Secondary Crashes Identification and Modeling along Highways in Utah

Zhao Zhang
Ph.D., Postdoctoral Researcher
Department of Civil and Environmental Engineering, University of Utah
100 Central Campus Dr., Suite 2000, Salt Lake City, UT, U.S., 84112
Email: zhao.zhang@utah.edu

Yaobang Gong
Ph.D., Postdoctoral Research Associate
Department of Civil and Environmental Engineering
University of Utah, Salt Lake City, UT, U.S., 84112
Email: yaobang.gong@utah.edu

Pan Lu, Ph.D.
Associate Professor
Upper Great Plains Transportation Institute
North Dakota State University, Fargo, ND, U.S., 58108
Email: pan.lu@ndsu.edu

Xianfeng (Terry) Yang
Ph.D., Assistant Professor
Department of Civil and Environmental Engineering, University of Maryland
1173 Glenn L. Martin Hall, College Park, MD, U.S., 20742
Email: xtyang@umd.edu

Word Count: 6000words + 6 Tables * 250 = 7500 words

Submitted to 102nd Transportation Research Board Annual Meeting for Presentation and Publication

Submission Date: August 1, 2022

ABSTRACT

The occurrence of secondary crashes on highways would bring many adverse effects, such as traffic congestion, air pollution, leading to more crashes. Accurate identification of secondary crashes is the basis for identifying contributing factors and contributing factors are the cornerstones for incident management system to find effective strategies to reduce the risk of secondary crash. However, secondary crash records are often not recorded correctly. To tackle this issue, this research aims to propose a hybrid method to accurately identify primary and secondary crashes. Based on the identified primary and secondary crashes, this study developed a binary logit model to find contributing factors of secondary crashes and construct a HOPIT model to analyze the crash injury patterns in primary and secondary crashes with identified data of primary and secondary crashes, respectively. This study provides a better understanding of contributing factors as well as crash injury patterns of secondary crashes.

Keywords: Crash identification, Secondary crashes, Crash injury severity, Binary logit model, Hierarchical ordered probit model

1. INTRODUCTION

Secondary crashes (SC) are typically defined as crashes that occur within the congested spatiotemporal boundaries of the region in which a primary crash occurred (*1*). It usually occurs within the spatial and temporal impact ranges of an existing primary crash (*2*). The occurrence of secondary crashes on highways would bring many adverse effects, such as traffic congestion, air pollution, leading to more crashes. Owens et al. (*1*) 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 (*3*, *4*), the identification of appropriate incident management strategies should be based on the understandings 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 freeway I-15 in the state of Utah from 2010 to 2020 are recorded as secondary crashes. Such data quality cannot meet the needs of developing effective incident management strategies.

In addition, the secondary crashes on freeways have not been well studied yet in Utah. 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 of secondary crashes, and examining the crash injury patterns of primary and secondary crashes.

Based on the discussion, this study aims to develop a hybrid method to effectively identify primary and secondary crashes from all crash records on freeways. Furthermore, the contributing factors of secondary crashes and crash injury patterns of primary and secondary crashes are analyzed by statistical models with identified primary and secondary crashes. This study could provide some basis for future research in the same field. The study results of this paper will provide some insightful findings to help transportation agencies build up a more effective incident management system to mitigate the secondary crashes on freeways.

The rest of the paper is organized as follows. Section 2 reviews existing studies related to secondary crash identification, contributing factors modeling, and crash injury severity modeling. The hybrid method for primary and secondary crashes identification, binary logit model, and HOPIT model are presented in Section 3. Section 4 conducts the results analysis. The last section summarizes the key findings and future research directions.

2. LITERATURE REVIEW

The static and dynamic methods are two popular approaches to identify secondary crashes. The static threshold methods assumes that the secondary crashes should happen within a spatial and temporal range of a primary crash. For example, Hirunyanitiwattana and Mattingly (*5*) 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 (*6*, *7*). 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 (*8*), incident progression curve (*3*), cumulative arrival and departure plots (*4*), and speed contour plot (*2*, *9*).

In the literature, the logit model has been widely implemented to identify the contributing factors of secondary crashes (*4*, *8*, *10–13*). 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 (*14*). 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 (*4*, *8*, *10–13*). 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, the secondary crashes are less likely to happen during the 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 are found to increase the risk of secondary crashes (Yang et al., 2014b).

For examining the crash injury severity, the discrete choice regression models are widely implemented to analyze crash injury severity. The discrete choice regression models can be further classified into (a) logit models, including nested models (*15*, *16*), multinomial logit models (*17–19*), Mixed logit models (*19–23*), and (b) probit models include ordered probit model with fixed and random parameters (*19*, *24–26*). 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 distinguished sets of independent variables for each crash injury severity result (*27*). 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. (*28*) and Savolainen et al. (*29*) pointed out that the ordered probit model cannot explain how the thresholds that are estimable parameters profoundly affect intermediate categories and the effect of the independent variables on the highest and lowest ordered discrete category probabilities, with the impact on the interior category probabilities. 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 (*27*).

3. METHODOLOGY

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 1 illustrates the flowchart of the hybrid method.

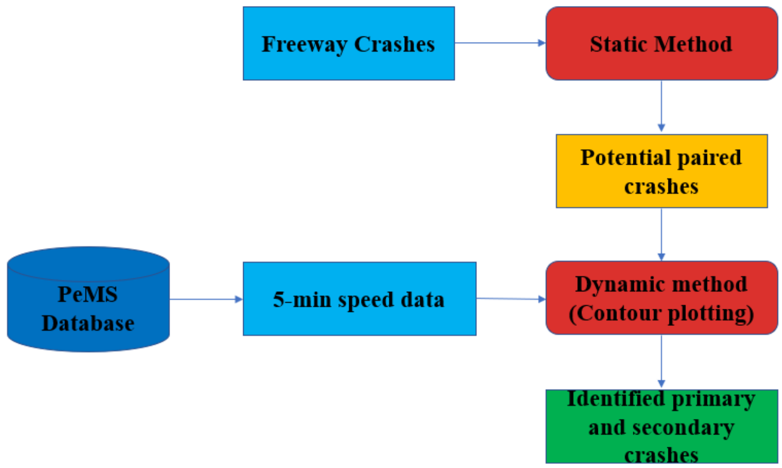

Figure 1 The flowchart of hybrid method for secondary crashes identification

Static Method

Firstly, 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.

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 congestions. 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 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.
- Constructing 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 3 presents an example of a speed contour plot for a prior crash. It can be clearly seen that congestions and queue formations.
- Subtracted the average speed over crash-free days to build a new contour plot, the effects of recurrent congestions can be eliminated. Figure 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 2 Illustration of downloading data for dynamic approach

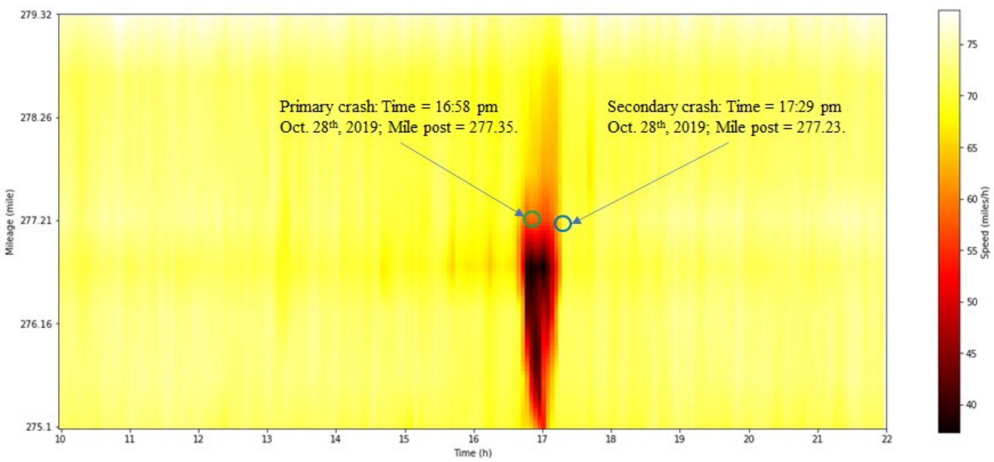

Figure 3 Speed contour plot of labeled secondary crash

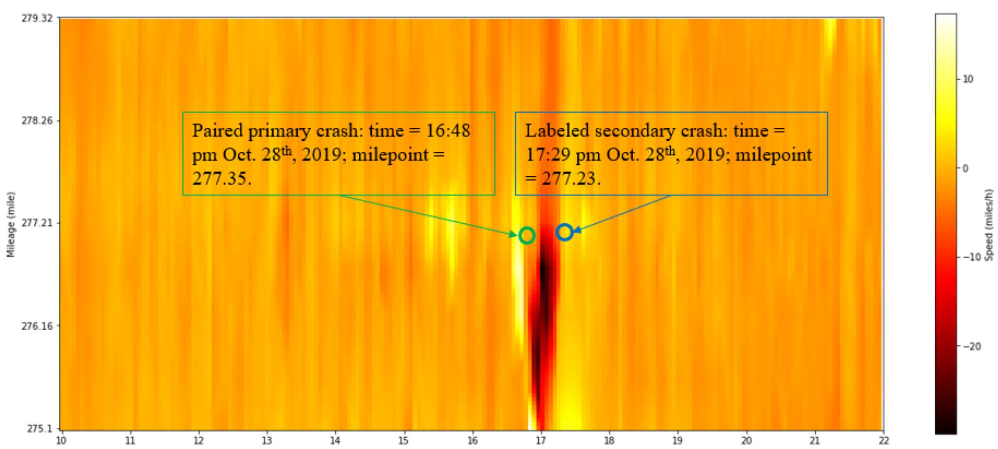

Figure 4 Subtracted Average Speed contour plot of labeled secondary crash

Binary Logit Model

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

$$P(Occurrence\ of\ secondary\ crash|crash) = P\ (primary\ crash|Crash) \quad (1)$$

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 indicates 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.

$$P_i = \frac{e^{\beta}}{1+e^{\beta}}, \beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad (2)$$

The logistic regression equation is approximately linear in the middle ranges and logarithmic at extreme values (*28*). A simple transformation of logistic regression is shown below:

$$(\frac{P_i}{1-P_i}) = e^{(\widehat{\beta_0}+\widehat{\beta_i}x_i)} = e^{\widehat{\beta_0}} e^{\widehat{\beta_i}x_i} \quad (3)$$

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:

$$(\frac{P_i}{1-P_i})^* = e^{\widehat{\beta_0}} e^{\widehat{\beta_i}(x_i+1)} = e^{\widehat{\beta_0}} e^{\widehat{\beta_i}x_i} e^{\widehat{\beta_i}} = (\frac{P_i}{1-P_i})\ e^{\widehat{\beta_i}} \quad (4)$$

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 $e^{\widehat{\beta_i}}$. The factor $e^{\widehat{\beta_i}}$ is the odds ratio and indicates the relative amount by which the odds of an outcome increases (odds ratio >1) or decreases (odds ratio <1) when the value of the corresponding independent variable increases by 1 unit.

Driver Injury Severity Modeling

In this study, hierarchical ordered probit (HOPIT) models are developed to identify significant casual factors and quantify their impacts on driver injury severities in primary and secondary crashes. To investigate the crash injury probabilities and 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 ordinal ranking of data. The unobserved variable *z* can be denoted as follows (*28*):

$$z_i = \beta \chi_i + \varepsilon_i \quad (5)$$

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 (*28*):

$$y_i = j \text{ if } \mu_{j-1} < z_i < \mu_j, j = 1, \ldots, J \quad (6)$$

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 (*28*) since the restrictions placed on how variables are believed to affect ordered discrete outcome probabilities,. 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 (*30*):

$$\mu_{i,j} = \mu_{i,j-1} + \exp(t_j + \boldsymbol{d}_j \boldsymbol{S}_i) \tag{7}$$

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 (*28*). 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 (*28*):

$$P(y = j) = \Phi(\mu_j - \beta\chi_i) - \Phi(\mu_{j+1} - \beta\chi_i) \tag{8}$$

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) (*28*). To address this problem, it can be calculated by marginal effects (*27*, *28*, *31*, *32*) using the following equation:

$$\frac{P(y = j)}{\partial \chi} = [\Phi(\mu_{j-1} - \beta\chi) - \Phi(\mu_j - \beta\chi)]\beta \tag{9}$$

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. **RESULTS ANALYSIS**

Primary and Secondary Crashes Identification

The experimental study was conducted on freeway I-15 in the state of Utah. Three-year (2017 to 2019) crash and traffic data were retrieved from Numetric database and Performance Measurement System (PeMS) managed by the Utah Department of Transportation (UDOT) respectively. 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.

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, we obtained 2,653 (97.95%) primary crashes, 2,953 (88.4%) secondary crashes, and 18,878 normal crashes were identified in the database. Table 1 presents the distribution of identified primary and secondary crashes from 2017 to 2019 and the percentage in the total crashes.

Table 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

Modeling Contributing Factors of Secondary Crash

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.) are 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.

As shown in Table 2, 12 variables (including young people, daylight, snow 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 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 occurred on weekends and in rural areas are more likely to lead to a secondary crash. The odds ratio represents the increase in the likelihood that a crash will lead to a secondary crash. For example, for a rear-end crash, there is almost an 89.3% increase in the likelihood that a crash leads to a secondary crash.

The correlation test is conducted to determine the correlations between variables. The autocorrelations of variables are presented in Figure 5, which indicate that there are no strong correlations between all candidate variables. The ROC curve is used to evaluate the predictive performance of different models (*33*). 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 6, the area under the ROC curve is 0.796, which indicates that the binary logit model has a good predictive performance.

Table 2 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 fix 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 fix 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			

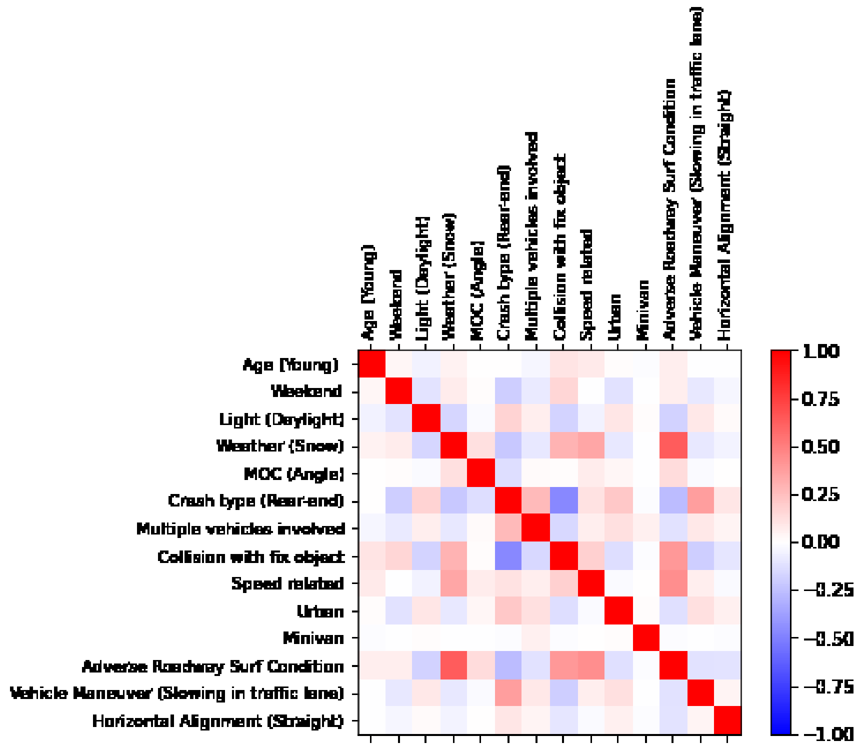

Figure 1 Variable correlation results in binary logit model

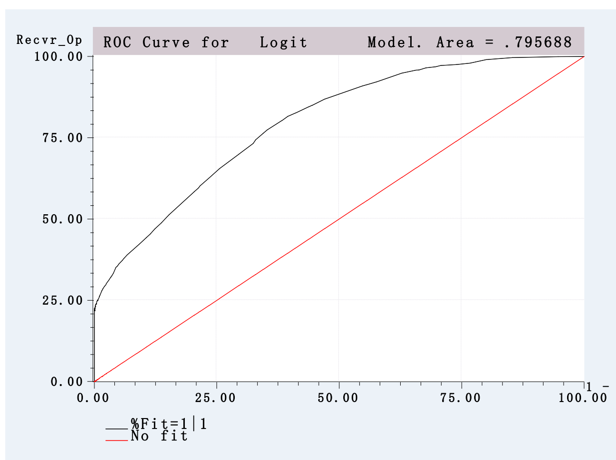

Figure 6 ROC curve for the binary logit model

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 (*34–36*). 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 7-8. The figures show that there is no strong relationship between all candidate variables.

Table 3 and Table 5 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 4 and Table 6, 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 are presented following.

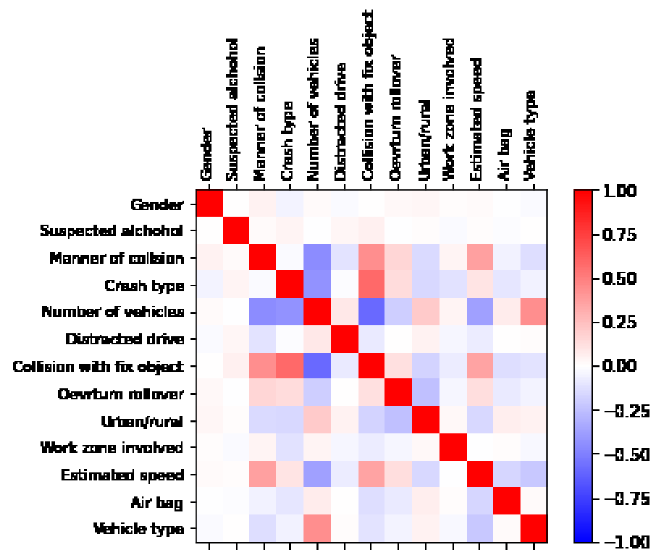

Figure 7 Variable correlation results in HOPIT model for primary crash

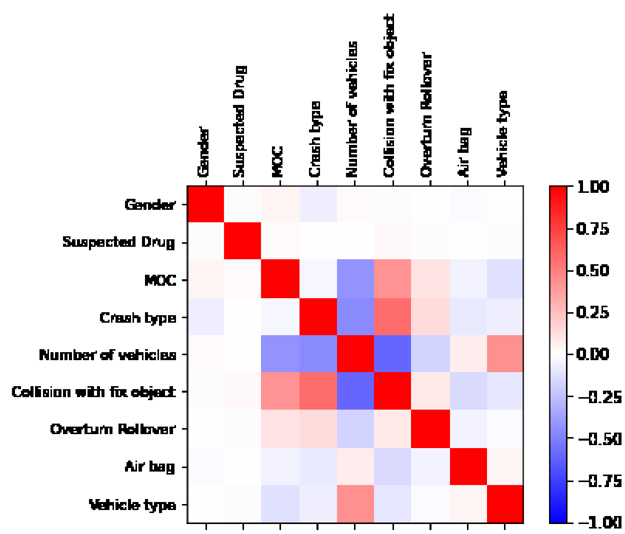

Figure 8 Variable correlation results in HOPIT model for secondary crash

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 coefficients 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 coefficients of 1.18. It is more likely to be involved in minor and severe injury primary crashes by 33.7% and 10.7%, respectively. It is perceptive that crashes with minor injuries and severe injuries are less likely to occur in rural areas (with a coefficient of -0.23). The 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 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 negative parameters -0.24. It may reduce the possibility of minor injury and severe injury in primary crashes.

According to the results of the HOPIT model developed for secondary crashes, female drivers are more likely to be involved in minor-injured 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 coefficients 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 severe 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 negative parameters 0.088. It may reduce the possibility of minor injury and severe injury in secondary crashes.

Table 3 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 Drive	0.23	0.11	2.02	0.04**
Collison with fix 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 4 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 Drive	-0.082	0.074	0.008
Collison with fix 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

Table 5 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 fix 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 6 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 fix 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. CONCLUSION AND FUTURE RESEARCH DIRENCTIONS

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 paper provided a preliminary analysis of traffic crash records and labeled secondary crash records in the UDOT's crash database. The results show that the accuracy of labeled secondary crash records is low. To tackle this issue, this paper 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. The experimental study results indicate that the proposed hybrid method can effectively identify the primary and secondary crashes from the database. The binary logit model finds the contributing factors of secondary crashes and the crash injury severity patterns are identified by HOPIT models with the identified data of primary and secondary crashes. Those findings could provide some insightful information to transportation agencies to find effective countermeasures to reduce the secondary crashes and reduce the injury severity of primary and secondary crashes on freeways.

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, road geometry information. This might illustrate future studies to overcome this challenge.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: X. Yang and Z. Zhang; data collection: Z. Zhang; analysis and interpretation of results: Z. Zhang, Y. Gong, and X. Yang; draft manuscript preparation: Z. Zhang, Y. Gong, and X. Yang. All authors reviewed the results and approved the final version of the manuscript.

REFERENCE

1. Owens, N., A. Armstrong, P. Sullivan, C. Mitchell, D. Newton, R. Brewster, and T.Trego. *Traffic Incident Management Handbook*. Washington, DC, 2010.

2. Xu, C., P. Liu, B. Yang, and W. Wang. Real-Time Estimation of Secondary Crash Likelihood on Freeways Using High-Resolution Loop Detector Data. *Transportation Research Part C: Emerging Technologies*, Vol. 71, 2016, pp. 406–418. https://doi.org/10.1016/j.trc.2016.08.015.

3. Sun, C. C., and V. Chilukuri. Dynamic Incident Progression Curve for Classifying Secondary Traffic Crashes. *Journal of Transportation Engineering*, Vol. 136, No. 12, 2010, pp. 1153–1158. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000187.

4. Zhan, C., A. Gan, and M. Hadi. Identifying Secondary Crashes and Their Contributing Factors. *Transportation Research Record*, No. 2102, 2009, pp. 68–75. https://doi.org/10.3141/2102-09.

5. Hirunyanitiwattana, Worapong, and S. P. Mattingly. Identifying Secondary Crash Characteristics for California Highway System. 2006.

6. Moore, J. E., G. Giuliano, and S. Cho. Secondary Accident Rates on Los Angeles Freeways. *Journal of Transportation Engineering*, Vol. 130, No. 3, 2004, pp. 280–285. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(280).

7. Zhan, C., L. Shen, M. A. Hadi, and A. Gan. Understanding the Characteristics of Secondary Crashes on Freeways. 2008.

8. Zhang, H., and A. Khattak. What Is the Role of Multiple Secondary Incidents in Traffic Operations? *Journal of Transportation Engineering*, Vol. 136, No. 11, 2010, pp. 986–997. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000164.

9. Yang, H., B. Bartin, and K. Ozbay. Mining the Characteristics of Secondary Crashes on Highways. *Journal of Transportation Engineering*, Vol. 140, No. 4, 2014, pp. 1–10. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000646.

10. Yang, H., K. Ozbay, and K. Xie. Assessing the Risk of Secondary Crashes on Highways. *Journal of Safety Research*, Vol. 49, No. February, 2014, pp. 143.e1-149. https://doi.org/10.1016/j.jsr.2014.03.007.

11. Mishra, S., M. M. Golias, A. A. Sarker, and A. Naimi. *Effect of Primary and Secondary Crashes: Identification, Visualiztion and Prediction*. 2017.

12. Wang, J., B. Liu, L. Zhang, and R. David R. Modeling Secondary Accidents Identified by Traffic Shock Waves. *Accident Analysis and Prevention*, Vol. 87, 2016, pp. 141–147. https://doi.org/10.1016/j.aap.2015.11.031.

13. Wang, J., W. Xie, B. Liu, S. Fang, and D. R. Ragland. Identification of Freeway Secondary Accidents with Traffic Shock Wave Detected by Loop Detectors. *Safety Science*, Vol. 87, 2016, pp. 195–201. https://doi.org/10.1016/j.ssci.2016.04.015.

14. Mitra, S., and S. Washington. On the Nature of Over-Dispersion in Motor Vehicle Crash Prediction Models. *Accident Analysis and Prevention*, Vol. 39, No. 3, 2007, pp. 459–468. https://doi.org/10.1016/j.aap.2006.08.002.

15. Lee, J., and F. Mannering. Impact of Roadside Features on the Frequency and Severity of Run-off-Roadway Accidents: An Empirical Analysis. *Accident Analysis and Prevention*, Vol. 34, No. 2, 2002, pp. 149–161. https://doi.org/10.1016/S0001-4575(01)00009-4.

16. Chang, L.-Y., and F. L. Mannering. Predicting Vehicle Occupancies from Accident Data: An Accident Severity Approach. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1635, No. 1, 2007, pp. 93–104. https://doi.org/10.3141/1635-13.

17. Shankar, V., and F. Mannering. An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity. *Journal of Safety Research*, Vol. 27, No. 3, 1996, pp. 183–194. https://doi.org/10.1016/0022-4375(96)00010-2.

18. Ulfarsson, G. F., and F. L. Mannering. Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents. *Accident Analysis and Prevention*, Vol. 36, No. 2, 2004, pp. 135–147. https://doi.org/10.1016/S0001-4575(02)00135-5.

19. Ye, F., and D. Lord. Comparing Three Commonly Used Crash Severity Models on Sample Size Requirements: Multinomial Logit, Ordered Probit and Mixed Logit Models. *Analytic Methods in Accident Research*, Vol. 1, 2014, pp. 72–85. https://doi.org/10.1016/j.amar.2013.03.001.

20. Li, Z., Q. Wu, Y. Ci, C. Chen, X. Chen, and G. Zhang. Using Latent Class Analysis and Mixed Logit Model to Explore Risk Factors on Driver Injury Severity in Single-Vehicle Crashes. *Accident Analysis and Prevention*, Vol. 129, No. March 2018, 2019, pp. 230–240. https://doi.org/10.1016/j.aap.2019.04.001.

21. Wu, Q., F. Chen, G. Zhang, X. C. Liu, H. Wang, and S. M. Bogus. Mixed Logit Model-Based Driver Injury Severity Investigations in Single- and Multi-Vehicle Crashes on Rural Two-Lane Highways. *Accident Analysis and Prevention*, Vol. 72, 2014, pp. 105–115. https://doi.org/10.1016/j.aap.2014.06.014.

22. Anastasopoulos, P. C., and F. L. Mannering. An Empirical Assessment of Fixed and Random Parameter Logit Models Using Crash- and Non-Crash-Specific Injury Data. *Accident Analysis and Prevention*, Vol. 43, No. 3, 2011, pp. 1140–1147. https://doi.org/10.1016/j.aap.2010.12.024.

23. Cerwick, D. M., K. Gkritza, M. S. Shaheed, and Z. Hans. A Comparison of the Mixed Logit and Latent Class Methods for Crash Severity Analysis. *Analytic Methods in Accident Research*, Vol. 3–4, 2014, pp. 11–27. https://doi.org/10.1016/j.amar.2014.09.002.

24. MCCARTHY, P. S., and S. MADANAT. Highway Accident Data Analysis: Alternative Econometric Methods. *Transportation Research Record*, No. 1467, 1994, pp. 44–49.

25. Fountas, G., and P. C. Anastasopoulos. Analysis of Accident Injury-Severity Outcomes: The Zero-Inflated Hierarchical Ordered Probit Model with Correlated Disturbances. *Analytic Methods in Accident Research*, Vol. 20, No. October, 2018, pp. 30–45. https://doi.org/10.1016/j.amar.2018.09.002.

26. Chen, C. C., C. C. Chen, and T. T. Hoang. Role of Concrete Confinement of Wide-Flange Structural Steel Shape in Steel Reinforced Concrete Columns under Cyclic Loading. *Engineering Structures*, Vol. 110, 2016, pp. 79–87. https://doi.org/10.1016/j.engstruct.2015.12.002.

27. Fountas, G., and P. C. Anastasopoulos. A Random Thresholds Random Parameters Hierarchical Ordered Probit Analysis of Highway Accident Injury-Severities. *Analytic Methods in Accident Research*, Vol. 15, 2017, pp. 1–16. https://doi.org/10.1016/j.amar.2017.03.002.

28. Washington, S. P., M. G. Karlaftis, and F. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall/CRC, 2011.

29. Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*, Vol. 43, No. 5, 2011, pp. 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025.

30. Greene, W. H., and D. A. Hensher. *Modeling Ordered Choices: A Primer*. Cambridge University Press, 2010.

31. Anastasopoulos, P. C., M. G. Karlaftis, J. E. Haddock, and F. L. Mannering. Household Automobile and Motorcycle Ownership Analyzed with Random Parameters Bivariate Ordered Probit Model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2279, No. 1, 2012, pp. 12–20. https://doi.org/10.3141/2279-02.

32. Russo, B. J., P. T. Savolainen, W. H. Schneider IV, and P. C. Anastasopoulos. Comparison of Factors Affecting Injury Severity in Angle Collisions by Fault Status Using a Random Parameters Bivariate Ordered Probit Model. *Analytic Methods in Accident Research*, Vol. 2, 2014, pp. 21–29. https://doi.org/10.1016/j.amar.2014.03.001.

33. Egan, J. P. Signal Detection Theory and ROC-Analysis. *Academic press*, 1975.

34. Hou, Q., X. Huo, J. Leng, and Y. Cheng. Examination of Driver Injury Severity in Freeway Single-Vehicle Crashes Using a Mixed Logit Model with Heterogeneity-in-Means. *Physica A: Statistical Mechanics and its Applications*, Vol. 531, 2019, p. 121760. https://doi.org/10.1016/j.physa.2019.121760.

35. Behnood, A., A. M. Roshandeh, and F. L. Mannering. Latent Class Analysis of the Effects of Age, Gender, and Alcohol Consumption on Driver-Injury Severities. *Analytic Methods in Accident Research*, Vol. 3–4, 2014, pp. 56–91. https://doi.org/10.1016/j.amar.2014.10.001.

36. Zhang, Z., R. Yang, Y. Yuan, G. Blackwelder, and X. Yang. Examining Driver Injury

Severity in Left-Turn Crashes Using Hierarchical Ordered Probit Models. *Traffic Injury Prevention*, Vol. 0, No. 0, 2021, pp. 1–6. https://doi.org/10.1080/15389588.2020.1841899.