Report No. UT-22.16

INVESTIGATION OF SECONDARY CRASH ALONG HIGHWAYS IN UTAH – PHASE I

Prepared For:

Utah Department of Transportation
Research & Innovation Division

Final Report
March 2022

Keeping Utah Moving
Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

DISCLAIMER

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT on the Technical Advisory Committee for helping to guide the research:

- Glenn Blackwelder
- Jeff Lewis
- Ivana Vladisavljevic
- Kaitlin Marousis
- Corey Coulam
- Rikki Sonnen
- Robert Chamberlin (consultant PM)

TECHNICAL REPORT ABSTRACT

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UT-22.16	N/A	N/A

4. Title and Subtitle
Investigation of Secondary Crash along Highways in Utah – Phase I

5. Report Date
March 2022

6. Performing Organization Code
N/A

7. Author(s)
Zhao Zhang, Yaobang Gong, and Xianfeng Terry Yang

8. Performing Organization Report No.
N/A

9. Performing Organization Name and Address
University of Utah
Department of Civil & Environmental Engineering
110 Central Campus Drive, Suite 2000
Salt Lake City, UT 84112

10. Work Unit No.
5H08427H

11. Contract or Grant No.
21-8225

12. Sponsoring Agency Name and Address
Utah Department of Transportation
4501 South 2700 West
P.O. Box 148410
Salt Lake City, UT 84114-8410

13. Type of Report & Period Covered
Final
Sept 2020 to March 2022

14. Sponsoring Agency Code
UT19.322

15. Supplementary Notes
Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration

16. Abstract
The occurrence of secondary crashes on highways has many adverse effects, such as traffic congestion, air pollution, and more crashes. Research suggests that secondary crashes account for about 20% of all crashes and 18% of all fatalities on US freeways (Owens et al., 2010). While improving incident management is one of the effective ways to reduce the risk of secondary crashes, the identification of appropriate strategies relies on the understanding of contributing factors to secondary crashes, which need accurate secondary crash data. However, only 63 secondary crashes that occurred on I-15 in the state of Utah from 2018-2019 are recorded in UDOT's crash database. Such data quality cannot meet the needs of developing effective incident management strategies. Therefore, the first goal of this project research is to develop an effective method to identify primary and secondary crashes from the crash database. Our research team proposes a hybrid method that can effectively identify primary and secondary crashes from all crash records on freeways. Based on the identified primary and secondary crashes, this project developed a binary logit model and found twelve contributing factors, including age of driver, snowy weather, angle collision, rear-end crash, multiple vehicles involved, collision with fixed objects, speed-related crash, minivan involved, adverse roadway surface condition, vehicle slowing in traffic lane, and roadway with straight alignment, are found to be positively associated with the occurrence of secondary crashes. Secondary crashes were also less likely to have occurred on weekends and in rural areas. Then the project implemented the HOPIT model to analyze the crash injury patterns in primary and secondary crashes and found several factors that may increase injury severity in primary crashes (e.g., suspected alcohol, angle collision, multiple vehicles involved, etc.) and secondary crashes (including suspected drugs, head-on collision, multiple vehicles involved, etc.), respectively.

17. Key Words	18. Distribution Statement	23. Registrant's Seal
Primary Crash, Secondary Crash, Contributing Factors, Crash Injury Severity	Not restricted. Available through: UDOT Research Division 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410 www.udot.utah.gov/go/research	N/A

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price	
Unclassified	Unclassified	95	N/A	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

SC	Secondary Crash
HOPIT	Hierarchical Ordered Probit Model
ITS	Intelligent Transportation Systems
ATMS	Advanced Traffic Management System
UDOT	Utah Department of Transportation
MI	Minor Injury
NI	No Injury
SI	Severe Injury
TAC	Technical Advisory Committee
PeMS	Performance Measurement System
PDO	Property Damage Only

EXECUTIVE SUMMARY

The occurrence of secondary crashes on highways has many adverse effects, such as traffic congestion, air pollution, and more crashes. Secondary crashes account for about 20% of all crashes and 18% of all fatalities on US freeways (Owens et al., 2010). While improving incident management is one of the effective ways to reduce the risks of secondary crashes, the identification of appropriate strategies relies on the understanding of contributing factors to secondary crashes, which need accurate secondary crash data. However, only 63 secondary crashes that occurred on I-15 in the state of Utah from 2018-2019 are recorded in UDOT's crash database. Such data quality cannot meet the needs of developing effective incident management strategies. The first goal of this project research is to develop an effective method to identify primary and secondary crashes from the crash database. Our research team proposed a hybrid method to identify primary and secondary crashes from UDOT's crash database. The study results indicate that the hybrid method could effectively identify primary and secondary crashes on freeways.

Based on the identified primary and secondary crash records, the second goal of this project is to apply a binary logit model to find the contributing factors of secondary crashes. Crash information (e.g., driver information, environmental characteristics, crash-specific information, etc.) was collected to build up the model. The modeling results indicate that twelve variables (including age of driver, snowy weather, characteristics of the primary crash [e.g., angle collision, rear-end crash, multiple vehicles involved, collision with fixed objects, speed-related crash, minivan involved], adverse roadway surface condition, vehicle slowing in traffic lane, and roadway with straight alignment) are found to be positively associated with the occurrence of secondary crashes, indicating that those factors will significantly increase the probability of the occurrence of secondary crashes. Only "weekend" and "rural" parameters are negatively associated with the probability of the occurrence of secondary crashes.

The third goal of this research is to implement the HOPIT model to analyze the crash injury patterns in primary and secondary crashes with the identified crash data and its collected information. The study results indicate that thirteen variables (including suspected alcohol, angle collision, multiple vehicles involved, etc.) and nine variables (including suspected drugs, head-

on collision, multiple vehicles involved, etc.) are significantly related to crash injury severity in primary and secondary crashes, respectively.

1.0 INTRODUCTION

1.1 Problem Statement

Secondary crashes (SC) are typically defined as crashes that occur within the congested spatiotemporal boundaries of the region in which a primary crash occurred (Owens et al., 2010 and (Xu et al., 2016). The occurrence of secondary crashes on highways has many adverse effects, such as traffic congestion, air pollution, and additional crashes. Owens et al. (Owens et al., 2010) reported that secondary crashes account for about 20% of all crashes and 18% of all fatalities on US freeways. While improving incident management is one of the effective ways to reduce the risk of secondary crashes (Sun and Chilukuri, 2010; Zhan et al., 2009) , the identification of appropriate incident management strategies should be based on the understanding of contributing factors to secondary crashes. To understand the contributing factors, accurate secondary crash data are needed. However, according to a preliminary study, only 390 crashes that occurred on I-15 in the state of Utah from 2010 to 2020 are recorded as secondary crashes (0.31%). Such data quality cannot meet the needs of developing effective incident management strategies.

In the state of Utah, secondary crashes on freeways have not been well studied yet. In recent decades, the development of intelligent transportation systems (ITS) has made transportation data easier to access, which offers the basis for secondary crash analysis. Notably, UDOT develops many databases for different research purposes, such as the Utah ClearGuide database, Freeway PeMS database, Numeric Crash Database, GIS-based crash database, etc. These databases offer the possibility to conduct research on identifying primary and secondary crashes from the crash database, finding the contributing factors to secondary crashes, and examining the crash injury patterns of primary and secondary crashes.

1.2 Objectives

The first goal of this project research is to develop a hybrid method that can effectively identify primary and secondary crashes from all crash records on freeways. The second goal of this project is to apply an appropriate statistical model to identify the contributing factors of secondary crashes. The third goal of this research is to implement the HOPIT model to analyze the crash injury patterns of primary and secondary crashes with the identified crash data. Strategically, it may help resolve potential conflicts between UDOT's Zero Fatalities and Optimize Mobility goals. The outcome of this research will provide some insightful findings to help UDOT build up a more effective incident management system to mitigate the secondary crashes on freeways in Utah.

1.3 Scope

This research project is divided into several phases, including preliminary investigation, data preparation, literature review, methodology, experimental study and results analysis, and conclusions. Each of these is described in the following subsections.

1.3.1 Preliminary Investigation

In the early stage of the project, the research team and the Technical Advisory Committee (TAC) members discussed the data availability for the project, reviewed the scope and schedule, conducted preliminary investigations on the expected project outcomes, and determined the potential risks associated with the project. This meeting included members of the University of Utah and engineers from UDOT.

1.3.2 Literature Review

In the third phase of this project, the research team conducted a comprehensive literature review of existing studies related to secondary crash identification, contributing factors modeling, and crash injury severity analysis. Results of this phase are presented in Chapter 2.

1.3.3 Data Preparation

In the second phase of this project, the research team conducted a preliminary data analysis of all crash data and manually verified the accuracy of labeled secondary crash records in the database. Results of this phase are presented in Chapter 3.

1.3.4 Methodology

In the fourth phase of this project, the research team developed primary and secondary crash identification methods, modeled secondary crash contributing factors, and conducted crash injury severity analysis. Results of this phase are presented in chapter 4.

1.3.5 Experimental Study and Results Analysis

In the fifth phase of this project, the research team conducted case studies and analyzed the results. The research team implemented the proposed hybrid method to identify the primary and secondary crashes in the database, a binary logit model for finding the contributing factors of secondary crashes, and the HOPIT model for examining the injury severity patterns in primary and secondary crashes. Results of this phase are presented in chapter 5.

1.3.6 Recommendations and Conclusions

In this phase, the research team summarized the key research findings and made recommendations for identifying secondary crashes and reducing the number of secondary crashes and crash injury severity. Conclusions and recommendations are provided in Chapter 5.

1.4 Outline of Report

This project report is organized with the following chapters:

- Introduction
- Literature Review
- Data Preparation
- Methodology
- Experimental Analysis
- Conclusions

2.0 LITERATURE REVIEW

2.1 Overview

The research team conducted a comprehensive literature review related to identifying secondary crashes, the contributing factors to secondary crashes, and injury severity patterns of the secondary crashes. This chapter summarizes the findings of existing studies in three areas: (1) existing methods for secondary crashes identification; (2) modeling contributing factors of secondary crashes; and (3) HOPIT model for crash injury severity analysis.

2.2 Existing Methods for Secondary Crash Identification

The static and dynamic methods are two popular approaches to identify secondary crashes. The static threshold methods assumes that secondary crashes should happen within a spatial and temporal range of a primary crash. For example, Hirunyanitiwattana and Mattingly (Hirunyanitiwattana et al., 2006) used the static thresholds of 1 h and 2 miles upstream of a primary crash to identify secondary crashes. Any crashes are determined as secondary crashes if they happen within 1 h and 2 miles after a primary crash. There are also other similar studies using static methods (Moore et al., 2004; Zhan et al., 2008). The disadvantage of the static method is the predetermined fixed spatial and temporal threshold. To overcome the limitation of the static method, a series of studies developed dynamic methods to identify secondary crashes, such as queue length estimations (Zhang and Khattak, 2010), incident progression curve (Sun and Chilukuri, 2010), cumulative arrival and departure plots (Zhan et al., 2009), and speed contour plot (Xu et al., 2016; Yang et al., 2014a).

2.3 Modeling Contributing Factors of Secondary Crashes

As shown in Table 2, the logit model has been widely implemented to identify the contributing factors of secondary crashes (Mishra et al., 2017; Wang et al., 2016a, 2016b; Yang et al., 2014b; Zhan et al., 2009; Zhang and Khattak, 2010). The logit model has many advantages. The error terms of dependent variables in the logit model do not need to be normally distributed. It has a superior ability to avoid overfitting problems (Mitra and Washington, 2007).

It also outperforms other models in dealing with an unbalanced sample, in which the number of one class is much larger than those of the other classes.

To reduce the risk of secondary crashes, the existing studies have been conducted to identify the relationship between the probability of secondary crashes and contributing factors, such as characteristics of the primary crash, traffic conditions, geometric information, weather conditions, and demographic information (Mishra et al., 2017; Wang et al., 2016a, 2016b; Yang et al., 2014b; Zhan et al., 2009; Zhang and Khattak, 2010). Based on the results of those studies, the collision type, occurrence time, number of vehicles involved, and crash duration were found to be significantly related to the likelihood of secondary crashes. In detail, secondary crashes are less likely to happen during off-peak hours or on the weekend (Yang et al., 2014b; Zhan et al., 2009). More vehicles involved in crashes increase the probability of secondary crashes (Mishra et al., 2017; Zhang and Khattak, 2010). In addition, rear-end crashes with longer durations (i.e., longer crash clearance time) are found to increase the risk of secondary crashes (Yang et al., 2014b).

Table 2.1 Contributing Factors of Secondary Crashes in Existing Studies

Author	**Model**	**Research findings**
Zhan et al. (2009)	Logit model	➢ Secondary crash risks at morning peak are higher than other times ➢ Longer duration crash increases secondary crash risks
Zhang and Khattak (2010)	Ordered logit model	➢ Longer duration crash increases secondary crash risks ➢ Larger number of involved vehicles increases secondary crash risks ➢ Curve segment leads to increased risks of secondary crashes
Yang et al. (2014b)	Logit model	➢ Secondary crash risks at peak hours are higher than those at peak-off hours ➢ Rear-end primary crashes are more likely to incur secondary crashes ➢ Longer duration crash increases secondary crash risks
Mishra et al. (2017)	Multinomial logit model	➢ Larger numbers of involved vehicles increase secondary crash risks ➢ Rear-end primary crashes are more likely to incur secondary crashes

		➢ Clear weather reduces the risks of secondary crashes ➢ Secondary crashes are more likely to occur on arterials than on freeways
Wang et al. (2016a, 2016b)	Logit model	➢ Clear weather reduces the risks of secondary crashes ➢ Larger speed of shockwave increases secondary crash risks ➢ Longer duration crash increases secondary crash risks

2.4 HOPIT Model for Examining the Crash Injury Severity

In the literature, discrete choice regression models are widely implemented to analyze crash injury severity. Discrete choice regression models can be further classified into (a) logit models, including nested models (Chang and Mannering, 2007; Lee and Mannering, 2002), multinomial logit models (Shankar and Mannering, 1996; Ulfarsson and Mannering, 2004; Ye and Lord, 2014), Mixed logit models (Anastasopoulos and Mannering, 2011; Cerwick et al., 2014; Li et al., 2019; Wu et al., 2014; Ye and Lord, 2014), and (b) probit models include ordered probit model with fixed and random parameters (Chen et al., 2016; Fountas and Anastasopoulos, 2018; MCCARTHY and MADANAT, 1994; Ye and Lord, 2014). Driver injury severities are often modeled as discrete injury severity outcomes (for instance, NI (no injury), MI (minor injury), and SI (severe injury)). Both ordered probit models and discrete choice models have their limitations in modeling discrete injury severity outcomes. These discrete outcome models with the flexibility of overlapping possible variables across the outcomes can estimate distinct sets of independent variables for each crash injury severity result (Fountas and Anastasopoulos, 2017). These models assume that the discrete outcomes are independent of each other and they cannot consider the ordinal nature of crash injury severity. In contrast, the ordered probit model assumes that the same independent variables have different influences on different crash injury severity outcomes, which enables the ability of the ordered probit model to account for the ordinal characteristics of crash injury severity. However, Washington et al. (Washington et al., 2011) and Savolainen et al. (Savolainen et al., 2011) pointed out that the ordered probit model cannot explain how the thresholds affect the independent variables on the highest and lowest ordered discrete category probabilities and impact on the interior category probabilities. The thresholds refer to the estimable parameters that profoundly affect intermediate categories. The hierarchical ordered probit (HOPIT) model can overcome this limitation. The thresholds in the

HOPIT model are always positive and ordered, as a function of unique explanatory parameters that do not necessarily affect the ordered probability outcomes directly (Fountas and Anastasopoulos, 2017).

2.5 Summary

In summary, this chapter provides a comprehensive review of current research about methods for secondary crash identification, modeling contributing factors of secondary crashes, and the HOPIT model for crash injury severity analysis. Based on the literature review, this project will develop a hybrid method for primary and secondary crash identification. A logit model will be developed to identify the contributing factors of secondary crashes, and HOPIT models will be applied to patterns of crash injury severity. The rest of the report is organized as follows: Chapter 3 introduces the data preparation for the project; Chapter 4 introduces the developed hybrid method, logit model, and HOPIT models; Chapter 5 conducts case studies and analyzes the results. Chapter 6 summarizes all the findings of this research.

3.0 DATA PREPARATION

3.1 Overview

The research team downloaded the records of all crashes and labeled secondary crashes on freeways from the Numetric database. Manual verification was conducted based on downloaded crash records. This chapter also summarizes the crash dataset and determines the range of crash data that will be used for the research. The remainder of this chapter is organized as follows: Section 3.2 introduces the Numetric database and downloaded crash dataset; Section 3.3 presents the manual verification of the accuracy of secondary crash records in the database; and Section 3.4 summarizes the findings.

3.2 Data Preparation

The Numetric database provides traffic crash data on local arterials and freeways in Utah. The interface of Numetric database is shown in Figure 1. In this project, we mainly retrieved records of crashes that occurred on freeways (e.g., I-15, I-215, I-80, etc.). 127,558 crash records were obtained from 2010 to 2020. The extracted crash characteristics included the date, time, crash severity, collision type, road surface conditions, weather conditions, lighting conditions, etc. Figure 2 shows the yearly distribution of crash frequencies. In addition, only 390 labeled SCs were found from 2010 to 2020 by filtering "Roadway Contributing Circumstances = Prior Crash" from the database, which, given other research on this topic, is likely significantly lower than the reality. Such data quality cannot meet the needs of developing effective incident management strategies. We also plotted the crash counts on I-15, I-80, and I-215 from 2017 to 2019. It shows that most crashes happened on I-15. So, crashes on I-15 from 2017 to 2019 have been used for the following case studies.

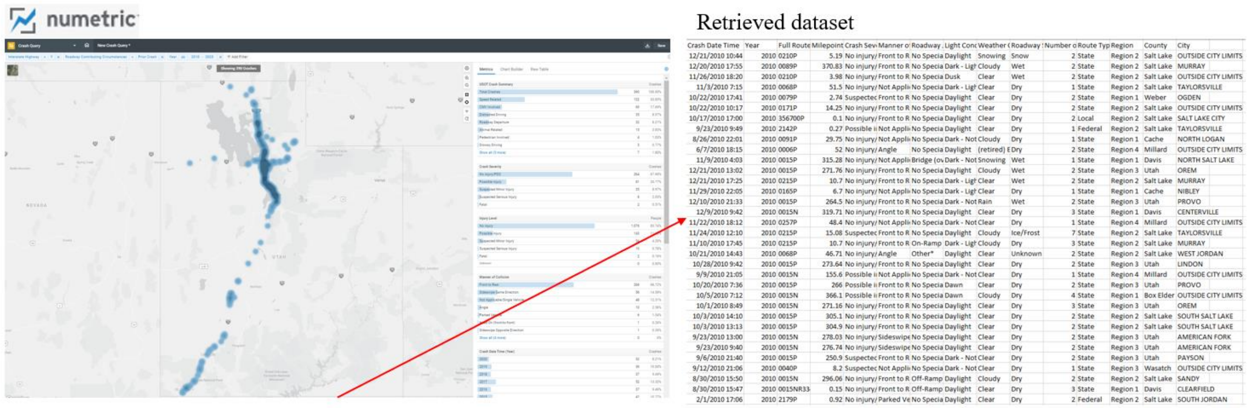


Figure 3.1 The interface of Numetric database

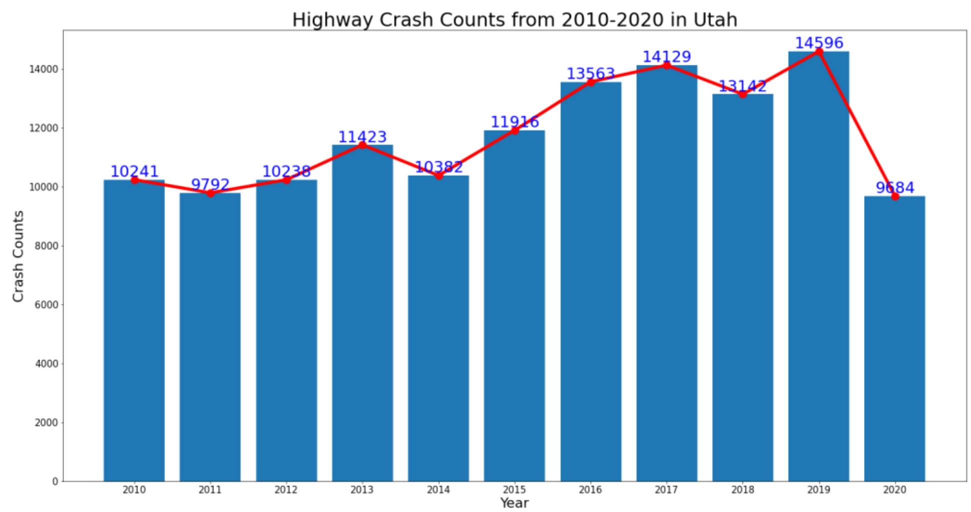


Figure 3.2 The distribution of crash records on highways in Utah

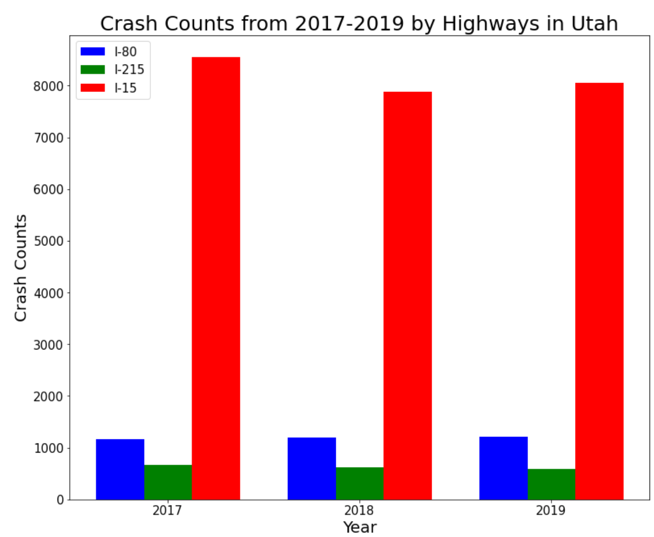


Figure 3.3 The crash counts on different freeways from 2017-2019

3.3 Manual Verification of the Accuracy of Secondary Crash Records in the Database

125 recorded secondary crashes from 2017-2019 were selected for manually verifying the accuracy of those records. The static method (temporospatial thresholds 1h and 2miles) was applied for finding the primary crashes. It is intuitive that there should be one or more crashes that occurred before the labeled secondary crashes. The record is accurate if we can find the primary crashes based on the time and location of the secondary crash extracted from the record. Otherwise, it is not accurate. Based on the records, there are three cases as shown in Figure 1.3: 1) Paired primary crash found; 2) No paired primary crash found; 3) Suspected paired primary crash found. Suspected paired primary crash found means that we could find the crashes that happened before the secondary crash record, but it was in the wrong direction. It may occur because the police officer registered the wrong information. The verification results are shown in Table 1.1. Only 56.8% of secondary crash records found the paired primary crashes. 36.8% of secondary crash records did not find the paired primary crashes. Hence, these conclusions could be reached:

- ❖ The accuracy of secondary crash records is low.
- ❖ The labeled secondary crash falls far below reality.
- ❖ An advanced algorithm for primary and secondary crashes identification will be needed.

Crash ID	Crash Date Time	Year	Full Route Name	Milepoint
1900534299	2/5/2019 12:16	2019	0015N	351.902
1900531409	2/5/2019 12:15	2019	0015N	351.827
1900532380	2/5/2019 12:15	2019	0015N	351.806
1900533420	2/5/2019 12:15	2019	0015N	351.858
1900533168	2/5/2019 12:14	2019	0015N	352.003
1900533274	2/5/2019 12:14	2019	0015N	351.906

(a) Paired primary crash found

1900590662	11/12/2019 10:00	2019	0015N	336.39
1900589970	11/12/2019 9:46	2019	0015N	337.835
1900592093	11/12/2019 8:54	2019	0015N	282.257
1900589507	11/12/2019 8:49	2019	0015N	63.007
1900590666	11/12/2019 8:45	2019	0015N	333.744

(b) No paired primary crash found:

1900530989	1/29/2019 14:35	2019	0015P	338.504
1900531025	1/29/2019 14:32	2019	0015P	339.9

(c) Suspected paired primary crash found:

Red = Labeled secondary crash; Green = Paired primary crash; Blue = No paired primary crash; Yellow = Suspected paired primary crash.

Figure 3.4 The cases of manual verification

Table 3.1 The Results of SC Manual Verification

	Number	Percentage
Paired primary crash found	71	56.8%
No paired primary crash found	46	36.8%
Suspected paired primary crash found	8	6.4%
Total	125	100%

3.4 Summary

In summary, this chapter provides a preliminary analysis of traffic crash records and labeled secondary crash records in the crash database. The results show that the accuracy of labeled secondary crash records is low. To tackle this issue, this study develops a hybrid method to identify the potential primary and secondary crashes from the database with all the traffic crash records that happened on I-15 from 2017 to 2019.

4.0 METHODOLOGY

4.1 Overview

As shown in the data preparation, the accuracy of labeled secondary crash records in the database is low. Therefore, this chapter presents a hybrid method for identifying potential primary and secondary crashes from the database. Based on the identified crash data, a binary logit model is developed to find the contributing factors of secondary crashes. In addition, we examined injury severity patterns of primary and secondary crashes using HOPIT models. The rest of this chapter is organized as follows: Section 4.2 introduces a hybrid method for primary and secondary crashes identification; Section 4.3 presents a binary logit model for modeling the contributing factors of secondary crashes; Section 4.4 introduces the HOPIT model for examining the injury severity patterns of primary and secondary crashes. Section 4.5 summarizes the findings.

4.2 Hybrid Method for Primary and Secondary Crashes Identification

To overcome the limitations of existing static and dynamic methods, this study used a hybrid method that combines the traditional static method (i.e., fixed temporospatial thresholds) and speed contour plot to identify primary and secondary crashes. The main idea of the hybrid method is to identify paired prior and secondary crash by fixed temporospatial thresholds and then validate it with the spatial and temporal impact range of a prior crash using real-time traffic flow data. Figure 4.1 illustrates the flowchart of the hybrid method.

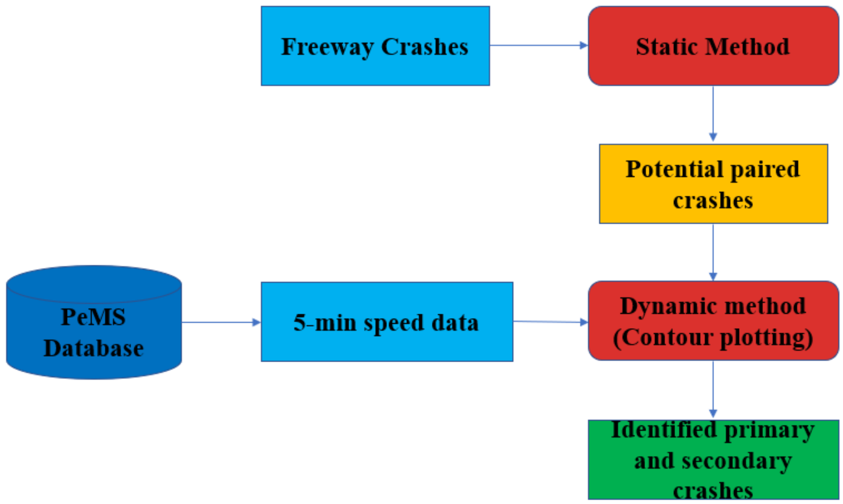


Figure 4.1 The flowchart of hybrid method for secondary crashes identification

4.2.1 Static Method

First, the static method is applied to obtain potential paired crashes. The basic logic in the static method is to use fixed temporospatial thresholds to identify paired prior and secondary crashes from the database. Based on the literature review, the fixed temporospatial thresholds of two miles and one hour are set up for the static method in this study.

4.2.2 Dynamic Method

After the potential paired crashes are filtered by the static method, the speed contour plot, one of the dynamic methods, is used to identify secondary crashes. The core logic is to determine the spatial and temporal impact range of a prior crash using real-time traffic flow data while accounting for the effects of recurrent congestion. A secondary crash is then identified if it is within the spatial and temporal impact range of this prior crash. The detailed procedure for implementing the dynamic method can be stated as follows:

- Identify the location of the labeled secondary crash (shown in Figure 4.2).
- Extract 5-min speed data from detectors upstream and downstream of the location of the labeled secondary crash.

- Implement traffic state estimation to obtain high-resolution data for plotting temporospatial speed contour.
- Construct a speed contour plot for a labeled secondary crash: speed data (between before and after 6 h of labeled secondary crash) from traffic detectors within about 2 miles of upstream and downstream. Figure 4.3 presents an example of a speed contour plot for a prior crash. Congestion and queue formation can be clearly seen.
- Subtract the average speed over crash-free days to build a new contour plot. With the average speed, the effects of recurrent congestion can be eliminated. Figure 4.4 presents an example of a subtracted average speed contour plot of a labeled secondary crash.
- The crashes were found as primary crashes if they happened in the same fixed temporospatial impact ranges of the labeled secondary crash.

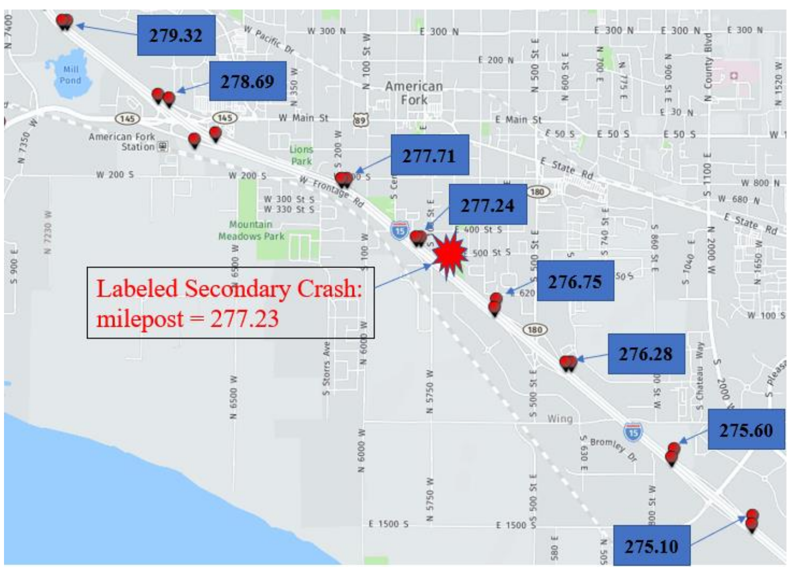


Figure 4.2 Illustration of downloading data for dynamic approach

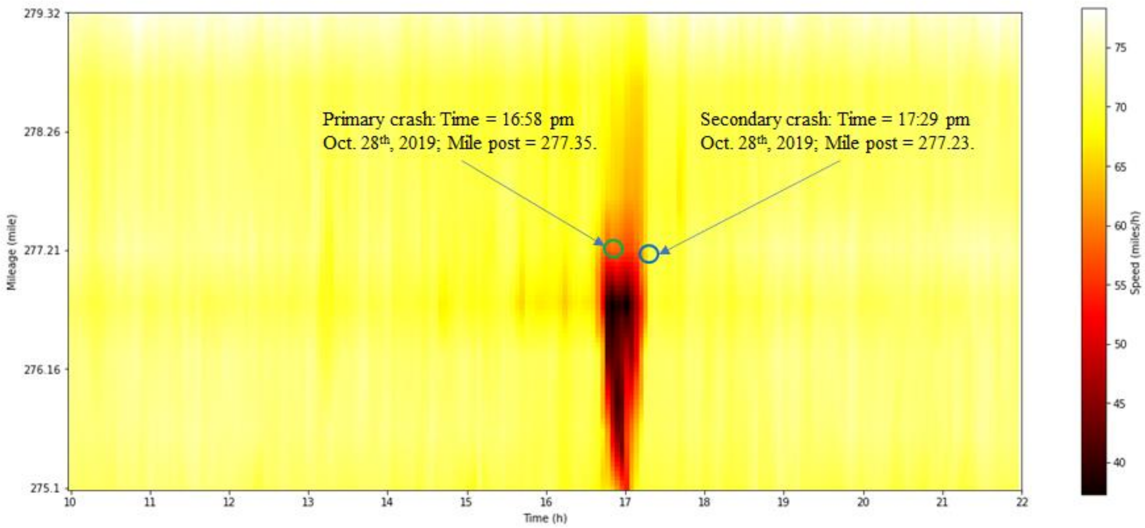


Figure 4.3 Speed contour plot of labeled secondary crash

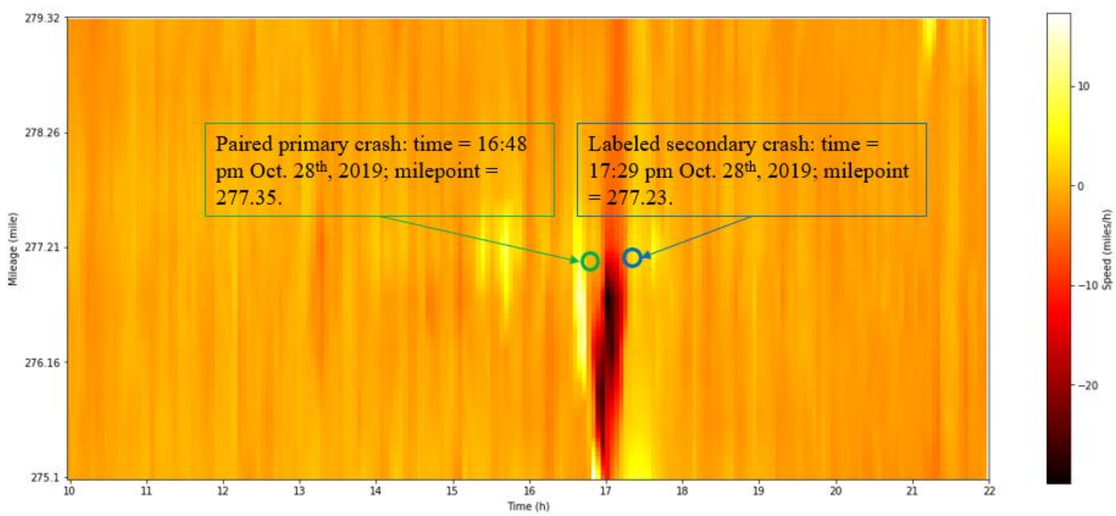


Figure 4.4 Subtracted Average Speed contour plot of labeled secondary crash

4.3 Binary Logit Model

The probability of the occurrence of a secondary crash, given that there is a crash, is equal to the probability of occurrence of the primary crash since the identified primary and secondary crashes are in pairs:

$$P(Occurrence\ of\ secondary\ crash|crash) = P\ (primary\ crash|Crash)$$

The binary logit model is used for modeling the probability of the occurrence of secondary crashes. In this project, the dependent variable of the logit model is the probability of the resulting outcome indicating the presence of a binary indicator variable coded as 1 (primary crash) or 0 (normal crash). The general form of the logistic model used in this project is presented in Equation 1.

$$P_i = \frac{e^{\beta}}{1 + e^{\beta}}, \beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

The logistic regression equation is approximately linear in the middle ranges and logarithmic at extreme values (Washington et al., 2011). A simple transformation of logistic regression is shown below:

$$(\frac{P_i}{1-P_i}) = e^{(\widehat{\beta_0}+\widehat{\beta_\iota}x_i)} = e^{\widehat{\beta_0}}e^{\widehat{\beta_\iota}x_i}$$

which shows that when the value of an explanatory variable increases by one unit, and all other variables are held constant, the probability ratio becomes:

$$(\frac{P_i}{1-P_i})* = e^{\widehat{\beta_0}}e^{\widehat{\beta_\iota}(x_i+1)} = e^{\widehat{\beta_0}}e^{\widehat{\beta_\iota}x_i}e^{\widehat{\beta_\iota}} = (\frac{P_i}{1-P_i})\, e^{\widehat{\beta_\iota}}$$

Thus, an increase in the independent variable x_i by one unit (all other factors held constant, which is typically only possible when multicollinearity does not exist), the odds $(\frac{P_i}{1-P_i})$ increase by the factor $e^{\widehat{\beta_\iota}}$. The factor $e^{\widehat{\beta_\iota}}$ is the odds ratio and indicates the relative amount by which the odds of an outcome increase (odds ratio >1) or decrease (odds ratio <1) when the value of the corresponding independent variable increases by 1 unit.

4.4 Hierarchical Ordered Probit (HOPIT) Model

In this project, the HOPIT models are developed to identify significant contributing factors and quantify their impacts on crash injury severities in primary and secondary crashes. To investigate the crash injury severity in primary and secondary crashes with an ordered probability setting, this study utilized ordered probability models by defining an unobserved variable z that can be used as a basis for modeling the ordinal ranking of data. The unobserved variable z can be denoted as follows (Washington et al., 2011):

$$z_i = \beta\chi_i + \varepsilon_i \quad (1)$$

where χ is a vector of explanatory variables determining the order for observation i; β is a vector of estimable parameters; and ε is a random disturbance. The observed ordinal data y, corresponding to the order of injury-severity outcomes for each observation, can be determined as below (Washington et al., 2011):

$$y_i = j \text{ if } \mu_{j-1} < z_i < \mu_j, j = 1, \dots, J \tag{2}$$

where, μ are threshold parameters; y and j represent ordered ranking of injury severity such as "no injury", "minor injury", and "severe injury."

The ordered probability results are fixed among the observations in the traditional ordered probit model. Not all ordinal data are best modeled using ordered probability models (Washington et al., 2011) since the restrictions that are placed on how variables are believed to affect ordered discrete outcome probabilities. The HOPIT model has the ability to solve this problem to some extent by allowing thresholds to be varied as a function of a set of explanatory parameters, which can be expressed as follows (Greene and Hensher, 2010):

$$\mu_{i,j} = \mu_{i,j-1} + \exp(t_j + \boldsymbol{d}_j\boldsymbol{S}_i) \tag{3}$$

where $\boldsymbol{S}$ are vectors of variables affecting the thresholds, $\boldsymbol{d}$ are vectors of estimable parameters for $\boldsymbol{S}$, and t is the intercept for each threshold. The threshold μ_0 is assumed to be zero, without loss of generality (Washington et al., 2011). The number of estimable thresholds is equal to the total crash severity level j – 2. In this study, the ordered probability of each crash severity level j of each observation can be determined by the following equation (Washington et al., 2011):

$$P(y = j) = \Phi(\mu_j - \beta\chi_i) - \Phi(\mu_{j+1} - \beta\chi_i) \tag{4}$$

where $P(y = j)$ is the probability of each crash injury severity level j; $\Phi(\cdot)$ represents the cumulative normal distribution; and μ_j and μ_{j+1} denote the upper and lower thresholds for outcome j.

The influence of each explanatory variable on the probability of each crash injury severity level cannot be captured by the parameter estimates (especially on the intermediate levels) (Washington et al., 2011). To address this problem, it can be calculated by marginal effects (Anastasopoulos et al., 2012; Fountas and Anastasopoulos, 2017; Russo et al., 2014; Washington et al., 2011) using the following equation:

$$\frac{P(y=j)}{\partial \chi} = [\Phi(\mu_{j-1} - \beta\chi) - \Phi(\mu_j - \beta\chi)]\beta \tag{5}$$

The marginal effects are computed at the sample mean of the explanatory variables and calculated using the average of β for random parameters. The marginal effects measure the change in the outcome probability of each ordered ranking, which is caused by a unit change in a continuous or ordinary explanatory variable.

4.5 Summary

This chapter elaborates the proposed methods for identifying the primary and secondary crashes from the database, modeling contributing factors of secondary crashes, and examining the crash injury severity patterns in primary and secondary crashes. First, a hybrid static and dynamic method was developed to identify the primary and secondary crashes from the crash database. Second, the binary logit model was used to model the relationship between secondary crashes and various contributing factors with the identified data. Lastly, the HOPIT model was introduced to examine the contributing factors on crash injury severity.

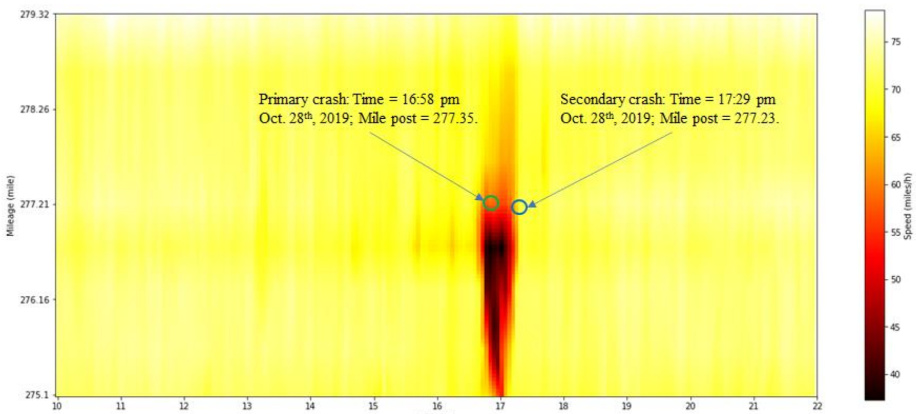
5.0 EXPERIMENTAL ANALYSIS

5.1 Overview

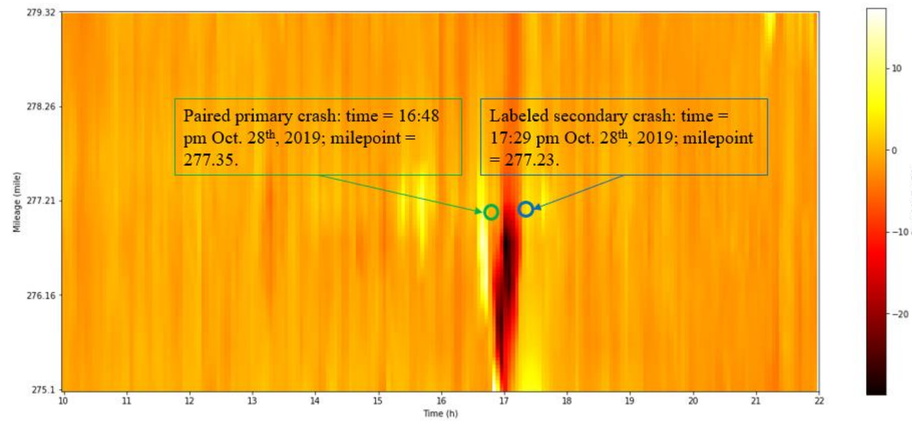
To test the proposed hybrid method for identifying primary and secondary crashes, this research selects the crashes that occurred on I-15 in Utah from 2017 to 2019 for experimental analysis. Then the contributing factors of secondary crashes are estimated by the binary logit model based on the identified primary and secondary crash dataset. In addition, the HOPIT models are constructed for examining the crash injury severity patterns in primary and secondary crashes.

5.2 Primary and Secondary Crashes Identification

The proposed hybrid method was applied to identify the primary and secondary crashes from all crash records in the database. The experimental study was conducted on I-15 in the state of Utah. Three-year (2017 to 2019) crash and traffic data were retrieved from the Numetric database and Performance Measurement System (PeMS) managed by the Utah Department of Transportation (UDOT). We used fixed temporospatial thresholds of two miles and one hour to filter the potential primary and secondary crashes. Then the dynamic approach is implemented to cross-check the accuracy of identified primary and secondary crashes. Figure 5.1 to Figure 5.4 show four typical cases of the dynamic approach.

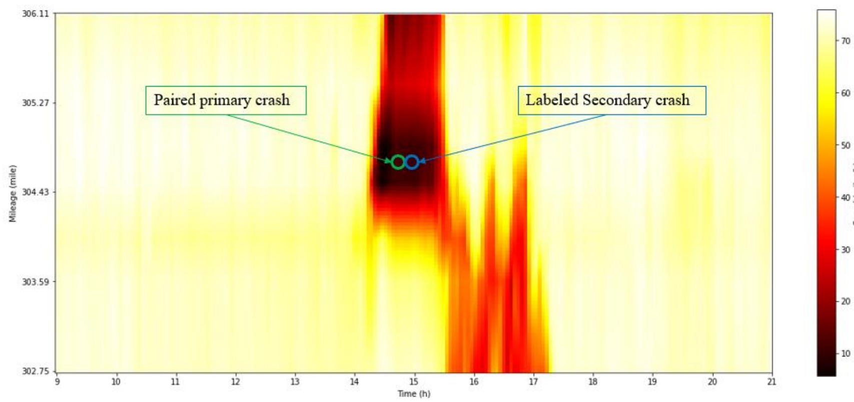


(a)

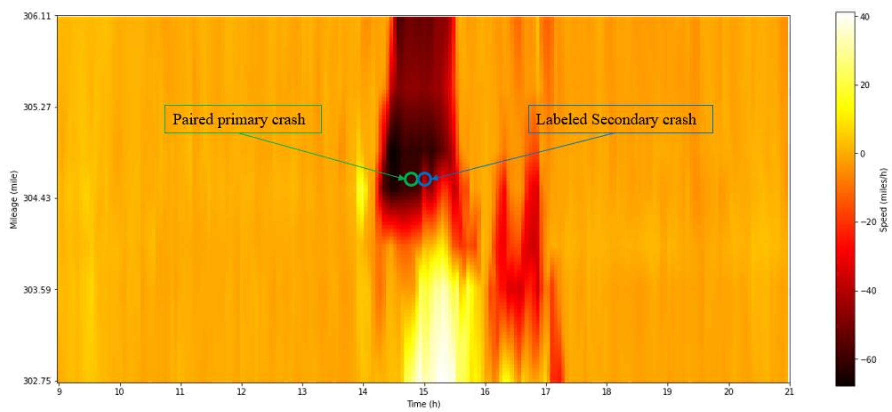


(b)

Figure 5.1 Labeled secondary crash: time = 17:29 pm Oct. 28th, 2019; MP = 277.231 (NB). Paired primary crash: time = 16:58 pm Oct. 28th, 2019; MP = 277.345 (NB).

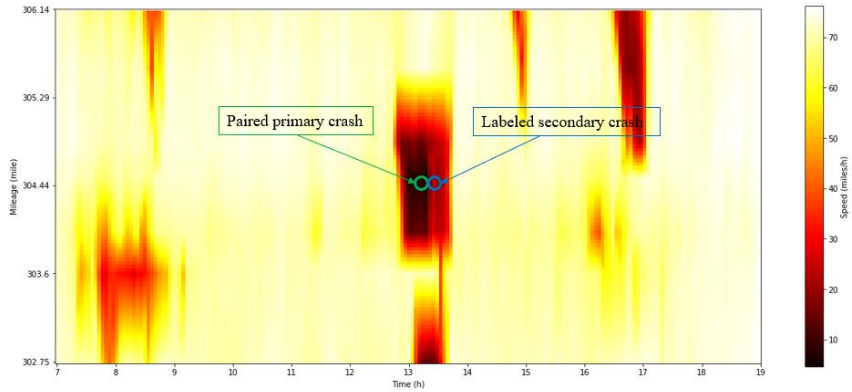


(a)

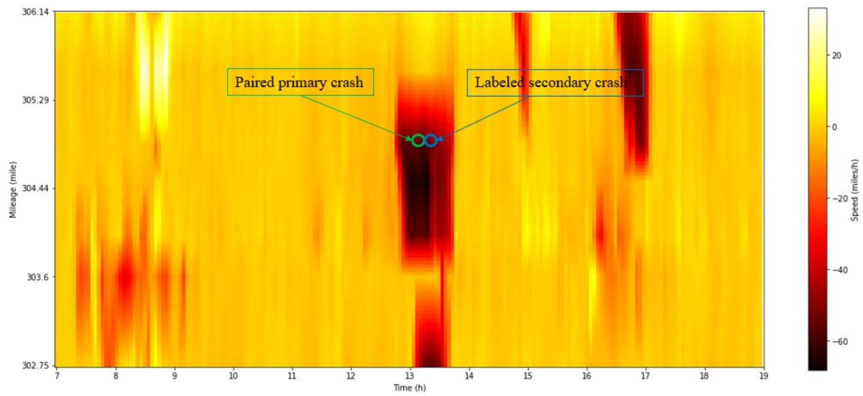


(b)

Figure 5.2 Labeled secondary crash: Time = 15:00 pm Sep. 13th, 2019; MP = 304.51(SB). Paired primary crash: Time = 14:49 pm Sep. 13th, 2019; MP = 304.22(SB).

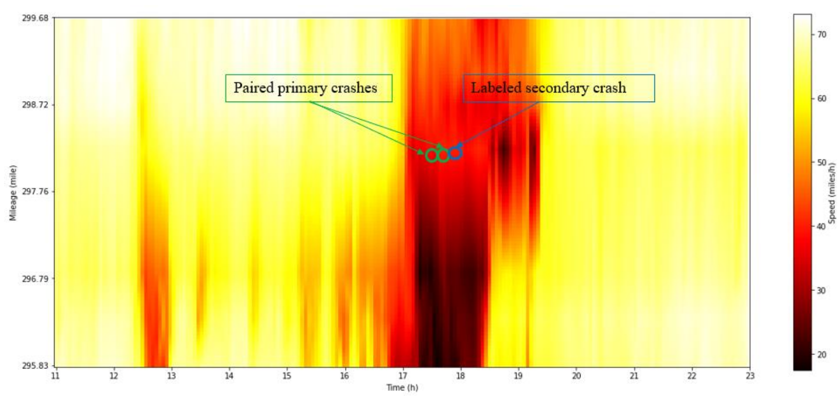


(a)

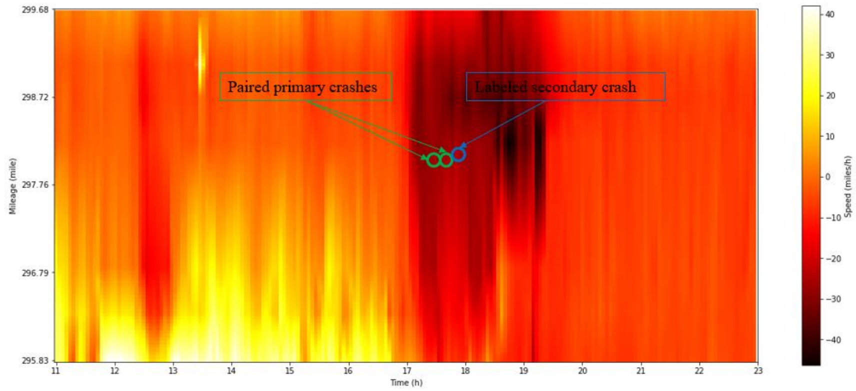


(b)

Figure 5.3 Labeled secondary crash: Time = 13:19 pm Apr. 11th, 2019; MP = 304.91 (NB). Paired primary crash: Time = 13:20 pm Apr. 11th, 2019; MP= 304.93 (NB).



(a)



(b)

Figure 5.4 Labeled secondary crash: Time = 17:37 pm Dec. 27th, 2018; MP = 298.04 (NB). Paired primary crash: Time = 17:37 pm Dec. 27th, 2018; MP = 297.99 (NB). Paired primary crash: Time = 17:37 pm Dec. 27th, 2018; MP= 297.99 (NB).

After we implemented the static method, 2,710 primary crashes and 3,341 secondary crashes are found. These identified primary and secondary crashes were validated by the hybrid method. Finally, 2,653 (97.95%) primary crashes, 2,953 (88.4%) secondary crashes, and 18,878 normal crashes were identified in the database. Table 5.1 presents the distribution of identified primary and secondary crashes from 2017 to 2019 and the percentage in the total crashes.

Table 5.1 Identified Primary, Secondary, and Normal Crashes by Hybrid Method

Time	Identified primary crash	Identified secondary crash	Identified normal crash	Total
2017	1049 (12.3%)	1181 (13.8%)	6349 (74.2%)	8549
2018	886 (11.2%)	960 (12.2%)	6042 (76.6%)	7888
2019	718 (8.5%)	812 (9.7%)	6877 (81.8%)	8407
Total	2,653	2,953	18,878	24484

5.3 Modeling Contributing Factors of Secondary Crashes

Based on identified prior and secondary crashes, detailed information (such as crash injury severity level, crash occurrence time, driver information, weather conditions, environmental conditions, roadway surface condition, location, etc.) is collected for each crash. 16,332 out of 18,878 normal crash records were used for model development in the next step, after removing incomplete records. A more detailed definition of variables of observations is presented in Table 5.2.

Table 5. 2 Variable Definition in Primary and Secondary Crash

Variable	Variable description
Crash injury severity	1. Property damage only (PDO); 2. Possible injury; 3. Suspected minor injury; 4. Suspected serious injury; 5. Fatal.
Age	Number of age.
Gender	1. Female; 2. Male.
Day of week	Mon – Sun: 1-7.
Suspected alcohol	N-0; Y-1.
Suspected drug	N-0; Y-1.
Light condition	1. Dark; 2. Dusk and dawn; 3. Daylight.
Weather Condition	1. Clear; 2. Cloudy/windy; 3. Fog; 4. Rain; 5. Snow (snow & sleet & hailing).
Manner of collision	1. Angle; 2. Front to rear; 3. Head-on; 4. Single crash; 5. Sideswipe same direction; 6. Other.
Crash type	1. Other; 2. Rear-end; 3. Roadway departure.
Roadway surface condition	1. Dry; 2. Ice; 3. Slush; 4. Wet; 5. Water; 6. Snow; 7. Other
Number of vehicles involved	1. Single vehicle; 2. Two vehicles; 3. Multiple vehicles.
Distracted driver	N-0; Y-1.
Drowsy driver	N-0; Y-1.
Animal related	N-0; Y-1.
Variable	Variable description
Collision with fix object	N-0; Y-1.
Overturn rollover	N-0; Y-1.
Speed related	N-0; Y-1.
Urban Rural	1. Rural; 2. Urban.
Work Zone Involved	N-0; Y-1.
Estimated speed	Number.
Post speed	Number.
Air Bag	1. No air bag; 2. Deployed; 3. Not deployed; 4. Unknown.
Seatbelt	1. No; 2. Yes; 3. Unknown.

Vehicle type	1. Passenger; 2. Pickup; 3. Minivan; 4. Other.
Adverse roadway surface condition	N-0; Y-1.
Zero Aggressive	N-0; Y-1.
Vehicle Maneuver	1. Changing lanes; 2. Slowing in traffic lane; 3. Stop in traffic lane;4. Straight ahead; 5. Other.
Horizontal Alignment	1. Curved; 2. Straight 3. Unknown.
Traffic Control Device	1. No control device; 2. Traffic control signal; 3. Yield sign; 4. HOV lane; 5, Other.
Roadway Junction Feature	1. No special feature; 2. Onramp; 3. Offramp; 4. Bridge. 5. Other.

As shown in Table 5.3, 12 variables (including young people, daylight, snowy weather, angle collision, rear-end crash, multiple vehicles involved, collision with fixed objects, speed-related crash, minivan, adverse roadway surface condition, vehicle slowing in traffic lane, and roadway with straight alignment) are found to be positively associated with the probability of secondary crashes, indicating that those factors will significantly increase the probability of the occurrence of secondary crashes. Only “Weekend” and “Rural” are negatively associated with the probability of the occurrence of secondary crashes, indicating that crashes occurring on weekends and in rural areas are less likely to lead to a secondary crash. The odds ratio represents the increase in likelihood that a crash will lead to a secondary crash. For example, for a rear-end crash, there is an almost 89.3% increase in the likelihood that a crash leads to a secondary crash.

Table 5.3 Modeling Results for Secondary Crash Risk Prediction

Variable	Estimated Parameter	*T-ratio*	*P-value*	95% Coefficient Interval
Constant	-5.19	-35.12	0.00^{***}	-5.482, -4.902
Age (Young)	0.10	1.73	0.08^{*}	-0.014, 0.216
Weekend	-0.35	-4.34	0.00^{***}	-0.513, -0.194
Light (Daylight)	2.40	18.07	0.00^{***}	2.137, 2.657
Weather (Snow)	0.83	8.30	0.00^{***}	0.636, 1.030

MOC (Angle)	0.50	3.26	0.00^{***}	0.200, 0.804
Crash type (Rear-end)	0.64	8.77	0.00^{***}	0.496, 0.781
Multiple vehicles involved	0.26	3.77	0.00^{***}	0.123, 0.390
Collision with fixed object	0.23	3.01	0.00^{***}	0.081, 0.386
Speed related	0.19	3.15	0.00^{***}	0.071, 0.303
Rural	-0.96	-7.85	0.00^{***}	-1.197, -0.719
Minivan	1.83	2.65	0.00^{***}	0.478, 3.176
Adverse Roadway Surf Condition	0.40	4.68	0.00^{***}	0.233, 0.568
Vehicle Maneuver (Slowing in traffic lane)	0.20	2.83	0.00^{***}	0.061, 0.335
Horizontal Alignment (Straight)	0.24	4.32	0.00^{***}	0.131, 0.348
	Odds Ratio			
Age (Young)	1.106			0.980, 1.233
Weekend	0.702			0.590, 0.814
Light (Daylight)	10.990			8.133, 13.847
Weather (Snow)	2.300			1.847, 2.752
MOC (Angle)	1.652			1.153, 2.151
Crash type (Rear-end)	1.893			1.623, 2.163
Multiple vehicles involved	1.292			1.120, 1.465
Collision with fixed object	1.264			1.071, 1.456
Speed related	1.206			1.065, 1.346
Rural	0.384			0.292, 0.475
Minivan	6.214			-2.171,14.600
Adverse Roadway Surf Condition	1.493			1.242, 1.743
Vehicle Maneuver (Slowing in traffic lane)	1.218			1.052, 1.385
Horizontal Alignment (Straight)	1.270			1.132, 1.408
Number of observations			18985	
Log-likelihood			-5078.60	

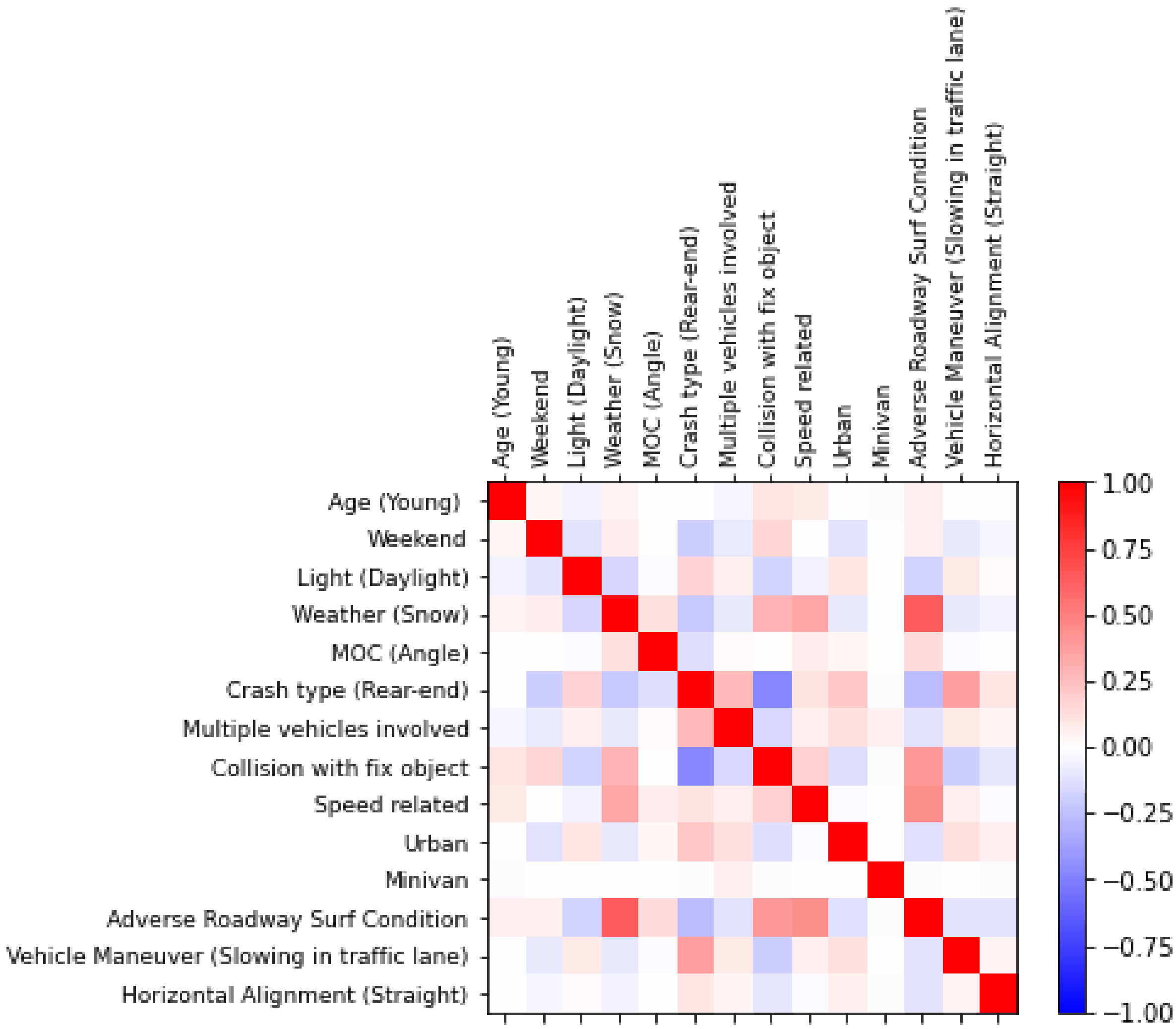


Figure 5.5 Variable correlation results in binary logit model

The correlation test is conducted to determine the correlation between variables. The autocorrelation of variables is presented in Figure 5.5, which indicates that there is no strong correlation between all candidate variables. The ROC curve is used to evaluate the predictive performance of different models (Egan, 1975). A model of binary outcome (primary crash = 1 and non-primary crash = 0) classifies an observation as an event if the predicted probability of the observation exceeds a pre-specified threshold. Otherwise, it will be classified as a non-event. The ROC curve was developed to evaluate the predictive performance of the developed secondary crash risk prediction model presented in Table 5.3. As shown in Figure 5.6, the area under the ROC curve is 0.796, which indicates that the binary logit model has good predictive performance.

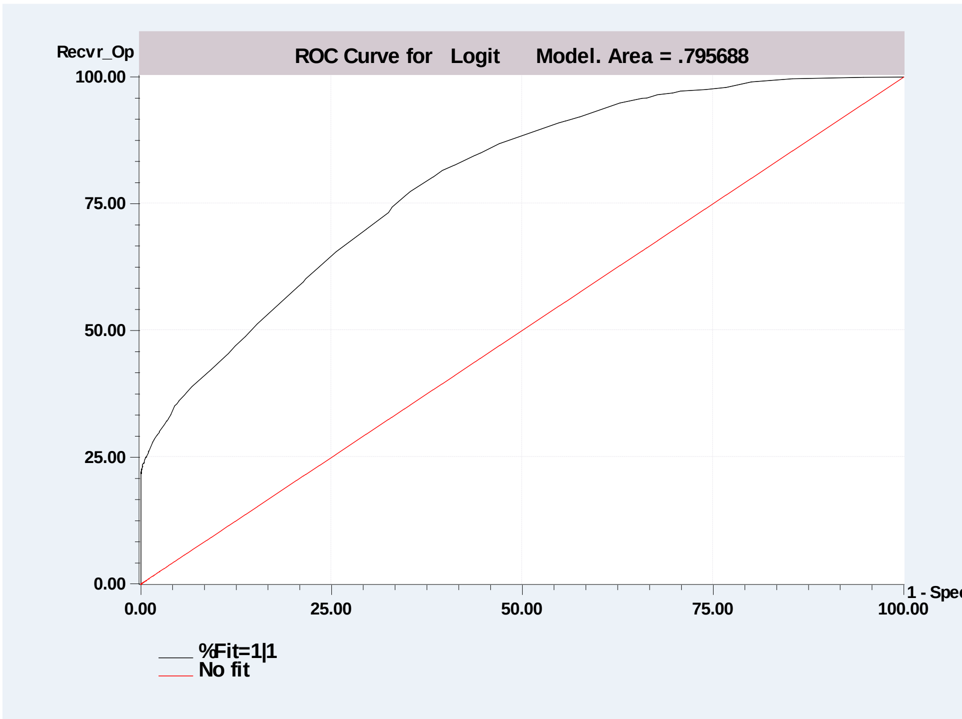


Figure 5.6 ROC curve for the binary logit model

5.4 Examining the Primary and Secondary Crash Injury Severity Patterns

The crash injury severity was grouped by five levels in the original UDOT dataset including no injury, possible injury, minor injury, severe injury, and fatal. In this study, possible and minor injuries are combined as the minor injury level, and severe injuries and fatal are combined as the severe injury level for yielding a statistically meaningful sample size. Hence, the driver injury severity is recategorized into three levels including NI (no injury), MI (Minor injury), and SI (severe injury) which is similar to existing studies (Behnood et al., 2014; Hou et al., 2019; Zhang et al., 2021). In this project, two HOPIT models were estimated for primary and secondary crashes. Before running the model, the correlation between variables was plotted to test the autocorrelation of variables and presented in Figure 5.7 – 5.8. The figures show that there is no strong relationship between all candidate variables. Table 5.4 and Table 5.6 show the estimated results of the HOPIT models for primary and secondary crashes.

Among 2,653 identified primary crashes, 1,835 (69.17%), 771 (29.06%), and 26 (1.77%) records were reported as no injury, minor injury, and severe injury, respectively. Among 2,953 identified secondary crashes, 2,159 (73.11%), 768 (26.01%), and 26 (0.88%) records were reported as no injury, minor injury, and severe injury, respectively. Thirteen variables are found to be significant in primary crashes. Nine variables are found to be significant in secondary crashes. All variables are statistically significant to explain the variations in the threshold. The negative coefficients of threshold covariates indicate an upward shift on the threshold parameter, and positive coefficients of threshold covariates indicate a downward shift on the threshold parameter. In Table 5.5 and Table 5.7, the marginal effects of each explanatory variable, related to the probability of a single crash that results in a severity outcome, are estimated for primary crash and secondary. More detailed result analyses and explanations follow.

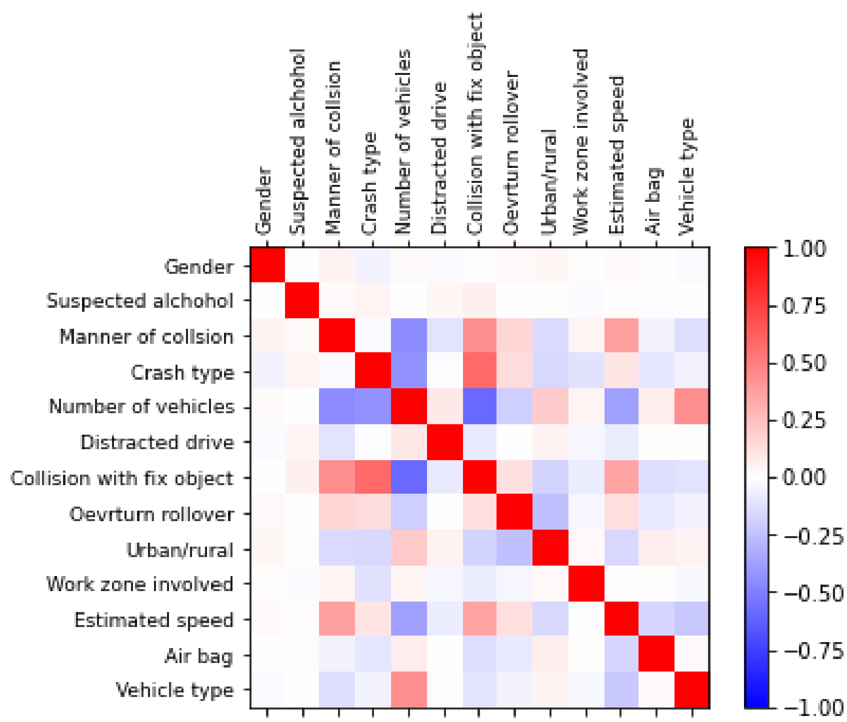


Figure 5.7 Variable correlation results in HOPIT model for primary crash

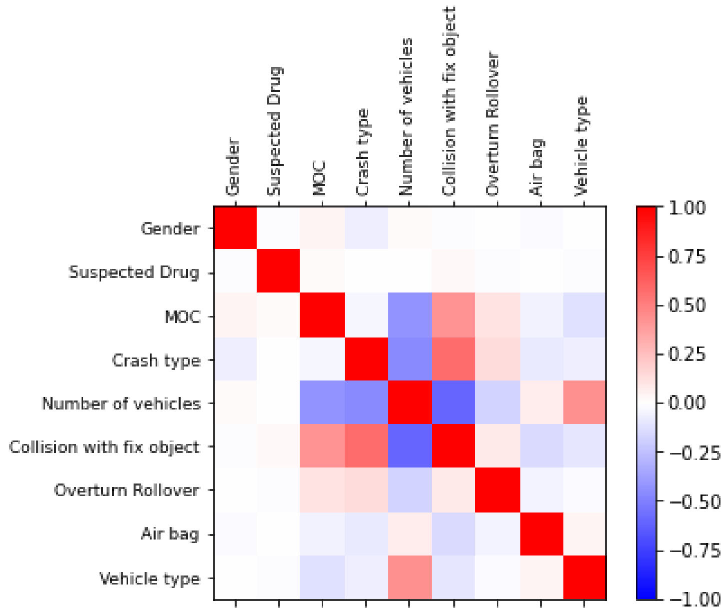


Figure 5.8 Variable correlation results in HOPIT model for secondary crash

For primary crashes, female drivers are more likely to be involved in severe primary crashes by 0.3%. It is reasonable that suspected alcohol is positively related to the crash severity, with the coefficients of 0.60. Drivers with suspected alcohol use are more prone to be involved in minor and severe-injury crashes, especially minor injury crashes. Compared with other collision types, the angle collision has a higher potential impact (coefficient = 0.89) on minor injury and severe injury in primary crashes. Vehicle-fixed-object crashes are 9.7% and 0.9% more likely to lead to minor injury and severe injury in primary crashes (relative to other types of crashes). The front-to-rear crash is positively significant in predicting the crash severity, with the coefficient of 0.54. Compared with single- and two-vehicle involved, multiple vehicles involved (coefficient = 0.60) can significantly increase the possibility of minor and severe injury. It is reasonable that distracted driving is positively related to the crash severity, with a coefficient of 0.23. The distracted driver is more prone to minor and severe injury crashes, especially minor injury crashes. Overturn vehicle crash is positively significant in predicting the crash injury severity, with the coefficient of 1.18. It is more likely to be involved in minor and severe injury primary crashes by 33.7% and 10.7%, respectively. Crashes with minor injuries and severe injuries are less likely to occur in rural areas (with a coefficient of -0.23). Work-zone-involved crashes are more likely to lead to minor injuries and severe injuries. It is reasonable that crashes with high speed have a positively significant impact on increasing the possibility of minor injuries and severe injuries, with a coefficient of 0.16. Primary crashes with airbags deployed are 19.7% and are 2.7% more likely to lead to minor injuries and severe injuries. Normally, the airbag deployed indicates that the vehicle is severely damaged, so the driver might get severely injured. Passenger vehicle type is found to be significant in primary crashes with a negative parameter of -0.24. It may reduce the possibility of minor injury and severe injury in primary crashes.

Table 5.4 Estimation Results for Primary Crash

Variable description	Estimated parameter	Standard error	*T-ratio*	*P-value*
Constant	-1.10	0.13	-8.42	0.00^{***}
Female driver	0.10	0.05	1.80	0.00^{*}
Suspected alcohol	0.60	0.24	2.53	0.01^{***}
Angle collision	0.89	0.12	7.46	0.00^{***}
Front-to-rear crash	0.54	0.07	7.18	0.00^{***}
Multiple vehicles involved	0.60	0.08	7.98	0.00^{***}
Distracted Driver	0.23	0.11	2.02	0.04^{**}
Collison with fixed object	0.30	0.07	4.11	0.00^{***}
Overturn	1.18	0.15	7.95	0.00^{***}
Rural	-0.23	0.12	-1.82	0.07^{*}
Work-zone involved	0.29	0.09	3.07	0.00^{***}
High speed	0.16	0.06	2.62	0.01^{***}
Air bag deployed	0.60	0.07	8.20	0.00^{***}
Passenger vehicle	-0.24	0.11	-2.26	0.02^{**}
Threshold parameter				
θ_1	0.42054	0.06298	6.68	0.00^{***}
Threshold covariates				
y_1	-0.20	0.09	-2.18	0.03^{**}
y_2	0.32	0.08	3.98	0.00^{***}
y_3	0.56	0.15	3.86	0.00^{***}
Summary statistics				
Number of observations		2653		
LL(0)		-1818.80		

LL(β)	-1615.39
AIC	3266.8
McFaden Pseudo R^2	0.11

***, **, * Significance at 1%, 5%, 10% level.

Table 5.5 Marginal Effects for Primary crash

Variable	**No injury**	**Minor injury**	**Severe injury**
Female driver	-0.034	0.031	0.003
Suspected alcohol	-0.227	0.196	0.031
Angle collision	-0.341	0.280	0.061
Front-to-rear crash	-0.182	0.167	0.015
Multiple vehicles involved	-0.220	0.195	0.025
Distracted Driver	-0.082	0.074	0.008
Collison with fixed object	-0.106	0.097	0.009
Overturn	-0.444	0.337	0.107
Rural	0.073	-0.068	-0.005
Work-zone involved	-0.105	0.095	0.010
High speed	-0.055	0.051	0.005
Air bag deployed	-0.224	0.197	0.027
Passenger vehicle	0.088	-0.080	-0.008

According to the results of the HOPIT model developed for secondary crashes, female drivers are more likely to be involved in minor-injury secondary crashes by 5.4%. Drivers with suspected drug use are more likely to be involved in crashes with minor and severe injuries, with a coefficient of 0.85. Drivers with suspected drug use are more prone to be involved in crashes with minor and severe injuries, especially minor-injury crash. Compared with other collision types, the head-on collision has a higher probability of leading to minor injuries and severe injuries, with a coefficient of 0.82. Vehicle-fixed object crashes are 6.0% and 0.2% more likely to lead to minor injuries and severe injuries in secondary crashes (relative to other types of crashes). The rear-end crash is positively related to crash severity, with the coefficient of 0.42.

Compared with single- and two-vehicle involved, multiple vehicles involved (coefficient = 0.48) can significantly increase the possibility of minor and severe injuries, especially for minor injuries. Overturn vehicle crashes may lead to higher crash injury severity, with a coefficient of 1.16. Drivers involved in overturn vehicle crashes are 38.0% or 5.5% more likely to get minor or severely injured, respectively. Secondary crashes with airbags deployed are 20.8% and 1.3% more likely to lead to minor injuries and severe injuries. Passenger vehicle type is found to be significant in primary crashes with a negative parameter of 0.088. It may reduce the possibility of minor injury and severe injury in secondary crashes.

Table 5.6 Estimation Results for Secondary Crash

Variable description	Estimated parameter	Standard error	*T-ratio*	*P-value*
Constant	-0.81	0.12	-6.49	0.00***
Female driver	0.17	0.05	3.35	0.00***
Suspected Drugs	0.85	0.29	2.94	0.00***
MOC (Head-on)	0.82	0.31	2.63	0.01***
Crash type (Rear-end)	0.42	0.07	6.42	0.00***
Multiple vehicles involved	0.48	0.08	6.23	0.00***
Collison with fixed object	0.19	0.07	2.55	0.01***
Overturn	1.16	0.16	7.28	0.00***
Air bag deployed	0.61	0.08	8.13	0.00***
Passenger vehicle	-0.43	0.11	-4.00	0.00***
Threshold parameter				
θ_1	0.76	.053	14.35	0.00***
Threshold covariates				
y_1	-0.33	0.12	-2.82	0.00***
y_2	-1.07	0.43	-2.48	0.01**
y_3	0.28	0.18	1.60	0.10*
Summary statistics				
Number of observations		2953		
LL(0)		1833.52		
LL(β)		-1667.21		

AIC	3362.4
McFaden Pseudo R^2	0.09

***, **, * Significance at 1%, 5%, 10% level.

Table 5.7 Marginal Effects for Secondary Crash

Variable	No injury	Minor injury	Severe injury
Female driver	-0.056	0.054	0.002
Suspected Drugs	-0.319	0.291	0.028
MOC (Head-on)	-0.308	0.282	0.026
Crash type (Rear-end)	-0.133	0.128	0.004
Multiple vehicles involved	-0.168	0.160	0.008
Collison with fixed object	-0.062	0.060	0.002
Overturn	-0.435	0.380	0.055
Air bag deployed	-0.221	0.208	0.013
Passenger vehicle	0.154	-0.146	-0.008

5.5 Summary

This chapter conducted case studies of the proposed hybrid method for primary and secondary crash identification, binary logit model for identifying the contributing factors of secondary crashes, and HOPIT models for examining the primary and secondary crashes injury severity patterns. Firstly, the study results indicate that the proposed hybrid method can effectively identify the primary and secondary crashes from the database. Secondly, the binary logit model finds the contributing factors of secondary crashes. Thirdly, the crash injury severity patterns are identified by HOPIT models.

6.0 CONCLUSIONS

6.1 Summary

Accurate identification of secondary crashes is the basis for identifying contributing factors, and contributing factors are the cornerstones for the incident management system to find effective strategies to reduce the risk of secondary crashes. This project provided a preliminary analysis of traffic crash records and labeled secondary crash records in UDOT's crash database. Results show that the accuracy of labeled secondary crash records is low. To tackle this issue, this project proposed a hybrid method to accurately identify primary and secondary crashes. Based on the identified crash data, the binary logit model was implemented for modeling the contributing factors. In addition, the HOPIT models were developed to examine the crash injury severity in identified primary and secondary crash datasets.

6.2 Findings

Only 125 secondary crashes are recoded from 2017-2019 in UDOT's crash database. Furthermore, the results of manual verification of the accuracy of labeled secondary crash records show that only 56.8% of secondary crash records found the paired primary crashes and 36.8% of secondary crash records did not find the paired primary crashes. Hence, the accuracy of labeled secondary crashes in the database is low.

The hybrid method was proposed to identify primary and secondary crashes from UDOT's crash database. The hybrid methods include static and dynamic components. After we implemented the static method, 2,710 primary crashes and 3,341 secondary crashes were found. Then the identified primary and secondary crashes were validated by the hybrid method. Finally, 2,653 (97.95%) primary crashes, 2,953 (88.4%) secondary crashes, and 18,878 normal crashes were identified in the database from 2017 to 2019. These results indicate that the proposed hybrid method could effectively identify the primary and secondary crashes from the database, which provides the possibility to further analyze the secondary crashes on freeways.

Based on the identified crash data, the binary logit model was conducted to examine the contributing factors of secondary crashes. The study results indicate that 12 variables (including young people, daylight, snowy weather, angle collision, rear-end crash, multiple vehicles involved, collision with fixed objects, speed-related crash, minivan, adverse roadway surface condition, vehicle slowing in traffic lane, and roadway with straight alignment) are found to be positively associated with the risk of secondary crashes, indicating that those factors will significantly increase the probability of the occurrence of secondary crashes. Only “weekend” and “rural” parameters are negatively associated with the probability of the occurrence of secondary crashes, indicating that crashes occurring on weekends and in rural areas are less likely to lead to a secondary crash. Those findings could provide some insightful information to deploy effective countermeasures to reduce secondary crashes on freeways.

In addition, the injury severity patterns of primary and secondary crashes were effectively estimated by HOPIT models. The study results indicate that 13 variables (including female driver, suspected alcohol, angle collision, multiple vehicles involved, etc.) and 9 variables (including female driver, suspected drugs, head-on collision, multiple vehicles involved, etc.) are significantly related to crash injury severity in primary and secondary crashes, respectively. Those findings could help UDOT to reduce the injury severity of primary and secondary crashes on freeways.

6.3 Limitations and Challenges

Although some insightful findings are presented in this research, there are some limitations, including: (1) more comprehensive and multi-source crash data should be utilized to improve the accuracy of primary and secondary crash identification; (2) more crash information should be collected to improve the modeling results of the binary logit model and HOPIT models; (3) there might be some confounding variables that need to be found, such as AADT and road geometry information. Future studies could focus on these issues to overcome this challenge.

REFERENCES

Anastasopoulos, P.C., Karlaftis, M.G., Haddock, J.E., Mannering, F.L., 2012. Household Automobile and Motorcycle Ownership Analyzed with Random Parameters Bivariate Ordered Probit Model. Transp. Res. Rec. J. Transp. Res. Board 2279, 12–20. https://doi.org/10.3141/2279-02

Anastasopoulos, P.C., Mannering, F.L., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accid. Anal. Prev. 43, 1140–1147. https://doi.org/10.1016/j.aap.2010.12.024

Behnood, A., Roshandeh, A.M., Mannering, F.L., 2014. Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities. Anal. Methods Accid. Res. 3–4, 56–91. https://doi.org/10.1016/j.amar.2014.10.001

Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Anal. Methods Accid. Res. 3–4, 11–27. https://doi.org/10.1016/j.amar.2014.09.002

Chang, L.-Y., Mannering, F.L., 2007. Predicting Vehicle Occupancies from Accident Data: An Accident Severity Approach. Transp. Res. Rec. J. Transp. Res. Board 1635, 93–104. https://doi.org/10.3141/1635-13

Chen, Cheng Cheng, Chen, Chien Chung, Hoang, T.T., 2016. Role of concrete confinement of wide-flange structural steel shape in steel reinforced concrete columns under cyclic loading. Eng. Struct. 110, 79–87. https://doi.org/10.1016/j.engstruct.2015.12.002

Egan, J.P., 1975. Signal detection theory and ROC-analysis. Acad. Press.

Fountas, G., Anastasopoulos, P.C., 2018. Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances. Anal. Methods Accid. Res. 20, 30–45. https://doi.org/10.1016/j.amar.2018.09.002

Fountas, G., Anastasopoulos, P.C., 2017. A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. Anal. Methods Accid. Res. 15, 1–16. https://doi.org/10.1016/j.amar.2017.03.002

Greene, W.H., Hensher, D.A., 2010. Modeling ordered choices: A primer. Cambridge University Press.

Hirunyanitiwattana, Worapong, Mattingly., S.P., 2006. Identifying secondary crash characteristics for California highway system, in: Presented at 85th Annual Meeting of the Transportation Research Board. Washington, D.C.

Hou, Q., Huo, X., Leng, J., Cheng, Y., 2019. Examination of driver injury severity in freeway single-vehicle crashes using a mixed logit model with heterogeneity-in-means. Phys. A Stat. Mech. its Appl. 531, 121760. https://doi.org/10.1016/j.physa.2019.121760

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis. Accid. Anal. Prev. 34, 149–161. https://doi.org/10.1016/S0001-4575(01)00009-4

Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., Zhang, G., 2019. Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. Accid. Anal. Prev. 129, 230–240. https://doi.org/10.1016/j.aap.2019.04.001

MCCARTHY, P.S., MADANAT, S., 1994. Highway Accident Data Analysis: Alternative Econometric Methods. Transp. Res. Rec. 44–49.

Mishra, S., Golias, M.M., Sarker, A.A., Naimi, A., 2017. Effect of primary and secondary crashes: identification, visualiztion and prediction, National Center for Freight and Infrastructure Research and Education (CFIRE), University of Wisconsin, Madison.

Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accid. Anal. Prev. 39, 459–468. https://doi.org/10.1016/j.aap.2006.08.002

Moore, J.E., Giuliano, G., Cho, S., 2004. Secondary accident rates on Los Angeles freeways. J. Transp. Eng. 130, 280–285. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(280)

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., T.Trego, 2010. Traffic incident management handbook. Washington, DC.

Russo, B.J., Savolainen, P.T., Schneider IV, W.H., Anastasopoulos, P.C., 2014. Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered probit model. Anal. Methods Accid. Res. 2, 21–29. https://doi.org/10.1016/j.amar.2014.03.001

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accid. Anal. Prev. 43, 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025

Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. J. Safety Res. 27, 183–194. https://doi.org/10.1016/0022-4375(96)00010-2

Sun, C.C., Chilukuri, V., 2010. Dynamic incident progression curve for classifying secondary traffic crashes. J. Transp. Eng. 136, 1153–1158. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000187

Ulfarsson, G.F., Mannering, F.L., 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accid. Anal. Prev. 36, 135–147. https://doi.org/10.1016/S0001-4575(02)00135-5

Wang, J., Liu, B., Zhang, L., David R., R., 2016a. Modeling secondary accidents identified by traffic shock waves. Accid. Anal. Prev. 87, 141–147. https://doi.org/10.1016/j.aap.2015.11.031

Wang, J., Xie, W., Liu, B., Fang, S., Ragland, D.R., 2016b. Identification of freeway secondary accidents with traffic shock wave detected by loop detectors. Saf. Sci. 87, 195–201. https://doi.org/10.1016/j.ssci.2016.04.015

Washington, S.P., Karlaftis, M.G., Mannering, F., 2011. Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.

Wu, Q., Chen, F., Zhang, G., Liu, X.C., Wang, H., Bogus, S.M., 2014. Mixed logit model-based driver injury severity investigations in single- and multi-vehicle crashes on rural two-lane highways. Accid. Anal. Prev. 72, 105–115. https://doi.org/10.1016/j.aap.2014.06.014

Xu, C., Liu, P., Yang, B., Wang, W., 2016. Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data. Transp. Res. Part C Emerg. Technol. 71, 406–418. https://doi.org/10.1016/j.trc.2016.08.015

Yang, H., Bartin, B., Ozbay, K., 2014a. Mining the characteristics of secondary crashes on highways. J. Transp. Eng. 140, 1–10. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000646

Yang, H., Ozbay, K., Xie, K., 2014b. Assessing the risk of secondary crashes on highways. J. Safety Res. 49, 143.e1-149. https://doi.org/10.1016/j.jsr.2014.03.007

Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. Anal. Methods Accid. Res. 1, 72–85. https://doi.org/10.1016/j.amar.2013.03.001

Zhan, C., Gan, A., Hadi, M., 2009. Identifying secondary crashes and their contributing factors. Transp. Res. Rec. 68–75. https://doi.org/10.3141/2102-09

Zhan, C., Shen, L., Hadi, M.A., Gan, A., 2008. Understanding the characteristics of secondary crashes on freeways, in: Presented at 87th Annual Meeting of the Transportation Research Board.

Zhang, H., Khattak, A., 2010. What is the role of multiple secondary incidents in traffic operations? J. Transp. Eng. 136, 986–997. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000164

Zhang, Z., Yang, R., Yuan, Y., Blackwelder, G., Yang, X., 2021. Examining driver injury severity in left-turn crashes using hierarchical ordered probit models. Traffic Inj. Prev. 0, 1–6. https://doi.org/10.1080/15389588.2020.1841899