between the two scenarios is greater than zero.

- Null Hypothesis, H_0: $\mu_1 - \mu_2 = 0$, OR $\mu_1 = \mu_2$
- Alternate Hypothesis, H_A: $\mu_1 - \mu_2 > 0$, OR $\mu_1 > \mu_1$
  where,
  μ_1 = mean number of conflicts at 0% penetration of CVs, and
  μ_2 = mean number of conflicts at *i*% penetration of CVs.

Tables 7-8 and 7-9 show the statistical test results at a 95% confidence level for the ILB and the OLB scenarios, respectively. Furthermore, the statistical test was done to check whether there is a significant difference in the number of conflicts between the ILB and OLB scenarios. The null hypothesis was that, for a given MPR of CVs, the mean difference between the average number of conflicts between the ILB and the respective OLB scenario is zero. The null hypothesis was tested against an alternate hypothesis that the mean difference between the average number of conflicts between the two scenarios is less than zero.

- Null Hypothesis, H_0: $\mu_i - \mu_o = 0$, OR $\mu_i = \mu_o$
- Alternate Hypothesis, H_A: $\mu_i - \mu_o < 0$, OR $\mu_i < \mu_o$
  where,
  μ_i = mean number of conflicts for the ILB scenario at *i*% penetration of CVs, and
  μ_o = mean number of conflicts for the OLB scenario at *i*% penetration of CVs.

For the ILB scenario, the difference in total conflicts and rear-end conflicts was statistically significant for every 25% increment in CV composition. Greater penetration of CVs was required to significantly reduce lane-change conflicts during the AM post-peak, PM peak, and PM post-peak hours. During low traffic volume (AM pre- and post-peak hours), under OLB scenarios, a significant reduction in total and rear-end conflicts was observed as the MPR of CVs increased. However, at high traffic volumes (AM peak and PM period), there was a significant reduction beyond 50% CV penetration. This shows that CV applications are a viable solution for reducing rear-end conflicts and lane-change conflicts. Even at low penetration of CVs, the results are more pronounced for low traffic volume situations than during congested periods.

Table 7-8: Summary of Paired *t*-Test Results for Number of Conflicts for ILB Scenarios

Period	CV Composition	N	Total Conflicts				Rear-end Conflicts				Lane-change Conflicts			
			Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant
AM Pre-peak hour	0%	5	574	125			495	126			79	3		
	25%	5	280	128	<.00318	YES	253	91	0.004	YES	63	10	0.004	YES
	50%	5	144	19	<.001	YES	86	16	<.001	YES	57	6	<.001	YES
	75%	5	112	28	<.001	YES	68	25	<.001	YES	42	14	<.001	YES
	100%	5	64	21	<.001	YES	35	11	<.001	YES	31	14	<.001	YES
AM Peak hour	0%	5	6259	313			5895	313			364	21		
	25%	5	5237	453	0.0016	YES	4912	439	0.002	YES	325	20	0.009	YES
	50%	5	3792	581	<.001	YES	3524	547	<.001	YES	268	35	<.001	YES
	75%	5	2057	505	<.001	YES	1870	491	<.001	YES	187	21	<.001	YES
	100%	5	643	25	<.001	YES	540	98	<.001	YES	103	15	<.001	YES
AM Post-peak hour	0%	5	3855	302			3653	290			203	14		
	25%	5	2738	16	<.001	YES	2540	15	<.001	YES	198	14	0.199	NO
	50%	5	1663	231	<.001	YES	1498	220	<.001	YES	165	11	0.02	NO
	75%	5	747	53	<.001	YES	635	45	<.001	YES	113	18	0.003	YES
	100%	5	96	5	<.001	YES	78	5	<.001	YES	17	5	<.001	YES
PM Pre-peak hour	0%	5	7626	273			7238	247			388	28		
	25%	5	6597	701	0.008	YES	6266	690	0.009	YES	331	11	0.001	YES
	50%	5	5191	302	<.001	YES	4875	297	<.001	YES	316	14	<.001	YES
	75%	5	4734	226	<.001	YES	4439	211	<.001	YES	295	32	<.001	YES
	100%	5	1343	84	<.001	YES	1160	84	<.001	YES	182	21	<.001	YES
PM Peak hour	0%	5	12948	634			12378	615			570	40		
	25%	5	11645	82	<.001	YES	11090	69	<.001	YES	556	14	0.2	NO
	50%	5	10785	468	<.001	YES	10234	450	<.001	YES	551	18	0.2	NO
	75%	5	10066	725	<.001	YES	9578	676	<.001	YES	488	49	0	YES
	100%	5	2675	229	<.001	YES	2500	255	<.001	YES	175	25	<.001	YES
PM Post-peak hour	0%	5	10830	420			10394	423			436	4		
	25%	5	9872	196	<.001	YES	9438	205	<.001	YES	434	35	0.455	NO
	50%	5	9537	609	0.002	YES	9108	595	0.002	YES	428	14	0.141	NO
	75%	5	8609	770	<.001	YES	8186	756	<.001	YES	422	19	0.081	NO
	100%	5	2500	383	<.001	YES	2204	354	<.001	YES	296	36	<.001	YES

Note: CV = Connected Vehicle; N = Number of Microsimulations Used to Generate Conflicts; SD = Standard Deviation.

Table 7-9: Summary of Paired *t*-Test Results for Number of Conflicts for OLB Scenarios

Period	CV Composition	N	Total Conflicts				Rear-end Conflicts				Lane-change Conflicts			
			Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant
AM Pre-peak hour	0%	5	2,897	136			2,727	163			170	3		
	25%	5	793	157	<.001	YES	721	96	<.001	YES	72	10	<.001	YES
	50%	5	257	19	<.001	YES	221	14	<.001	YES	37	5	<.001	YES
	75%	5	199	32	<.001	YES	178	27	<.001	YES	21	14	<.001	YES
	100%	5	70	15	<.001	YES	56	6	<.001	YES	14	14	<.001	YES
AM Peak hour	0%	5	8,273	413			7,891	413			382	28		
	25%	5	8,064	589	0.2674	NO	7,686	570	0.266	NO	379	26	0.428	NO
	50%	5	7,227	523	0.003977	YES	6,856	492	0.003	YES	371	31	0.294	NO
	75%	5	4,441	595	<.001	YES	4,092	580	<.001	YES	349	25	0.044	YES
	100%	5	317	91	<.001	YES	257	87	<.001	YES	60	14	<.001	YES
AM Post-peak hour	0%	5	6,096	423			5,809	363			287	13		
	25%	5	3,189	18	<.001	YES	3,013	17	<.001	YES	176	1	<.001	YES
	50%	5	1,736	295	<.001	YES	1,576	264	<.001	YES	160	11	<.001	YES
	75%	5	1,217	62	<.001	YES	1,076	50	<.001	YES	140	24	<.001	YES
	100%	5	105	5	<.001	YES	85	4	<.001	YES	21	5	<.001	YES
PM Pre-peak hour	0%	5	14,930	437			14,300	358			630	34		
	25%	5	14,441	1051	0.183	NO	13,900	1020	0.216	NO	541	13	<.001	YES
	50%	5	12,432	484	<.001	YES	11,901	506	<.001	YES	531	20	<.001	YES
	75%	5	7,919	408	<.001	YES	7,433	359	<.001	YES	486	42	<.001	YES
	100%	5	1,342	42	<.001	YES	1,160	75	<.001	YES	182	40	<.001	YES
PM Peak hour	0%	5	18,400	1331			17,500	1230			900	60		
	25%	5	17,933	139	0.2	NO	17,337	109	0.388	NO	596	20	<.001	YES
	50%	5	17,407	739	0.1	NO	16,655	630	0.104	NO	752	23	<.001	YES
	75%	5	12,298	1485	<.001	YES	11,492	1284	<.001	YES	806	80	0	YES
	100%	5	5,520	46	<.001	YES	5,115	122	<.001	YES	405	40	<.001	YES
PM Post-peak hour	0%	5	17,160	883			16,500	847			660	6		
	25%	5	16,832	334	0.2	NO	16,214	325	0.25	NO	618	53	0.059	NO
	50%	5	16,170	961	0.064	NO	15,609	833	0.066	NO	561	18	<.001	YES
	75%	5	11,636	1578	<.001	YES	10,909	1437	<.001	YES	727	31	<.001	YES
	100%	5	3,843	77	<.001	YES	3,454	170	<.001	YES	389	56	<.001	YES

Note: CV = Connected Vehicle; N = Number of Microsimulations Used to Generate Conflicts; SD = Standard Deviation.

Table 7-10: Summary of Paired *t*-Test Results Comparing Conflicts for ILB and OLB Scenarios

Period	CV Composition	N	Total Conflicts				Rear-end Conflicts				Lane-change Conflicts			
			Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant
AM Pre-peak hour	0%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	25%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.093	NO
	50%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	75%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.024	YES
	100%	5	N/A	N/A	0.311	NO	N/A	N/A	0.003	YES	N/A	N/A	0.046	YES
AM Peak hour	0%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.146	NO
	25%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.003	YES
	50%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	75%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	100%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
AM Post-peak hour	0%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	25%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	50%	5	N/A	N/A	0.336	NO	N/A	N/A	0.312	NO	N/A	N/A	0.259	YES
	75%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.038	YES
	100%	5	N/A	N/A	0.008	YES	N/A	N/A	0.024	YES	N/A	N/A	0.142	NO
PM Pre-peak hour	0%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	25%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	50%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	75%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	100%	5	N/A	N/A	0.489	NO	N/A	N/A	0.498	NO	N/A	N/A	0.5	NO
PM Peak hour	0%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	25%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.003	YES
	50%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	75%	5	N/A	N/A	0.008	YES	N/A	N/A	0.009	YES	N/A	N/A	<.001	YES
	100%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
PM Post-peak hour	0%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	25%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	50%	5	N/A	N/A	0.002	YES	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES
	75%	5	N/A	N/A	<.001	YES	N/A	N/A	0.003	YES	N/A	N/A	<.001	YES
	100%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.007	YES

Note: CV = Connected Vehicle; N = Number of Microsimulations Used to Generate Conflicts; SD = Standard Deviation.

Detour Strategy

The analysis considered both the AM and PM periods with either an innermost (ILB) or outermost (OLB) lane blockage. The analysis used travel time as a criterion to advise traffic to take a detour. Surprisingly, even at high traffic volumes (during the PM peak period), it rarely happened that travel time in the mainline traffic was greater than using a detour. Thus, less than 2% of all vehicles received a detour advisory message during the entire analysis period. This is because the interchanges along the study corridor are spaced further apart (about two miles) to enhance mobility along the mainline.

Additionally, the detour route has a lower posted speed limit (35 mph) and four signalized intersections that the traffic must navigate through before re-entering the freeway. Consequently, it takes much longer for traffic to use the detour to get back to the mainline. This finding is in line with a previous study, which concluded that a detour strategy could potentially be beneficial when incident on the mainline involves multiple lane closures and takes over 30 minutes to clear (Chou and Miller-Hooks, 2011). Overall, in the present study, the safety evaluation using the SSAM indicated a more or less similar conflict reduction pattern for both *with* and *without* detour scenarios. These results necessitate further analysis of the use of CVs and a detour advisory when an incident results in the blockage of two lanes.

7.3.2 Scenario with Two Lanes Blocked

The corridor segment used for analyzing scenarios with a single lane blockage was also used to analyze scenarios with two lanes blocked. The incident location was consistent with scenarios in which a single lane was closed. Two outer lanes, out of the three northbound lanes, were closed. The study analyzed the blockage of outer lanes since the results from a single lane closure showed higher traffic conflicts for the outer lane blockage. The following subsections discuss the conflicts resulting from scenarios in which the two outermost lanes were blocked, both *with* and *without* detour advisory.

Traffic Conflicts During AM Peak Period

Three morning peak hours consisting of low, high, and moderate traffic volume were chosen to analyze scenarios in which two outer lanes are closed. In all three analysis periods, an incident that blocks two lanes resulted in a rapid formation of queues upstream of the incident location. The rate of queue formation was a function of the analysis period when the incident occurred; the upstream queues were found to be most severe during the peak period, with fewer queues during the pre-peak hour. This could be due to the difference in traffic volumes among the analyzed periods. Unlike scenarios where a single lane was closed, more vehicles experienced longer travel times using the Turnpike mainline than if they opted for the Lyons Road (parallel arterial) detour. This finding is consistent with previous research (Chou and Miller-Hooks, 2011).

Total Conflicts: Surprisingly, during the peak period, for both *with* and *without* detour advisory, 25% CV penetration increased total conflicts (Figure 7-8). This could be due to fewer CVs in the network that receive less cooperation from conventional vehicles while the CVs adjusted their driving behavior for safety benefits. A reduction in total conflicts was observed from 50%

penetration of CVs in scenarios *with* detour advisory, a slightly higher reduction than those *without* the advisory. With full MPR of CVs, a decrease in the total conflicts of about 20% and 15% for *with* and *without* detour advisory, respectively, was observed.

For pre- and post-peak scenarios, with full CV market penetration, a reduced number of total conflicts of up to 71% and 64% for *with* and *without* detour advisory, respectively, was observed. The number of conflicts were reduced, even at 25% MPR of CVs, during these scenarios, as shown in Table 7-11. A consistent trend in conflict decline was observed for every 25% increment of CV composition. The following section gives a detailed discussion on conflict reduction, per specific conflict type.

Rear-end Conflicts: As expected, in all scenarios, rear-end conflicts were more prominent than lane-change conflicts. The blocking of two lanes disrupted the traffic on a high-speed facility, i.e., freeway, which resulted in long queues, creating many stop-and-go situations for vehicles. Consequently, more rear-end conflicts were observed than during the blockage of a single lane, as observed in a previous study (Atamo, 2012). As shown in Figure 7-8, greater reduction in rear-end conflicts was observed when the detour advisory was disseminated under the increased penetration of CVs. This could be due to the reduction of traffic approaching the incident scene after the detour advisory. Moreover, early speed and lane-change advisories to CVs help drivers adjust their driving behavior as they approach the scene.

During the peak period, queue dissipation to normal traffic conditions was not observed during the analysis period. A similar finding was observed by Pulugurtha and Mahanthi (2016). Greater conflict reduction (about 25%) was observed in the transition from 25% to 50% penetration of CVs for both *with* and *without* detour scenarios. The difference in conflict reduction between *with* and *without* detour advisory was less for the transitions to higher CV penetration rates. This is because, during the peak period, Lyons Road also experience high traffic. Thus, only few CVs received detour advisory.

For pre-peak scenarios, conflict reduction was observed even at 25% MPR of CVs. The conflict reduction was 55% and 49% for *with* and *without* detour advisory. The parallel arterial used as a detour had less traffic during the pre-peak hour. Thus, more CVs received the detour advisory and increased the safety benefit, even at lower penetration. A similar trend in the reduction of rear-end conflicts was observed during the post-peak hour.

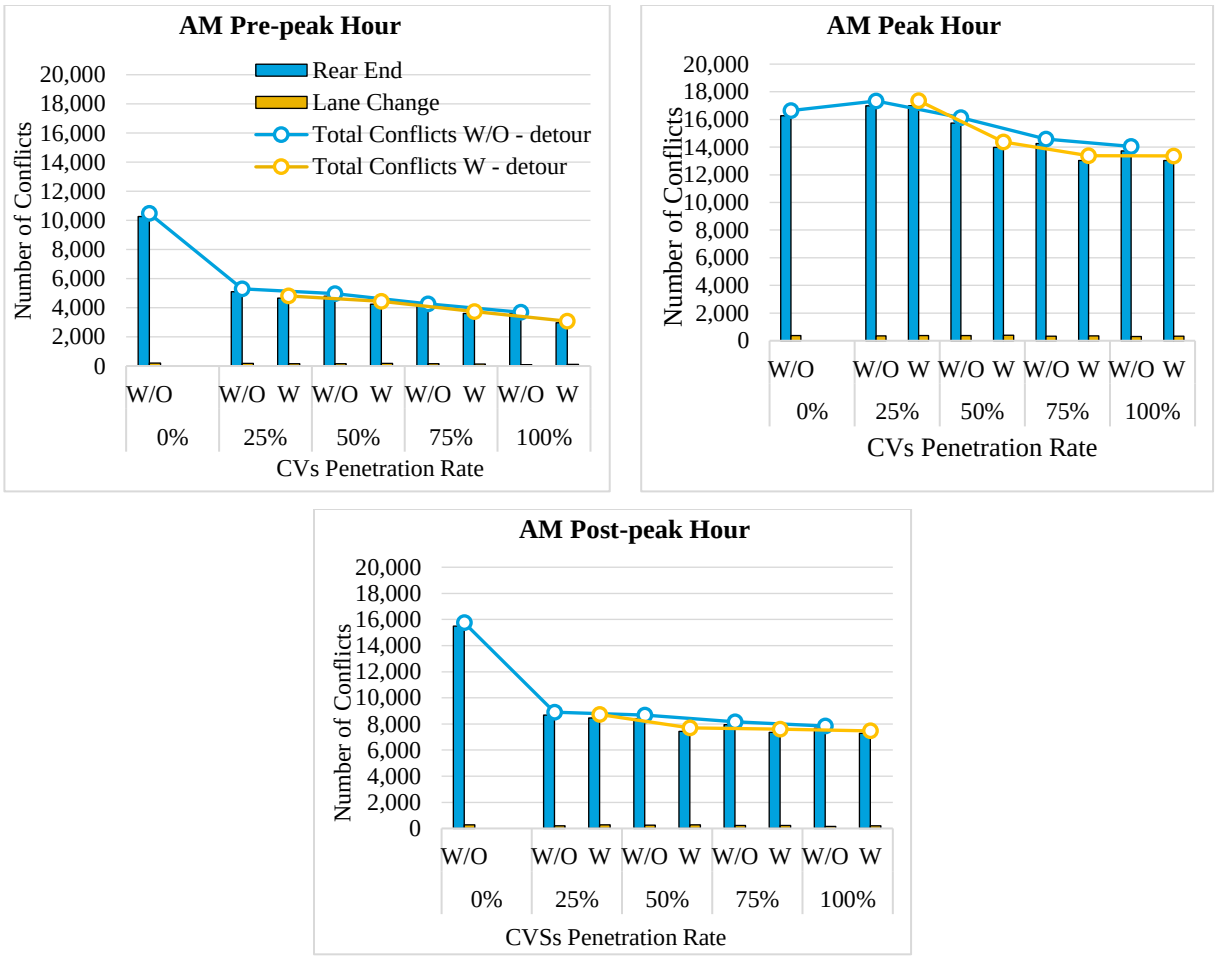


Figure 7-8: Traffic Conflicts For Scenarios With Two Outer Lanes Blocked (AM Peak Period)

Lane-change Conflicts: In all the scenarios, fewer lane-change conflicts were observed than rear-end conflicts, as shown in Figure 7-8. As discussed in the analysis of a single lane closure, CVs in congested periods experienced less cooperation from conventional vehicles during lane-changing maneuvers. Thus, there was an increase in lane-change conflicts with lower CV penetration in high traffic volume, as shown in Table 7-11. Moreover, more conflicts were observed for scenarios with the detour advisory than those *without* the advisory. The reason could be due to an increase in conflicts as CVs changed lanes to access the detour route.

Table 7-11: Percent Change in Conflicts for Two Outer Lane Block Scenarios (*with* and *without* Detour Advisory)

TWO OUTER LANE BLOCKED - *WITHOUT* DETOUR ADVISORY												
AM Pre-peak Hour					AM Peak Hour				AM Post Hour			
Total Conflicts												
	Initial Composition (%)				Initial Composition (%)				Initial Composition (%)			
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-49.45%				4.11%				-43.56%			
50	-52.61%	-6.23%			-3.15%	-6.97%			-44.97%	-2.50%		
75	-59.26%	-19.41%	-14.05%		-12.38%	-15.84%	-9.53%		-48.11%	-8.07%	-5.71%	
100	-64.79%	-30.34%	-25.70%	-13.56%	-15.62%	-18.96%	-12.88%	-3.70%	-50.24%	-11.83%	-9.57%	-4.10%
Rear-End Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-50.25%				4.36%				-43.85%			
50	-53.23%	-5.98%			-3.23%	-7.27%			-45.61%	-3.12%		
75	-59.88%	-19.36%	-14.23%		-12.42%	-16.08%	-9.50%		-48.60%	-8.46%	-5.51%	
100	-65.05%	-29.76%	-25.28%	-12.89%	-15.62%	-19.15%	-12.81%	-3.66%	-50.41%	-11.68%	-8.83%	-3.52%
Lane-Change Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-10.65%				-6.64%				-27.23%			
50	-22.24%	-12.98%			0.42%	7.57%			-9.65%	24.16%		
75	-29.09%	-20.64%	-8.80%		-10.73%	-4.38%	-11.11%		-20.82%	8.81%	-12.36%	
100	-51.81%	-46.06%	-38.02%	-32.04%	-15.72%	-9.73%	-16.08%	-5.59%	-40.63%	-18.42%	-34.29%	-25.02%
TWO OUTER LANE BLOCKED - *WITH* DETOUR ADVISORY												
AM Pre-peak Hour					AM Peak Hour				AM Post Hour			
Total Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-54.08%				4.29%				-44.63%			
50	-57.63%	-7.73%			-13.61%	-17.17%			-51.10%	-11.69%		
75	-64.25%	-22.14%	-15.62%		-19.67%	-22.97%	-7.01%		-51.80%	-12.95%	-1.43%	
100	-70.63%	-36.05%	-30.69%	-17.86%	-19.73%	-23.03%	-7.08%	-0.08%	-52.61%	-14.41%	-3.09%	-1.68%
Rear-End Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-54.69%				4.36%				-45.39%			
50	-58.60%	-3.91%			-14.10%	-17.69%			-51.94%	-12.00%		
75	-64.93%	-10.24%	-15.29%		-19.92%	-23.27%	-6.78%		-52.45%	-12.93%	-1.05%	
100	-71.06%	-16.37%	-30.09%	-17.48%	-19.94%	-23.29%	-6.80%	-0.02%	-53.00%	-13.94%	-2.20%	-1.17%
Lane-Change Conflicts												
%CVs	0	25	50	75	0	25	50	75	0	25	50	75
25	-24.14%				1.01%				-2.45%			
50	-10.27%	18.30%			7.33%	6.26%			-4.25%	-1.85%		
75	-30.99%	-9.02%	-23.09%		-8.55%	-9.46%	-14.79%		-15.71%	-13.59%	-11.96%	
100	-49.81%	-33.83%	-44.07%	-27.27%	-10.52%	-11.41%	-16.63%	-2.15%	-30.84%	-29.10%	-27.77%	-17.95%
					Conflict reduction				Conflict increase			

Statistical Analysis

Table 7-12 shows the statistical test results at a 95% confidence level for the *with* and *without* detour advisory scenarios, respectively. The statistical test was performed to check whether there is a significant difference in the number of conflicts between the *with* and *without* detour advisory scenarios. The null hypothesis was that, for a given MPR of CVs, the mean difference between the average number of conflicts between the two scenarios is zero. It was tested against an alternate hypothesis that the mean difference between the average number of conflicts between the two scenarios is greater than zero.

- Null Hypothesis, H_0: $\mu_i - \mu_o = 0$, OR $\mu_i = \mu_o$
- Alternate Hypothesis, H_A: $\mu_i - \mu_o > 0$, OR $\mu_i > \mu_o$
  where,
  μ_i = mean number of conflicts for the *without* detour advisory at *i*% penetration of CVs
  μ_o = mean number of conflicts for the *with* detour advisory at *i*% penetration of CVs

For the periods with less traffic volume compared to that of the peak hour, i.e., pre- and post-peak hours, both *with* and *without* detour advisory, the reduction in total conflicts and rear-end conflicts was statistically significant for every 25% increment in CV composition. Greater penetration of CVs (about 75%) was required to significantly reduce lane-change conflicts during the post-peak hour. On the other hand, during high traffic volume (i.e., peak hour), a significant reduction in total and rear-end conflicts was observed in the transition from 50% to 75% of CV composition for scenarios *without* the detour advisory. For scenarios *with* detour advisory, a significant reduction was observed in the transition from 25% to 50% CVs. In regards to lane-change conflicts, a less significant benefit was observed during the peak hour.

When comparing scenarios *with* and *without* detour advisory, no significant reduction was observed in the number of rear-end conflicts using the detour strategy during the peak hour. Also, with full penetration of CVs, the number of conflicts reduced by adopting the detour advisory was not statistically significant. As the CVs take the detour, a point is reached where the travel time using the detour becomes longer than not taking the detour. Thus, no more CVs get diverted. Except for lane-change conflicts, other scenarios showed a statistical significance of using the detour to reduce traffic conflicts, as shown in Table 7-12.

Table 7-12: Summary of Paired *t*-Test Results for Number of Conflicts for Two Lanes Blocked Scenarios (*with* and *without* Detour Advisory)

Period	CV MPR	N	Total Conflicts				Rear-end Conflicts				Lane-change Conflicts			
			Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant
***Without* Detour Advisory**														
AM Pre-peak hour	0%	5	10,474	422			10,264	404			210	24		
	25%	5	5,294	229	<.001	YES	5,106	233	<.001	YES	188	7	0.04	YES
	50%	5	4,964	153	<.001	YES	4,801	156	<.001	YES	164	15	0.003	YES
	75%	5	4,267	280	<.001	YES	4,118	282	<.001	YES	149	18	0.001	YES
	100%	5	3,688	199	<.001	YES	3,587	200	<.001	YES	101	9	<.001	YES
AM Peak hour	0%	5	16,649	1319			16,272	1283			377	49		
	25%	5	17,334	1043	0.195	NO	16,982	1025	0.181	NO	352	23	0.168	NO
	50%	5	16,125	1387	0.279	NO	15,747	1372	0.274	NO	378	22	0.475	NO
	75%	5	14,587	1224	0.017	YES	14,251	1209	0.017	YES	336	25	0.071	NO
	100%	5	14,048	699	0.002	YES	13,730	702	0.002	YES	317	20	0.019	YES
AM Post-peak hour	0%	5	15,759	588			15,482	586			278	14		
	25%	5	8,895	133	<.001	YES	8,693	134	<.001	YES	202	23	<.001	YES
	50%	5	8,672	223	<.001	YES	8,421	235	<.001	YES	251	25	0.034	YES
	75%	5	8,177	297	<.001	YES	7,957	307	<.001	YES	220	19	<.001	YES
	100%	5	7,842	496	<.001	YES	7,677	497	<.001	YES	165	17	<.001	YES
***With* Detour Advisory**														
AM Pre-peak hour	0%	5	10,474	422			10,264	404			210	24		
	25%	5	4,810	397	<.001	YES	4,651	376	<.001	YES	160	27	0.007	YES
	50%	5	4,438	170	<.001	YES	4,249	175	<.001	YES	189	7	0.045	YES
	75%	5	3,745	306	<.001	YES	3,600	297	<.001	YES	145	13	<.001	YES
	100%	5	3,076	274	<.001	YES	2,971	267	<.001	YES	106	21	<.001	YES

Note: CV = Connected Vehicle; N = Number of Microsimulations Used to Generate Conflicts; SD = Standard Deviation.

Table 7-12 (Cont'd): Summary of Paired t-Test Results for Number of Conflicts for Two Lanes Blocked Scenarios (With and Without Detour Advisory)

Period	CV MPR	N	Total Conflicts				Rear-end Conflicts				Lane-change Conflicts			
			Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant	Mean	SD	*p*-value	Significant
AM Peak hour	0%	5	16,649	1319			16,272	1283			377	49		
	25%	5	17,363	470	0.144	NO	16,982	457	0.139	NO	380	26	0.442	NO
	50%	5	14,382	471	0.003	YES	13,978	480	0.003	YES	404	16	0.135	NO
	75%	5	13,374	159	<.001	YES	13,030	164	<.001	YES	344	11	0.096	NO
	100%	5	13,364	791	<.001	YES	13,027	769	<.001	YES	337	32	0.086	NO
AM Post-peak hour	0%	5	15,759	588			15,482	586			278	14		
	25%	5	8,726	123	<.001	YES	8,455	120	<.001	YES	271	15	0.237	NO
	50%	5	7,706	375	<.001	YES	7,440	366	<.001	YES	266	10	0.082	NO
	75%	5	7,596	149	<.001	YES	7,362	147	<.001	YES	234	4	<.001	YES
	100%	5	7,468	84	<.001	YES	7,276	84	<.001	YES	192	18	<.001	YES
Comparison Between *Without* and *With* Detour Advisory														
AM Pre-peak hour	25%	5	N/A	N/A	0.023	YES	N/A	N/A	0.025	YES	N/A	N/A	0.025	YES
	50%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.005	YES
	75%	5	N/A	N/A	0.011	YES	N/A	N/A	0.011	YES	N/A	N/A	0.349	NO
	100%	5	N/A	N/A	0.002	YES	N/A	N/A	0.002	YES	N/A	N/A	0.347	YES
AM Peak hour	25%	5	N/A	N/A	0.478	NO	N/A	N/A	0.5	NO	N/A	N/A	0.052	NO
	50%	5	N/A	N/A	0.014	YES	N/A	N/A	0.013	YES	N/A	N/A	0.034	YES
	75%	5	N/A	N/A	0.03	YES	N/A	N/A	0.028	YES	N/A	N/A	0.258	NO
	100%	5	N/A	N/A	0.093	NO	N/A	N/A	0.085	NO	N/A	N/A	0.142	NO
AM Post-peak hour	25%	5	N/A	N/A	0.035	YES	N/A	N/A	0.009	YES	N/A	N/A	<.001	YES
	50%	5	N/A	N/A	<.001	YES	N/A	N/A	<.001	YES	N/A	N/A	0.124	NO
	75%	5	N/A	N/A	0.002	YES	N/A	N/A	0.002	YES	N/A	N/A	0.07	NO
	100%	5	N/A	N/A	0.068	NO	N/A	N/A	0.057	NO	N/A	N/A	0.019	NO

Note: CV = Connected Vehicle; N = Number of Microsimulations Used to Generate Conflicts; SD = Standard Deviation.

7.4 Summary

This chapter presented the findings on efforts to quantify the potential benefits of CV applications in mitigating secondary crashes on freeways. The CV applications that utilize V2I communication, including speed advisory, lane-change advisory, and detour advisory, were explored. Advisory messages were disseminated to vehicles equipped with OBUs through V2I communication.

The study area included a 7.8-mile, 6-lane (3-lanes in each direction) road segment on SR-91. This site was chosen over other segments along the Florida's Turnpike due to its relatively high number of crashes in 2016-2019, based on observations from the Signal Four Analytics database. Furthermore, a 4.2-mile section of Lyons Road, with two lanes in each direction, was considered for detouring purposes. Compared to other possible detour routes along the Turnpike corridor, the selected detour is one of the sections with the closest consecutive interchanges (~ two miles).

A microscopic simulation approach was used to evaluate the effectiveness of CV applications to mitigate secondary crashes. The VISSIM microscopic simulation software was used to demonstrate traffic flow characteristics under the CV environment. As the first step, a calibrated VISSIM model was developed for the study area. The model mimics geometric and traffic characteristics for the selected corridor. Furthermore, incident and V2I communication were modeled using the V2X module in the VISSIM software through the COM API. This module simulates wireless communication and data exchange within the connected environment.

The following simulation scenarios were considered: blockage of either the inner lane, the outer lane, or the two outer lanes to represent a primary incident. The scenarios also considered the effect of varying traffic volumes by analyzing traffic data during AM and PM peak periods. Since full market penetration of CVs is not anticipated soon, a sensitivity analysis was performed by considering varying MPRs of CVs. In each scenario, the safety evaluation was performed in the SSAM software by importing trajectory files from VISSIM to analyze the associated traffic conflicts. The change in the number of simulated conflicts, considered to be the surrogate measure of secondary crashes, was used to evaluate the safety benefits of the CV applications. Finally, a statistical analysis was conducted to assess the significance of CVs in mitigating secondary crashes using traffic conflicts as a proxy for secondary crashes.

Deployment of CV applications was found to result in up to 98% reduction in traffic conflicts. The greater reduction was observed in less congested traffic, even with low penetration of CVs. Greater MPR was required to achieve significant conflict reduction during congested periods. It was also observed that incidents that block the outer lane result in more traffic conflicts than incidents that block the inner lane. Thus, with the same market penetration of CVs, more conflicts were reduced when the inner lane was blocked than when the outer lane was blocked. This could be due to less restrictions to merge and diverge when the inner lane was closed, compared to the scenario when the outer lane was blocked. Additionally, more conflicts were observed when two lanes were closed, compared to single lane closures. The detour advisory was found to be significant for incidents that block multiple lanes.

CHAPTER 8
SUMMARY AND CONCLUSIONS

Traffic incidents are the primary source of non-recurring congestion. In addition to affecting roadway operations, traffic congestion resulting from an incident exposes other vehicles to the risk of being involved in additional incidents, typically referred to as secondary crashes. Secondary crashes are non-recurrent in nature and lead to reduced freeway capacity, increased delay, and decreased travel time reliability. Not only do secondary crashes affect traffic operations, but they also impose risk on the safety of road users. Incident management agencies have therefore been investing a substantial amount of resources in devising strategies to mitigate secondary crashes. Accurate identification of secondary crashes is the first and the most crucial step in devising strategies to mitigate their occurrence. Nonetheless, the identification of secondary crashes is not a straightforward process. The definition itself is subjective, and identifying secondary crashes depends on the accuracy of the approach used to estimate the incident impact area.

Following the identification of secondary crashes, the next step towards developing strategies to mitigate secondary crashes is to explore the causal relationship between secondary crashes and potential explanatory variables. There are also some difficulties involved in identifying factors influencing the occurrence of secondary crashes: infrequent nature of secondary crashes, selection of the most important variables, and identification of variable correlation.

The primary goal of this research was to develop a comprehensive approach to identify and mitigate secondary crashes on the Florida's Turnpike System in real time. The research goal was achieved through the following objectives:

1. Investigate ways to accurately identify secondary crashes on the Florida's Turnpike System using real-time traffic and incident data.
2. Identify factors that influence the occurrence of secondary crashes on the Florida's Turnpike System.
3. Develop an algorithm that predicts the likelihood of secondary crashes in real time.
4. Explore the potential of CV applications in mitigating secondary crashes.

8.1 Secondary Crash Identification

Accurate estimation of the primary incident spatiotemporal impact area is essential and imperative for mitigating secondary crashes. A data-driven approach was developed to better estimate the primary incident impact area and identify secondary crashes within the impacted area. The developed approach considered how the queue, initiated by the incident, grows and dissipates upstream of the incident. This approach is able to estimate the spatial and temporal impact ranges of primary incidents while accounting for the effects of traffic flow conditions.

Traffic incidents from the SunGuide® database and high-resolution speed data from HERE Technologies were used to estimate the impact area of a primary incident. These data were collected from January 2014 to June 2019. The study area, which is located in Florida, included a 97-mile section of the Florida's Turnpike Mainline and TE, a 48-mile adjoining corridor. The Mainline study corridor consisted of a 69-mile NTM and a 28-mile STM.

The analysis was based on 322,259 traffic incidents that occurred along the study corridors between January 2014 and June 2019. Overall, 4,549 secondary crashes in the upstream direction of the primary incident were identified from 3,977 primary incidents. The identified secondary crashes on the upstream direction of the primary incident accounted for 1.4% of the 322,259 incidents. This is equivalent to 5.7 secondary crashes per mile per year.

Descriptive statistics of the secondary crashes indicated that 93% of the secondary crashes occurred within two hours after the occurrence of the primary incidents. Spatially, 47% of the secondary crashes occurred within two miles from the primary incident. Overall, 40% of secondary crashes occurred within two hours of the onset of a primary incident and within two miles upstream of the primary incident, the most commonly considered spatiotemporal threshold. The following are some of the key characteristics of the primary incidents and secondary crashes:

- Only 3% of secondary crashes occurred between midnight and 5:00 AM, whereas 85% occurred during morning and evening peak hours. Specifically, 33% of secondary crashes occurred during the morning peak (i.e., 6:00 AM - 10:00 AM) while the remaining 52% occurred during the evening peak (i.e., 2:00 PM - 8:00 PM). The highest proportion of primary incidents (13%) occurred between 4:00 PM and 5:00 PM, while the highest proportion of secondary crashes (13%) occurred an hour after the primary incident, i.e., between 5:00 PM and 6:00 PM.

- The proportion of normal incidents and secondary crashes was much higher on weekdays than on weekends. Compared to other days of the week, Friday was found to experience the highest proportion of secondary crashes (20%).

- While secondary crashes were found to occur on Mondays and Fridays, normal incidents were found to occur primarily on weekdays (i.e., Monday through Friday). Only 13% of secondary crashes were found to occur on weekends.

- As expected, traffic incidents involving towing and/or EMS resulted in longer incident clearance durations, as they tend to require more time to be cleared. Approximately 94% of normal incidents were cleared within 90 minutes, while 82% of primary incidents were cleared within 90 minutes. Likewise, 94% of traffic incidents that did not involve EMS were cleared within 90 minutes, while only 64% of traffic incidents that involved EMS were cleared within 90 minutes. The longer clearance time of the primary incidents could be considered to have contributed to the occurrence of secondary crashes.

- The severity of primary incidents was found to be one of the factors that influence the occurrence of secondary crashes. About 9% of primary incidents were moderate/severe while only 1% of normal incidents were moderate/severe. Besides the severity of primary incidents, the number of responding agencies, percentage of lanes closed, incident type, and incidents that required towing and/or EMS were also considered to be good indicators of incident severity. About 99% of normal incidents did not result in lane closure, while 21% of primary incidents resulted in a lane closure. Only 10% of normal incidents were identified as crashes, while 47% of primary incidents were crashes. About 13% of primary

incidents required towing, while only 3% of normal incidents required towing. Similarly, a higher percentage of incidents involving EMS resulted in secondary crashes (11%). While only 28% of normal incidents involved more than one responding agency, 51% of primary incidents and 55% of secondary crashes involved more than one responding agency. These statistics indicate that the severity of primary incidents influences the occurrence of secondary crashes.

- Compared to normal incidents (2%), a higher proportion of primary incidents (13%) occurred during cloudy/foggy/rainy conditions. Similarly, a higher percentage of primary incidents (11%) and secondary crashes (18%) occurred on wet surface conditions. These statistics imply that inclement weather conditions and adverse road surface conditions are among the factors that increase the probability of secondary crashes.

In practice, the developed approach can be easily implemented considering that its algorithm does not require much computational effort except for establishing the speed profiles for normal traffic conditions. Notably, these profiles are established once and can be used for a prolonged time (up to a year). This method can be used by the incident management officials while generating standard reports on a monthly, quarterly, and yearly basis. With additional programming work and the availability of access to real-time traffic and incident data, this method could be utilized to accurately identify potential secondary crashes in real time.

8.2 Factors Influencing the Occurrence of Secondary Crashes

Identifying risk factors that influence the likelihood of secondary crashes is critical to the development and implementation of efficient and resilient traffic management strategies. The LASSO penalized logistic regression model, fitted using the bootstrap resampling approach, was used to identify risk factors that influence the risk of secondary crashes. Traffic flow, incident, temporal, weather, and roadway geometric attributes were considered as potential factors that may influence the likelihood of secondary crashes. The proposed model is considered to improve the predictive accuracy of the secondary crash risk model because it accounts for the asymmetric nature of secondary crashes, performs variable selection, and removes correlated variables.

The study area included a 28 miles STM section. Data used were collected from the following sources: HERE Technologies (traffic flow attributes), SunGuide® database (incident, temporal, and weather attributes), RCI database, Google Earth Pro software, and Google Maps (roadway geometric attributes). These data were collected for a period of five and a half years from January 2014 through June 2019.

The following factors were found to significantly influence the likelihood of secondary crashes:

- A unit increase in the average prevailing speed reduces the likelihood of a secondary crash by 11%. In other words, a unit decrease in average prevailing speed was accompanied with an 89% increase in the risk of secondary crashes.
- A unit increase in the standard deviation of speed before the occurrence of the incident reduced the risk of a secondary crash by 7%.

- Hazard-related incidents were 29% more likely to result in secondary crashes compared to vehicle problem-related incidents.
- Crashes were 70% more likely to result in secondary crashes than vehicle problem-related incidents.
- Incidents attended to by more than one responding agency were 28% more likely to result in secondary crashes compared to incidents attended by only one responding agency.
- Moderate/severe incidents were twice as likely to result in secondary crashes compared to minor incidents.
- Incidents that occurred on wet road surface conditions were twice more likely to result in secondary crashes than those on dry surface conditions.
- Compared to off-peak hours, incidents that occurred during morning peak hours were 25% more likely to result in secondary crashes
- Incidents with diverge influence area within their impact area were 89% more likely to result in secondary crashes.
- Incidents with merge influence area within their impact area were 29% less likely to cause secondary crashes.
- A unit increase in shoulder width was associated with a 7% decrease in the likelihood of a secondary crash.
- Incidents with a curved segment within their impact area increased the risk of secondary crashes by 34%.
- Incidents with a vertical curve segment within their impact area increased the risk of secondary crashes by 127%.

As can be inferred from the study findings, the occurrence of secondary crashes is influenced by incident severity and how quickly the incident is cleared. To prevent the risk of secondary crash occurrence, traffic management strategies should be developed to accelerate the dissipation of the queue upstream of the potential primary incident. Warnings could be sent to drivers approaching a primary crash scene in real time through various means, including DMSs, information sharing technologies such as the *Waze* application, and emerging technologies such as CVs, allowing them to take necessary precautions (such as detour or drive with caution) to avoid being involved in a secondary crash. Furthermore, when the conditions associated with a high likelihood of secondary crashes prevail, responding agencies could be better prepared to respond to secondary crashes if they were to occur. These strategies will help to potentially reduce the frequency and severity of secondary crashes.

8.3 Real-time Secondary Crash Likelihood Prediction: Proof-of-concept

Mitigating the risk of secondary crashes is a crucial goal for effective traffic incident management. An algorithm was developed as a *proof-of-concept* to predict the likelihood of secondary crashes in real time. The study area included a 28-mile section of SR-91. The algorithm consists of three main parts, the first one being *Internal Storage* database, which include:

- historical databases,
- statistical model equation with the secondary crash likelihood parameters, and
- other potential attributes required to predict the likelihood of secondary crashes but are not collected in real time, i.e., roadway geometric characteristics.

The second part of the algorithm consists of *Backend Programs* for collecting, parsing, and saving incident, traffic, and rainfall data in real time. One of the algorithm applications continuously accesses the Florida's Turnpike SunGuide® database through VPN every two minutes and ping new incidents. The *RealtimeSpeed* application functions to retrieve, process, and save real-time speed data from the HERE real-time flow XML feed. This process is implemented for each Traffic Message Channel within the study corridor every minute. The *RealtimeRainfall* application retrieves real-time rainfall information from the NEXRAD Level-II network hosted in the "noaa-nexrad-level2" Amazon S3 bucket. The rainfall intensity values are retrieved every 4-6 minutes.

The information from the *Internal Storage* database and *Real-time Data Backend Programs* are then combined to predict the likelihood of secondary crashes. The secondary crash prediction process makes the third and final part, referred to as the *Secondary Crash Prediction Program*. This program is a two-fold process that focuses on (1) estimating the impact area of an incident and (2) predicting the likelihood of secondary crashes.

In summary, this algorithm could be used to develop ATMS to proactively prevent secondary crashes. Through this algorithm, first responders will be more vigilant and better prepared in case secondary crashes occur. In addition, motorists upstream of the primary incident could be warned about the potential for secondary crashes. Warnings could be sent to drivers approaching a potential primary incident scene in real time through various means, including DMSs, information sharing technologies such as *Waze* application, and the emerging technologies such as connected vehicles, giving them an opportunity to take necessary precautions (such as detour and/or drive with caution) to avoid being involved in a crash.

8.4 Potential of CV Applications in Mitigating Secondary Crashes

In light of the potential benefits offered by CV applications and the safety concerns associated with secondary crashes, the potential benefits of CV applications in mitigating secondary crashes was quantified. Since CVs are not yet fully operational, a sensitivity analysis was conducted at varying MPRs. The study leverages the application of CV technologies that utilize V2I communication, including speed advisory, lane-change advisory, and detour advisory. Information was disseminated to vehicles equipped with OBUs through V2I communication.

The study area included a 7.8-mile, 6-lane (3-lanes in each direction) road segment on SR-91. The freeway segment is in Broward County and crosses four roadways: Sample Road, Copans Road, Coconut Creek Road, and Atlantic Boulevard. The interchanges are 1 to 2 miles apart along the study corridor, with access to the Turnpike at each interchange except the Copans Road crossing. This site was chosen over other segments along the Florida's Turnpike due to its relatively high number of crashes in 2016-2019, based on observations from the Signal Four Analytics database. Furthermore, a 4.2-mile section of Lyons Road, with two lanes in each direction, was considered for detouring purposes. Compared to other possible detour routes along the Turnpike corridor, the selected detour is one of the sections with the closest consecutive interchanges (~ two miles).

A microscopic simulation approach was used to evaluate the effectiveness of CV applications to mitigate secondary crashes. The VISSIM microscopic simulation software was used to demonstrate traffic flow characteristics under the CV environment. As the first step towards

achieving the study objective, a calibrated VISSIM model was developed for the study area. The model mimics geometric and traffic characteristics for the selected corridor. Furthermore, incident and V2I communication were modeled using the V2X module in the VISSIM software through the COM API. This module simulates wireless communication and data exchange within the connected environment.

The study designed various simulation scenarios, including blockage of either the inner lane, the outer lane, or the two outer lanes to represent a primary incident. It also considered the effect of varying traffic volumes by analyzing traffic data during AM and PM peak periods. Since full market penetration of CVs is not anticipated soon, the study performed a sensitivity analysis by considering varying MPRs of CVs. In each scenario, the safety evaluation was performed in the SSAM software by importing trajectory files from VISSIM to analyze the associated traffic conflicts. The change in the number of simulated conflicts, considered to be the surrogate measure of secondary crashes, was used to evaluate the safety benefits of the CV applications. Finally, a statistical analysis was conducted to assess the significance of CVs in mitigating secondary crashes using traffic conflicts as a proxy for secondary crashes.

Deployment of CV applications was found to result in up to 98% reduction in traffic conflicts. Greater reduction was observed in less congested traffic, even with low penetration of CVs. Greater MPR was required to achieve significant conflict reduction during congested periods. It was also observed that incidents that block the outer lane result in more traffic conflicts than incidents that block the inner lane. Thus, with the same market penetration of CVs, more conflicts were reduced when the inner lane was blocked than when the outer lane was blocked. This could be due to fewer restrictions to merge and diverge when the inner lane was closed, compared to the scenario when the outer lane was blocked. Additionally, more conflicts were observed when two lanes were closed, compared to single lane closures. The detour advisory was found to be significant for incidents that block multiple lanes.

The adopted methodology used traffic conflicts from the microscopic simulation as a proxy for the risk of secondary crashes. However, with the current state-of-the-art, there is no direct quantitative link between simulated traffic conflicts and the number of possible secondary crashes. Future research could investigate and establish the connection between the simulated traffic conflicts and secondary crashes. Parameters based on previous studies were used while modeling incident conditions. It is recommended that future work should consider calibrating the VISSIM model based on traffic conditions within the incident impact duration.

REFERENCES

Algamal, Z. Y., and Lee, M. H. (2015a). "Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification". *Computer in Biology and Medicine*, 67, 136–145.

Algamal, Z. Y., and Lee, M. H. (2015b). "Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification". *Expert Systems with Applications*, 42(23), 9326–9332.

Amin-Naseri, M., Chakraborty, P., Sharma, A., Gilbert, S.B., Hong, M. (2018). "Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze". *Transportation Research Record*, 2672(43), 34–43.

AMS (American Meteorological Society). (n.d.). Glossary of meteorology. Retrieved June 17, 2019 from https://glossary.ametsoc.org/wiki/Rain.

Andrew, L. (2019). *Investigating the Effects of Rainfall on Traffic Operations on Florida Freeways*. Jacksonville, Florida: University of North Florida.

Atamo, M. A. (2012). *Safety Assessment of Freeway Merging and Diverging Influence Areas Based on Conflict Analysis of Simulated Traffic*. Denver, Colorado: University of Colorado Denver.

Barr, J. (2018). New AWS public data set – real-time and archived NEXRAD weather data | Amazon Web Services. Amazon, Amazon. Retrieved May 15, 2019 from aws.amazon.com/blogs/aws/new-aws-public-data-set-real-time-and-archived-nexrad-weather-data/.

Baykal-Gürsoy, M., Xiao, W., Ozbay, K. (2009). "Modeling traffic flow interrupted by incidents". *European Journal of Operational Research*, 195(1), 127–138.

Cabbagestalk, S. (2017). All Lanes Open; Tractor-Trailer Carrying Hazardous Materials Overturns. WCBD News, August 28, 2017.

Carson, J. L. (2010). *Best Practices in Traffic Incident Management*. Washington, D.C.: Federal Highway Administration.

Chang, G. L., & Rochon, S. (2011). *Performance Evaluation and Benefit Analysis For CHART*. Hanover, Maryland: Maryland Department of Transportation.

Chatterjee, K., Hounsell, N. B., Firmin, P. E., and Bonsall, P. W. (2002). "Driver response to variable message sign information in London". *Transportation Research Part C: Emerging Technologies*, 10(2), 149–69.

Chimba, D., and Kutela, B. (2014). “Scanning secondary derived crashes from disabled and abandoned vehicle incidents on uninterrupted flow highways”. *Journal of Safety Research*, 50, 109–116.

Chiu, Y-C., and Huynh, N. (2007). “Location configuration design for dynamic message signs under stochastic incident and ATIS scenarios”. *Transportation Research Part C: Emerging Technologies*, 15(1), 33–50.

Chou, C. S., & Miller-Hooks, E. (2010). “Simulation-based secondary incident filtering method”. *Journal of Transportation Engineering*, 136(8), 746-754.

Chung, Y. (2013). “Identifying Primary and Secondary Crashes from Spatiotemporal Crash Impact Analysis”. *Transportation Research Records*, 2386(1), 62–71.

Dougald, L. E., Goodall, N. J., and Venkatanarayana, R. (2016). *Traffic Incident Management Quick Clearance Guidance and Implications* (FHWA/VTRC 16-R9). Virginia: Transportation Research Council.

Dunn Engineering Associates. (2006). *Alternate Route Handbook* (FHWA-HOP-06-092). Washington, D.C.: Federal Highway Administration.

FDOT (Florida Department of Transportation). (2017). *TSM&O 2017 Strategic Plan*. Tallahassee, Florida: Florida Department of Transportation.

FDOT. (2014). *Traffic Analysis Handbook 2014: A Reference for Planning and Operations*. Tallahassee, Florida: Florida Department of Transportation.

FDOT. (2016). *Roadway Characteristic Inventory (RCI): Features and Characteristics Handbook*. Tallahassee, Florida: Florida Department of Transportation.

FDOT. (2018). *Standard Operating Guidelines*. Tallahassee, Florida: Florida Department of Transportation.

FDOT. (n.d.). *FDOT Open Data Hub*. Retrieved May 28, 2020 from https://gis-fdot.opendata.arcgis.com/.

FDOT: Transportation Statistics Office. (2016). *Roadway Characteristic Inventory (RCI): Features and Characteristics Handbook*. Tallahassee, Florida: Florida Department of Transportation.

Florida’s Turnpike Enterprise. (2017). *SW 10th Street PD&E VISSIM model calibration Florida’s Turnpike to I-95* (FPN 43989-1 March).

Gettman, D., and Head, L. (2003). “Surrogate safety measures from traffic simulation models”. *Transportation Research Records*, 1840, 104–115.

Ghori, M. R., Zamli, K. Z., Quosthoni, N., Hisyam, M., and Montaser, M. (2018). "Vehicular Ad-Hoc Network (VANET): Review." *Proceedings of the 2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, May 11-12, 2018. Bangkok Thailand.

Glotzbach, G. (2014). *The Waze Connection*. Tallahassee, Florida.

Golob, T. F., and Recker, W. W. (2003). "Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions". *Journal of Transportation Engineering*, 129(4), 342–353.

Goodall, N. J. (2017). "Probability of secondary crash occurrence on freeways with the use of private-sector speed data". *Transportation Research Record*, 2635, 11–18.

Green, E. R., Pigman, J. G., Walton, J. R., & McCormack, S. (2012). "Identification of secondary crashes and recommended countermeasures to ensure more accurate documentation". *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, January 22-26, 2012. Transportation Research Board, Washington, D.C.

Grumert, E., and Tapani, A. (2012). "Impacts of a cooperative variable speed limit system". *Procedia Social and Behavioral Sciences*, 43(43), 595–606.

Guo, J., Liu, Z., Huang, W., Wei, Y., and Cao, J. (2018). "Short-term traffic flow prediction using fuzzy information granulation approach under different time intervals". *IET Intelligent Transport Systems*, 12(2), 143-150.

Habtemichael, F. G., Santos, L. D. P. (2013). "Sensitivity analysis of VISSIM driver behavior parameters on safety of simulated vehicles and their interaction with operations of simulated". *Proceedings of the 92nd Annual Meeting of the Transportation Research Board*, January 13-17, 2013. Transportation Research Board, Washington, D.C.

Hagen, T. (2017). Avoid Secondary Crashes & Improve Safety with Proper Training [WWW Document]. Retrieved September 16, 2020 from https://www.zolldata.com/blog/avoid-secondary-crashes-and-improve-safety-with-proper-training.

Harding, J., Powell, G., Yoon, R., Fikentscher, J., Doyle, C., Sade, D., Lukuc, M., Simons, J. and Wang, J. (2014). *Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application* (DOT HS 812 014). Washington, D.C.: National Highway Traffic Safety Administration.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer.

Haule, H. J., Alluri, P., Sando, T., Raihan, A. M. (2020). "Investigating the impact of rain on crash-clearance duration". *Journal of Transportation Engineering Part A: Systems*, 146(11), 04020130. HCM (Highway Capacity Manual). (2016). *Highway Capacity Manual*, Sixth Edition.

Hirunyanitiwattana, W., and Mattingly, S. P. (2006). "Identifying secondary crash characteristics for California highway system". *Proceedings of the 85th Annual Meeting of the Transportation Research Board*, January 22-26, 2006. Transportation Research Board, Washington, D.C.

Imani, H. N. (2019). *The use of real-time connected vehicles and HERE data in developing an automated freeway incident detection algorithm*. Jacksonville, Florida: University of North Florida.

Imprialou, M. I. M., Orfanou, F. P., Vlahogianni, E. I., and Karlaftis, M. G. (2013). "Methods for defining spatiotemporal influence areas and secondary incident detection in freeways". *Journal of Transportation Engineering*, 140(1), 70-80.

INRIX. (2019). INRIX: Congestion Costs Each American 97 hours, $1,348 A Year [WWW Document]. Retrieved July 11, 2019 from http://inrix.com/press-releases/scorecard-2018-us/.

Jalayer, M., Baratian-Ghorghi, F., and Zhou, H. (2015). "Identifying and characterizing secondary crashes on the Alabama state highway systems". *Advances in Transportation Studies*, 37, 129–140.

Karlaftis, M. G., Latoski, S. P., Richards, N. J., and Sinha, K. C. (1999). "ITS impacts on safety and traffic management: an investigation of secondary crash causes". *Journal of Intelligent Transportation Systems*, 5(1), 39-52.

Kassambara, A. (2017). *Machine Learning Essentials: Practical Guide in R*, First Edit. ed. STHDA.

Kenney, J. B. (2011). "Dedicated short-range communications (DSRC) standards in the United States". *Proceedings of IEEE*, 99(7), 1162–1182.

Khattak, A., Wang, X., and Zhang, H. (2009). "Are incident durations and secondary incidents interdependent?". *Transportation Research Record*, 2099, 39-49.

Khattak, A., Wang, X., and Zhang, H. (2012). "Incident management integration tool: dynamically predicting incident durations, secondary incident occurrence and incident delays". IET Intelligent Transport Systems, 6(2), 204-214.

Kidando, E., Kitali, A. E., Lyimo, S. M., Sando, T., Moses, R., Kwigizile, V., and Chimba, D. (2019). "Applying probabilistic model to quantify influence of rainy weather on stochastic and dynamic transition of traffic conditions". *Journal of Transportation Engineering Part A: Systems*, 145(5), 04019017.

Kitali, A. E., Alluri, P., Sando, T., Haule, H., Kidando, E., and Lentz, R. (2018). "Likelihood estimation of secondary crashes using Bayesian complementary log-log model". *Accident Analysis and Prevention*, 119: 58–67.

Kitali, A. E., Alluri, P., Sando, T., Wu, W. (2019b). "Identification of secondary crash risk factors using penalized logistic regression model". *Transportation Research Record*, 2673(11):901-914.

Kitali, A. E., Kidando, E., Sando, T., Moses, R., and Ozguven, E. E. (2017). "Evaluating aging pedestrian crash severity with Bayesian complementary log–log model for improved prediction accuracy". *Transportation Research Record*, 2659(1):155-163.

Kitali, A., Alluri, P., and Sando, T. (2021). "Influence of Incident Spatiotemporal Estimation Method in Secondary Crash Identification," *Proceedings of the 100th Annual Meeting of the Transportation Research Board*, January 2021. Transportation Research Board, Washington, D.C.

Kitali, A.E., Alluri, P., Sando, T., and Lentz, R. (2019a). "Impact of primary incident spatiotemporal influence thresholds on the detection of secondary crashes". *Transportation Research Record*, 2673(10), 271–283.

Kopitch, L., and Saphores, J. D. M. (2011). "Assessing effectiveness of changeable message signs on secondary crashes". *Proceedings of the 90th Annual Meeting of the Transportation Research Board*, January 23-27, 2011. Transportation Research Board, Washington, D.C.

Latoski, S. P., Pal, R., and Sinha, K. C. (1999). "Cost-effectiveness evaluation of Hoosier Helper freeway service patrol". *Journal of Transportation Engineering*, 125(5), 429-438.

Li, J., Guo, J., Wijnands, J.S., Yu, R., Xu, C., and Stevenson, M. (2020). "Assessing injury severity of secondary incidents using support vector machines". *Journal of Transportation Safety and Security*, 1–20.

Li, Z., Li, Y., Liu, P., Wang, W., and Xu, C. (2014). "Development of a variable speed limit strategy to reduce secondary collision risks during inclement weathers". *Accident Analysis and Prevention*, 72, 134–145.

Liu, Y., Kim, W., and Chang, G-L. (2012). "Decision model for justifying the benefits of detour operation under non-recurrent congestion." *Journal of Transportation Engineering*, 139(1), 40–49.

Lou, Y., Yin, Y., and Lawphongpanich, S. (2011). "Freeway service patrol deployment planning for incident management and congestion mitigation". *Transportation Research Part C: Emerging Technologies*, *19*(2), 283–295.

Mahmud, S. M. S., Ferreira, L., Hoque, M. S., and Tavassoli, A. (2019). "Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment". *IATSS Research*, 43(1), 27–36.

Marshall, D., Wu, X., Brown, T., and Boyle, L. N. (2017). "Connected vehicle auditory alerts: one size doesn't fit all scenarios". *Proceedings of the 96th Annual Meeting of the Transportation Research Board, January 08-12, 2017*. Transportation Research Board, Washington, D.C.

Martin, J. L., and Wu, D. (2018). “Pedestrian fatality and impact speed squared: Cloglog modeling from French national data”. *Traffic Injury Prevention*, 19(1), 94-101.

Menard, G., and N. Torelli. (2014). “Training and Assessing Classification Rules with Imbalanced Data”. *Data Mining and Knowledge Discovery*, 28, 92–122.

Mishra, S., Golias, M., Sarker, A., and Naimi, A. (2016). *Effect of primary and secondary crashes: identification, visualization, and prediction research report* (CFIRE 09-05). Madison, Wisconsin: Wisconsin Department of Transportation.

Montes, Ca., Faquir, T., Hapney, TJ., and Birriel, E. (2008). *Guidelines for the Use of Dynamic Message Signs on The Florida State Highway System*. Tallahassee, Florida: Florida Department of Transportation.

Moore, J. E., Giuliano, G., and Cho, S. (2004). “Secondary accident rates on Los Angeles freeways”. *Journal of Transportation Engineering*, 130(3), 280-285.

Motamed, M., and Machemehl, R. (2014). *Real time Freeway Incident Detection* (SWUTC/14/600451-00083-1). College Station, Texas: Southwest Region University Transportation Center.

Mounce, J. M., Ullman, G., Pesti, G., and Pezoldt, V. (2007). *Guidelines for the Evaluation of Dynamic Message Sign Performance* (FHWA/TX-07/0-4772-1). Austin, TX: Texas Department of Transportation.

NCHRP (National Cooperative Highway Research Program). (2014). Guidance for implementation of traffic incident management performance measurement/ performance measurement for traffic incident management programs/ Florida (webpage). Retrieved Feb. 12, 2018, from http://nchrptimpm.timnetwork.org/?page_id=79.

NOAA (National Oceanic and Atmospheric Administration). (n.d.). National Doppler radar sites. Retrieved April 29, 2019 from https://www.ncdc.noaa.gov/nexradinv/map.jsp.

O'Laughlin, J., and Smith, A. (2002). “Operational issues discussion paper on “incident management operations: Top five issues”. *Proceedings of the National Conference on TrafficIncident Management: A Road Map to the Future*, March 11-13, 2002. American Association of State Highway and Transportation Officials, Washington, D.C.

Ou, J., Xia, J., Wang, Y., Wang, C., and Lu, Z. (2020). “A data-driven approach to determining freeway incident impact areas with fuzzy and graph theory-based clustering”. *Computer-Aided Civil and Infrastructure Engineering*, 35, 178–199.

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., and Trego, T. (2010). *Traffic Incident Management Handbook* (FHWA-HOP-10-013). Washington, D.C.: Federal Highway Administration, Office of Transportation Operations.

Paikari, E., Tahmasseby, S., and Far, B. (2014). "A simulation-based benefit analysis of deploying connected vehicles using dedicated short range communication". *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium*, Dearborn, Michigan, USA, 980-985.

Park, H., and Haghani, A. (2016b). "Real-time prediction of secondary incident occurrences using vehicle probe data". *Transportation Research Part C: Emerging Technologies*, 70, 69–85.

Park, H., and Smith, B. L. (2012). "Investigating benefits of intellidrive in freeway operations: lane changing advisory case study." *Journal of Transportation Engineering* 138(9): 1113–22.

Park, H., Gao, S., Haghani, A., Samuel, S., and Knodler, M. A. (2017). "Sequential interpretation and prediction of secondary incident probability in real time". *Proceedings of the 96th Annual Meeting of the Transportation Research Board*, January 8-12, 2017. Transportation Research Board, Washington, D.C.

Park, H., Haghani, A. (2016a). "Use of Clustering model and adjusted boxplot model for identification of secondary incidents". *Proceedings of the 95th Annual Meeting of the Transportation Research Board*, January 10-14, 2016. Transportation Research Board, Washington, D.C.

Park, H., Haghani, A., Samuel, S., and Knodler, M. A. (2018). "Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion." *Accident Analysis and Prevention*, 112, 39–49.

Pearce, V., and Subramaniam, S. (1998). *Incident Management: Detection, Verification, and Traffic Management* (FHWA-RD-JPO-034). Washington, DC: U.S. Department of Transportation.

Pei, X., Sze, N.N., Wong, S.C., and Yao, D. (2016). "Bootstrap Resampling Approach to Disaggregate Analysis of Road Crashes in Hong Kong". *Accident Analysis and Prevention*, 95, 512–520.

PTV. (2020). *PTV Vissim 2020 User Manual*. Karlsruhe, Germany.

Pulugurtha, S. S., Mahanthi, B. S. S. (2016). "Assessing spatial and temporal effects due to a crash on a freeway through traffic simulation". *Case Studies on Transport Policy*, 4(2), 122–132.

Rahman, S., Abdel-Aty, M., and Wang, L. (2018). "Understanding the highway safety benefits of different approaches of connected vehicles in reduced visibility conditions". *Transportation Research Record*, 2672(19), 91–101.

Raub, R. (1997). "Occurrence of secondary crashes on urban arterial roadways". *Transportation Research Record*, 1581, 53–58.

Riggins, G., Bertini, R., Ackaah, W., and Margreiter, M. (2016). "Evaluation of driver compliance to displayed variable advisory speed limit systems: comparison between Germany and the U.S." *Transportation Research Procedia*, 15, 640–51.

Saffarzadeh, M., Nadimi, N., Naseralavi, S., and Mamdoohi, A. R. (2013). “A general formulation for time-to-collision safety indicator”. *Proceedings of the Institution of Civil Engineers*, 166(5), 294–304.

Sando, T., Alluri, P., Chuan, C., Haule, H., Kitali, A., Lentz, R., and Huq, A. (2018). *Evaluation of Incident Response Improvements for Statewide Application: Learning from the New Regional Traffic Management Center in Jacksonville, Florida* (BDV34-977-10). Tallahassee, Florida: Florida Department of Transportation.

Sarker, A.A., Paleti, R., Mishra, S., Golias, M. M., Freeze, P. B. (2017). “Prediction of secondary crash frequency on highway networks”. *Accident Analysis and Prevention*, 98, 108–117.

SAS Institute Inc. (2019). *SAS Visual Statistics 8.5 Procedures*. Cary, NC.

Sen, S., Chowdhury, T., Mitra, A., and Roy, S. K., (2007). *Recent Advances in Transportation Engineering*. Springer Singapore. doi:10.1007/978-981-13-7162-2

Shahdah, U., Saccomanno, F., and Persaud, B. (2015). “Application of traffic microsimulation for evaluating safety performance of urban signalized intersections”. *Transportation Research Part C: Emerging Technologies*, 60, 96–104.

Smith, B. L., & Ulmer, J. M. (2003). “Freeway traffic flow rate measurement: Investigation into impact of measurement time interval”. *Journal of Transportation Engineering*, 129(3), 223-229.

Soloka, M. (2019). *Simulation Exploration of The Potential of Connected Vehicles in Mitigating Secondary Crashes*. Jacksonville, Florida: University of North Florida.

Songchitruksa, P., Bibeka, A., Lin, L. (Irene), and Zhang, Y. (2016). *Incorporating Driver Behaviors into Connected and Automated Vehicle Simulation* (Atlas-2016-13). Ann Arbor, Michigan: Advancing Transportation Leadership and Safety (ATLAS) Center.

Sun, C. C., and Chilukuri, V. (2010). “Dynamic incident progression curve for classifying secondary traffic crashes”. *Journal of Transportation Engineering*, 136(12), 1153-1158.

Sun, C., and Chilukuri, V. (2006). “The use of dynamic incident progression curve for classifying secondary accidents”. *Proceedings of the 85th Annual Meeting of the Transportation Research Board*, January 22-26, 2006. Transportation Research Board, Washington, D.C.

Tian, Y., Chen, H., Truong, D. (2016). “A case study to identify secondary crashes on interstate highways in Florida by using geographic information systems”. *Advances in Transportation Studies*, 2, 103–112.

Tibshirani, R. (1996). “Regression Shrinkage and Selection via the Lasso”. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

TRB (Transportation Research Board). (2016). *Highway Capacity Manual 6th Edition: A Guide for Multimodal Mobility Analysis*. Washington D.C.

Vlahogianni, E. I., Karlaftis, M. G., and Orfanou, F. P. (2012). "Modeling the effects of weather and traffic on the risk of secondary incidents". *Journal of Intelligent Transportation Systems*, 16(3), 109-117. 112.

Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C., & Halkias, B. M. (2010). "Freeway operations, spatiotemporal-incident characteristics, and secondary-crash occurrence". *Transportation Research Record*, (2178), 1-9.

Wang, J., Boya, L., Lanfang, Z., and Ragland, D. R. (2016). "Modeling secondary accidents identified by traffic shock waves". *Accident Analysis and Prevention*, 87, 141-147.

Wang, J., Liu, B., Fu, T., Liu, S., and Stipancic, J. (2019). "Modeling when and where a secondary accident occurs". *Accident Analysis and Prevention*, 130, 160–166.

WSDOT (Washington State Department of Transportation). (2019). *Active Traffic and Demand Management - How ATDM symbols work*. https://www.wsdot.wa.gov/travel/operations-services/active-traffic-management/how-atdm-symbols-work.

Xu, C., Liu, P., Yang, B., and Wang, W. (2016). "Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data". *Transportation Research Part C: Emerging Technologies*, 71, 406-418.

Xu, C., Xu, S., Wang, C., and Li, J. (2019). "Investigating the factors affecting secondary crash frequency caused by one primary crash using zero-inflated ordered probit regression". *Physica A: Statistical Mechanics and Its Applications*, 524, 121–129.

Yang, H., Bartin, B., and Ozbay, K. (2014a). "Mining the Characteristics of Secondary Crashes on Highways". *Journal of Transportation Engineering*, 140(4), 04013024.

Yang, H., Ozbay, K., and Xie, K. (2014b). "Assessing the risk of secondary crashes on highways". *Journal of Safety Research*, *49*, 143.e1-149.

Yang, H., Ozbay, K., Morgul, E., Bartin, B., and Xie, K. (2014c). "Development of online scalable approach for identifying secondary crashes". *Transportation Research Record*, 2470, 24-33.

Yang, H., Wang, Z., and Xie, K. (2017). "Impact of connected vehicles on mitigating secondary crash risk". *International Journal of Transportation Science and Technology*, *6*(3), 196–207.

Yang, H., Wang, Z., Xie, K., Ozbay, K., and Imprialou, M. (2018). "Methodological evolution and frontiers of identifying, modeling and preventing secondary crashes on highways". *Accident Analysis and Prevention*, 117, 40-54.

Yue, L., Abdel-Aty, M., and Wang, L. (2018). "Assessment of the safety benefits of connected vehicle technologies." *Proceedings of the 97th Annual Meeting of the Transportation Research Board*, January 07-11, 2018. Transportation Research Board, Washington, D.C.

Zhan, C., Gan, A., and Hadi, M. (2009). "Identifying secondary crashes and their contributing factors". *Transportation Research Record*: 2102, 68-75.

Zhan, C., Shen, L., Hadi, M. A., and Gan, A. (2008). "Understanding the characteristics of secondary crashes on freeways". *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, January 13-17, 2008. Transportation Research Board, Washington, D.C.

Zhang, H., and Khattak, A. (2010a). "What is the role of multiple secondary incidents in traffic operations?". *Journal of Transportation Engineering*, 136(11), 986–997.

Zhang, H., and Khattak, A. (2010b). "Analysis of cascading incident event durations on urban freeways". *Transportation Research Record*, 2178, 30–39.

Zheng, D., Chitturi, M. V., Bill, A. R., and Noyce, D.A. (2014). "Secondary crash identification on a large-scale highway system". *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, January 12-16, 2014. Transportation Research Board, Washington, D.C.

Zhou, H. (2008). "Evaluation of route diversion strategies using computer simulation." *Journal of Transportation Systems Engineering and Information Technology*, 8(1), 61–67.

Zou, H. (2006). "The adaptive lasso and its oracle properties". *Journal of the American Statistical Association*, 101(476), 1418–1429.