International Journal of Transportation Science and Technology 10 (2021) 110–120

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

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

Exploring truck driver-injury severity at intersections considering heterogeneity in latent classes: A case study of North Carolina

Li Song, Wei (David) Fan *

USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE), Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, Charlotte, NC 28223-0001, United States

ARTICLE INFO

Article history:
Received 23 July 2020
Received in revised form 1 November 2020
Accepted 18 December 2020
Available online 31 December 2020

Keywords:
Truck-involved crashes
Intersection, severity analysis
Latent class analysis
Ordered logit

ABSTRACT

The fatal rate of truck-involved crashes is increasing and crashes become more severe than passenger vehicles in recent years. Much research has been dedicated to exploring the truck crash factors while scarce research focused on the intersection scenarios. This study investigates the factors that affect the severity level of truck-involved crashes at cross- and T-intersections. Due to the unobserved heterogeneity inherent in crash data, latent class analysis is firstly conducted to divide the crash dataset into relatively homogeneous clusters. Considering the ordinal feature of the severities, general ordered logit models are subsequently developed to further explore the specific factors within each cluster. This study uses the North Carolina's truck-involved crash at intersection data during 2005 to 2017 from the Highway Safety Information System (HSIS). The estimated parameters and associated marginal effects are combined to interpret the impact of the significant variables within specific clusters. Many factors are found to contribute to the severities, and T-intersection is found to be safer than cross-intersection. For driving behaviors, followed too closely, disregarded signs, disregarded signals, failed to yield, and exceeded speed are found to be top five factors that increase the crash severity at intersections. These results indicate that distraction and speed limits violation always result in severe injury for humans involved in the truck crashes at the intersections. The results of this research provide more reliable analysis for the impact factors of truck-involved crashes at intersections to engineering practitioners and researchers.

1. Introduction

The truck-involved crashes have suffered more severe injury compared to passenger vehicles crashes in recent years. According to U.S. Department of Transportation (USDOT) statistics, from 2016 to 2017, the number of large truck-involved fatal crashes increased 9.6 % from 4251 to 4657, while passenger vehicle in fatal crashes has decreased 1.4%. Meanwhile, the fatal crash rate per 100 million vehicle miles involving large trucks reached 1.42 and increased 5.2% from 2016 to 2017, and it was 1.38 times compared to the passenger vehicle (USDOT, 2019). In this case, a large number of research stud-

Peer review under responsibility of Tongji University and Tongji University Press.
* Corresponding author.
E-mail address: wfan7@uncc.edu (W. (David) Fan).

https://doi.org/10.1016/j.ijtst.2020.12.006
2046-0430/

ies have been dedicated to investigations into the factors that impact the truck-involved crash severity. In considering that different crash scenarios were a result of different influence factors, truck-involved crashes were specifically drawn into the circumstances of human characteristics (Bernard and Mondy, 2016; Osman et al., 2018), roadway attributes (Ahmed et al., 2018; Wu et al., 2019a), location (Khorashadi et al., 2005; Uddin and Huynh, 2017), crash characteristics (Azimi et al., 2020), vehicle characteristics (Uddin and Huynh, 2018), time (Anderson and Dong, 2017; Behnood and Mannering, 2019) and environment (Uddin and Huynh, 2017). Intersections have more complex traffic circumstances and may cause more severe and frequent crash injury compared to roadways (FHWA, 2004; Zhu and Srinivasan, 2011). Some studies mentioned intersection/non-intersection (Wu et al., 2019b; Zhu and Srinivasan, 2011) or signal/non-signal control (Anderson and Dong, 2017; Chen and Chen, 2011) as location or control type variables. However, scarce research on truck-involved crashes that specifically considers the cross- and T-intersection scenarios has been conducted. Hence, it is important to explore the factors that contribute to the truck-involved severity at cross- and T-intersections.

Even though many studies investigated the truck-involved crashes under specific conditions, there still remain many unobserved factors that impact the crash severity and result in heterogeneity within the dataset. Research neglected the data's heterogeneity might generate wrong parameter estimations and conclusions (Song et al., 2020). Recently, many clustering methods, such as k-means method (Mohamed et al., 2013), support vector method (Chen et al., 2016; Mokhtarimousavi, 2019) and latent class analysis (Li and Fan, 2019; Liu and Fan, 2020; Mohamed et al., 2013; Sivasankaran and Balasubramanian, 2020), were used to minimize the heterogeneity within the crash severity dataset. Mohamed et al. (2013) combined multinomial logit and ordered probit model with k-means and latent class analysis method, and their results confirmed that clustering the crash severity dataset into homogeneous clusters helps better identify factors that would otherwise have been hidden without data segmentation. Since latent class analysis (LCA) is a model-based method which could guarantee the homogeneity within the cluster based on statistical criterions, many traffic crash severity studies recently implemented LCA for data segmentation (Fountas et al., 2018; Liu and Fan, 2020). Hence, this paper uses latent class analysis to separate the dataset into groups which have the largest homogeneity within the same group.

For truck-involved crash severities, which are typically discrete in nature, a variety of ordered or unordered logit/probit methods were implemented to conduct severity impact factors analysis, such as binary/multinomial logit model (Khorashadi et al., 2005), Bayesian binary/multinomial logit model (Dong et al., 2017), mixed logit model (Anderson and Dong, 2017; Chen and Chen, 2011; Hou et al., 2019), ordered logit models (Osman et al., 2016), ordered probit model (Uddin and Huynh, 2018), partial proportional odds model (Li and Fan, 2019) and random parameter ordered logit model (Azimi et al., 2020).

For unordered model, Dong et al. (2017) used Bayesian multinomial logit and negative binomial model to investigate factors that affect large truck–involved crash frequency and severity based on 2006 to 2010 Tennessee data. Seat belt usage, light condition, and terrain type were found to have significant effects only on the crash severity. In order to explore the heterogeneity within the data, Anderson and Dong (2017) applied a mixed logit model to estimate heavy-vehicle crash factors based on 2004 to 2014 Minnesota data from HSIS database. Results showed that time-of-week needs to be considered separately for safety analyses. Uddin and Huynh (2017) employed a mixed logit model to study the impacts of different lighting conditions on truck-involved crash severity in Ohio's rural and urban areas during 2009 to 2013 from HSIS. The heterogeneous results indicated the importance of dividing data into different scenarios for more specific and accurate results.

The ordinal nature of crash severity (which usually increases from non-injury to fatal) violates the independence assumption of the response variable for unordered model (Derr, 2013). Hence, many ordinary models were applied for better investigating the impact of the severity level. Chen et al. (2015) developed a hierarchical Bayesian random intercept model to analyze the factors affecting rural truck-involved crash severity in New Mexico from 2010 to 2011. Results indicated the existence of cross-level interaction effects between severity levels. Hassan et al. (2015) developed ordered probit and structural equation models to investigate truck crash severity' factors and to study the impact of truck road based on the Abu Dhabi data between 2007 and 2013. Results indicated that the likelihood of truck crashes involving fatalities was 35% higher on truck roads than that on mixed-vehicle roads. Azimi et al. (2020) constructed a random parameter ordered logit model to detect potential sources of heterogeneity within large truck rollover crashes based on Florida's 2007 to 2016 data. Results showed significant variation within observations and the heterogeneous impacts on the severities. Osman et al. (2016) compared ordered models with unordered models to analyze truck crash severity in work zones of Minnesota from HSIS. Results showed that ordered logit has better model fitness than unordered models. By considering the ordinal feature of the crash severities in the ordered model and investigating the heterogeneity characteristics of the crash data, this paper combines the LCA with ordered logit model to further explore the factors that affect the truck-involved crash severity.

2. Data description

In this paper, a total number of 18,346 truck-involved crashes data at cross- and T-intersections in North Carolina from 2005 to 2017 are obtained from Highway Safety Information System (HSIS) database after data screening and cleaning. The independent variable, severity, is the most severe injury in any crash and is categorized into three levels according to severity characteristics, proportion and literature (Behnood and Mannering, 2019; Chen et al., 2015; Chen and Chen, 2011; Uddin and Huynh, 2018). The data consist of 3.5% fatal and incapacitating injury (FI), 31.4% non-incapacitating and possible injury (NP), and 65.1% no injury (N). 24 independent variables are selected based on the human, roadway, location, environment, time and control characteristics from the crash observations and are categorized into total 90 dummy variables. Table 1 shows the description of each variable as well as the number of observations at each severity level.

For dependent variable, the "no injury" is selected as the reference level. Also, for each independent variable, the first category is selected as the base (marked in bold). Dummy variables (0 and 1) are created for all independent variables where 1 represents the appearance of independent variable in a crash record and 0 represents the opposite. All the methodology procedures are developed in the SAS 9.4 software.

3. Methodology

3.1. Latent class analysis

Latent class analysis (LCA) is a statistical clustering method for identifying unobservable, or latent, subgroups among data with categorical and/or continuous variables by calculating the conditional probabilities that variables take on certain values (Chang et al., 2019; Lanza and Rhoades, 2013; Liu and Fan, 2020; McLachlan and Peel, 2004). As the number of the clusters of LCA could be decided by some statistical criterions, there is no need to predefine the number of the clusters. Also, there is no need to standardize different types of variables, including counts, continuous, categorical, and nominal variables (Lanza and Rhoades, 2013).

In the crash data, continuous variables are divided into discrete categories for describing the case more specifically. When classifying the crash dataset with j categorical variables into N classes, the probability of response is defined as follows:

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

where γ_n represents the membership probability for latent class cluster n (n = 1, 2, …, N). Suppose that each crash case i contains J variables, let Y_i be the result of the case i for J categorical variables, andY_i = 1, 2, …r_j. $\rho_{j,r_j|n}^{I(y_j=r_j)}$ indicates the probability that case i has attribute r_j, conditional on latent class membership m. $I(y_j = r_j)$ denotes an indicator function that equals 1 when the $y_j = r_j$ and 0 otherwise (Lanza and Rhoades, 2013; Liu and Fan, 2020).

The appropriate cluster number N could be determined by some statistical criteria which denote the goodness-of-fit of the results. This paper adopts the commonly used Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), Consistent Akaike Information Criterion (CAIC), and entropy-based measure (Li and Fan, 2019). The entropy-based measure is an averaged weighted case's posterior probabilities of membership, ranging between 0 and 1, and closing to 1 indicates a better clustering (McLachlan and Peel, 2004). The smaller values of the AIC, BIC, and CAIC indicates a better clustering result. However, there might not exist the minimum values for these information criteria. Hence, the percentage reduction in criteria between different cluster numbers is considered.

3.2. Ordered logit models

Ordered logit model (OLM) is a cumulative-logit model for ordinal responses (McKelvey and Zavoina, 1975). The OLM calculates the cumulative probabilities based on thresholds. Suppose that the response variable has natural order, Y = 1, 2, …, M. The corresponding probabilities are $\{\pi_1, \pi_2, \ldots, \pi_M\}$, and a cumulative probability of a response less than m is:

$$P(Y \le m) = \sum\nolimits_{i=1}^{m} \pi_i \tag{2}$$

Then, the cumulative logit link function with the linear predictors is defined as:

$$log\left(\frac{P(Y \le m)}{1 - P(Y \le m)}\right) = log\left(\frac{\pi_1 + \cdots + \pi_m}{\pi_{m+1} + \cdots + \pi_M}\right) = \alpha_m + \beta \boldsymbol{X}_j' \tag{3}$$

where α_m is the constant variable for response level m. β represents the coefficient of variables $\boldsymbol{X}_j'$. This measures how likely the response Y is to be in or below m versus Y *is* higher than m.

Thus, the cumulative probabilities are given by:

$$P(Y \le m) = \frac{e^{\alpha_m + \beta \boldsymbol{X}_j'}}{1 + e^{\alpha_m + \beta \boldsymbol{X}_j'}} \tag{4}$$

since β is constant based on the proportional odds assumption, the curves of cumulative probabilities plotted against x are parallel (Derr, 2013).

Table 1
Descriptive statistics of independent variables of the truck-involved crashes at intersections.

Factor	Variable	ID	Description	Total	Severity level		
					KI[a]	NP[b]	**N**[c]
	severity		Injury type	18,346	634	5764	11,948
Human	drv_sex	**1**	**Male**	17,765	620	5571	11,574
		2	Female	581	14	193	374
	drv_age	**1**	**<=25**	1372	41	455	876
		2	26–45	8506	293	2673	5540
		3	46–65	7577	263	2362	4952
		4	>=66	891	37	274	580
	drv_rest	**1**	**None restraint**	473	36	213	224
		2	With belt	17,553	588	5469	11,496
		3	Other restraints	320	10	82	228
	alcflag	**1**	**Not detect**	17,997	589	5592	11,816
		2	Drink or drug	349	45	172	132
	contrib1	**1**	**Unknown/none**	1280	11	229	1040
		2	Disregarded signs	291	31	145	115
		3	Disregarded signals	706	37	338	331
		4	Exceeded speed	307	11	130	166
		5	Failure to Reduce Speed	1972	25	762	1185
		6	Improper turn	1341	5	185	1151
		7	Improper Lane use	305	12	74	219
		8	Improper lane change	251	1	31	219
		9	Failed to Yield	1634	51	735	848
		10	Inattention	1172	9	238	925
		11	Improper backing	541	4	41	496
		12	Followed too closely	115	1	39	75
		13	Equipment defect	177	4	46	127
		14	Other	8254	432	2771	5051
Roadway	no_lanes	**1**	**<=2**	9858	396	3292	6170
		2	3 and 4	7164	214	2118	4832
		3	>=4	1324	24	354	946
	rd_surf	**1**	**Dry**	15,871	566	4994	10,311
		2	Wet	2288	66	713	1509
		3	Water, ice, snow, slush	187	2	57	128
	rd_curve	**1**	**Straight**	17,297	581	5402	11,314
		2	Curve	1049	53	362	634
	rd_grad	**1**	**Level**	14,556	499	4498	9559
		2	Grade	2842	96	946	1800
		3	Hillcrest	725	30	230	465
		4	Bottom	223	9	90	124
	rd_pave	**1**	**Concrete**	165	3	60	102
		2	Smooth asphalt	12,266	403	3849	8014
		3	Coarse asphalt	5915	228	1855	3832
	rd_conf	**1**	**One-way, not divided**	371	3	90	278
		2	Two-way, not divided	12,320	447	3947	7926
		3	Two-way, divided	5655	184	1727	3744
	rte_type	**1**	**Interstate**	218	1	47	170
		2	US route	6448	230	2096	4122
		3	NC route	5160	240	1781	3139
		4	Secondary	6520	163	1840	4517
	func_cls	**1**	**Principle arterial**	7529	230	2318	4981
		2	Minor arterial	5182	175	1607	3400
		3	Collector	3875	182	1344	2349
		4	Local	1760	47	495	1218
location	rururb	**1**	**Rural**	9586	483	3383	5720
		2	Urban	8760	151	2381	6228
	locality	**1**	**Farms, woods, pastures**	5256	302	2050	2904
		2	Residential	2812	111	915	1786
		3	Commercial	9966	210	2707	7049
		4	Institutional	144	7	45	92
		5	Industrial	168	4	47	117
	loc_type	**1**	**Four-way intersection**	11,428	405	3645	7378
		2	T-intersection	6918	229	2119	4570
	terrain	**1**	**Flat**	4715	221	1658	2836
		2	Rolling	12,672	380	3856	8436

(*continued on next page*)

Table 1 (*continued*)

Factor	Variable	ID	Description	Total	Severity level		
					KI[a]	NP[b]	**N**[c]
		3	Mountainous	959	33	250	676
Environ-ment	weather1	**1**	**Clear**	13,889	497	4381	9011
		2	Cloudy	3052	102	922	2028
		3	Rain	1179	27	378	774
		4	Snow, sleet, hall, freezing rain	88	0	29	59
		5	Fog, smog, smoke	138	8	54	76
	light	**1**	**Daylight**	15,795	499	4848	10,448
		2	Dusk, down	533	22	183	328
		3	Dark light	858	30	277	551
		4	Dark	1160	83	456	621
Time	hour	**1**	**6:00 – 11:59**	7475	246	2347	4882
		2	12:00 – 17:59	8462	266	2570	5626
		3	18:00 – 23:59	1789	76	600	1113
		4	0:00 – 5:59	620	46	247	327
	month	**1**	**3–5**	4641	168	1458	3015
		2	6–8	4646	155	1419	3072
		3	9–11	3325	124	1066	2135
		4	12–2	5734	187	1821	3726
Control	access	**1**	**No access**	14,604	543	4721	9340
		2	Partial control	1780	36	481	1263
		3	Full control	1962	55	562	1345
	trf_cntl	**1**	**No control**	1224	32	386	806
		2	Signs	5244	289	1793	3162
		3	Signals	9861	237	2821	6803
		4	Double Yellow Line, No Passing Zone	2017	76	764	1177
	spd_limt	**1**	**<=35 mph**	5228	84	1328	3816
		2	36–55 mph	12,890	542	4388	7960
		3	56–70 mph	228	8	48	172

Note: bold was set for the base of the categorical variables.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury.
[c] N - No injury (set as base).

3.3. Marginal effect

It is important to know that the sign of the β in OLM does not always denote the direction of variable's effect on the whole utility function, also could not determine the changing direction of severity probability outcomes, especially for those Y in the middle of the order (Derr, 2013; Song and Fan, 2020). For evaluating the impacts of significant variables, marginal effect analysis is conducted. As all variables X are categorized and coded as dummy variables in this study, the marginal effect can be calculated as follows:

$$E^{P_{mi}}_{X_{mij}} = \frac{1}{n}\sum_{i=1}^{n}\left[P_{mi}(X_{mij}=1) - P_{mi}(X_{mij}=0)\right] \tag{5}$$

where P_{mi} denotes the probability of case i with response level m, and P_{mi} is calculated when X_{ijk}, the j^{th} binary indicator variable, changes from 0 to 1, respectively. Each variable's marginal effect is the average value of all cases.

4. Results and discussions

4.1. Latent class analysis results

Based on the LCA with cluster numbers ranging from 1 to 15, the results of AIC, BIC, and CAIC are shown in Fig. 1. The results of three criteria decrease with the increase in cluster number while the percentage decrease of three criteria drops to less than 1% after 4 clusters. Also, 4 clusters' entropy value is 0.9 which is close to 1. This indicates that 4 is the satisfying cluster number for the whole dataset. Hence, the truck-involved crash data at intersections are divided into 4 clusters.

As shown in Table 2, the variables (in bold) whose proportions in the dataset are significantly larger than other categorical variables (larger than 50%) are selected for featuring the latent class subsets (Li and Fan, 2019). It should be noted that the clustered subsets still have all categorical variables (i.e., the variables in bold are only used to denote the latent class). With the distribution results, cluster 1 could be referred to crashes occurred on urban, 1 or 2 lanes, one-way, not divided minor arterial with signal control and 35 mph speed limits. Cluster 2 can be specified as urban, 3 or 4 lanes principle arterial with

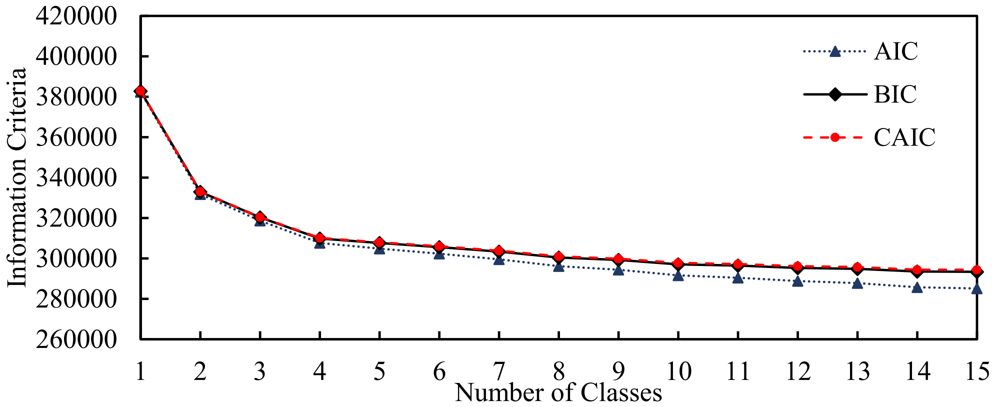

Fig. 1. Information criteria of AIC, BIC and CAIC for different cluster numbers.

Table 2
Distributions of featured variables (bold) based on the Latent class analysis.

Variable	NO.	Meaning	Cluster 1	Cluster 2	Cluster 3	Cluster 4
		Total	4877	5005	6087	2377
rururb	1	Rural	0.179833	0.048026	**0.999122**	**0.999943**
	2	Urban	**0.820167**	**0.951974**	0.000878	0.000057
no_lanes	1	1 or 2	**0.614199**	0.083566	**0.990308**	0.192945
	2	3 or 4	0.36365	**0.68713**	0.009475	**0.778813**
func_cls	1	Principle arterial	0.13495	**0.922053**	0.063831	**0.769756**
	2	Minor arterial	**0.563449**	0.07758	0.260518	0.19499
rd_conf	2	One-way, not divided	**0.735021**	0.366834	**0.982576**	0.399565
spd_limt	1	<=35 mph	**0.651863**	0.298678	0.072947	0.048352
	2	36 to 55mph	0.348136	**0.677925**	**0.927052**	**0.906137**
trf_cntl	3	Signal	**0.628636**	**0.860694**	0.157072	**0.634377**

signal control and 36 to 55mph speed limits. Cluster 3 can be identified as rural, 1 or 2 lanes, one-way, not divided with 36 to 55mph speed limits. Cluster 4 can be described as rural 3 or 4 lanes principle arterial with signal control and 36 to 55mph speed limits.

4.2. Latent class ordered logit model results

All variables are included in the original OLMs for 4 clusters, and this paper uses 95% confidence level as the criterion to select significant variables for better describing the crash causes. The Chi-square test is used to determine whether to drop or remain the variables. The results of the significant variables' coefficients are shown in the Tables 3 and 4.

4.3. Marginal effect results

Marginal effects are used to describe the variable's impact on the probability of the crash severity since the coefficient could not denote this effect directly. The results are shown in the Tables 5 and 6. It is found that even the same variable would have different impacts on the crash severity in different clusters. Also, more specific impacts of the variable to the crash severity (especially for FI injury) are discussed as follows:

4.3.1. Human characteristics

According to the Tables 5 and 6, female drivers are found more likely to suffer fatal and incapacitating injuries (FI) and non-incapacitating and possible injury (NP) (marginal effects: 3.79% and 8.83%) than male in cluster 4. As cluster 4 represents crashes mostly occurred in principle arterials with speed limits of 35–55 mph, results indicate that females are more vulnerable during a high speed truck crash, which is similar to the results in (Chen and Chen, 2011; Osman et al., 2018). Also, drivers with belts are much safer than those without restraints in clusters 2, 3 and 4, which show a 2.16–5.77% reduction in FI injury and 11.14–15.05% reduction in NP injury. The same conclusions could also been made in (Dong et al., 2017; Osman et al., 2018). In addition, for a driver who was drunk or on drugs, the injury of both FI injury (marginal effects from 2.8% to 7.33%) and NP injury (marginal effects from 11.97% to 25.15%) increase significantly in all clusters, similar results could be drawn in (Dong et al., 2017; Liu and Fan, 2019).

Table 3
Ordered logit models' significant independent variables in clusters 1 and 2.

Parameter	Description	Cluster = 1		Cluster = 2	
		Estimate	S.E.	Estimate	S.E.
Intercept_1		−4.5586**	0.1841	−3.3378**	0.2879
Intercept_2		−1.2876**	0.144	−0.121	0.2725
alcflag_2	Drink or drug	1.1119**	0.2553	1.3606**	0.2284
drv_rest_2	With belt	–	–	−0.8067**	0.2357
drv_rest_3	Other restraints	–	–	−0.7475*	0.3064
contrib1_2	Disregarded signs	1.2183**	0.3513	1.6551**	0.4704
contrib1_3	Disregarded signals	1.2983**	0.1776	1.5699**	0.1415
contrib1_4	Exceeded speed	1.134**	0.3745	1.0294**	0.319
contrib1_5	Failure to reduce speed	0.9278**	0.1336	0.9204**	0.1104
contrib1_6	Improper turn	−0.4675**	0.1578	−0.5234**	0.1753
contrib1_8	Improper lane change	–	–	−0.7336*	0.303
contrib1_9	Failed to yield	0.9888**	0.1353	0.946**	0.1536
contrib1_11	Improper backing	−1.0544**	0.3121	−1.2095**	0.4306
contrib1_12	Followed too closely	0.7972*	0.3671	–	–
contrib1_14	Other	0.6817**	0.0955	0.6058**	0.0905
rd_grad_2	Grade	0.2078*	0.0944	–	–
rd_grad_4	Bottom	0.7478*	0.2981	–	–
rururb_2	Urban	0.1866*	0.0912	–	–
locality_3	Commercial	−0.2942**	0.0743	−0.3317**	0.1122
loc_type_2	T-intersection	−0.176*	0.0815	–	–
hour_2	12:00–17:59	−0.1739*	0.0681	–	–
hour_4	0:00–5:59	0.4599*	0.2232	0.6664**	0.1543
trf_cntl_3	Signals	−0.2531**	0.0825	−0.2076*	0.091
spd_limt_2	36–55 mph	0.1488*	0.07	–	–
Spd_limt_3	50–70 mph	–	–	−1.0713**	0.2809
No. of observation		4877		5005	
Log Likelihood (intercept only)		−3106.691		−3416.51	
Log Likelihood at convergence		−2931.048		−3217.77	

Note: Confident Level: *for 5%, **for 1%.

For driver behaviors that contribute to the crash injury, this paper and relevant literature show that behaviors including disregarded signs, disregarded signals (Azimi et al., 2020), exceeded speed(Khorashadi et al., 2005), failure to reduce speed (Chen and Chen, 2011), improper lane use (Chen and Chen, 2011), failed to yield (Anderson and Dong, 2017; Bernard and Mondy, 2016), inattention, followed too closely (Bernard and Mondy, 2016) and equipment defect (Chen and Chen, 2011) all result in the increase of the truck crash severity compared to the circumstance of no contributions. Especially in cluster 3, in which observations mostly occurred at rural non-signal intersections, factors including followed too closely, disregarded signs, disregarded signals, failed to yield, and exceeded speed, contribute to 26.41%, 21.55%, 12.94%, 12.70% and 12.69% increase of FI injury respectively. Behaviors such as improper turn (Chen et al., 2015; Chen and Chen, 2011; Uddin and Huynh, 2018), improper lane change (Hassan et al., 2015; Khorashadi et al., 2005) and improper backing could slightly reduce the severity level of the crashes. For example, in cluster 2, improper turn, lane change, and backing could reduce the probability of FI injury by 0.78%, 0.99% and 1.34% respectively. Those results indicate that slow speed and not direct head-on crashes might not always result in severe injury for truck-involved crash at intersections.

4.3.2. Roadway and location characteristics

For crashes at gradient road, compared to level road in the cluster 1, road with grade increase 0.33% and 3.57% of the FI and NP injury, same results could be found in (Azimi et al., 2020; Chen et al., 2015). Also, crashes at bottom segment could increase 1.58% and 13.74% of the FI and NP injury. Meanwhile, in cluster 1, crashes occurred in urban areas lead to 0.26% and 3.06% increase of the FI and NP injury than those in the rural areas, and same results are shown in (Park et al., 2017). For truck-involved crash occurred in the commercial area compared to farm and green land, marginal effects show a reduction of the severity in all clusters. This might be caused by the lower speed limits and better infrastructure in commercial locations. For crashes that took place at T-intersections in clusters 1 and 3, they have lower injury level than those at cross- intersections. In cluster 3, T-intersection reduced 1.78% and 6.26% of the FI and NP injury. This might be the result of the reduction of conflict points, especially the vertical crossing points at T-intersections. It is also noted that crashes occurred in the mountainous terrain reduce 1.57% and 6.36% of the FI and NP injury compared to flat area in the cluster 3. Compared to the interstate route type, crashes took place in the US route, NC route and secondary route increase the injury level in clusters 1, 3 and 4. Meanwhile, crashes occurred in the minor arterial and collector road compared to principle arterial could also increase 1.28% and 4.23% of the FI and NP injury in cluster 3.

4.3.3. Environment and time characteristics

Compared to the clear weather, raining weather decreases the severity level in cluster 3. This might be caused by the lower travel speed and higher alertness of the driver in raining condition. Similar conclusions could also be found in

Table 4
Ordered logit models' significant independent variables in clusters 3 and 4.

Parameter	Description	Cluster = 3		Cluster = 4	
		Estimate	S.E.	Estimate	S.E.
Intercept_1		−3.8412**	0.2077	−3.8818**	0.3722
Intercept_2		−1.1132**	0.2012	−1.311**	0.3643
alcflag_2	Drink or drug	0.777**	0.1689	0.9114**	0.2605
drv_rest_2	With belt	−0.7096**	0.1274	−0.7636**	0.2328
drv_sex_2	Female	–	–	0.5442*	0.2731
contrib1_2	Disregarded signs	2.0217**	0.1743	1.4624**	0.4807
contrib1_3	Disregarded signals	1.4442**	0.283	1.3679**	0.173
contrib1_4	Exceeded speed	1.4419**	0.1826	–	–
contrib1_5	Failure to reduce speed	1.1191**	0.1412	0.7922**	0.159
contrib1_7	Improper lane use	1.1113**	0.1986	–	–
contrib1_9	Failed to yield	1.5325**	0.1334	1.1373**	0.1664
contrib1_10	Inattention	0.5121**	0.1943	–	–
contrib1_11	Improper backing	−0.6802**	0.2465	–	–
contrib1_12	Followed too closely	2.2286**	0.5546	–	–
contrib1_13	Equipment defect	0.6405*	0.286	–	–
contrib1_14	Other	1.2969**	0.1192	0.8986**	0.1207
rte_type_2	US route	0.3301**	0.0834	1.132**	0.2682
rte_type_3	NC route	0.4012**	0.0638	1.2184**	0.2825
rte_type_4	Secondary	–	–	1.2128**	0.3364
func_cls_2	Minor arterial	0.2489**	0.0799	–	–
func_cls_3	Collector	0.1379*	0.0698	–	–
locality_3	Commercial	−0.268**	0.0817	−0.2238**	0.0845
loc_type_2	T-intersection	−0.3617**	0.0551	−0.223*	0.0902
hour_4	0:00 – 5:59	–	–	0.4053*	0.2026
terrain_3	Mountainous	−0.3687**	0.1235	–	–
weather1_3	Rain	−0.3176**	0.1169	–	–
light_4	dark	0.1783*	0.0845	–	–
spd_limit_2	36–55 mph	0.2469*	0.1122	–	–
No. of observation		6087		2377	
Log Likelihood (intercept only)		−5116.76		−2037.94	
Log Likelihood at convergence		−4775.81		−1952.48	

Note: Confident Level: * for 5%, ** for 1%.

Table 5
Marginal effects of the independent variables in clusters 1 and 2.

Variable		Cluster = 1 (%)			Cluster = 2 (%)		
Severity	Description	FI[a]	NP[b]	N[c]	FI[a]	NP[b]	N[c]
alcflag_2	Drink or drug	2.80	20.78	−23.58	4.81	25.15	−29.96
drv_rest_2	With belt	–	–	–	−2.16	−15.05	17.21
drv_rest_3	Other restraints	–	–	–	−1.02	−11.51	12.53
contrib1_2	Disregarded signs	3.29	22.64	−25.94	7.01	29.07	−36.08
contrib1_3	Disregarded signals	3.54	24.05	−27.60	5.82	28.66	−34.48
contrib1_4	Exceeded speed	2.94	21.08	−24.02	3.16	19.22	−22.38
contrib1_5	Failure to reduce speed	2.02	16.96	−18.99	2.42	16.93	−19.35
contrib1_6	Improper turn	−0.56	−7.28	7.85	−0.78	−8.55	9.33
contrib1_8	Improper lane change	–	–	–	−0.99	−11.35	12.34
contrib1_9	Failed to yield	2.21	18.18	−20.40	2.69	17.65	−20.34
contrib1_11	Improper backing	−0.98	−14.06	15.04	−1.34	−16.73	18.07
contrib1_12	Followed too closely	1.74	14.66	−16.40	–	–	–
contrib1_14	Other	1.07	11.37	−12.44	1.22	10.52	−11.74
rd_grad_2	Grade	0.33	3.57	−3.90	–	–	–
rd_grad_4	Bottom	1.58	13.74	−15.31	–	–	–
rururb_2	Urban	0.26	3.06	−3.31	–	–	–
locality_3	Commercial	−0.45	−5.05	5.50	−0.70	−6.04	6.75
loc_type_2	T-intersection	−0.25	−2.91	3.16	–	–	–
hour_2	12:00–17:59	−0.25	−2.91	3.16	–	–	–
hour_4	0:00–5:59	0.84	8.22	−9.06	1.65	12.49	−14.14
trf_cntl_3	Signals	−0.38	−4.30	4.67	−0.41	−3.72	4.14
spd_limt_2	36–55 mph	0.22	2.51	−2.74	–	–	–
spd_limt_3	50–70 mph	–	–	–	−1.27	−15.37	16.64

Note: The base was set at when the variable was denoted with the number of 1.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury.
[c] N - No injury.

Table 6
Marginal effects of the independent variables in clusters 3 and 4.

Variable		Cluster = 3 (%)			Cluster = 4 (%)		
Severity	Description	FI[a]	NP[b]	N[c]	FI[a]	NP[b]	N[c]
alcflag_2	Drink or drug	5.23	11.97	−17.20	7.33	13.60	−20.92
drv_rest_2	With belt	−4.60	−11.14	15.74	−5.77	−11.88	17.65
drv_sex_2	Female	–	–	–	3.79	8.83	−12.62
contrib1_2	Disregarded signs	21.55	18.10	−39.65	14.86	16.84	−31.70
contrib1_3	Disregarded signals	12.94	17.25	−30.19	12.40	18.01	−30.42
contrib1_4	Exceeded speed	12.69	17.38	−30.07	–	–	–
contrib1_5	Failure to reduce speed	8.38	15.32	−23.70	5.85	12.10	−17.95
contrib1_7	Improper lane use	8.66	15.17	−23.83	–	–	–
contrib1_9	Failed to yield	12.70	19.13	−31.83	9.50	16.09	−25.59
contrib1_10	Inattention	3.12	8.17	−11.29	–	–	–
contrib1_11	Improper backing	−2.53	−11.61	14.14	–	–	–
contrib1_12	Followed too closely	26.41	15.37	−41.78	–	–	–
contrib1_13	Equipment defect	4.14	10.01	−14.15	–	–	–
contrib1_14	Other	7.06	19.65	−26.72	5.33	14.56	−19.89
rte_type_2	US route	1.79	5.52	−7.32	5.45	17.80	−23.25
rte_type_3	NC route	2.03	6.88	−8.92	10.06	16.72	−26.78
rte_type_4	Secondary	–	–	–	10.98	15.88	−26.85
func_cls_2	Minor arterial	1.28	4.23	−5.51	–	–	–
func_cls_3	Collector	0.68	2.35	−3.03	–	–	–
locality_3	Commercial	−1.21	−4.61	5.82	−1.24	−3.90	5.13
loc_type_2	T-intersection	−1.78	−6.26	8.04	−1.20	−3.89	5.09
hour_4	0:00–5:59	–	–	–	2.65	6.76	−9.41
terrain_3	Mountainous	−1.57	−6.36	7.93	–	–	–
weather1_3	Rain	−1.38	−5.46	6.85	–	–	–
light_4	dark	0.93	3.02	−3.95	–	–	–
spd_limit_2	36–55 mph	1.10	4.25	−5.36	–	–	–

Note: The base was set at when the variable was denoted with the number of 1.
[a] FI - Fatal and incapacitating injury.
[b] NP - Non-incapacitating and possible injury.
[c] N - No injury.

(Khorashadi et al., 2005; Liu and Fan, 2020). For the traveling time, compared to the morning time from 6:00 to 11:59, afternoon from 12:00 to 17:59 could slightly decrease the injury level in the cluster 1. In addition, for those traveling during the late night from 0:00 am to 5:59 am, result shows an increase of the severity level in clusters 1, 2 and 4. Similar conclusions could be drawn in (Li and Fan, 2019; Uddin and Huynh, 2017). Also, driving in the dark without roadside light could increase 0.93 % and 3.02% of the FI and NP injury compared to the condition of daylight in cluster 3, similar conclusion could also be found in (Uddin and Huynh, 2018).

4.3.4. Control characteristics

Compared to the no control condition, locations with signals control could slightly decrease the injury level in clusters 1 and 2, and similar conclusions were drawn in (Anderson and Dong, 2017; Chen and Chen, 2011). The speed limit is set according to the functional class of the roadway. And the speed limit of the crash in the middle of the intersection is determined by the largest speed limit of the approaches. It is noted that speed limits within 36 to 55 mph increased the injury level compared to the speed limits of less than 35 mph in clusters 1 and 3, the result is also in line with (Liu and Fan, 2020). While, for speed limits within 50 to 70 mph, results showed a decrease of the injury level in the cluster 2, which also shows heterogeneous impact of the factors between each cluster.

5. Transferability test

To test the transferability of the factors in four latent class clustered crashes, likelihood ratio tests are applied according to (Washington et al., 2011).

$$X^2 = -2\left[LL(\boldsymbol{\beta}_{total}) - \sum_i^n LL(\boldsymbol{\beta}_{latentclassi})\right] \tag{6}$$

where $LL(\boldsymbol{\beta}_{total})$ is the log-likelihood at the convergence of a model containing the converged parameters based on the total data. $LL(\boldsymbol{\beta}_{latentclassi})$ denotes the log-likelihood at the convergence of a model containing the converged parameters based on the latent class i data. All models utilized the same variables based on the whole dataset model. The degrees of freedom are calculated by the summation of the number of estimated parameters in all latent class models minus the number of estimated parameters in the whole dataset model. The X^2 is χ^2 distributed with the null hypothesis that the parameters for all models are the same.

Based on the ordered logit models, the log-likelihood value for the whole dataset model and four latent class models are −12984, −2928, −3220, −4782, and −1974, respectively. The value of X^2 is 213 with 89 degrees of freedom. This gives a 99.9% confidence level to reject the null hypothesis and indicates significant distinctions (or instability) between the factors of four latent class clustered crashes with the whole dataset model. Meanwhile, latent class models provide some specific factors which are not identified in the whole dataset model. All these results indicate better model performance and accuracy to segment the whole dataset according to the latent class analysis.

6. Conclusions

This study aims to investigate and identify the contributing factors effecting crash severities of truck-involved crash at intersections among heterogeneous clusters. Four ordered logit models are constructed based on the latent class analysis of the data between 2005 and 2017 in North Carolina from HSIS. Different clusters have different significant variables and coefficient values, which indicate the heterogeneities between specific clusters and proved the superiority of this combined model for obtaining more specific results of the contributing factors. Also, marginal effects are analyzed for better interpreting the impacts of variable variation, which gave further insights into mitigating the truck-involved crash severity at intersections.

Based on the ordered logit models, many categorical factors and the impact of these factors are analyzed. Though heterogeneous results exist between different clusters (e.g., speed limits), factors such as female, without belt, drunk or in drug, urban, grade or bottom segment, NC (state) and secondary route, late night, dark without light, minor arterial and collector road could increase the injury level. While factors such as rain, signal control, commercial, T-intersection, mountainous terrain, and afternoon could reduce the severity level. Also, driver behaviors are analyzed and followed too closely, disregarded signs, disregarded signals, failed to yield and exceeded speed are the five most contributing factors to increase crash severity, while behaviors cause indirect and low-speed crash such as improper turn, lane change, and backing, are found to reduce the severity level.

The results can give insights to engineers and planners for further modifying the transportation regulations and infrastructure management (e.g., amending the regulations for speed limits, belt enforcement, signal control and grade design at specific locations). In this paper, the latent class analysis and the ordered logit models were conducted independently. Though it is much easier to estimate crash severities in independent models, combining latent classes and crash severity models could be more beneficial to model performances and inferences (Xiong and Mannering, 2013), and this is worth exploring in the future.

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 by 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).

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, 106–113. https://doi.org/10.1016/j.aap.2018.04.011.

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, 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,. https://doi.org/10.1016/j.aap.2019.105355 105355.

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,. https://doi.org/10.1016/j.amar.2019.100102 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, 49–58. https://doi.org/10.3141/2585-06.

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, 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.

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

Derr, B., 2013. Ordinal Response Modeling with the LOGISTIC Procedure. SAS Glob. Forum, 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, 1–9. https://doi.org/10.1061/JTEPBS.0000060.

FHWA, 2004. Signalized Intersections: Informational Guide. https://doi.org/10.1109/CVPR.2005.217.
Fountas, G., Anastasopoulos, P.C., Mannering, F.L., 2018. Analysis of vehicle accident-injury severities: a comparison of segment- versus accident-based latent class ordered probit models with class-probability functions. Anal. Methods Accid. Res. 18, 15–32. https://doi.org/10.1016/j.amar.2018.03.003.
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.
Hou, Q., Huo, X., Leng, J., Cheng, Y., 2019. Examination of driver injury severity in freeway single-vehicle crashes using a mixed logit model with heterogeneity-in-means. Phys. A Stat. Mech. its Appl. 531,. https://doi.org/10.1016/j.physa.2019.121760 121760.
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, 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, 157–168. https://doi.org/10.1007/s11121-011-0201-1.
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, 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, 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,. https://doi.org/10.1016/j.aap.2019.105388 105388.
McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. 4, 103–120. https://doi.org/10.1080/0022250X.1975.9989847.
McLachlan, G.J., Peel, D., 2004. Finite Mixture Models. John Wiley & Sons. https://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.
Mokhtarimousavi, S., 2019. A time of day analysis of pedestrian-involved crashes in California: investigation of injury severity, a logistic regression and machine learning approach using HSIS data. Inst. Transp. Eng. ITE J. 89, 25–33.
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.
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.
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, 305–314. https://doi.org/10.1016/j.aap.2017.06.009.
Sivasankaran, S.K., Balasubramanian, V., 2020. Exploring the severity of bicycle – vehicle crashes using latent class clustering approach in India. J. Safety Res. 72, 127–138. https://doi.org/10.1016/j.jsr.2019.12.012.
Song, L., Fan, W., 2020. Combined latent class and partial proportional odds model approach to exploring the heterogeneities in truck-involved severities at cross and T-intersections. Accid. Anal. Prev. 144. https://doi.org/10.1016/j.aap.2020.105638.
Song, L., Li, Y., Fan, W., (David), Wu, P., 2020. Modeling pedestrian-injury severities in pedestrian-vehicle crashes considering spatiotemporal patterns: insights from different hierarchical Bayesian random-effects models. Anal. Methods Accid. Res. 28,. https://doi.org/10.1016/j.amar.2020.100137 100137.
Uddin, M., Huynh, N., 2018. Factors influencing injury severity of crashes involving HAZMAT trucks. Int. J. Transp. Sci. Technol. 7, 1–9. https://doi.org/10.1016/j.ijtst.2017.06.004.
Uddin, M., Huynh, N., 2017. Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. Accid. Anal. Prev. 108, 44–55. https://doi.org/10.1016/j.aap.2017.08.009.
USDOT, 2019. Large Truck and Bus Crash Facts 2017, Federal Motor Carrier Safety Administration.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Statistical and Econometric Methods for Transportation Data Analysis, second ed. CRC Press.
Wu, P., Meng, X., Song, L., 2019a. A novel ensemble learning method for crash prediction using road geometric alignments and traffic data. J. Transp. Saf. Secur. 1–19. https://doi.org/10.1080/19439962.2019.1579288.
Wu, P., Meng, X., Song, L., Zuo, W., 2019b. Crash risk evaluation and crash severity pattern analysis for different types of urban junctions: fault tree analysis and association rules approaches. Transp. Res. Rec. J. Transp. Res. Board 2673, 403–416. https://doi.org/10.1177/0361198118822817.
Xiong, Y., Mannering, F.L., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: a finite-mixture random-parameters approach. Transp. Res. Part B Methodol. 49, 39–54. https://doi.org/10.1016/j.trb.2013.01.002.
Zhu, X., Srinivasan, S., 2011. Modeling occupant-level injury severity: an application to large-truck crashes. Accid. Anal. Prev. 43, 1427–1437. https://doi.org/10.1016/j.aap.2011.02.021.