Accident Analysis and Prevention 144 (2020) 105638

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

# Combined latent class and partial proportional odds model approach to exploring the heterogeneities in truck-involved severities at cross and T-intersections

Li Song, Wei Fan (David)*

*Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, EPIC Building, 9201 University City Boulevard, Charlotte, NC 28223-0001, United States*

ARTICLE INFO

*Keywords:*
Severity analysis
Truck crashes
Intersection
Latent class clustering
Partial proportional odds model

ABSTRACT

Although the fatal rate of passenger vehicle-involved crashes has decreased in the United States, the fatal rate of truck-involved crashes has increased. This has, in recent years, become a more severe problem than that caused by passenger vehicle-involved crashes. More studies need to be conducted in order to investigate factors that impact the severity of truck-involved crashes within specific scenarios. This study identifies and evaluates the factors that affect the severity of the truck-involved crashes at cross and T-intersections in North Carolina from 2005 to 2017. A latent class clustering for data segmentation is implemented to mitigate unobserved heterogeneity inherent in the crash data. Four partial proportional odds models, which include fixed and unfixed parameters, are developed considering the heterogeneous and ordinal nature inherent in severities. Estimated parameters and marginal effects are further investigated for better interpreting the impacts. Results show heterogeneous explanatory variables and associated coefficients for different classes and severity levels, which indicate the superiority of this combined approach to obtaining more specific factors and accurate coefficients that are estimated in different scenarios. Many factors are found to contribute to the severities, and crossroad scenarios are found to be more severe than T-intersections. The top five driving behaviors at intersections that contribute to the severity include disregarded signs, improper lane use, followed too closely, ignored signals, and failure to yield. These behaviors arouse a necessity to amend the traffic laws and strengthen drivers' education while giving further insights to engineering practitioners and researchers.

## 1. Introduction

Truck-involved crashes always result in more severe injuries in comparison to passenger vehicles even if they have lower average traveling speeds. According to the U.S. Department of Transportation statistics, the proportion of trucks only made up 4.9 % of the total vehicles registered in 2017, while they completed 10.3 % of the vehicles' total mileage traveled (USDOT, 2019). Truck drivers are more likely to suffer fatigue when driving because of the long duration. In addition, trucks need more time to brake and are more likely to roll over because of their size and weight. These factors all cause truck driving to be more hazardous than those of passenger cars.

In the United States, the number of passenger vehicles resulting in fatal injury decreased by 1.4 % in 2017 due to implemented measurements that improve driving safety. In contrast, the fatal crashes involving large trucks increased by 9.6 % from 4251 in 2016 to 4657 in 2017 (USDOT, 2019). In addition, fatality rates of large truck-involved crashes per 100 million vehicle miles increased by 5.2 % and reached 1.42 in 2017, which is 1.38 times the rate of crashes involving passenger vehicles (USDOT, 2019). In summary, the deteriorative trends and outcomes of truck-involved crashes deserve more attention and research to explore the truck-involved crash factor's impacts, providing a better guidance to prevent and mitigate the truck-involved crashes.

Different crash scenarios might be caused by different factors, and those factors also have heterogeneous impacts on different crash severity levels. In addition, the severity levels, from non-injury to fatal, have ordinal features. These issues require a framework that can consider the natural order of the severity and heterogeneity inherent in the explanatory variables and severity levels at the same time. In this paper, a combined framework of latent class and partial proportional odds

* Corresponding author.
*E-mail addresses:* lsong1@uncc.edu (L. Song), wfan7@uncc.edu (W. Fan).

https://doi.org/10.1016/j.aap.2020.105638
Received 13 February 2020; Received in revised form 15 May 2020; Accepted 4 June 2020
Available online 26 June 2020
0001-4575/ 

models is constructed to explore the factors and their impact on the truck-involved crash severity level at cross and T-intersections in North Carolina during 2005 to 2017.

## 2. Literature review

By considering different crash outcomes that are caused by different factors, many studies investigated truck-involved crash severity in specific scenarios, which are summarized in Table 1. These scenarios could be mainly divided into seven categories, including locations (rural and urban, work zones), crash types (run-off-road, rollover), human characteristics (gender, age, driver and occupant), roadway function (highways, truck-only and mixed road), vehicle characteristics (single-vehicle and multi-vehicles, hazmat truck), time (day and week), and environment (light and weather). However, few studies specifically focused on exploring the factors of truck-involved crashes that occurred at cross and T-intersections. Past truck-involved research set intersection/non-intersection (Khorashadi et al., 2005; Zhu and Srinivasan, 2011) or signal/non-signal control (Anderson and Dong, 2017; Chen and Chen, 2011) as explanatory variables but did not specifically focus on the truck-involved crashes at intersections. Since intersections have more complex traffic conditions and conflict points than normal roadways, traffic at intersections is more vulnerable and therefore likely to suffer more frequent and serious crashes (FHWA, 2004; Zhu and Srinivasan, 2011). This paper is dedicated to investigating the truck-involved crashes at cross and T- intersections.

For studies focused on truck-involved crash scenario, unobserved heterogeneity remained in the research data because many factors were not included or cannot be directly observed. Recent studies have realized the importance of counting the unobserved heterogeneity because neglecting it could result in incorrect model results and conclusions (Li and Fan, 2019). Two main approaches to mitigating the heterogeneity of the data were proposed. The first is clustering the crash data into more specific groups (Mohamed et al., 2013), and the second is constructing an analytical model which could estimate observation-specific variations of the variables (Behnood and Mannering, 2019). The clustering methods divide the data based on some homogenous features, which can increase the heterogeneity between subgroups. However, heterogeneity remains within the subgroups. Several clustering methods were implemented in crash severity analysis, including classification tree (Bernard and Mondy, 2016), support vector machine (Chen et al., 2016), k-means (Mohamed et al., 2013; Nitsche et al., 2017), k-medoids (Iranitalab and Khattak, 2017), nearest neighbor classification (Iranitalab and Khattak, 2017), block clustering (Rahimi et al., 2019), and latent class analysis/clustering (Chang et al., 2019; Iranitalab and Khattak, 2017; Li and Fan, 2019; Liu and Fan, 2020; Mohamed et al., 2013). Rahimi et al. (2019) implemented a block clustering method to segment large heterogeneous crash records into meaningful subgroups. Results showed a 95.72 % average degree of homogeneity for the selected blocks/subgroups.

Latent class clustering (LCC) is a statistical model-based method that could maximize the heterogeneity between the crash data subsets. Iranitalab and Khattak (2017) compared the crash severity prediction performance of four statistical and machine learning models, and results proved that LCC improved the performance of the multinomial logit model. Mohamed et al. (2013) combined k-means and LCC with multinomial logit and ordered probit models for crash severity analysis in New York. Results indicated that LCC combined with the ordered probit model outperforms other models, and crash data segmentation contributed to better model estimations and comprehensions. Hence, this paper implements LCC to cluster crash data for reducing the impact of unobserved heterogeneity within the crash data.

Many of the previous studies adopted discrete-outcome models to investigate truck-involved crash severity as shown in Table 1. Among them, statistic-based methods such as logit/probit models have been widely used because of their good performance in parameter calibration and outcome interpretation for discrete data (Behnood and Mannering, 2019). It is worth noting that crash severities, which increase from non-

**Table 1**
Summary of the methods used in truck-involved crash severity analysis in specific scenarios.

| Model | Specific scenarios | Data | NO. | Literature |
|---|---|---|---|---|
| ***Unordered logit/probit Model*** | | | | |
| Multinomial logit | Urban & rural | 1997–2000 California | 17,372 | Khorashadi et al. (2005) |
| Mixed logit | Rural highways, SV & MV | 1991 – 2000 illinois | 19,741 | Chen and Chen (2011) |
| | Rural & urban, SV & MV | 2010 – 2012 Alabama | 8171 | Islam et al. (2014) |
| | Urban & time of day | 2006 – 2010 Texas | 11,560 | Pahukula et al. (2015) |
| | Time of week | 2001 – 2014 Minnesota | 23,218 | Anderson and Dong (2017) |
| | Light condition, rural and urban | 2009 – 2013 Ohio | 41,461 | Uddin and Huynh (2017) |
| | Run-off-road, lighted & dark | 2007 – 2013 Oregon | 2486 | Al-Bdairi et al. (2018) |
| | Time of day | 2010 – 2017 Los Angeles | 5737 | Behnood and Mannering (2019) |
| Mixed logit & latent class method | – | 2007 – 2012 Iowa | 17,608 | Cerwick et al. (2014) |
| Bayesian inference binary logit | Rural highways | 2002 – 2011 Wyoming | 160,613 | Ahmed et al. (2018) |
| Bayesian estimated multinomial logit | – | 2006 – 2010 Tennessee | 5650 | Dong et al. (2017) |
| ***Ordered logit/probit Model*** | | | | |
| Multinomial logit, Nested logit, ordered logit, generalized ordered logit | Work zones | 2003 – 2012 Minnesota | 2881 | Osman et al. (2016) |
| Ordered probit | Truck-only & mixed road | 2007 – 2013 Abu Dhabi | 1426 | Hassan et al. (2015) |
| Cross-classified multilevel ordered logit | Company & region related | 2010 – 2014 South Korea | 86,622 | Park et al. (2017) |
| Heteroskedastic ordered probit | – | 2001 – 2003 17 USA States | 1894 | Lemp et al. (2011) |
| Fixed- and random parameters ordered probit | Hazmat truck | 2005 – 2011 California | 1173 | Uddin and Huynh (2018) |
| Random parameter ordered probit | Occupant level | 2001 – 2003 17 USA States | 918 | Zhu and Srinivasan (2011) |
| | Run-off-road | 2007 – 2013 Orgeon | 13,364 | Al-Bdairi and Hernandez (2017) |
| Random parameter generalized ordered probit | Age groups | 2012 – 2015 Minnesota | 6247 | Osman et al. (2018) |
| Spacial generalized ordered probit | SV & MV | 2008 – 2012 New York city | 4504 | Zou et al. (2017) |
| Random parameter ordered logit | Rollover | 2007 – 2016 Florida | 3418 | Azimi et al. (2020) |
| Mixed logit & random parameter ordered logit | Weather & SV | 2009 – 2011 Nebraska | 1721 | Naik et al. (2016) |
| ***Other Models*** | | | | |
| Classification and regression tree | – | 2005 – 2006 Taiwan | 1620 | Chang and Chien (2013) |
| Decision tree | Driver gender | 2002 – 2012 Missouri | 30,904 | Bernard and Mondy (2016) |
| Hierarchical Bayesian random intercept | Rural | 2010 – 2011 New Mexico | 5398 | Chen et al. (2015) |

*Note: SV denotes single-vehicle, MV denotes multi-vehicles.

injury to fatal, have inherently ordinal features. This ordinal nature violates the independence of the irrelevant alternatives assumption for unordered models. Accounting for this, a proportional odds model (POM), also named order logit model, which segmented the response level based on the thresholds of the cumulative probability was proposed (McCullagh, 1980). POM assumes that explanatory variables have the same slope at different levels (i.e. parallel lines assumption), which makes it easier for parameter estimation and interpretation. However, this assumption is always violated in practice (Derr, 2013). Thus, a generalized ordered logit model (GOLM) which has unfixed slopes for parameters in different response levels was proposed (McCullagh, 1980). This model mitigates the heterogeneity of variables in different response levels to some extent. Osman et al. (2016) compared the performance of unordered models (multinomial logit and nested logit) and ordered models (POM and GOLM) in analyzing truck-involved crash severity in Minnesota's work zones. Results showed that GOLM outperformed others with better model fitness. However, GOLM needs vast amounts of data to estimate parameters and could not certainly guarantee the order of the response level (Derr, 2013). Past research has noted that models addressing the heterogeneity are statistically superior (Behnood and Mannering, 2019). Mannering and Bhat (2014) discussed that the heterogeneity inherent in the crash observations due to observed or unobserved variables could result in biased parameter estimation and incorrect inferences. However, the random parameter model can also complicate the model results and make them hard to interpret. The heterogeneous results are also undesirable to be used in practice since some findings are specific to some observations. The partial proponent odds model (PPO), which could both include fixed and unfixed parameters, is introduced in this paper to maintain the ordinal nature of severity levels and explore the heterogeneities of explanatory variables in different severity levels. Sasidharan and Menéndez (2014) used PPO to investigate pedestrian severity and results indicated that PPO outperformed multinomial logit and ordered logit models. In summary, both ordinal nature and heterogeneities within datasets and severity levels are considered in this paper, and a combined LCC and PPO approach framework is used to investigate the severity of the truck-involved crashes at cross and T-intersections.

## 3. Methodology

### 3.1. Latent class clustering

Latent class clustering (LCC) is a statistical model-based method that can divide the dataset into mutually exclusive subsets by maximizing the heterogeneity between classes (Lanza et al., 2007). LCC identifies categorical latent/unobservable variables by calculating the class membership probabilities and conditional probabilities that variables take on certain class memberships (Chang et al., 2019; Liu and Fan, 2019; McLachlan and Peel, 2004). The LCC does not need to predefine the number of clusters since LCC can determine it according to some statistical criteria. Meanwhile, LCC does not require standardized variables including counts, continuous, categorical, and nominal variables (Lanza and Rhoades, 2013).

In this paper, all continuous variables are categorized into discrete variables for specifically describing the scenarios. Suppose that the LCC model classifies the whole dataset with $j$ discrete independent variables into $N$ classes. The probability of response can be calculated as:

$$P(Y_i = y) = \sum_{n=1}^{N} \gamma_n \prod_{j=1}^{J} \prod_{rj=1}^{R_j} \rho_{j,\,rj|n}^{I(y_j=r_j)} \tag{1}$$

where $\gamma_n$ denotes the membership probability for latent class cluster $n$ ($n = 1,\ 2,\ \ldots,\ N$). It is assumed that each observation $i$ contains $J$ variables, $Y_i$ denotes the result of the observation $i$ for $J$ categorical variables, and $Y_i = 1,\ 2,\ \ldots\ r_j$. $\rho_{j,\,rj|n}^{I(y_j=r_j)}$ represents the item-response probability that case $i$ has attribute $r_j$, conditioned on latent class membership $n$. $\rho$ indicates the correspondence between observed and unobserved classes. $I(y_j = r_j)$ denotes the indicator function that equals to 1 when $y_j = r_j$, and 0 otherwise (Lanza and Rhoades, 2013; Liu and Fan, 2020). $\gamma_n$ is a multinomial logit model that could be presented as:

$$\gamma_n = P(Class = n) = \frac{e^{\beta_{0n}+\beta_{1n}\boldsymbol{x_i}}}{1 + \sum_{k=1}^{N-1} e^{\beta_{0k}+\beta_{1k}\boldsymbol{x_i}}} \tag{2}$$

where $\gamma_n$ denotes the probability when class is $n$, and class $N$ is set as the reference class. $x_i$ is the $i$th category variable. $\beta$ is the coefficient estimated.

To determine a suitable cluster number, this paper adopts the commonly used criteria including Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), Consistent Akaike Information Criterion (CAIC), and entropy-based measures (EM) (Liu and Fan, 2020).

$$AIC = -2\log(SSE/n) + 2k \tag{3}$$

$$BIC = -2\log(SSE/n) + log(n)k \tag{4}$$

$$CAIC = -2\log(SSE/n) + [\log(n) + 1]k \tag{5}$$

$$EM = -\sum_{i=1}^{n} p_i \ln(p_i) \tag{6}$$

where $SSE$ denotes the error sum of squares. $k$ represents the number of parameters to be estimated. $i$ means the number of observations, $i = 1$, 2, …, $n$. $p_i$ is the probability of the possible value of observation $x_i$.

The EM indicates the data quality because including different values of the variable from the original dataset denotes that the subgroup has more information. The value of EM ranging within 0 and 1 and closing to 1 denotes a better clustering result (McLachlan and Peel, 2004). AIC, BIC, and CAIC are all the penalized-likelihood criteria which denote the error between the estimated and true likelihood function (Lanza and Rhoades, 2013). Hence, a better clustering result has smaller values of the AIC, BIC, and CAIC. However, it is noted that increasing the cluster number might not always reach the minimum values for these information criteria. Hence, an appropriate number of clusters is determined by the combination analysis of these four criteria.

### 3.2. Partial proportional odds model

Partial proportional odds model (PPO) is a hybrid of the proportional odds model and general ordered logit (Peterson and Harrell, 1990). The PPO includes the parameters that both satisfy and violate the parallel lines assumption. Suppose that the response $Y$ has a natural sequence, $Y = 1,2,\ \ldots,\ J$. The PPO divides the cumulative probabilities into different categories by the thresholds. The corresponding cumulative probabilities are $\{\ \pi_1,\ \pi_2,\ldots,\ \pi_J\ \}$. The cumulative probability for $Y$ less than $j$ is denoted as:

$$P(Y \le j) = \sum_{m=1}^{j} \pi_m \tag{7}$$

Based on the cumulative logit link function with the linear predictors of both proportional odds parameters $\boldsymbol{X}$ and general ordered parameters $\boldsymbol{Z}$, the logit function for PPO is defined as:

$$log\left(\frac{P(Y \le j)}{1 - P(Y \le j)}\right) = log\left(\frac{\pi_1+\ldots+\pi_j}{\pi_{j+1}+\ldots+\pi_J}\right) = \alpha_j + \boldsymbol{X}'\beta + \boldsymbol{Z}'\gamma_j \tag{8}$$

where $\alpha_j$ is the constant variable for response level $j$. $\boldsymbol{X}_j^{'}$ denotes the proportional odds parameters with the same slope coefficient $\beta$. $\boldsymbol{Z}'$ represents the general logit parameters with different coefficient $\gamma_j$. This logit function measures the odds of the probability when $Y$ is to be in or below $j$ versus the probability when $Y$ is higher than j.

The cumulative probability function for PPO can be calculated as:

$$P(Y \le j) = \frac{e^{\alpha_j + X'\beta + Z'\gamma_j}}{1 + e^{\alpha_j + X'\beta + Z'\gamma_j}} \quad (9)$$

To construct a PPO model, the proportional odds model is firstly built to obtain the basic significant variables. Then the ordered logit model is constructed as the base for the PPO model. The Wald Chi-square test and graphical assessment are used to test the parallel lines assumptions. All modelling processes are developed based on the SAS 9.4 software.

### 3.3. Marginal effect

The interpretation of the proportional odds parameter $\beta$ for the value and the sign are independent of the response function. However, the interpretation of the unfixed parameters $\gamma_j$ needs to be combined with the response functions (Derr, 2013). To interpret the results of PPO with category variables, especially when $Y$ is in the middle of the severity levels, marginal effects are commonly used to calculate the changing direction and changing value of severity probability outcomes (Derr, 2013; Li and Fan, 2019). As category variables are set into dummy variables with 1 when event happens and 0 otherwise, the marginal effect is calculated as:

$$E_{X_{ijk}}^{P_{ij}} = \frac{1}{n}\sum_{i=1}^{n} [P_{ij}(X_{ijk} = 1) - P_{ij}(X_{ijk} = 0)] \quad (10)$$

where $P_{ij}$ denotes the probability of case $i$ with severity level $j$, and $P_{ij}$ is calculated when the $k$th independent variable $X_{ijk}$ changes from 0 to 1. The marginal effect for each independent variable $X_{ijk}$ is the average difference value of all observations' $P_{ij}$.

## 4. Data description

Data used in this paper are obtained from the Highway Safety Information System (HSIS), which include 18,346 effective truck-involved crash observations at cross and T-intersections in North Carolina between 2005 and 2017. The most severe crash severity level during the truck-involved crash is set as the response variable and is classified into three sequence levels according to crash rate and severity features. The same classification criterion could be found in (Behnood and Mannering, 2019; Chen et al., 2015; Chen and Chen, 2011; Uddin and Huynh, 2018). The statistic descriptions of the explanatory variables are shown in Table 2. The whole data includes 3.46 % Fatal and Incapacitating injury (FI), 31.42 % Non-incapacitating and Possible injury (NP), and 65.13 % No injury (N). It is noted that the classification of peak hours and non-peak hours (i.e. morning peak hours, noon hours, evening peak hours, and evening hours) for private vehicle crashes could not be directly used for the truck crashes. For the total truck crash frequency, Fig. 1 shows that there has no significant difference between morning/evening peak hours (i.e. 6:00 – 9:00 and 17:00 – 19:00) with the noon hours (i.e. 10:00 – 16:00). Also, it is noted that the fatal and incapacitating crash frequency shows a slight decrease during the noon hours. Hence, this paper categorizes the crash hours into four different periods according to the frequency of the total crashes and FI injuries. Behnood and Mannering (2019) also classified the daytime of the truck crash into the morning period (6:00–11:59 A.M.) and the afternoon period (12:00–5:59 P.M.) by statistical and empirical testing. All explanatory variables are classified into the human, roadway, location, environment, time and control categories, and are dummied (set as 1 when the variable's situation happened and 0 otherwise) according to statistical and classification features. Also, the first category (shown in bold) of each explanatory variable is set as the base in the PPO model. Meanwhile, no injury (N) is select as the reference level for the response variable of the PPO model.

## 5. Results and discussions

### 5.1. Latent class clustering results

The class number in the LCC is set from 1 to 15, and Fig. 2 shows the results of AIC, BIC, CAIC, and the Entropy value for each class number. The results of AIC, BIC, CAIC all decrease correspondingly with the increase of the class numbers. Meanwhile, three criteria's percentage deductions drop to less than 1% after 4 classes. The entropy value for 4 classes reaches 0.9 (which is close to 1) and therefore denotes a good separation of the clustering results. These results suggest that 4 is an applicable class number, and is set as the class number for the truck-involved crash data at cross and T-intersections in this paper.

As shown in Table 3, the variables in bold are selected as the characteristic variables for sub-datasets because of the significant distribution differences between classes (Li and Fan, 2019). It should be noted that classes did not only have different features but also have some overlapped features. According to the proportion results, in class 1, 82.02 % of the crashes occurred in urban areas, 70.47 % in commercial areas, 61.42 % occurred in 1 or 2 lanes, 73.5 % on one way but not divided road, 62.86 % with the signal control, and 65.19 % with the speed limits of less than 35 miles per hour (mph). Hence, class 1 can be denoted as a scenario under which crashes occurred in the urban commercial area, 1 or 2 lanes, one-way and not divided, with signal control and 35 mph speed limits. Class 2 can be defined as a scenario in which urban, 3 or 4 lanes, controlled with signal and speed limits between 36–55 mph. Class 3 can be described as a scenario of rural, 1 or 2 lanes, one-way and not divided with speed limits of between 36–55mph. Class 4 can be specified as a scenario of rural, 3 or 4 lanes principal arterial with signal control, and speed limits between 36–55 mph.

### 5.2. Proportional odds model results

All variables are firstly estimated in four basic proportional odds models (POM) to obtain the basic significant variables for each class specifically, and a 5% confidence level is set as the standard when selecting significant variables. After that, Chi-square tests are used to determine whether the variable is suitable for the general ordered model. Secondly, four general ordered logit models (GOLM) are constructed as the reference for PPO models. Finally, four specific PPO models are constructed including both POM variables and GOLM variables. All log-likelihood values at convergence are large than log-likelihood values for constant only model for each class and are larger compared to those in the basic POM and GOLM. All results are shown in Tables 4a–4d for each class respectively. It is found that different class has different significant variables. In addition, within the same class, the same variables have varied parameters at different severity levels. These results show that significant heterogeneity exist both within and between the classes.

### 5.3. Model comparison and selection

The AIC and the BIC, which could be used to assess the goodness of fit and the simplicity of the model, are commonly used as the model selection criteria, and the model with the smallest value is preferred. In this regard, the PPO model is compared with two commonly used logit models (i.e. Multinomial Logit (MNL) and Generalized Ordered Logit (GOL)), the results of AIC, BIC and the number of significant variables identified are shown in Table 5.

Compared to the GOL model, the PPO model identifies more significant variables in all classes. Meanwhile, the AIC and BIC values of the PPO model are smaller than those of the GOL model. These results further indicate that the PPO model is better than the GOL since the PPO has less AIC and BIC values even though it estimated more parameters. It is also noted that PPO has slightly fewer variables than the

**Table 2**
Statistics of explanatory variables in truck-involved crashes at cross and T- intersections.

| Variable | | Description | Total | Severity level (%) | | |
|---|---|---|---|---|---|---|
| | | Injury type | No. | FI[a] | NI[b] | N[c] |
| Severity | | | 18,346 | 634(3.46 %) | 5764(31.42 %) | 11,948(65.13 %) |
| Driver characteristics | | | | | | |
| Gender | **1** | **Male*** | 17,765 | 620(3.49 %) | 5571(31.36 %) | 11,574(65.15 %) |
| | 2 | Female | 581 | 14(2.41 %) | 193(33.22 %) | 374(64.37 %) |
| Age | 1 | < = **25*** | 1372 | 41(2.99 %) | 455(33.16 %) | 876(63.85 %) |
| | 2 | 26 – 45 | 8506 | 293(3.44 %) | 2673(31.42 %) | 5540(65.13 %) |
| | 3 | 46 – 65 | 7577 | 263(3.47 %) | 2362(31.17 %) | 4952(65.36 %) |
| | 4 | > =66 | 891 | 37(4.15 %) | 274(30.75 %) | 580(65.1 %) |
| Restraint | **1** | **None restraint*** | 473 | 36(7.61 %) | 213(45.03 %) | 224(47.36 %) |
| | 2 | With belt | 17,553 | 588(3.35 %) | 5469(31.16 %) | 11,496(65.5 %) |
| | 3 | Other restraint | 320 | 10(3.13 %) | 82(25.63 %) | 228(71.25 %) |
| Alcohol/drug | **1** | **Not detect*** | 17,997 | 589(3.27 %) | 5592(31.07 %) | 11,816(65.66 %) |
| | 2 | Drunk or drug | 349 | 45(12.89 %) | 172(49.28 %) | 132(37.82 %) |
| Contribution | **1** | **Unknown/none*** | 1280 | 11(0.86 %) | 229(17.89 %) | 1040(81.25 %) |
| | 2 | Disregarded sign | 291 | 31(10.65 %) | 145(49.83 %) | 115(39.52 %) |
| | 3 | Disregarded signals | 706 | 37(5.24 %) | 338(47.88 %) | 331(46.88 %) |
| | 4 | Exceeded speed | 307 | 11(3.58 %) | 130(42.35 %) | 166(54.07 %) |
| | 5 | Failure to reduce speed | 1972 | 25(1.27 %) | 762(38.64 %) | 1185(60.09 %) |
| | 6 | Improper turn | 1341 | 5(0.37 %) | 185(13.8 %) | 1151(85.83 %) |
| | 7 | Improper lane use | 305 | 12(3.93 %) | 74(24.26 %) | 219(71.8 %) |
| | 8 | Improper lane change | 251 | 1(0.4 %) | 31(12.35 %) | 219(87.25 %) |
| | 9 | Failure to yield | 1634 | 51(3.12 %) | 735(44.98 %) | 848(51.9 %) |
| | 10 | Inattention | 1172 | 9(0.77 %) | 238(20.31 %) | 925(78.92 %) |
| | 11 | Improper backing | 541 | 4(0.74 %) | 41(7.58 %) | 496(91.68 %) |
| | 12 | Followed too closely | 115 | 1(0.87 %) | 39(33.91 %) | 75(65.22 %) |
| | 13 | Equipment defect | 177 | 4(2.26 %) | 46(25.99 %) | 127(71.75 %) |
| | 14 | Other | 8254 | 432(5.23 %) | 2771(33.57 %) | 5051(61.19 %) |
| Roadway characteristics | | | | | | |
| No. of lanes | **1** | < = **2*** | 9858 | 396(4.02 %) | 3292(33.39 %) | 6170(62.59 %) |
| | 2 | 3 and 4 | 7164 | 214(2.99 %) | 2118(29.56 %) | 4832(67.45 %) |
| | 3 | > =4 | 1324 | 24(1.81 %) | 354(26.74 %) | 946(71.45 %) |
| Road surface | **1** | **Dry*** | 15,871 | 566(3.57 %) | 4994(31.47 %) | 10,311(64.97 %) |
| | 2 | Wet | 2288 | 66(2.88 %) | 713(31.16 %) | 1509(65.95 %) |
| | 3 | Water, ice, snow, slush | 187 | 2(1.07 %) | 57(30.48 %) | 128(68.45 %) |
| Road curve | **1** | **Straight*** | 17,297 | 581(3.36 %) | 5402(31.23 %) | 11,314(65.4 %) |
| | 2 | Curve | 1049 | 53(5.05 %) | 362(34.51 %) | 634(60.44 %) |
| Road gradient | **1** | **Level*** | 14,556 | 499(3.43 %) | 4498(30.9 %) | 9559(65.67 %) |
| | 2 | Grade | 2842 | 96(3.38 %) | 946(33.29 %) | 1800(63.34 %) |
| | 3 | Hillcrest | 725 | 30(4.14 %) | 230(31.72 %) | 465(64.14 %) |
| | 4 | Bottom | 223 | 9(4.04 %) | 90(40.36 %) | 124(55.61 %) |
| Road pave | **1** | **Concrete*** | 165 | 3(1.82 %) | 60(36.36 %) | 102(61.82 %) |
| | 2 | Smooth asphalt | 12,266 | 403(3.29 %) | 3849(31.38 %) | 8014(65.34 %) |
| | 3 | Coarse asphalt | 5915 | 228(3.85 %) | 1855(31.36 %) | 3832(64.78 %) |
| Road configuration | **1** | **One-way, not divided*** | 371 | 3(0.81 %) | 90(24.26 %) | 278(74.93 %) |
| | 2 | Two-way, not divided | 12,320 | 447(3.63 %) | 3947(32.04 %) | 7926(64.33 %) |
| | 3 | Two-way, divided | 5655 | 184(3.25 %) | 1727(30.54 %) | 3744(66.21 %) |
| Route type | **1** | **Interstate*** | 218 | 1(0.46 %) | 47(21.56 %) | 170(77.98 %) |
| | 2 | Us route | 6448 | 230(3.57 %) | 2096(32.51 %) | 4122(63.93 %) |
| | 3 | Nc route | 5160 | 240(4.65 %) | 1781(34.52 %) | 3139(60.83 %) |
| | 4 | Secondary | 6520 | 163(2.5 %) | 1840(28.22 %) | 4517(69.28 %) |
| Functional class | **1** | **Principal arterial*** | 7529 | 230(3.05 %) | 2318(30.79 %) | 4981(66.16 %) |
| | 2 | Minor arterial | 5182 | 175(3.38 %) | 1607(31.01 %) | 3400(65.61 %) |
| | 3 | Collector | 3875 | 182(4.7 %) | 1344(34.68 %) | 2349(60.62 %) |
| | 4 | Local | 1760 | 47(2.67 %) | 495(28.13 %) | 1218(69.2 %) |
| Location characteristics | | | | | | |
| Rural or urban | **1** | **Rural*** | 9586 | 483(5.04 %) | 3383(35.29 %) | 5720(59.67 %) |
| | 2 | Urban | 8760 | 151(1.72 %) | 2381(27.18 %) | 6228(71.1 %) |
| Locality | **1** | **Farms, woods, pastures*** | 5256 | 302(5.75 %) | 2050(39 %) | 2904(55.25 %) |
| | 2 | Residential | 2812 | 111(3.95 %) | 915(32.54 %) | 1786(63.51 %) |
| | 3 | Commercial | 9966 | 210(2.11 %) | 2707(27.16 %) | 7049(70.73 %) |
| | 4 | Institutional | 144 | 7(4.86 %) | 45(31.25 %) | 92(63.89 %) |
| | 5 | Industrial | 168 | 4(2.38 %) | 47(27.98 %) | 117(69.64 %) |
| Intersection | **1** | **Four-way intersection*** | 11,428 | 405(3.54 %) | 3645(31.9 %) | 7378(64.56 %) |
| | 2 | T-intersection | 6918 | 229(3.31 %) | 2119(30.63 %) | 4570(66.06 %) |
| Terrain | **1** | **Flat*** | 4715 | 221(4.69 %) | 1658(35.16 %) | 2836(60.15 %) |
| | 2 | Rolling | 12,672 | 380(3%) | 3856(30.43 %) | 8436(66.57 %) |
| | 3 | Mountainous | 959 | 33(3.44 %) | 250(26.07 %) | 676(70.49 %) |
| Environment characteristics | | | | | | |
| Weather | **1** | **Clear*** | 13,889 | 497(3.58 %) | 4381(31.54 %) | 9011(64.88 %) |
| | 2 | Cloudy | 3052 | 102(3.34 %) | 922(30.21 %) | 2028(66.45 %) |
| | 3 | Rain | 1179 | 27(2.29 %) | 378(32.06 %) | 774(65.65 %) |
| | 4 | Snow, sleet, hall, freezing rain | 88 | 0(0%) | 29(32.95 %) | 59(67.05 %) |
| | 5 | Fog, smog, smoke | 138 | 8(5.8 %) | 54(39.13 %) | 76(55.07 %) |

(*continued on next page*)

**Table 2** (*continued*)

| Variable | | Description | Total | Severity level (%) | | |
|---|---|---|---|---|---|---|
| Light | **1** | **Daylight*** | 15,795 | 499(3.16 %) | 4848(30.69 %) | 10,448(66.15 %) |
| | 2 | Dusk, down | 533 | 22(4.13 %) | 183(34.33 %) | 328(61.54 %) |
| | 3 | Dark light | 858 | 30(3.5 %) | 277(32.28 %) | 551(64.22 %) |
| | 4 | Dark | 1160 | 83(7.16 %) | 456(39.31 %) | 621(53.53 %) |
| ***Time*** | | | | | | |
| Hour | **1** | **6 am −11 am*** | 7475 | 246(3.29 %) | 2347(31.4 %) | 4882(65.31 %) |
| | 2 | 12 pm - 5 pm | 8462 | 266(3.14 %) | 2570(30.37 %) | 5626(66.49 %) |
| | 3 | 5 pm - 12 am | 1789 | 76(4.25 %) | 600(33.54 %) | 1113(62.21 %) |
| | 4 | 12 am −5 am | 620 | 46(7.42 %) | 247(39.84 %) | 327(52.74 %) |
| Month | 1 | **3 - 5*** | 4641 | 168(3.62 %) | 1458(31.42 %) | 3015(64.96 %) |
| | 2 | 6 - 8 | 4646 | 155(3.34 %) | 1419(30.54 %) | 3072(66.12 %) |
| | 3 | 9 - 11 | 3325 | 124(3.73 %) | 1066(32.06 %) | 2135(64.21 %) |
| | 4 | 12 - 2 | 5734 | 187(3.26 %) | 1821(31.76 %) | 3726(64.98 %) |
| Control characteristics | | | | | | |
| Access | **1** | **No access*** | 14,604 | 543(3.72 %) | 4721(32.33 %) | 9340(63.96 %) |
| | 2 | Partial control | 1780 | 36(2.02 %) | 481(27.02 %) | 1263(70.96 %) |
| | 3 | Full control | 1962 | 55(2.8 %) | 562(28.64 %) | 1345(68.55 %) |
| Traffic control | **1** | **No control*** | 1224 | 32(2.61 %) | 386(31.54 %) | 806(65.85 %) |
| | 2 | Sign | 5244 | 289(5.51 %) | 1793(34.19 %) | 3162(60.3 %) |
| | 3 | Signals | 9861 | 237(2.4 %) | 2821(28.61 %) | 6803(68.99 %) |
| | 4 | Double yellow line, no passing zone | 2017 | 76(3.77 %) | 764(37.88 %) | 1177(58.35 %) |
| Speed limits | **1** | **< = 35 mph*** | 5228 | 84(1.61 %) | 1328(25.4 %) | 3816(72.99 %) |
| | 2 | 36 - 55 mph | 12,890 | 542(4.2 %) | 4388(34.04 %) | 7960(61.75 %) |
| | 3 | 56 - 70 mph | 228 | 8(3.51 %) | 48(21.05 %) | 172(75.44 %) |

Note: bold and noted with * denotes the base of the explanatory variables.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury.
[c] N – No injury. (set as reference level).

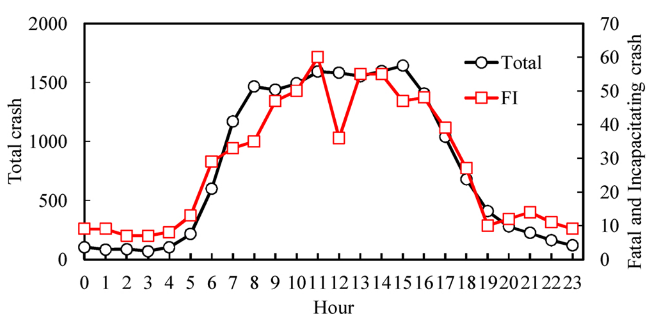


**Fig. 1.** The frequency of the total truck crash and FI injury crash per hour.

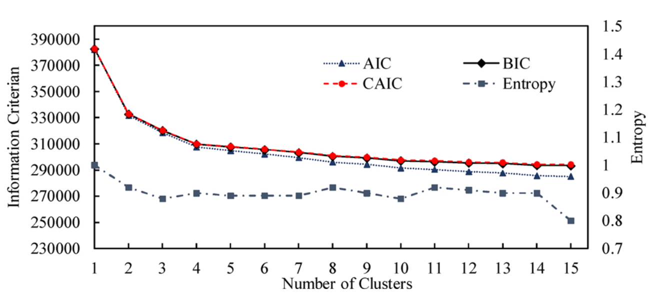


**Fig. 2.** Latent class results of AIC, BIC CAIC, and Entropy value for different class numbers.

MNL model while PPO has smaller BIC than MNL. The reason for this might be that the PPO considers the ordinal nature of the severity level for the crash data, which requires a stricter condition than the unordered MNL model. This finding is in accord with the previous result of the comparison between PPO, GOL, and MNL (Gong et al., 2016).

### 5.4. Marginal effects

The PPO model includes both fixed and unfixed parameters in the model. When the traditional ordered model is used, it is hard to obtain the impact of the explanatory variable on the middle severity level directly. As all explanatory variables are categorized for a better explanation of the scenarios and factors, marginal effects are introduced to better describe the impact of factors on the severity levels. All results are shown in Tables 6a and 6b . Variations of the same explanatory variable's marginal effect are found between different classes. Specific analysis of these impacts and variations, especially for the impact on the FI injury, are investigated as follows:

#### 5.4.1. Human and crash characteristics

As shown in Tables 6a and 6b, factors of human characteristics mainly included driver's gender, whether with the belt, whether drunk or on drugs, and driving behaviors. As shown in class 4, female drivers suffered more fatal and incapacitating (FI) injury (+0.033) and more non-incapacitating and possible (NP) injury (+0.098) compare to male drivers. One possible reason for this might be the physical difference between female and male drivers, and similar results could be found in (Chen and Chen, 2011; Osman et al., 2018). Also, results in classes 2, 3 and 4 indicate that drivers with the belt and other restraint could significantly reduce the probability of FI injury (decreased from -0.022 to -0.062) and NP injury (decreased from -0.11 to -0.15). These results reflect the significant and necessary requirement for restraint such as driving with the belt (see also Dong et al., 2017; Osman et al., 2018). Additionally, results in all classes indicate that drivers who were drunk or on drugs could significantly increase the probability of FI injury (increased from +0.026 to +0.108) and NP injury (increased from +0.043 to +0.231). Drivers could be out of control in intoxicated situation, which may result in dangerous driving behaviors. Similar results can also be reached in (Dong et al., 2017; Liu and Fan, 2020).

This paper categorizes driving behaviors which contribute to the truck-involved crashes into 14 categories based on the characteristics and the proportion of the observations. Compared to the condition of no contributing factors, many driving behaviors are found to be contributed to the increase of truck-involved crash severity at intersections in this paper and relevant literatures, including disregarded signs and signals (Azimi et al., 2020), exceeded speed limits (Khorashadi et al., 2005), failure for speed reduction (Chen and Chen, 2011), improper

**Table 3**
Distributions of featured variables (bold) based on the Latent class clustering.

| Variable | | Meaning | Class 1 | Class 2 | Class 3 | Class 4 |
|---|---|---|---|---|---|---|
| | | Total | 4877(26.58 %) | 5005(27.28 %) | 6087(33.18 %) | 2377(12.96 %) |
| No. of lanes | 1 | 1 or 2 | **2995**(61.42 %) | 418(8.36 %) | **6028**(99.03 %) | 459(19.29 %) |
| | 2 | 3 or 4 | 1774(36.37 %) | **3439**(68.71 %) | 58(0.95 %) | **1851**(77.88 %) |
| Rural/urban | 1 | rural | 877(17.98 %) | 240(4.8 %) | **6082**(99.91 %) | **2377**(99.99 %) |
| | 2 | urban | **4000**(82.02 %) | **4765**(95.2 %) | 5(0.09 %) | 0(0.01 %) |
| Locality | 3 | Commercial | **3437**(70.47 %) | **4609**(92.1 %) | 828(13.61 %) | 1079(45.41 %) |
| Speed limits | 1 | < = 35 mph | **3179**(65.19 %) | 1495(29.87 %) | 444(7.29 %) | 115(4.84 %) |
| | 2 | 36–55 mph | 1698(34.81 %) | **3393**(67.79 %) | **5643**(92.71 %) | **2154**(90.61 %) |
| Functional class | 1 | Principal Arterial | 658(13.5 %) | **4615**(92.21 %) | 389(6.38 %) | **1830**(76.98 %) |
| Road configuration | 2 | One-Way, Not Divided | **3585**(73.5 %) | 1836(36.68 %) | **5981**(98.26 %) | 950(39.96 %) |
| Traffic control | 3 | Signal | **3066**(62.86 %) | **4308**(86.07 %) | 956(15.71 %) | **1508**(63.44 %) |

lane (Chen and Chen, 2011), failure to yield (Anderson and Dong, 2017; Bernard and Mondy, 2016), inattention during driving, following too closely (Bernard and Mondy, 2016), and driving defect equipment (Chen and Chen, 2011). Among all the classes, significantly increasing severity level could be found in class 3, and class 3 can be referred to as a scenario of rural, 1 or 2 lanes, one-way and not divided with speed limits of between 36–55mph. Ignored signs, improper lane use, too closely car following, ignored signals, and failure to yield are the top five most contributing behaviors that result in a probability increase of the FI injury by 0.326, 0.215, 0.212, 0.174 and 0.138, respectively. Since these driving behaviors that contribute to the injury severity are mainly caused by the wrong operation and violation of traffic laws of the drivers, there is a need to strengthen the traffic laws and drivers' education, such as raising the penalties for running the red light and ignoring the signs.

Some behaviors also result in the deduction of the crash severity, such as improper turning (Chen et al., 2015; Chen and Chen, 2011; Uddin and Huynh, 2018), improper lane changing (Hassan et al., 2015; Khorashadi et al., 2005), and improper backing. As shown in class 2, improper backing, turning and lane-changing could reduce the FI injury by -0.014, -0.01 and -0.008, and reduce the NP injury by -0.174, -0.116 and -0.088 respectively. Since these driving behaviors are more likely to be manipulated under the low-speed circumstances and more likely to cause side collisions, these factors are less likely to result in FI and NP injury for truck-involved crashes at intersections.

### 5.4.2. Roadway and location characteristics

Factors of roadway and location characteristics mainly included road gradient, route type, road class, intersection type, urban or rural, land use type and terrain features. Compared to the level road in class 1, gradient road increases +0.003 of the FI injury and +0.036 of the NP injury. Similar results can be reached in (Azimi et al., 2020; Chen et al., 2015). Meanwhile, the bottom segment increases the truck-involved crash severity level by +0.015 of the FI injury and +0.125 of the NP injury. Compared to the interstate route, truck-involved crashes at intersections connected with US (national route), NC (state route) and secondary route all added to the severity level in classes 3 and 4. For example, in class 4, the US route, NC route, and secondary route increase the FI injury by +0.078, +0.08, and +0.054, respectively. Meanwhile, compared to the principal road, crashes occur in the minor arterial road increased FI injury by +0.005 and +0.026 of the and NP injury in class 3.

Compared to the crossroad scenario, truck crashes that occurred at the T-intersections had lower severity levels both in classes 1 and 3. For example, T-intersection reduces -0.009 and -0.063 of the FI and NP injury respectively in class 3. This might be caused by the decrease of

**Table 4a**
PPO model's significant variable coefficients for class 1..

| Level | | All level | | FI[a] | | NP[b] | |
|---|---|---|---|---|---|---|---|
| Variable | Interpretation | Coef. | z-value | Coef. | z-value | Coef. | z-value |
| Intercept | | | | −5.3675** | −15.603 | −1.2893** | −8.929 |
| alcflag_2 | Drink or drug | | | 1.2922** | 2.646 | 1.05** | 3.983 |
| contrib1_2 | Disregarded sign | 1.2612** | 3.594 | | | | |
| contrib1_3 | Disregarded signals | | | 2.1348** | 3.624 | 1.298** | 7.255 |
| contrib1_4 | Exceeded speed | 1.1679** | 3.119 | | | | |
| contrib1_5 | Failure to reduce speed | 0.9444** | 7.069 | | | | |
| contrib1_6 | Improper turn | −0.4691** | −2.977 | | | | |
| contrib1_9 | Failed to yield | 1.0097** | 7.457 | | | | |
| contrib1_11 | Improper backing | −1.0571** | −3.392 | | | | |
| contrib1_12 | Followed too closely | 0.8105* | 2.216 | | | | |
| contrib1_14 | Other | | | 2.1468** | 6.531 | 0.6587** | 6.890 |
| rd_grad_2 | Grade | 0.2077* | 2.203 | | | | |
| rd_grad_4 | Bottom | 0.7408* | 2.473 | | | | |
| rururb_2 | Urban | 0.1847* | 2.023 | | | | |
| locality_3 | Commercial | | | −0.6557** | −2.762 | −0.2815** | −3.768 |
| Int_type_2 | T-intersection | −0.1759* | −2.156 | | | | |
| hour_2 | 12 pm-5 pm | −0.1725* | −2.529 | | | | |
| hour_4 | 12 am-5 am | 0.4549* | 2.021 | | | | |
| trf_cntl_3 | Signals | −0.2519** | −3.046 | | | | |
| spd_limt_2 | 36 – 55 mph | 0.1488* | 2.126 | | | | |

Note: No. of observation: 4877; Log-likelihood for constant only: -3106.691; Log-likelihood at convergence: -2914.581; log-likelihood ratio test: $\chi^2 = 384.2 > \chi^2(0.05,23) = 35.17$; Confidence Level: * for 5%, ** for 1%.

[a] FI - Fatal and incapacitating injury.

[b] NP - Non-incapacitating and possible injury (No injury is set as reference level).

**Table 4b**
PPO model's significant variable coefficients for class 2.

| Level Variable | Interpretation | All level Coef. | z-value | FI[a] Coef. | z-value | NP[b] Coef. | z-value |
|---|---|---|---|---|---|---|---|
| Intercept | | | | −4.0682** | −11.088 | −0.1142 | −0.420 |
| alcflag_2 | Drink or drug | | | 1.3791** | 3.423 | 1.2829** | 5.295 |
| drv_rest_2 | With belt | −0.8041** | −3.426 | | | | |
| drv_rest_3 | Other | −0.7384* | −2.415 | | | | |
| contrib1_2 | Disregarded sign | 1.7137** | 3.638 | | | | |
| contrib1_3 | Disregarded signals | | | 2.2575** | 5.841 | 1.5629** | 10.846 |
| contrib1_4 | Exceeded speed | | | 2.1585** | 2.794 | 0.9886** | 3.057 |
| contrib1_5 | Failure to reduce speed | 0.933** | 8.451 | | | | |
| contrib1_6 | Improper turn | −0.5251** | −2.999 | | | | |
| contrib1_8 | Improper lane change | −0.7375* | −2.424 | | | | |
| contrib1_9 | Failed to yield | | | 1.6846** | 3.626 | 0.9359** | 6.054 |
| contrib1_11 | Improper backing | −1.2175** | −2.831 | | | | |
| contrib1_14 | Other | | | 1.5865** | 5.393 | 0.5878** | 6.481 |
| locality_3 | Commercial | −0.3338** | −2.986 | | | | |
| hour_4 | 12 am-5 am | | | 1.0191** | 3.040 | 0.6166** | 3.883 |
| trf_cntl_3 | Signals | −0.2054* | −2.255 | | | | |
| spd_limt_3 | 56−70 mph | −1.0683** | −3.818 | | | | |

Note: No. of observation: 5005; Log-likelihood for constant only: -3416.508; Log-likelihood at convergence: -3208.842; log-likelihood ratio test: $\chi^2 = 415.3 > \chi^2(0.05,22) = 33.92$; Confidence Level: * for 5%, ** for 1%.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury (No injury is set as reference level).

conflict points, especially the vertical crossing points, at the T-intersections. FHWA (2004) also introduced separating four-leg intersection into two T-intersections to improve safety performance. Compared to crashes that took place in rural areas, urban areas could result in +0.003 and +0.032 increase of the FI and NP injury in class 1, a similar result could be drawn from (Park et al., 2017). Also, it is noted that rural areas are also found to result in severer injury in (Ulak et al., 2017), since the lack of emergency resources (or the accessibility to hospitals) is highly related to the injury severity outcomes. Additionally, compared to farm and green land, truck-involved crashes occurred in the commercial land show a reduction of the severity level in all classes (decreased FI injury from -0.007 to -0.011). This might be caused by lower speed limits, entrance time limitation and better transport infrastructures in commercial areas. What's more, compared to crashes that occurred in the flat area, mountainous terrain reduces -0.01 and -0.062 of the FI and NP injury respectively in class 3. This might be caused by the lower travel speed and higher concertation in mountainous terrain.

#### 5.4.3. *Environment and temporal characteristics*

Factors of the environment and time characteristics mainly included raining weather, light condition and traveling period. Compared to clear weather, the raining condition reduces the severity by -0.009 of FI injury and -0.052 of the NP injury in class 3. Similar conclusions could

**Table 4c**
PPO model's significant variable coefficients for class 3.

| Level Variable | Interpretation | All level Coef. | z-value | FI[a] Coef. | z-value | NP[b] Coef. | z-value |
|---|---|---|---|---|---|---|---|
| Intercept | | | | −4.7252** | −15.233 | −1.0314** | −5.073 |
| alcflag_2 | Drink or drug | | | 0.6557* | 2.479 | 0.774** | 4.184 |
| drv_rest_2 | With belt | | | −0.7755** | −3.515 | −0.6977** | −5.130 |
| contrib1_2 | Disregarded sign | | | 3.1717** | 11.504 | 1.9296** | 10.631 |
| contrib1_3 | Disregarded signals | | | 2.2775** | 4.095 | 1.4599** | 4.974 |
| contrib1_4 | Exceeded speed | | | 1.9596** | 4.769 | 1.4629** | 7.886 |
| contrib1_5 | Failure to reduce speed | 1.1555** | 8.178 | | | | |
| contrib1_7 | Improper lane use | | | 2.5347** | 6.864 | 1.0431** | 5.184 |
| contrib1_9 | Failed to yield | 1.6141** | 12.028 | | | | |
| contrib1_10 | Inattention | 0.5118** | 2.635 | | | | |
| contrib1_11 | Improper backing | −0.6965** | −2.818 | | | | |
| contrib1_12 | Followed too closely | 2.4949** | 4.323 | | | | |
| contrib1_13 | Equipment defect | 0.6505* | 2.259 | | | | |
| contrib1_14 | Other | | | 2.4854** | 12.261 | 1.2507** | 10.475 |
| rte_type_2 | US route | 0.3367** | 4.028 | | | | |
| rte_type_3 | NC route | | | 0.6144** | 5.124 | 0.4163** | 6.715 |
| func_cls_2 | Minor arterial | 0.1556* | 2.435 | | | | |
| locality_3 | Commercial | −0.2711** | −3.318 | | | | |
| lnt_type_2 | T-intersection | | | −0.3166** | −2.658 | −0.3671** | −6.509 |
| terrain_3 | Mountainous | −0.3807** | −3.073 | | | | |
| weather1_3 | Rain | −0.3165** | −2.707 | | | | |
| light_4 | dark | 0.1762* | 2.083 | | | | |
| spd_limt_2 | 36−55 mph | 0.2505* | 2.231 | | | | |

Note: No. of observation: 6087; Log-likelihood for constant only: -5116.763; Log-likelihood at convergence: -4738.879; log-likelihood ratio test: $\chi^2 = 755.8 > \chi^2(0.05,31) = 44.98$; Confidence Level: * for 5%, ** for 1%.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury (No injury is set as reference level).

**Table 4d**
PPO model's significant variable coefficients for class 4.

| Level Variable | Interpretation | All level Coef. | z-value | FI[a] Coef. | z-value | NP[b] Coef. | z-value |
|---|---|---|---|---|---|---|---|
| Intercept | | | | −4.5236** | −7.810 | −1.3034** | −3.490 |
| alcflag_2 | Drink or drug | | | 1.3446** | 3.983 | 0.649* | 2.330 |
| drv_rest_2 | With belt | | | −0.9339* | −2.493 | −0.7385** | −2.931 |
| drv_sex_2 | Female | 0.5626* | 2.094 | | | | |
| contrib1_2 | Disregarded signals | 1.5825** | 3.140 | | | | |
| contrib1_3 | Exceeded speed | | | 1.9015** | 4.072 | 1.4063** | 7.839 |
| contrib1_5 | Failure to reduce speed | | | 1.0061* | 1.955 | 0.8146** | 5.082 |
| contrib1_9 | Failed to yield | | | 1.9774** | 4.343 | 1.1062** | 6.511 |
| contrib1_14 | Other | | | 1.9027** | 4.732 | 0.8484** | 6.994 |
| rte_type_2 | US route | 1.121** | 4.224 | | | | |
| rte_type_3 | NC route | 1.2097** | 4.319 | | | | |
| rte_type_4 | Secondary | 1.196** | 3.585 | | | | |
| locality_3 | Commercial | −0.2235** | −2.648 | | | | |
| loc_type_2 | T-intersection | −0.217* | −2.406 | | | | |
| hour_4 | 12 am − 5 am | 0.4057* | 2.027 | | | | |

Note: No. of observation: 2377; Log-likelihood for constant only: -2037.936; Log-likelihood at convergence: -1941.496; log-likelihood ratio test: $\chi^2 = 192.8 > \chi^2(0.05,20) = 31.41$; Confidence Level: * for 5%, ** for 1%.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury (No injury is set as reference level).

**Table 5**
Model comparison and selection for four latent classes.

| | | Multinomial Logit | | | Generalized Ordered Logit | | | Proportional Odds Logit | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Class | NO. of obs. | AIC | BIC | NO. of var. | AIC | BIC | NO. of var. | AIC | BIC | NO. of var. |
| 1 | 4877 | 5926 | 6088 | 25 | 5904 | 6040 | 21 | 5879 | 6041 | 25 |
| 2 | 5005 | 6458 | 6640 | 28 | 6472 | 6589 | 18 | 6466 | 6622 | 24 |
| 3 | 6087 | 9541 | 9769 | 34 | 9602 | 9769 | 25 | 9544 | 9765 | 33 |
| 4 | 2377 | 3929 | 4062 | 23 | 3937 | 4029 | 16 | 3927 | 4054 | 22 |

**Table 6a**
Marginal effects of explanatory variables for class 1 and class 2.

| Variable | Description | Class 1 | | | Class 2 | | |
|---|---|---|---|---|---|---|---|
| | Severity | FI[a] | NP[b] | N[c] | FI[a] | NP[b] | N[c] |
| Human and Crash Characteristics | | | | | | | |
| alcflag_2 | Drink or drug (vs. Not detect) | 0.034 | 0.164 | −0.198 | 0.051 | 0.231 | −0.282 |
| drv_rest_2 | With belt (vs. None) | | | | −0.022 | −0.150 | 0.172 |
| drv_rest_3 | Other restraint | | | | −0.011 | −0.116 | 0.127 |
| contrib1_2 | Disregarded signs (vs. Unknown) | 0.034 | 0.201 | −0.235 | 0.076 | 0.295 | −0.371 |
| contrib1_3 | Disregarded signals | 0.085 | 0.155 | −0.240 | 0.123 | 0.219 | −0.342 |
| contrib1_4 | Exceeded speed | 0.030 | 0.188 | −0.218 | 0.118 | 0.097 | −0.215 |
| contrib1_5 | Failure to reduce speed | 0.021 | 0.153 | −0.174 | 0.027 | 0.173 | −0.200 |
| contrib1_6 | Improper turn | −0.006 | −0.082 | 0.088 | −0.008 | −0.088 | 0.096 |
| contrib1_8 | Improper lane change | | | | −0.010 | −0.116 | 0.126 |
| contrib1_9 | Failed to yield | 0.023 | 0.163 | −0.186 | 0.067 | 0.134 | −0.202 |
| contrib1_11 | Improper backing | −0.010 | −0.175 | 0.185 | −0.014 | −0.174 | 0.188 |
| contrib1_12 | Followed too closely | 0.017 | 0.107 | −0.125 | | | |
| contrib1_14 | Other | 0.033 | 0.417 | −0.450 | 0.036 | 0.080 | −0.116 |
| Roadway and Location Characteristics | | | | | | | |
| rd_grad_2 | Grade (vs. Level) | 0.003 | 0.036 | −0.039 | | | |
| rd_grad_4 | Bottom | 0.015 | 0.125 | −0.140 | | | |
| rururb_2 | Urban (vs. Rural) | 0.003 | 0.032 | −0.035 | | | |
| locality_3 | Commercial (vs. Green lands) | −0.010 | −0.043 | 0.053 | −0.007 | −0.061 | 0.068 |
| Int_type_2 | T-intersection (vs. Cross) | −0.003 | −0.030 | 0.033 | | | |
| Environment and Temporal Characteristics | | | | | | | |
| hour_2 | 12 pm – 5 pm (vs. 6 am-11 am) | −0.003 | −0.030 | 0.033 | | | |
| hour_4 | 12 am – 5 am | 0.008 | 0.078 | −0.086 | 0.032 | 0.099 | −0.131 |
| Traffic Control Characteristics | | | | | | | |
| trf_cntl_3 | Signals (vs. None) | −0.004 | −0.044 | 0.048 | −0.004 | −0.037 | 0.041 |
| spd_limt_2 | 36 − 55 mph (vs. < = 35) | 0.002 | 0.026 | −0.028 | | | |
| spd_limt_3 | 50 − 70 mph | | | | −0.013 | −0.155 | 0.168 |

Note: The base is set at 1 st category, which is shown in the brackets.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury.
[c] N - No injury.

**Table 6b**
Marginal effects of the explanatory variables for class 3 and class 4.

| Variable | Description | Class 3 | | | Class 4 | | |
|---|---|---|---|---|---|---|---|
| | Severity | FI[a] | NP[b] | N[c] | FI[a] | NP[b] | N[c] |
| Human and Crash Characteristics | | | | | | | |
| alcflag_2 | Drink or drug (vs. Not detect) | 0.026 | 0.140 | −0.166 | 0.108 | 0.043 | −0.151 |
| drv_rest_2 | With belt (vs. None) | −0.032 | −0.117 | 0.148 | −0.062 | −0.110 | 0.172 |
| drv_sex_2 | Female (vs. Male) | | | | 0.033 | 0.098 | −0.130 |
| contrib1_2 | Disregarded signs (vs. Unknown) | 0.326 | 0.078 | −0.404 | 0.140 | 0.209 | −0.348 |
| contrib1_3 | Disregarded signals | 0.174 | 0.139 | −0.313 | 0.181 | 0.136 | −0.317 |
| contrib1_4 | Exceeded speed | 0.138 | 0.176 | −0.314 | 0.069 | 0.118 | −0.187 |
| contrib1_5 | Failure to reduce speed | 0.057 | 0.187 | −0.244 | | | |
| contrib1_7 | Improper lane use | 0.215 | 0.009 | −0.224 | | | |
| contrib1_9 | Failed to yield | 0.095 | 0.247 | −0.342 | 0.182 | 0.071 | −0.254 |
| contrib1_10 | Inattention | 0.019 | 0.086 | −0.106 | | | |
| contrib1_11 | Improper backing | −0.016 | −0.108 | 0.123 | | | |
| contrib1_12 | Followed too closely | 0.212 | 0.282 | −0.494 | | | |
| contrib1_13 | Equipment defect | 0.026 | 0.112 | −0.138 | | | |
| contrib1_14 | Other | 0.090 | 0.154 | −0.244 | 0.099 | 0.092 | −0.191 |
| Roadway and Location Characteristics | | | | | | | |
| rte_type_2 | US route (vs. Inter-sate) | 0.011 | 0.058 | −0.069 | 0.078 | 0.172 | −0.250 |
| rte_type_3 | NC route | 0.021 | 0.064 | −0.085 | 0.080 | 0.184 | −0.264 |
| rte_type_4 | Secondary | | | | 0.054 | 0.186 | −0.241 |
| func_cls_2 | Minor arterial (vs. Principal) | 0.005 | 0.026 | −0.031 | | | |
| locality_3 | Commercial (vs. Green land) | −0.009 | −0.046 | 0.055 | −0.011 | −0.040 | 0.051 |
| Int_type_2 | T-intersection (vs. Cross) | −0.009 | −0.063 | 0.072 | −0.010 | −0.039 | 0.049 |
| terrain_3 | Mountainous (vs. Flat) | −0.010 | −0.062 | 0.071 | | | |
| Environment and Temporal Characteristics | | | | | | | |
| hour_4 | 12 am −5 am (vs. 6 am −11 am) | | | | 0.022 | 0.071 | −0.094 |
| weather1_3 | Rain (vs. Clear) | −0.009 | −0.052 | 0.060 | | | |
| light_4 | Dark (vs. Daylight) | 0.006 | 0.030 | −0.036 | | | |
| Traffic Control Characteristics | | | | | | | |
| spd_limit_2 | 36 - 55 mph (vs. < = 35) | 0.008 | 0.043 | −0.050 | | | |

Note: The base is set at 1 st category, which is shown in the brackets.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury.
[c] N - No injury.

also be drawn from (Khorashadi et al., 2005; Liu and Fan, 2020). This might be caused by the high alertness of the driver and low travel speed during the raining weather. In addition, compared to daylight condition, driving in the dark (without roadside light) adds up to +0.006 and +0.03 of the FI and NP injury in class 3, a similar result could be referred to (Uddin and Huynh, 2018). For the traveling period, compared to the morning period from 6:00 am to 11:59 am, the afternoon from 12:00 pm to 5:59 pm shows a slight decrease of the severity in class 1 (decreased FI injury for -0.003 and NP injury for -0.03). However, for those traveling during the late-night period between 0:00 am and 5:59 am, marginal results show an increase of FI injury (from +0.008 to +0.032) and NP injury (from +0.071 to +0.1) in classes 1, 2, and 4. Similar conclusions could be drawn from (Li and Fan, 2019; Uddin and Huynh, 2017), which might be caused by the dark environment, high speed and fatigue of the truck driver in the middle night.

#### 5.4.4. Traffic control characteristics

Factors of traffic control characteristics mainly included signal control and speed limits. Compared to those without control, locations controlled by signals decrease -0.004 to -0.004 of the FI injury and -0.037 to -0.044 of the NP injury respectively in classes 1 and 2. Similar results could be found in (Anderson and Dong, 2017; Chen and Chen, 2011). Compared to the speed limits within 35 mph, marginal effects show heterogeneous results. Truck-involved crash severity increase in classes 1 and 3 under speed limits within 36–55 mph, this is also in line with (Liu and Fan, 2020). However, results in class 2 show a decrease of the FI injury by -0.013 and NP injury by -0.155 under speed limits within 50–70 mph, which indicates heterogeneous impacts of the factors under different classes. This reduction might be caused by truck's stable driving behaviors in high-speed road, i.e., keeping constant speed in the right side of the road and rarely changing lanes.

## 6. Conclusions

This study dedicates to explore and investigate the factors that impact truck-involved crash severities at cross and T-intersections. To mitigate the heterogeneity within the North Carolina crash data between 2005 and 2017 from the HSIS database, a latent class clustering that maximizes the heterogeneity between clusters is implemented. Four partial proportional odds models are constructed considering the ordinal nature of the severity levels. Results show different significant variables and coefficient values between classes, and some variables in the same class also have different coefficients estimated in different severity levels. These results indicate that significant heterogeneities existed both within and between classes, and reveal the superiority of this combined methodology which could obtain more specific factors and accurate coefficients for different scenarios. Marginal effects are calculated for better interpreting the impacts of factors, especially for those providing further insights into reducing the severity of the truck-involved crash at intersections.

Based on the significant variables obtained from basic proportional odds models, both fixed and unfixed parameters in different severity levels of the PPO models are analyzed under specific latent class scenarios. In addition, heterogeneous results are found between different classes and severity levels. In summary, factors increase the severity levels including female driver, drunk or on drugs, gradient road, bottom segment, US (national route), NC (state route), secondary route, minor arterial, urban areas, driving in the dark without roadside light, late-night period between 0:00 am and 5:59 am, and speed limits within 36–55 mph. Meanwhile, other factors decrease the severity level including with belt and other restraint, T-intersections, commercial area, mountainous terrain, raining, afternoon from 12:00 to 17:59, signals control, and speed limits within 50–70 mph. For driver behaviors, disregarded signs and signals, exceeded speed limits, failure for speed reduction, improper lane use, failure to yield, inattention during driving, following too closely, and driving defect equipment are found to increase the crash severity. In contrast, driving behaviors under low speed and mainly cause indirect crash such as improper turning, lane changing, and backing are found to reduce the severity.

Overall, this study explores the factors and their influence on truck-involved crash severity at intersections. Ordinal features of the severity level and heterogeneity of the factors within severity level and between classes are considered. Many results, such as driving behaviors, can arouse the attention to amend the traffic laws, modify the transportation facilities and strengthen driving education. In practice, it is better to consider both safety impacts and traffic volume at the intersections when determining whether or not to set signals. Drunk driving and without belt are found to result in severer injury outcomes in all classes, which requires stricter roadside belt inspection, alcohol test, and penalty. Also, traditional speed limits are mainly determined by the road level and function, and there is a need to set dynamic speed limits considering the traffic volume and safety conditions of the roadways. Additionally, some site-specific characteristics, such as land use and roadway characteristics, could help to develop specific countermeasures for the roadway alignment modification.

The transferability of some heterogeneous impacts of the factors across the classes and site-specific characteristics also need to be further investigated. It is noted that the class 1 and class 2 are mainly located in commercial areas, thus the factors identified are specific to these areas. Also, 70 % of the crashes in this study occurred in the rolling areas, which is the main terrain type in NC. The spatial impacts of the surrounding areas on the crash injury severity also need to be further investigated. In summary, future research could explore the cause of the heterogeneities within variables and between classes, and study the correlation between different factors, as well as include external data for a better understanding of the impact and the transferability of the factors.

## CRediT authorship contribution statement

**Li Song:** Conceptualization, Methodology, Software, Validation, Writing - original draft, Visualization, Investigation. **Wei Fan:** Data curation, Supervision, Writing - review & editing.

## Declaration of Competing Interest

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors want to express their deepest gratitude for the financial support from the United States Department of Transportation, University Transportation Center through the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE) at The University of North Carolina at Charlotte (Grant Number: 69A3551747133).

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2020.105638.

## References


Ahmed, M.M., Franke, R., Ksaibati, K., Shinstine, D.S., 2018. Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. Accid. Anal. Prev. 117 (May), 106–113. https://doi.org/10.1016/j.aap.2018.04.011. 2015.

Al-Bdairi, N.S.S., Hernandez, S., 2017. An empirical analysis of run-off-road injury severity crashes involving large trucks. Accid. Anal. Prev. 102, 93–100. https://doi.org/10.1016/j.aap.2017.02.024.

Al-Bdairi, N.S.S., Hernandez, S., Anderson, J., 2018. Contributing factors to run-off-road crashes involving large trucks under lighted and dark conditions. J. Transp. Eng. Part A Syst. 144 (1), 1–9. https://doi.org/10.1061/JTEPBS.0000104.

Anderson, J.C., Dong, S., 2017. Heavy-vehicle driver injury severity analysis by time of week: a mixed logit approach using HSIS crash data. ITE J. (Institute Transp. Eng. 87 (9), 41–49.

Azimi, G., Rahimi, A., Asgari, H., Jin, X., 2020. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. Accid. Anal. Prev. 135 (April), 105355. https://doi.org/10.1016/j.aap.2019.105355. 2019.

Behnood, A., Mannering, F., 2019. Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes. Anal. Methods Accid. Res. 23, 100102. https://doi.org/10.1016/j.amar.2019.100102.

Bernard, J.M., Mondy, C.M., 2016. Correlation of driver gender with injury severity in large truck crashes in Missouri. Transp. Res. Rec. 2585 (2585), 49–58. https://doi.org/10.3141/2585-06.

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

Chang, L.Y., Chien, J.T., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Saf. Sci. 51 (1), 17–22. https://doi.org/10.1016/j.ssci.2012.06.017.

Chang, F., Xu, P., Zhou, H., Chan, A.H.S., Huang, H., 2019. Investigating injury severities of motorcycle riders: a two-step method integrating latent class cluster analysis and random parameters logit model. Accid. Anal. Prev. 131, 316–326. https://doi.org/10.1016/j.aap.2019.07.012.

Chen, F., Chen, S., 2011. Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. Accid. Anal. Prev. 43 (5), 1677–1688. https://doi.org/10.1016/j.aap.2011.03.026.

Chen, C., Zhang, G., Tian, Z., Bogus, S.M., Yang, Y., 2015. Hierarchical Bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. Accid. Anal. Prev. 85, 186–198. https://doi.org/10.1016/j.aap.2015.09.005.

Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z., 2016. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accid. Anal. Prev. 90, 128–139. https://doi.org/10.1016/j.aap.2016.02.011.

Derr, B., 2013. Ordinal Response Modeling With the LOGISTIC Procedure. SAS Glob. Forum, pp. 1–20.

Dong, C., Dong, Q., Huang, B., Hu, W., Nambisan, S.S., 2017. Estimating factors contributing to frequency and severity of large truck-involved crashes. J. Transp. Eng. Part A Syst. 143 (8), 1–9. https://doi.org/10.1061/JTEPBS.0000060.

FHWA, 2004. Signalized Intersections: Informational Guide. https://doi.org/10.1109/CVPR.2005.217.

Gong, L., Fan, W., Washing, E.M., 2016. Modeling severity of single vehicle run-off-road crashes in rural areas : model comparison and selection. Can. J. Civ. Eng. 43 (6), 493–503. https://doi.org/10.1139/cjce-2015-0449.

Hassan, H.M., Albusaeedi, N.M., Garib, A.M., Al-Harthei, H.A., 2015. Exploring the nature and severity of heavy truck crashes in Abu Dhabi, United Arab Emirates. Transp. Res. Rec. 2517, 1–9. https://doi.org/10.3141/2517-01.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108 (August), 27–36. https://doi.org/10.1016/j.aap.2017.08.008.

Islam, S., Jones, S.L., Dye, D., 2014. Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama. Accid. Anal. Prev. 67, 148–158. https://doi.org/10.1016/j.aap.2014.02.014.

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. Accid. Anal. Prev. 37 (5), 910–921. https://doi.org/10.1016/j.aap.2005.04.009.

Lanza, S.T., Rhoades, B.L., 2013. Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. Prev. Sci. 14 (2), 157–168. https://doi.org/10.1007/s11121-011-0201-1.

Lanza, S.T., Collins, L.M., Lemmon, D.R., Schafer, J.L., 2007. PROC LCA: a SAS procedure for latent class analysis. Struct. Equ. Model. 14 (4), 671–694. https://doi.org/10.1080/10705510701575602.

Lemp, J.D., Kockelman, K.M., Unnikrishnan, A., 2011. Analysis of large truck crash severity using heteroskedastic ordered probit models. Accid. Anal. Prev. 43 (1), 370–380. https://doi.org/10.1016/j.aap.2010.09.006.

Li, Y., Fan, W., 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. Accid. Anal. Prev. 131 (April), 284–296. https://doi.org/10.1016/j.aap.2019.07.008.

Liu, P., Fan, W., 2019. Modeling head-on crash severity on ncdot freeways: a mixed logit model approach. Can. J. Civ. Eng. 46 (6), 322–328. https://doi.org/10.1139/cjce-2018-0262.

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. Accid.

Anal. Prev. 135 (November), 105388. https://doi.org/10.1016/j.aap.2019.105388. 2019.
Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22. https://doi.org/10.1016/j.amar.2013.09.001.
McCullagh, P., 1980. Regression models for ordinal data. J. R. Stat. Soc. Ser. B 42 (2), 109–127. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x.
McLachlan, G.J., Peel, D., 2004. Finite Mixture Models. John Wiley & Sonshttps://doi.org/10.1146/annurev-statistics-031017-100325.
Mohamed, M.G., Saunier, N., Miranda-Moreno, L.F., Ukkusuri, S.V., 2013. A clustering regression approach: a comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal. Canada. Saf. Sci. 54, 27–37. https://doi.org/10.1016/j.ssci.2012.11.001.
Naik, B., Tung, L.W., Zhao, S., Khattak, A.J., 2016. Weather impacts on single-vehicle truck crash injury severity. J. Safety Res. 58, 57–65. https://doi.org/10.1016/j.jsr.2016.06.005.
Nitsche, P., Thomas, P., Stuetz, R., Welsh, R., 2017. Pre-crash scenarios at road junctions: a clustering method for car crash data. Accid. Anal. Prev. 107, 137–151. https://doi.org/10.1016/j.aap.2017.07.011.
Osman, M., Paleti, R., Mishra, S., Golias, M.M., 2016. Analysis of injury severity of large truck crashes in work zones. Accid. Anal. Prev. 97, 261–273. https://doi.org/10.1016/j.aap.2016.10.020.
Osman, M., Mishra, S., Paleti, R., 2018. Injury severity analysis of commercially-licensed drivers in single-vehicle crashes: accounting for unobserved heterogeneity and age group differences. Accid. Anal. Prev. 118, 289–300. https://doi.org/10.1016/j.aap.2018.05.004.
Pahukula, J., Hernandez, S., Unnikrishnan, A., 2015. A time of day analysis of crashes involving large trucks in urban areas. Accid. Anal. Prev. 75, 155–163. https://doi.org/10.1016/j.aap.2014.11.021.
Park, H.C., Kim, D.K., Kho, S.Y., Park, P.Y., 2017. Cross-classified multilevel models for severity of commercial motor vehicle crashes considering heterogeneity among companies and regions. Accid. Anal. Prev. 106 (April), 305–314. https://doi.org/10.1016/j.aap.2017.06.009.
Peterson, B., Harrell, F.E., 1990. Partial proportional odds models for ordinal response variables. Appl. Stat. 39 (2), 205. https://doi.org/10.2307/2347760.
Rahimi, A., Azimi, G., Asgari, H., Jin, X., 2019. Clustering approach toward large truck crash analysis. Transp. Res. Rec. 2673 (8), 73–85. https://doi.org/10.1177/0361198119839347.
Sasidharan, L., Menéndez, M., 2014. Partial proportional odds model - an alternate choice for analyzing pedestrian crash injury severities. Accid. Anal. Prev. 72, 330–340. https://doi.org/10.1016/j.aap.2014.07.025.
Uddin, M., Huynh, N., 2017. Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. Accid. Anal. Prev. 108 (March), 44–55. https://doi.org/10.1016/j.aap.2017.08.009.
Uddin, M., Huynh, N., 2018. Factors influencing injury severity of crashes involving HAZMAT trucks. Int. J. Transp. Sci. Technol. 7 (1), 1–9. https://doi.org/10.1016/j.ijtst.2017.06.004.
Ulak, M.B., Kocatepe, A., Ozguven, E.E., Horner, M.W., Spainhour, L., 2017. Geographic information system–based spatial and statistical analysis of severe crash hotspot accessibility to hospitals. Transp. Res. Rec. 2635 (1), 90–97. https://doi.org/10.3141/2635-11.
USDOT, 2019. Large Truck and Bus Crash Facts 2017. Federal Motor Carrier Safety Administration.
Zhu, X., Srinivasan, S., 2011. Modeling occupant-level injury severity: an application to large-truck crashes. Accid. Anal. Prev. 43 (4), 1427–1437. https://doi.org/10.1016/j.aap.2011.02.021.
Zou, W., Wang, X., Zhang, D., 2017. Truck crash severity in New York city: an investigation of the spatial and the time of day effects. Accid. Anal. Prev. 99, 249–261. https://doi.org/10.1016/j.aap.2016.11.024.