Accident Analysis and Prevention 151 (2021) 105953

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

A Motorcyclist-Injury Severity Analysis: A Comparison of Single-, Two-, and Multi-Vehicle Crashes Using Latent Class Ordered Probit Model

Jing Li, Shouen Fang, Jingqiu Guo *, Ting Fu, Min Qiu

The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, China

A R T I C L E I N F O

Keywords:
Motorcycle crash
Injury severity
Latent Class Clustering

A B S T R A C T

Motorcycle crashes increasingly become a high proportion of the overall motorized vehicle fatalities. However, limited research has been conducted to compare the injury severity of single-, two- and multi-vehicle crashes involving a motorcycle. This study aims to investigate the effects of rider characteristics, road conditions, pre-crash situations, and crash features on motorcycle severities with respect to different numbers of vehicles involved. The crash data used was obtained through a comprehensive Motorcycle Crash Causation Study (MCCS) by the Federal Highway Administration. An anatomic injury severity indicator, the New Injury Severity Score (NISS), is utilized to calculate a total score as the sum of squared the abbreviated injury scale scores of each of the rider's three most severe injuries. A hybrid approach integrating Latent Class Clustering (LCC) and Ordered Probit (OP) models was used to uncover the unobserved heterogeneity and to explore the major factors which significantly affect the injury severities resulting from single-, two- and multi-vehicle crashes involving a motorcycle. The results show that the significant differences in severity exist between different numbers of vehicles involved. More importantly, they also indicate dividing motorcycle crashes into homogeneous classes before modelling helps to discover insightful information. Pre-speed of the motorcycle is found to be a main factor associated with serious and critical injuries in most types of crashes. Findings of the study provide specific and insightful countermeasures targeting at the contributing factors of motorcycle crashes.

1. Introduction

Motorcycles are more vulnerable and less visible than passenger motor vehicles. The number of injuries and fatalities in motorcycle crashes has severe consequences both socially and economically. More than 4800 motorcyclists were killed in crashes involving motorcycles each year during the past decades in the United States (NHTSA 2019). Moreover, the number of motorcycle fatalities increased substantially in recent years, i.e., from 4469 fatalities in 2009 to 5138 fatalities in 2018 – an increase of 15% in the USA. Many studies have accordingly been conducted to examine those major contributing factors of motorcycle crash severity (Quddus et al. 2002, Alnawmasi and Mannering 2019, Ding et al. 2019). Collectively, a wide variety of variables such as weather conditions, motorcyclist-related issues, road conditions are found to be associated with injury severity of motorcyclists. For example, (Cheng et al. 2017) used a full Bayesian hierarchical approach to examine the relationship between weather and severity outcomes based on the 5-point (KABCO) injury classification scale for motorcycle crashes. The result shows that temperature decreases the possibility of fatal injury in motorcycle crashes; Vajari et al. (2020) used a multinomial logit model to investigate the impact of intersection features on injury severity in motorcycle crashes, which indicates that specific treatment measures of T-intersections decrease the possibility of fatal injury.

However, majority of the motorcycle crash studies does not distinguish the influence of different sub-groups with respect to the number of vehicles involved on the crash risk and severity. A few exceptions have noticed and analyzed motorcycle crashes that were classified into one or two types of sub-groups (Haque et al. 2012, Adinegoro et al. 2015). For example, based on police-reported crash data, Savolainen and Mannering (2007) used probabilistic models to investigate the factors that significantly influence the motorcyclist's injury severity in single-vehicle and multi-vehicle crashes, respectively, and showed that the motorcyclist age significantly impacts on the injury severity in crashes. Relying on police-reported data, Haque et al. (2012) also examined the impacts of various factors on risks of multi-vehicle motorcycle crashes by log-linear models, with key findings showing that riding at night time might decrease the motorcycle's level of safety.

* Corresponding author.
E-mail address: guojingqiu@hotmail.com (J. Guo).

https://doi.org/10.1016/j.aap.2020.105953
Received 26 May 2020; Received in revised form 3 October 2020; Accepted 6 December 2020
Available online 29 December 2020

The aforementioned studies did exhibit insightful results. However, due to lack of detailed motorcycle crash data, police-reported crash datasets were often used, which leads to potential problems. Firstly, the detailed information supposed to be correlated with motorcycle injury outcomes may not be available in the typical police report, which may lead to bias in the analysis. For example, the motorcycle speed before a crash, the multiple injuries on different parts of a motorcyclist's body, and the pre-crash operational conditions of motorcycle, all of which may be key risk factors of single-, two-, multi-vehicle motorcycle crashes.

The second problem associated with the police-reported crash data is related to the injury classification scales. Three injury classification scales are widely used in previous studies: 5-point (KABCO) scale (Geedipally and Lord 2010, Kitali et al. 2020), the Abbreviated Injury Scale (Tarko et al. 2010), and the Injury Severity Score (Wali et al. 2019). In the KABCO scale, which is the most popular method, there are typically five categories: type K (killed or fatal), type A (Incapacitating injury), type B (Non-incapacitating injury), type C (possible injury) and type O (no injury). However, it is hard to find out information about multiple injuries on different parts of a motorcyclist's body, while most motorcycle crashes cause injuries on multiple parts of a body. Meanwhile, compared with motor vehicle crashes, injuries caused by motorcycle crashes tend to be under-reported by the police (Amoros et al. 2006, Loo and Tsui 2007). The Abbreviated Injury Scale (AIS) assigns a severity score to each injury by a scale from 1 to 6. However, the AIS scale also has limitations of lacking sensitivity towards the possible multiple injuries to the motorcyclist. The Injury Severity Score (ISS) provides an overall score for the AIS scores of each of the rider's three most severely injured body parts (Baker et al. 1974), and only considers one injury of each body part. Multi-serious injuries may occur in the same body part, but ISS only considers one injury of them. Therefore, less severe ones occurring on other parts of a body are included in the ISS instead of more serious injuries of the same body part, which may lead to injuries being overlooked or underestimated. The New Injury Severity Score (NISS) is designed to counter the above problems by producing a total score as a sum of squared the AIS scores of the three most severe injuries regardless of which parts of a body the injuries occur on (Osler et al. 1997). The NISS displayed better discrimination power and goodness-of-fit than the ISS in previous studies (Lavoie et al. 2004, Eid and Abu-Zidan 2015).

Moreover, another challenging issue in this field is unobserved heterogeneity due to some factors which are not observed by the researchers (Mannering et al. 2016), which may cause biased estimations of parameters and further draw potentially incorrect conclusions (Li and Fan 2019). So far, few exiting studies have addressed heterogeneity when analyzing the injury severity of the motorcycle crash. Two modelling methods have been commonly applied in such analysis: random parameter models (Malyshkina and Mannering 2009, Flask et al. 2014), and latent class models (Liu and Fan 2020). Assumptions of parametric distribution for random parameter modelling are required, which may be a limitation for capturing unobserved heterogeneity (Alnawmasi and Mannering 2019). In contrast, latent class approaches need not a distributional assumption and deal with heterogeneity by finding the unobserved sub-groups of the overall population.

The main objective of this study is to explore various factors that contribute to the injury severities of motorcyclists involving in the single-, two-, multi-vehicle motorcycle crashes. To the best of our knowledge, it is probably the first attempt in the context of severity analysis for the motorcycle crash. In order to counter the shortcomings of injury classification scales commonly used before, this study uses a unique method by calculating an anatomic injury severity indicator, the NISS for injury scaling. The Motorcycle Crash Causation Study (MCCS), a sophisticated motorcycle crash data maintained by FHWA, has been selected. The whole dataset is first divided into subsets of single-, two-, and multi-vehicle motorcycle crashes based on the number of vehicles involved. Then, a hybrid model with latent class cluster (LCC) and Ordered Probit (OP) is developed to address the heterogeneity problem and to identify the significant factors for each sub-group.

2. Data Collection and Description

The dataset employed specifically in the study is the motorcycle crash data of 2011-2016 in Orange County, California, which were collected in the MCCS of the Federal Highway Administration. MCCS has been regarded as the most comprehensive motorcycle crash data in America in more than 30 years (FHWA, 2017). Variables include motorcycle rider factors, road conditions, traffic flow, and geometric characteristics, motorcycle pre-crash situations (i.e., speed before a crash), and crash-related issues (i.e., the number of involved vehicles and injury severities), all of which are extensively detailed. In order to collect precise injury data, investigators in the MCCS also collected reliable information through interviews with motorcyclists for details of all injuries, patient release forms from private physicians, and autopsy reports. Therefore, compared to the police-reported dataset with the aforementioned shortcomings, the crash data collected from the MCCS are more precise and elaborate.

A total of 322 injury cases are available in the MCCS. As shown in Fig. 1, the crash data were classified as single-vehicle (n = 75, 23.3%), two-vehicle (n = 222, 68.9%) and multi-vehicle (n = 25, 7.8%) motorcycle crash. Among the motorcyclists, males were involved in a much higher percent (95.0%) of the crashes than females (5.0%). Meanwhile, 54 motorcyclists (16.8%) have permanent physical impairments or chronic diseases (i.e., hearing reduction and cardiovascular condition). In terms of weather conditions, most crashes (235, 73%) occurred when temperature records are in the range of 58 to 78 degrees Fahrenheit. As far road geometry is concerned, 32.9% crashes occurred at road segments related to horizontal alignment design; 43.8% crashes occurred at road segments related to vertical alignment condition.

3. Methodology

3.1. Injury Classification Scale-NISS

The MCCS dataset has detailed information on body parts of a rider in each record, which enables us to apply the NISS to assess injury severity. A two-stage method is adopted to calculate the NISS for each injury motorcyclist. Each injury is classified based on body parts, including head, face, neck, thorax, abdomen, spine, upper extremity, lower extremity, and other body parts, while the severity of each injury is coded through an AIS score in the MCCS. Detailed process is that the AIS assigns a score to each injury by a scale from 1 to 6 on the basis of severity, that is 1 - minor injury, 2 - moderate injury, 3 - serious injury, 4 - severe injury, 5 - critical injury, and 6 - maximum. A maximum AIS score of 6 indicates a currently untreatable injury, such as a most severe injury to the organ. Past study shows survival rate clearly decreased with the AIS severity. When an AIS score equals 5, the mean survival rate is

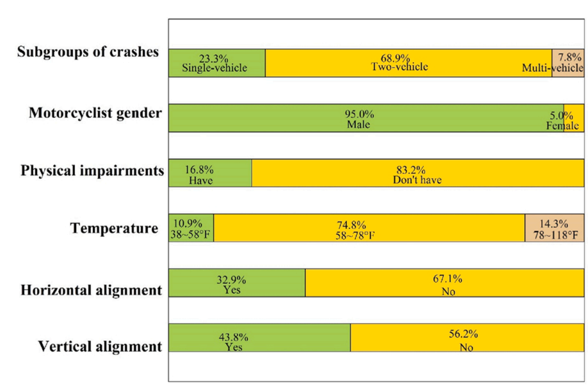


Fig. 1. Description of motorcycle crashes data.

60.4%, which sharply drops to 21.0% with an AIS score of 6 (Gennarelli and Wodzin 2006).

Having established the AIS scores, the NISS value is calculated as the sum of squared the three highest AIS scores regardless of body parts. Mathematically, $NISS = a^2 + b^2 + c^2$ where *a, b, c* is the AIS scores of the three most serious injuries. For instance, as shown in Table 1, the AIS scores of the three most serious injuries of the motorcyclist are 4, 3 and 1. Accordingly $\mathrm{NISS} = 4^2 + 3^2 + 1^2 = 26$. The NISS values range from 1 to 75. If any of the three AIS values is 6, which means maximum injury, the NISS automatically comes to 75. Because an AIS score of 6 is a currently untreatable injury, which means further medical care for saving human lives probably useless and further medical care may be discontinued for a motorcyclist. Thus, the maximum value of the $\mathrm{NISS} = 5^2 + 5^2 + 5^2 = 75$. Injury severities in this study are classified based on the NISS score, including minor injury (NISS = 1-3), moderate injury (NISS = 4-8), serious injury (NISS = 9-15), severe injury (NISS = 16-25), and critical injury (NISS = 25-75). For example, the NISS score of 26 is classified as critical injury since $26 \in [25, 75]$. A more detailed classification protocol can be found in (Stevenson et al. 2001).

3.2. Latent Class Clustering

The Latent Class Clustering (LCC) is a statistical method that assumes the whole database can be divided into subgroups by a latent categorical variable (Lanza and Rhoades 2013). The LCC is based on heuristics to maximize the homogeneity of samples within classes and the heterogeneity between classes (Hair et al. 1998). Compared to other types of clustering methods, the LCC does have some notable advantages (De Oña et al. 2013, Liu and Fan 2020). Firstly, the LCC does not require to specify a priori number of classes. The most suitable number of classes can be determined by different statistical criteria; Secondly, the LCC can use various types of variables (frequencies, categorical, continuous, nominal variables, and a combination of the above types) without prior standardization, which would cause a bearing on the results. Thirdly, probability classifications can be made by subsequent membership probability that is estimated by the maximum likelihood method.

Assume there are N latent classes and M characteristics of motorcycle crash i. The parameter γ_c is the membership probability of latent class c. The vector $\boldsymbol{Y_i} = Y_{i1}, \ldots, Y_{iM}$ denotes response of motorcycle crash i to M characteristics. The possible values of Y_{im} are $1, \ldots, \gamma_m$. n_i denotes the latent class membership of crash i, $n_i = 1, 2, \ldots, N$. $I\ (y_m = h)$ is the indicator function that equals 1 when the attribute of m equals to h and 0, otherwise. Parameters ρ denote the probability of crash i' response conditional on membership of latent class. Therefore, the probability of crash i' s response is (Lanza et al. 2015):

$$P(Y_i = y) = \sum_{n=1}^{N} \gamma_n \prod_{m=1}^{M} \prod_{h=1}^{\gamma_m} \rho_{m,h|n}^{I(y_m=h)} \tag{1}$$

For the most suitable number of classes and the goodness-of-fit, the information criteria, including the Bayesian Information Criteria (BIC), the Akaike Information Criterion (AIC) and the Consistent Akaike Information Criterion (CAIC) and entropy are commonly used. Previous research show that the BIC is more reliable than the AIC and the CAIC in finding the best number of classes (Nylund et al. 2007). Therefore, this study uses the BIC to identify the most suitable number of classes. The number of classes that minimizes the BIC indicates the best results.

3.3. Ordered Probit Model

The Ordered Probit (OP) model has been widely used in developing a mechanism for an ordinal response (Ivan and Konduri 2018). The previous study shows that the OP model requires the smallest sample size compared to the multinomial logit model and the mixed logit model (Ye and Lord 2014), which suits well the size of the study's dataset. In this study, the OP model is used to identify the model structure for injury severity of motorcyclists, and generally uses a latent (unobserved) variable y_i^* to measure injury severity outcomes, as shown in Eq. (2).

$$y_i^* = \boldsymbol{\beta X_i} + \varepsilon_i,\ i = 1, 2, \cdots, n \tag{2}$$

where y_i^* is a latent variable determining the injury severity of crash i; $\boldsymbol{X_i}$ is a vector of observed explanatory variables for crash i; $\boldsymbol{\beta}$ is a vector of the coefficients; and ε_i is the random error term which follows a standard normal distribution.

In Eq. (2), the observed dependent variable $\boldsymbol{y_i}$ is determined by Eq. (3).

$$Y_i = \begin{cases} 1,\ if\ y_i^* \le \alpha_1\ (minor\ injury) \\ 2,\ if\ \alpha_1 < y_i^* \le \alpha_2\ (moderate\ injury) \\ 3,\ if\ \alpha_2 < y_i^* \le \alpha_3\ (\text{serious injury}) \\ 4,\ if\ \alpha_3 < y_i^* \le \alpha_4\ (\text{severe injury}) \\ 5,\ if\ \ y_i^* > \alpha_4 (critical\ injury) \end{cases} \tag{3}$$

where $\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4,\ \alpha_5$ are the threshold values for categories of injury severities and are unknown parameters to be estimated. The parameters of the OP model are estimated by the method of maximum likelihood.

The probability of injury severity for crash i belongs to each category is shown as Eq. (4) after given the value of $\boldsymbol{X_i}$.

$$\begin{cases} Prob(y_i = 1) = \Phi(\alpha_1 - \boldsymbol{\beta X_i}) \\ Prob(y_i = 2) = \Phi(\alpha_2 - \boldsymbol{\beta X_i}) - \Phi(\alpha_1 - \boldsymbol{\beta X_i}) \\ Prob(y_i = 3) = \Phi(\alpha_3 - \boldsymbol{\beta X_i}) - \Phi(\alpha_2 - \boldsymbol{\beta X_i}) \\ Prob(y_i = 4) = \Phi(\alpha_4 - \boldsymbol{\beta X_i}) - \Phi(\alpha_3 - \boldsymbol{\beta X_i}) \\ Prob(y_i = 5) = 1 - \Phi(\alpha_5 - \boldsymbol{\beta X_i}) \end{cases} \tag{4}$$

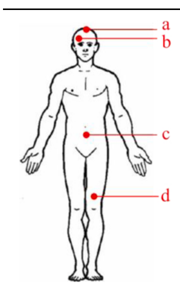
where $\Phi(\cdot)\Phi(\cdot)$ represents the probability function which follows standard normal distribution.

4. Results and Discussion

4.1. Data descriptive statistics

Variables employed in this study and their descriptive statistics are presented in Table 2a. Based on the scope of this research, 322 observations were selected from the total 351 motorcycle injury crashes in MCCS since the injury data of the other 29 cases have missing values. Since more than 1000 variables are available in MCCS, given the focus of this study, we have carefully selected vital variables. Rider characteristics, road conditions, pre-crash, and crash attributes are considered in our study. Specifically, rider characteristics include physical impairments and retroreflective parts on motorcycle. Environmental features included are lighting condition, temperature, horizontal and vertical

Table 1
Example of the ISS score and the NISS score.

Illustration	Injury	AIS score	Body part
	a	4	Head
	b	3	Head
	c	1	Abdomen
	d	1	Lower extremity
$NISS\ score = a^2 + b^2 + c^2 = 4^2 + 3^2 + 1^2 = 26$			

Note: (1) The AIS score includes 1 - minor injury, 2 - moderate injury, 3 - serious injury, 4 - severe injury, 5 - critical injury, 6 - maximum. (2) The illustration is referenced from the final report of the MCCS (FHWA, 2017).

Table 2a
Study variables and descriptive statistics of the data used in this study.

Variable	Description	Scale	Mean	SD
Rider characteristics				
PImpair	Permanent Physical Impairments or not	1- yes 0- no	0.366	0.482
MRetro	Motorcycle have retroreflective parts or not	1- yes 0- no	0.720	0.449
Road conditions				
LightCon	Was lighting condition good or not	1- yes 0- no	0.276	0.447
TDensity	Traffic Density	0-no other traffic 1-light traffic 2-moderate traffic 3-heavy traffic, traffic moving 4-heavy traffic, congested roadway	1.765	0.956
Temp	Ambient Temperature (Fahrenheit)	continuous	68.710	9.375
RJunction	Relation to Junction or not	1- yes 0- no	0.694	0.461
RFJunction	Relation to Four-leg Junction or not	1- yes 0- no	0.297	0.457
RHoriz	Relation to horizontal alignment or not	1- yes 0- no	0.331	0.471
RVerti	Relation to vertical alignment or not	1- yes 0- no	0.442	0.497
Radius	Radius of horizontal Curvature (Feet)	continuous	323.836	556.180
TDivided	Trafficway divided	1-yes 0-no	0.031	0.174
LWideth	Lane Width (Feet)	continuous	13.625	4.683
Pre-crash situations				
Pspeed	Pre-speed of motorcycle prior to crash(mile/h)	continuous	36.243	16.189
OLeft	Did the other vehicle turn left prior to precipitating event	1- yes 0- no	0.114	0.318
PStraight	Pre-crash motion of motorcycle prior to precipitating event is going straight or not	1- yes 0- no	0.614	0.487
Crash features				
Truck	The type of the other vehicle is truck or not	1- yes 0- no	0.165	0.371
Distance	Distance from point of impact to Motorcycle point of rest (Feet)	continuous	2.425	1.647
Week	Weekday or not	1-weekday; 0-weekend	0.671	0.470
Day	Night or daytime	0-Night 1-Daytim	0.693	0.461

alignment, the radius of horizontal curvature, junction, lane width and separation, traffic density. Further, for pre-crash attributes, speed of the motorcycle before crash, pre-crash motions of the motorcycle are included. Crash attributes are the other vehicle's type being truck or not, being weekend or not, being daytime or not, and distance from point of impact to Motorcycle point of rest.

In order to minimize the estimation bias introduced by the multicollinearity among explanatory variables, the correlation coefficient was conducted for each pair of variables (for two continous variables, Pearson correlation coefficient was used; for two discrete variables,Phi coefficient was used; for a continuous and discrete variable, Point-biserial correlation coefficient was used), as shown in Fig. 2. If two variables are highly correlated, the pair will not be placed in the model at the same time during the estimations. A single dependent variable (*Injury Severity*) is modeled as a function of multiple explanatory variables. As mentioned in Section 3, the injury severity score is calculated based on the NISS. Among the 322 observations, the mean of the NISS is 27.98, with a standard deviation of 31.99. Details of the NISS distribution are shown in Table 2b: 75 (23.2%) Single-vehicle motorcycle crashes (SingleC), 222 (68.9%) two-vehicle motorcycle crashes (TwoC), and 25 (7.8%) multi-vehicle motorcycle crashes (MultiC). Across all crash categories, there are 103 (32.0%) minor injuries, 52 (16.2%) moderate injuries, 50 (15.5%) serious injuries, 11 (3.4%) severe injuries, and 106 (32.9%) critical injuries.

4.2. Results from the Latent Class Clustering

Due to heterogeneity in crash data that may not be discovered by police-reports, it is not easy to identify and evaluate elements that affect the injury severity of motorcycles in such crash cases. By using of the latent class clustering (LCC) method and the MCCS, the LCC is used to identify the latent classes in the total dataset. The LCC model is estimated by using Mplus software. All candidate variables mentioned in Table 2a were used in the LCC for each level based on the number of involved vehicles.

The LCC does not require the number of clusters to be predetermined. In order to determine a suitable number of clusters, the number of clusters ranging from 1 to 6 were tried in the estimation and the corresponding BIC are plotted in Fig. 3. For SingleC, the one-class model has the lowest BIC value (973). The two-class model has the second lowest BIC value (974). But according to the bootstrap p-value of two-class model, which is smaller than 0.0001, the two-class model fits significantly better than the one-class model. For TwoC, the two-class model has the lowest BIC value. Therefore, the SingleC and TwoC database are separated into two classes for further analysis based on the latent class membership. For MultiC, the BIC of the one-class model is less than all other five models as well as MultiC has a small sample size. Thus, we decide not to divide the MultiC database.

After ascertaining the number of classes or clusters, the next step is profiling the categories in each class for each variable. More specifically, that is identifying the variables that show differentiation of conditional probability between classes. Two types of variables need to be removed from the LCC model (De Oña et al. 2013): Firstly, The highest value of probability is found in the same category of one variable for all classes. Here is an example, for the variable Truck of TwoC's class 1, the highest value of probability is found in category 1, which is the same as class 2. Therefore, this variable does not help in differentiating the classes and further is removed from the LCC model. Secondly, the same distribution of the probability values is obtained between each category of one variable. For example, for class 1 of TwoC, the probabilities of category 1 and 2 of the variable " PImpair" are 83.3% and 16.7%, and class 2 has the same value. Therefore, it does not show the classes' characterization.

The final LCC models are described on the basis of the characterization of specific variables that differ between the classes (Yu et al. 2017, Adanu et al. 2018). Fig. 4 shows the distribution of injury severities in each class for SingleC, TwoC and MultiC. Table 3 presents the distribution of crashes in each cluster of SingleC and TwoC across selected featured variables, which are significantly different in the specific class. Interestingly, two classes of SingleC and TwoC show consistency based on the characterization of variables PStraight and RHoriz.

4.3. Results from Ordered Probit model

Ordered Probit (OP) models are developed to identify the variables that are statistically significant in terms of contributing to the crash severity of motorcyclists. An OP model that included all observations is built initially to ascertain the merit of identifying predefined subgroups by the number of involved vehicles. Then the OP models were estimated for each subgroup based on the number of involved vehicles. Finally, in

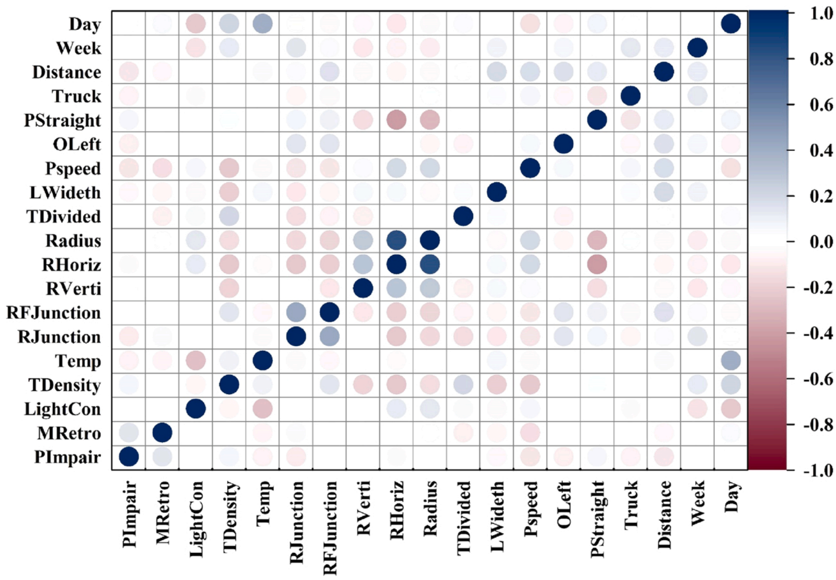


Fig. 2. Correlation matrix.

Table 2b
Distribution of motorcycle crashes with respect to injury severity.

Sub-groups of motorcycle crash	Minor	Moderate	Serious	Severe	Critical	*Total*
Single-vehicle crashes	20	10	11	2	32	75
Two-vehicle crashes	74	39	35	8	66	222
Multi-vehicle crashes	9	3	4	1	8	25
Total	103	52	50	11	106	322

order to determine the advantage of the LCC model over without the LCC model, the OP models for each latent class of SingleC and TwoC were also developed. The results of all models are presented in Tables 4a and 4b.

Model results of the whole database and the subgroups based on the number of involved vehicles (SingleC, TwoC, and MultiC) are presented in Table 4a. Without separating the database based on the number of involved vehicles, the OP model for the whole database cannot show some variables to significantly affect the injury severity, such as junction, week, and daytime. Thus, it is beneficial to identify predefined subgroups by the number of involved vehicles.

Model results of each class of SingleC and TwoC are shown in Table 4b. Since the exist of heterogeneity, the OP models for the combined databases of SingleC and TwoC ignore the impact of some variables to the injury severity of motorcyclists, such as retroreflective parts on motorcycle, traffic density and temperature. Furthermore, the sums of the log-likelihood values for two classes' models of Single and TwoC are respectively bigger than the log-likelihood of models based on whole SingleC and TwoC databases. Therefore, it is beneficial to separate the

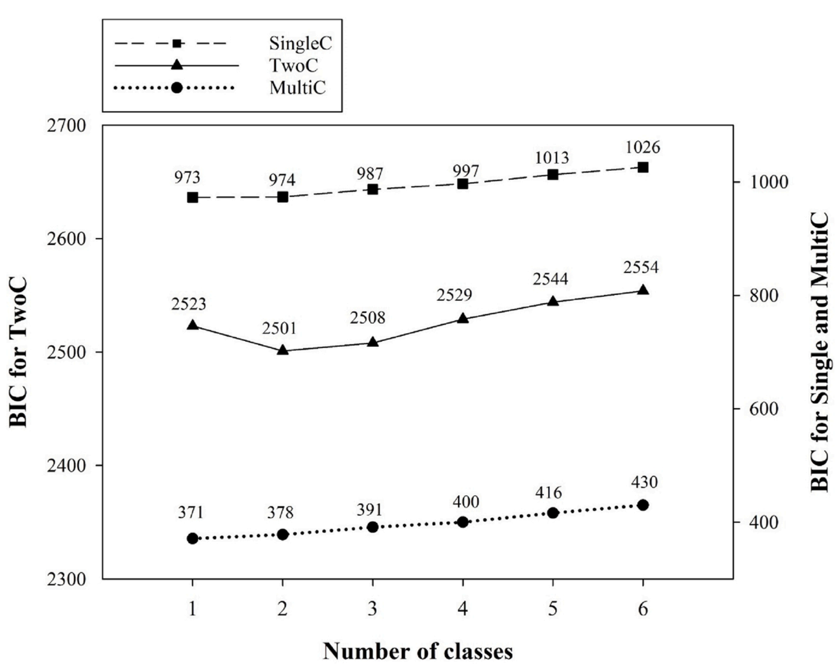


Fig. 3. The BIC values of SingleC, TwoC, and MultiC for different number of classes.

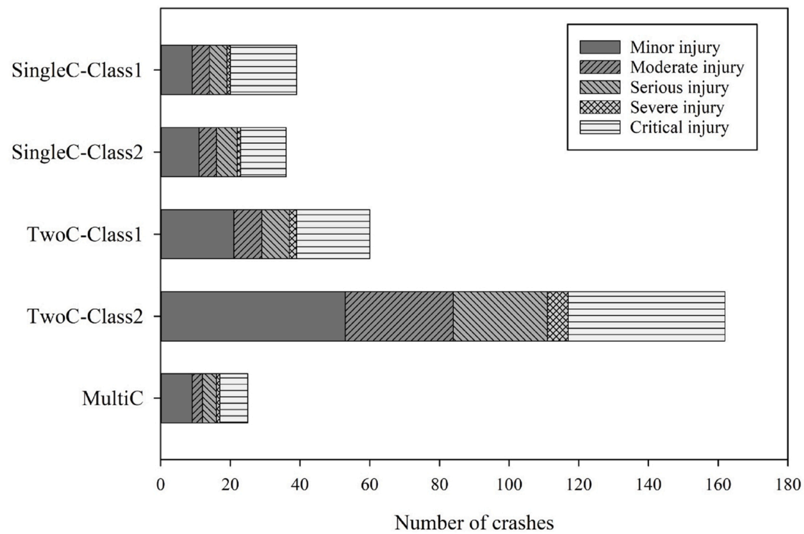


Fig. 4. Distribution of motorcycle crashes in each class.

Table 3
Variables of each class for SingleC and TwoC.

Variable	Value	Class 1(%)	Class 2(%)
Latent class clustering for SingleC			
Number of cases		39	36
RHoriz	0-Level	0	**100**
	1-Relation to horizontal alignment	**100**	0
PStraight	0-Oher motion	**76.4**	34.5
	1-Go straight	23.6	**65.5**
Day	0-Night	51.6	38.4
	1-Day time	48.4	61.
Latent class clustering for TwoC			
Variable	Value	Class 1(%)	Class 2(%)
Number of cases		60	162
RHoriz	0-level	8.0	**95.6**
	1-Relation to horizontal alignment	**92.0**	4.4
PStraight	0-other motion	**69.3**	21.9
	1-go straight	30.7	**78.1**
RVerti	0-straight line	29.1	65.0
	1-Relation to vertical alignment	**70.9**	35.0

TwoC database to homogeneity classes by the LCC model. Consequently, the two-class models of SingleC, the two-class models of TwoC, and the MultiC model are chosen for further analysis. Table 5 shows the marginal effect of the significant variables of the above models. The discussion of the results would be based on the marginal effects.

4.4. Marginal effect analysis

Marginal effects are utilized to show the effects of all significant variables on the injury severity of motorcyclists. Table 5 compares the marginal effects of each variable in sub-models, including class 1 of single-vehicle crashes (SC1), class 2 of single-vehicle crashes (SC2), class 1 of two-vehicle crashes (TC1), class 2 of two-vehicle crashes (TC2) and multi-vehicle crashes (MC). This section will discuss variables that significantly affect injury severities of motorcyclists, especially on critical injury.

(1) Rider and motorcycle characteristics

Motorcyclists who have permanent physical impairments (PImpair) are more likely to suffer critical injuries for TC1and TC2 (the marginal effects being +0.332 and +0.196, abbreviated as +0.332 and +0.196, the same below). PImpair may result in impaired judgment and slow reaction during driving, which includes infirmity, arthritis, senility, diabetes, neurological, cardiovascular condition, and hearing reduction or loss. The finding is in line with previous research (Duddu et al. 2018, Wali et al. 2019). Thus, enhancing the physical examination of motorcyclists and restriction of motorcyclists' physical condition may be beneficial in reducing the injury severity of motorcyclists in two-vehicle motorcycle crashes. For motorcycles that have retroreflective parts, motorcyclists are less likely to suffer critical injuries in SC2, TC2 and MC (-0.753, -0.126, and -0.821). According to Table 3, 100% crashes of SC2 happened in a level road. 95.6% crashes of TC2 happened in a level road. This is probably because the retroreflective motorcycles are eye-catching and other vehicles are easy to notice them. Therefore, retroreflective parts on motorcycles could help in reducing the injury severity of motorcyclists in crashes when riding in level roads. This result also indicates the benefit of latent class clustering that divides the database to homogeneity classes and provides insightful information about the crash database.

(2) Road conditions

Riding in good lighting conditions (i.e., bright daylight) decreases the probability of critical injury for TC1 (-0.405). According to Table 3, 92.0% crashes of TC1 happened in horizontal alignment of roads. One possible reason for this result is the good lighting conditions make good visions for both motorcyclists and drivers of the other vehicle on the horizontal alignment of roads. Riding in heavier traffic density also decreases the probability of critical injury for TC1(-0.196), which might be attributed to the possibility that complex riding environment, combined by the heavy traffic density and horizontal alignment of roads, makes motorcyclist generally more cautious. Another interesting finding is that the rise of the temperature slightly decreases the likelihood of critical injury in TC1 (-0.017), which is probably because motorcyclist protective clothing and helmet are less tolerable in the relatively high temperature (Bogerd et al. 2011, Shaheed et al. 2013). Another possible reason is the effect of the riding season (May, June, July, and August). The temperature in the heart of the riding season is relatively high than the not-riding season. The previous study shows single motorcycle crashes occurring in the heart of riding season are less likely to be critical injuries (Shaheed and Gkritza 2014). Multi-vehicle motorcycle crashes are found to be severer for motorcyclists near an intersection (+0.150). One possible reason is that complex road characteristics make drivers of other vehicles take no notice of motorcycles. By contrast, motorcyclists are less likely to be critical injuries in SC1 when crashes occurred near a

Table 4a
OP model for WholeC, SingleC, TwoC, and MultiC.

	WholeC[1]		SingleC		TwoC		MultiC	
variables	Coef.[2]	S.E.[3]	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Rider characteristics								
PImpair	0.40*	0.21			0.61***	0.23		
MRetro	−0.48***	0.17			−0.31*	0.18	−3.04***	1.00
Road conditions								
LightCon	−0.34*	0.18			−0.36**	0.18		
RJunction							1.63*	0.90
RFJunction			−0.81**	0.41				
RVerti			−0.84**	0.36			2.55***	0.95
Radius							−0.003**	0.001
TDivided	−0.91**	0.44						
LWideth	0.06***	0.02			0.06***	0.02		
Pre-crash situations								
Pspeed	0.02***	0.01	0.02**	0.01	0.02***	0.01	0.08***	0.02
Crash features								
Truck	0.42**	0.21			0.53**	0.23		
Distance			0.29**	0.12				
Week			−0.85**	0.37				
Day			−0.73**	0.35				
LL[4]	−288		−63		−258		−16	
Prob > chi^2	0.0000		0.0005		0.0000		0.0000	
	LR chi2(7) = 42.2		LR chi2(6) = 23.9		LR chi2(6) = 34.72		LR chi2(5) = 32.92	
Pseudo R^2	0.0682		0.1587		0.0631		0.5123	

[1] WholeC: model based on the whole crash database; SingleC: model based on the database of single-vehicle motorcycle crashes; TwoC: model based on the database of two-vehicle crashes involved a motorcycle; MuitiC: model based on the database of multi-vehicle motorcycle crashes.
[2] Coef: coefficient; *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.
[3] S.E.: standard error.
[4] LL: log likelihood.

Table 4b
OP model for each class of SingleC and TwoC.

	SingleC-class1[1]		SingleC-class2		TwoC-class 1		TwoC-class 2	
variables	Coef.[2]	S.E.[3]	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Rider characteristics								
PImpair					0.87*	0.52	0.54**	0.27
MRetro			−2.33**	1.07			−0.36*	0.21
Road conditions								
LightCon					−1.37***	0.42		
TDensity					−0.59**	0.30		
Temp					−0.05**	0.02		
RFJunction	−1.71**	0.87						
RVerti	−1.54***	0.58	3.15**	1.41			−0.33*	0.20
TDivided							−0.86**	0.48
LWideth			−0.29***	0.11	0.11**	0.05	0.05**	0.02
Pre-crash characteristics								
Pspeed			0.08***	0.03	0.03**	0.01	0.02***	0.01
OLeft					−1.16*	0.68		
PStraight	−1.29**	0.63						
Crash features								
Truck							0.55**	0.26
Distance	0.51***	0.19	1.03***	0.38				
Week	−1.51***	0.52						
Day	−0.81*	0.47	−1.65**	0.78				
LL[4]	−36		−18		−55		−193	
Prob > chi2	0.0103		0.0002		0.0001		0.0010	
	LR chi2(6) = 16.75		LR chi2(6) = 26.81		LR chi2(7) = 31.14		LR chi2(7)24.22	
Pseudo R2	0.1868		0.4223		0.2209		0.0591	

[1] WholeC: model based on the whole crash database; SingleC: model based on the database of single-vehicle motorcycle crashes; TwoC: model based on the database of two-vehicle motorcycle crashes; MuitiC: model based on the database of multi-vehicle motorcycle crashes.
[2] Coef: coefficient; *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.
[3] S.E.: standard error.
[4] LL: log likelihood.

four-leg junction (-0.411), which is line with the previous research (Savolainen and Mannering 2007). According to Table 3, 100% crashes of SC1 happened near horizontal alignment. One possible reason is that four-leg junction and horizontal alignment make motorcyclists more cautious for the perceived higher risk. Motorcyclists are more likely to be critical injuries when crashes occurred on vertical alignment for SC2 and MC (+0.885, +0.561), which is agrees with the previous study (Savolainen and Mannering 2007). By contrast, the vertical alignment decreases the risk of critical injury for SC1 and TC2 (-0.557, -0.107). The higher radius of horizontal alignment is more likely to decrease the

Table 5
Marginal Effects of Variables.

Variable	SingleC-Class 1					SingleC-Class 2				
	Mi[1]	Mo	Sr	Sv	Cr	Mi[1]	Mo	Sr	Sv	Cr
Rider and motorcycle characteristics										
MRetro*						0.196	0.351	0.204	0.003	−0.753
Road conditions										
RFjunction*	0.589	−0.008	−0.136	−0.034	−0.411					
RVerti*	0.317	0.179	0.064	−0.003	−0.557	−0.295	−0.383	−0.197	−0.010	0.885
LWideth						0.045	0.062	−0.007	−0.014	−0.086
Pre-crash situations										
Pspeed						−0.012	−0.017	0.002	0.004	0.023
PStraight*	0.399	0.083	−0.053	−0.022	−0.406					
Crash features										
Distance	−0.125	−0.06	−0.008	0.005	0.193	−0.163	−0.224	0.026	0.050	0.310
Week*	0.340	0.167	0.041	−0.007	−0.540					
Day*	0.199	0.098	0.012	−0.008	−0.301	0.294	0.275	−0.044	−0.62	−0.463
Variable	TwoC-Class 1					TwoC-Class 2				
	Mi[1]	Mo	Sr	Sv	Cr	Mi[1]	Mo	Sr	Sv	Cr
Rider and motorcycle characteristics										
PImpair*	−0.262	−0.051	−0.020	0.011	0.322	−0.163	−0.046	0.005	0.008	0.196
MRetro*						0.119	0.025	−0.011	−0.007	−0.126
Road conditions										
LightCon*	0.490	0.018	−0.062	−0.040	−0.405					
TDensity	0.214	0.019	−0.019	−0.018	−0.196					
Temp	0.018	0.002	−0.002	−0.002	−0.017					
RVerti*						0.115	0.016	−0.016	−0.007	−0.107
TDivided*						0.328	−0.020	−0.074	−0.022	−0.212
LWideth	−0.040	−0.004	0.004	0.003	0.036	−0.019	−0.003	0.002	0.001	0.018
Pre-crash situations										
Pspeed	−0.013	−0.001	0.001	0.001	0.012	−0.006	−0.001	0.001	0.000	0.006
OLeft*	0.438	−0.025	−0.101	−0.043	−0.270					
Crash features										
Truck*						−0.166	−0.047	0.005	0.008	0.200
Variable	MultiC									
	Mi[1]	Mo	Sr	Sv	Cr					
Rider and motorcycle characteristics										
MRetro*	0.323	0.418	0.122	−0.043	−0.821					
Road conditions										
RJunction*	−0.492	−0.042	0.303	0.082	0.150					
RVerti*	−0.448	−0.334	0.113	0.109	0.561					
Radius	0.001	0.001	−0.001	−0.0002	−0.0004					
Crash features										
Pspeed	−0.017	−0.015	0.014	0.006	0.012					

[1] Mi: minor injury; Mo: moderate injury; Sr: serious injury; Sv: severe injury; Cr: critical injury.
* dy/dx is for discrete change of dummy variable from 0 to 1.

critical injuries for MC (-0.0004). Divided traffic ways decrease the probability of critical injury for TC2 (-0.107), which is consist with the previous research (Haque et al. 2012). One possible reason is that interactions of opposing traffic may be detrimental on motorcyclists' attention on undivided traffic ways. The increasing lane width is associated with a decrease in critical injury severity for SC2 (-0.086). it is also associated with an increase in critical injury severity for TC1 and TC2 (+0.036, +0.018).

In summary, the results indicate different impacts of road conditions among databases. Difference countermeasures can then be concluded accordingly for each type of crashes.

(3) Pre-crash situations

Higher pre-speed can increase the likelihood of critical injury in most types of crashes, except for SC1. Specifically, the probabilities of critical injury can be increased by 2.3% for SC2, 1.2% for TC1, 0.6% for TC2 and 1.2% for MC. This finding is consistent with existing research (Shankar and Mannering 1996, Shaheed and Gkritza 2014), which is logical because higher pre-speed gives less reaction time for the motorcyclist in an emergency. Thus, optimizing the speed limit of road and traffic law enforcement could be useful in reducing the injury severity of motorcyclists. When other vehicle turns left before crashes, the motorcyclist is found to be a less critical injury in TC1 (-0.27). When motorcycle go straight before crashes, the motorcyclist is found to be a less critical injury in SC1 (-0.406).

(4) Crash features

It is found that motorcycle-truck crashes are more likely to be critical injuries (+0.2 for TC2), which is intuitive and in line with past research (Waseem et al. 2019). The variable Distance (the distance from the point of impact to motorcycle point of rest) shows positive significance in the critical injury outcome of motorcyclist for SC1 and SC2 (+0.193, 0.31), which indicates that single crashes with long distance are more likely to result in critical injuries. This finding is consistent with the past study (Wali et al. 2019). It is also found that the weekday's single crash occurring on horizontal alignment is less likely to result in a critical injury compared to the weekend one (-0.54 for SC1), which agrees with the previous study (Shaheed and Gkritza 2014). Results also indicate that the effect of daytime decreases critical injury severities for SC1 and SC2 (-0.301, -0.463).

5. Conclusions

The principal aim of this study is to investigate the differences of the variables that are statistically significant in impacting injury severity of motorcyclists among latent classes. The motorcycle crash database in from the Motorcycle Crash Causation Study (MCCS) is used in the study. In comparison with studies based on the police-reported database and measuring injury severity by KABCO, we use an anatomic injury severity indicator, named the New Injury Severity Score (NISS), to scale the injury severity. A hybrid approach integrating the Latent Class Clustering and Ordered Probit (OP) models was proposed to account for the unobserved heterogeneity and to explore the major factors which are statistically significant to affect the injury severities of motorcyclists.

The results indicate that it is beneficial to split the whole database into single-, two-, and multi-vehicle crashes. The significant differences in motorcyclist injury severity exist between different number of involved vehicles. The results also show that sub-models based on latent classes provide better goodness-of-fit than the model based on the whole dataset, and some variables are statistically significant only in specific classes. Therefore, latent class clustering analysis can provide more insightful information. Pre-speed of the motorcycle is the main factor associated with serious and critical injury on most types of crashes. According to the possible safety-oriented policies and strategies suggested in the study, it shows that optimizing speed limit of road and traffic law enforcement could be useful in reducing the injury severity of motorcyclists. Results also shows some variables found to be more likely to increase injury severity, such as permanent physical impairments, truck, low temperature, bad lighting condition. It requires motorcyclists to use safety-oriented equipment when riding a motorcycle, such as retroreflective motorcycle.

The study still has some limitations. First, the data used in this study are of a relatively small sample size and only include crashes from 2011 to 2016, which may not be representative for the national situation. This would be desirable for future work to use a more extensive database that includes crashes after 2016 and cases in other areas of the United States, which could allow researchers to understand more comprehensively the effects of various variables. Secondly, though most of the results of this research are line with past studies on injury severity analysis of motorcycle, some variables show different influence on injury severity of motorcyclists, which is probably caused by differences in the riding environment and needs further investigation. Thirdly, though the MCCS only includes non-fatal injury crashes, most of the crashes involved a motorcycle result in injuries instead of fatalities (NHTSA, 2015). Therefore, the MCCS data represent most of the motorcycle crashes in the U.S. Availability of comprehensive injury data (including fatality and injury) in the future can provide an opportunity for more in-depth analysis of motorcyclist safety.

CRediT authorship contribution statement

Jing Li: Conceptualization, Methodology, Writing - original draft, Data curation. **Shouen Fang:** Conceptualization, Methodology, Funding acquisition. **Jingqiu Guo:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Ting Fu:** Methodology, Writing - review & editing. **Min Qiu:** Supervision, Writing - review & editing.

Declaration of Competing Interest

All authors declare there is no conflict of interest.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant [2018YFB1600500, 2019YFE0108300] and the Fundamental Research Funds for the Central Universities. The data for this study were provided by the Federal Highway Administration (FHWA). The authors would like to thank the editor and reviewers for their insightful comments to improve the quality of this paper.

References

Adanu, E.K., Hainen, A., Jones, S., 2018. Latent class analysis of factors that influence weekday and weekend single-vehicle crash severities. Accident Analysis & Prevention 113, 187–192.

Adinegoro, Y., Haworth, N., Debnath, A.K., 2015. Characteristics of road factors in multi and single vehicle motorcycle crashes in queensland.

Alnawmasi, N., Mannering, F., 2019. A statistical assessment of temporal instability in the factors determining motorcyclist injury severities. Analytic methods in accident research 22, 100090.

Amoros, E., Martin, J.-L., Laumon, B., 2006. Under-reporting of road crash casualties in france. Accident Analysis & Prevention 38 (4), 627–635.

Baker, S.P., O'neill, B., Haddon Jr., W., Long, W.B., 1974. The injury severity score: A method for describing patients with multiple injuries and evaluating emergency care. Journal of Trauma and Acute Care Surgery 14 (3), 187–196.

Bogerd, C.P., Rossi, R.M., Brühwiler, P.A., 2011. Thermal perception of ventilation changes in full-face motorcycle helmets: Subject and manikin study. Annals of occupational hygiene 55 (2), 192–201.

Cheng, W., Gill, G.S., Sakrani, T., Dasu, M., Zhou, J., 2017. Predicting motorcycle crash injury severity using weather data and alternative bayesian multivariate crash frequency models. Accident Analysis & Prevention 108, 172–180.

De Oña, J., López, G., Mujalli, R., Calvo, F.J., 2013. Analysis of traffic accidents on rural highways using latent class clustering and bayesian networks. Accident Analysis & Prevention 51, 1–10.

Ding, C., Rizzi, M., Strandroth, J., Sander, U., Lubbe, N., 2019. Motorcyclist injury risk as a function of real-life crash speed and other contributing factors. Accident Analysis & Prevention 123, 374–386.

Duddu, V.R., Penmetsa, P., Pulugurtha, S.S., 2018. Modeling and comparing injury severity of at-fault and not at-fault drivers in crashes. Accident Analysis & Prevention 120, 55–63.

Eid, H.O., Abu-Zidan, F.M., 2015. New injury severity score is a better predictor of mortality for blunt trauma patients than the injury severity score. World journal of surgery 39 (1), 165–171.

Flask, T., Schneider Iv, W.H., Lord, D., 2014. A segment level analysis of multi-vehicle motorcycle crashes in ohio using bayesian multi-level mixed effects models. Safety science 66, 47–53.

Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of poisson–gamma models. Accident Analysis & Prevention 42 (4), 1273–1282.

Gennarelli, T.A., Wodzin, E., 2006. Ais 2005: A contemporary injury scale. Injury 37 (12), 1083–1091.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., 1998. Multivariate data analysis Prentice hall Upper Saddle River, NJ.

Haque, M.M., Chin, H.C., Debnath, A.K., 2012. An investigation on multi-vehicle motorcycle crashes using log-linear models. Safety science 50 (2), 352–362.

Ivan, J.N., Konduri, K.C., 2018. Crash severity methods. Safe mobility: Challenges, methodology and solutions. Emerald Publishing Limited.

Kitali, A.E., Kidando, E., Alluri, P., Sando, T., Salum, J.H., 2020. Modeling severity of motorcycle crashes with dirichlet process priors. Journal of Transportation Safety & Security 1–22.

Lanza, S.T., Dziak, J.J., Huang, L., Wagner, A.T., Collins, L.M., 2015. Lca stata plugin users' guide (version 1.2). The Methodology Center, Penn State, University Park.

Lanza, S.T., Rhoades, B.L., 2013. Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. Prevention Science 14 (2), 157–168.

Lavoie, A., Moore, L., Lesage, N., Liberman, M., Sampalis, J.S., 2004. The new injury severity score: A more accurate predictor of in-hospital mortality than the injury severity score. Journal of Trauma and Acute Care Surgery 56 (6), 1312–1320.

Li, Y., Fan, W.D., 2019. Modelling severity of pedestrian-injury in pedestrian-vehicle crashes with latent class clustering and partial proportional odds model: A case study of north carolina. Accident Analysis & Prevention 131, 284–296.

Liu, P., Fan, W., 2020. Exploring injury severity in head-on crashes using latent class clustering analysis and mixed logit model: A case study of north carolina. Accident Analysis & Prevention 135, 105388.

Loo, B.P., Tsui, K., 2007. Factors affecting the likelihood of reporting road crashes resulting in medical treatment to the police. Injury prevention 13 (3), 186–189.

Malyshkina, N.V., Mannering, F.L., 2009. Markov switching multinomial logit model: An application to accident-injury severities. Accident Analysis & Prevention 41 (4), 829–838.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic methods in accident research 11, 1–16.

Nhtsa, 2019. National center for statistics and analysis. 2018 fatal motor vehicle crashes: Overview. National Highway Traffic Safety Administration, Washington, DC pp. DF-1.6.

Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. Structural equation modeling: A multidisciplinary Journal 14 (4), 535–569.

Osler, T., Baker, S.P., Long, W., 1997. A modification of the injury severity score that both improves accuracy and simplifies scoring. Journal of Trauma and Acute Care Surgery 43 (6), 922–926.

Quddus, M.A., Noland, R.B., Chin, H.C., 2002. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety research 33 (4), 445–462.

Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single-and multi-vehicle crashes. Accident Analysis & Prevention 39 (5), 955–963.

Shaheed, M.S., Gkritza, K., 2014. A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research 2, 30–38.

Shaheed, M.S.B., Gkritza, K., Zhang, W., Hans, Z., 2013. A mixed logit analysis of two-vehicle crash severities involving a motorcycle. Accident Analysis & Prevention 61, 119–128.

Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of safety research 27 (3), 183–194.

Stevenson, M., Segui-Gomez, M., Lescohier, I., Di Scala, C., Mcdonald-Smith, G., 2001. An overview of the injury severity score and the new injury severity score. Injury Prevention 7 (1), 10–13.

Tarko, A.P., Bar-Gera, H., Thomaz, J., Issariyanukula, A., 2010. Model-based application of abbreviated injury scale to police-reported crash injuries. Transportation research record 2148 (1), 59–68.

Vajari, M.A., Aghabayk, K., Sadeghian, M., Shiwakoti, N., 2020. A multinomial logit model of motorcycle crash severity at australian intersections. Journal of Safety Research.

Wali, B., Khattak, A.J., Ahmad, N., 2019. Examining correlations between motorcyclist's conspicuity, apparel related factors and injury severity score: Evidence from new motorcycle crash causation study. Accident Analysis & Prevention 131, 45–62.

Waseem, M., Ahmed, A., Saeed, T.U., 2019. Factors affecting motorcyclists' injury severities: An empirical assessment using random parameters logit model with heterogeneity in means and variances. Accident Analysis & Prevention 123, 12–19.

Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. Analytic methods in accident research 1, 72–85.

Yu, R., Wang, X., Abdel-Aty, M., 2017. A hybrid latent class analysis modeling approach to analyze urban expressway crash risk. Accident Analysis & Prevention 101, 37–43.