Journal of Safety Research 76 (2021) 184–196

Contents lists available at ScienceDirect

Journal of Safety Research

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

Exploring pedestrian injury severities at pedestrian-vehicle crash hotspots with an annual upward trend: A spatiotemporal analysis with latent class random parameter approach

Li Song [a], Wei (David) Fan [a,*], Yang Li [a], Peijie Wu [b]

[a] USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE), Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, EPIC Building, Room 3366, 9201 University City Boulevard, Charlotte, NC 28223-0001, United States

[b] School of Transportation Science and Engineering, Harbin Institute of Technology, No. 73, Huanghe Road, Nangang District, Harbin, China

ARTICLE INFO

Article history:
Received 10 August 2020
Received in revised form 1 September 2020
Accepted 9 December 2020
Available online 29 December 2020

Keywords:
Pedestrian
Injury severity
Spatiotemporal analysis
Latent class clustering
Random parameter logit model

ABSTRACT

Introduction: With the increasing trend of pedestrian deaths among all traffic fatalities in the past decade, there is an urgent need for identifying and investigating hotspots of pedestrian-vehicle crashes with an upward trend. *Method:* To identify pedestrian-vehicle crash locations with aggregated spatial pattern and upward temporal pattern (i.e., hotspots with an upward trend), this paper first uses the average nearest neighbor and the spatial autocorrelation tests to determine the grid distance and the neighborhood distance for hotspots, respectively. Then, the spatiotemporal analyses with the Getis-Ord Gi* index and the Mann-Kendall trend test are utilized to identify the pedestrian-vehicle crash hotspots with an annual upward trend in North Carolina from 2007 to 2018. Considering the unobserved heterogeneity of the crash data, a latent class model with random parameters within class is proposed to identify specific contributing factors for each class and explore the heterogeneity within classes. Significant factors of the pedestrian, vehicle, crash type, locality, roadway, environment, time, and traffic control characteristics are detected and analyzed based on the marginal effects. *Results:* The heterogeneous results between classes and the random parameter variables detected within classes further indicate the superiority of latent class random parameter model. *Practical Applications:* This paper provides a framework for researchers and engineers to identify crash hotspots considering spatiotemporal patterns and contribution factors to crashes considering unobserved heterogeneity. Also, the result provides specific guidance to developing countermeasures for mitigating pedestrian-injury at pedestrian-vehicle crash hotspots with an upward trend.

1. Introduction

Compared to other entities in traffic crashes, pedestrians are more vulnerable to suffer severe injuries. According to one report from the National Highway Safety Administration (NHTSA, 2019), in the United States there were 5,977 pedestrian fatalities in traffic crashes in 2017. From 2008 to 2017, the percentage of pedestrian deaths in total traffic fatalities has constantly increased from 12% to 16%. In recent years, more and more efforts have been put into investigating contributing factors of the pedestrian injury severity at specific hazardous locations (Anderson, 2009; Dai, 2012). Meanwhile, existence of the temporal variation and tendency of the pedestrian crash data might affect the model result in different ways (Behnood & Mannering, 2016), and neglecting the fundamental temporal features could result in erroneous conclusions (Mannering, 2018). One previous research study has identified the instability of different time scales among the pedestrian crashes, and the annual variation mainly shows an increasing/decreasing trend (Dai, 2012). Hence, there is an urgent need to develop a proper approach to identifying the contributing factors at crash hotspots with annual uptrends.

Previous studies have applied several methods to explore the factors to crash severity. A detailed review was summarized in (Mannering & Bhat, 2014), and this review also pointed out that the heterogeneity inherent in the crash observations could result in biased parameter estimations and incorrect inferences. To obtain more accurate and specific model results, it is important to investigate the pedestrian injury severity by considering the heterogeneity both within and between the pedestrian crash observations.

* Corresponding author.
E-mail address: wfan7@uncc.edu (Wei (David) Fan).

https://doi.org/10.1016/j.jsr.2020.12.008
0022-4375/

To identify specific contributing factors and provide guidance for improving the deteriorative tendency of pedestrian-vehicle crashes at hotspots, this paper uses the spatiotemporal trend analysis with the Getis-Ord Gi* index and the Mann-Kendall trend test to explore the annual spatial clustering and the temporal tendency of pedestrian-vehicle crashes in North Carolina from 2007 to 2018. Meanwhile, the grid distance interval and the neighborhood distance are determined by the average nearest neighbor and the spatial autocorrelation test, respectively. Then, a sequential process by combining latent class clustering with random parameter logit approaches are used to identify contributing factors considering the heterogeneity within and between the classes.

2. Literature review

2.1. Crash locations considering spatiotemporal patterns

To identify aggregated/high-frequency traffic crash locations, point pattern analyses, such as the kernel density estimation (KDE) (Ouni & Belloumi, 2018) and Getis-Ord Gi* index (Songchitruksa & Zeng, 2010), were commonly used in previous studies. However, KDE is not feasible for the statistical significance test and the density pattern will certainly be influenced by the choice of bandwidth (Plug, Xia, & Caulfield, 2011). Hence, the Getis-Ord Gi* index, which is a statistical-based test for high/low value clusters, was then deployed to identify the spatial patterns in several studies. Ulak, Kocatepe, Ozguven, Horner, and Spainhour (2017) employed the Getis-Ord Gi* index to identify hotspots with the optimized neighborhood distance that was determined by the Global Moran's I test. Results showed that the accessibility to hospitals of hotspots is one of the major reasons for severe injuries. Considering that the Getis-Ord Gi* and the global Moran's I could identify spatial patterns from local and global perspectives, respectively (Blazquez, Picarte, Calderón, & Losada, 2018), this paper employs the Getis-Ord Gi* for hotspots identification and utilizes the global Moran's I to provide a reference for the neighborhood distance. Meanwhile, the average nearest neighbor (ANN) was employed to calculate the distance interval between traffic crashes (Yalcin & Sebnem Duzgun, 2015).

For the temporal trend analysis, the Mann-Kendall trend test, which is a statistical-based non-parametric rank correlation analysis method, has been widely used in previous studies (Gudes, Varhol, Sun, & Meuleners, 2017; Wang & Chan, 2016). Gudes et al. (2017) evaluated temporal patterns of the hot/cold spot regions of the heavy-vehicle crashes by the Mann-Kendall trend test. Results showed inconsistency of temporal patterns in hotspots over time. With such analyses, the temporal tendency of hotspots could be further investigated.

2.2. Identification of injury-severity factors considering unobserved heterogeneity

As summarized in Table 1, statistics-based methods, such as an ordered/unordered response model with a logit/probit link function, have been widely used because of their good performance in parameter calibration and outcome interpretation (Mannering & Bhat, 2014). Moreover, to avoid biased parameters estimation and incorrect inferences caused by the unobserved heterogeneity, random parameter models, which can potentially capture unobserved heterogeneity by allowing parameters to vary across observations, were proposed (Mannering & Bhat, 2014). Abay (2013) compared the pedestrian severity outcomes with ordered logit, mixed ordered logit, multinomial logit, and mixed logit. The result revealed that mixed models can accommodate flexible variable effects to some extent while fixed-parameters injury severity models underestimated the effect of some important behavioral attributes of the crashes.

For a heterogeneity-based data segmentation approach employed in pedestrian injury severities, Table 1 also shows many sequential processes of combining the Latent Class Clustering (LCC) with other models, such as the Multinomial Logit model (MNL) (Sun, Sun, & Shan, 2019), Partial Proportional Odds model (PPO) (Li & Fan, 2019a), and Mixed Logit Model (Behnood & Mannering, 2016). Iranitalab and Khattak (2017) compared the crash severity prediction performance of the LCC and k-means clustering with the MNL and three machine learning methods. Results indicated that LCC could well improve the performance of the multinomial logit model. Behnood and Mannering (2016) analyzed differing injury-severity levels sustained by pedestrians in Chicago using both latent class and mixed logit models, which better accounts for unobserved heterogeneity compared to conventional models. Hence, a random parameter model (mixed logit model), which accounts for the heterogeneity across the observations, is considered after the implementation of LCC.

3. Methodology

3.1. Spatiotemporal analysis

3.1.1. Spatiotemporal trend analysis

The basic idea of conducting the spatiotemporal trend analysis is to first divide the map into square bins with a specific distance interval and time interval. The Getis-Ord Gi* index (Getis & Ord, 2010) and the Mann-Kendall test (Kendall & Gibbons, 1990; Mann, 1945) are used to investigate spatial hot/cold (i.e., aggregating of high/low values) pattern and the temporal tendency of these patterns, respectively. The formula of Getis-Ord Gi* index is:

$$G_i^* = \frac{\sum_{j=1}^{n}\omega_{i,j}x_j - \overline{X}\sum_{j=1}^{n}\omega_{i,j}}{SD(x_j)\sqrt{\frac{n}{n-1}\sum_{j=1}^{n}\omega_{i,j}^2 - \frac{1}{n-1}\left(\sum_{j=1}^{n}\omega_{i,j}\right)^2}} \quad (1)$$

where x_j represents the attribute value of the jth bin. $\omega_{i,j} = 1$ if the jth bin is within the spatiotemporal neighborhood distance of the ith bin and 0 otherwise. n denotes the total number of bins within the spatiotemporal neighborhood distance. $\overline{X}$ is the mean value for x_j. $SD(x_j)$ means the standard deviation for x_j. G_i^* is also a Z-score. When the p-value is statistical significance, $G_i^* > 0$ represents a clustering pattern of the high values (hotspot), $G_i^* = 0$ denotes a random pattern of the values, and $G_i^* < 0$ means a clustering pattern of the lower values (cold spot).

Then the Mann-Kendall trend test is performed on every location/grid with data within a specified time interval. For the Gi* value within each time interval $\{N_t : t = 1, 2, \cdots, T\}$, the trend test statistic S is:

$$S = \sum_{i=1}^{T-1}\sum_{j=i+1}^{T} a_{ij} \quad (2)$$

$$a_{ij} = sign(N_j - N_i) = \begin{cases} 1 N_i < N_j \\ 0 N_i = N_j \\ -1 N_i > N_j \end{cases} \quad (3)$$

where a_{ij} is a symbolic variable which counts the rank/trend of the Getis-Ord Gi* index.

The null hypothesis for S is zero, which means no trend in the values over time. Based on the variance of the values in the bin time series, Z statistic is used for the statistical significance test.

Table 1
Summary of methodological approaches in pedestrian injury severity studies.

Model	Specific scenario	Year	Location	Data	Literature
Multinomial logit model (MNL)	–	2005–2012	North Carolina	3,553	(Chen & Fan, 2019)
Partial proportional odds model (PPO)	pedestrian	2007–2014	North Carolina	10,875	(Li & Fan, 2019b)
Support vector machine and MNL	time of day	2010–2014	California	8,573	(Mokhtarimousavi, 2019)
Binary logistic regression and tree-based models	–	2014–2016	Changsha, China	791	(Hu, Wu, Huang, Peng, & Liu, 2020)
Classification and regression tree with random forest approach.	weather	2013	Britain	14,174	(Li, Ranjitkar, Zhao, Yi, & Rashidi, 2017)
Extracted rules from Bayesian networks	urban and suburban	2009–2011	Jordan	21,852	(Mujalli, Garach, López, & Al-Rousan, 2019)
Considering unobserved heterogeneity					
Mixed logit model	signalized and non-signalized locations	2008–2010	Florida	7,630	(Haleem et al., 2015)
Mixed logit model	–	1997–2000	North Carolina	5,808	(Kim et al., 2010)
Random-parameter (mixed) logit	–	2002–2006	New York City	4,666	(Aziz et al., 2013)
Artificial neural network and random parameter ordered response models	day of week	2010–2014	California	10,146	(Mokhtarimousavi, Anderson, Azizinamini, & Hadi, 2020)
Ordered logit, mixed ordered logit, multinomial logit, mixed logit	–	1998–2009	Denmark	4,952	(Abay, 2013)
Ordered logit model, generalized ordered logit model, and latent class ordered logit model	–	2002–2006	New York City	4,701	(Yasmin et al., 2014)
Latent class clustering and MNL	whole and each cluster	2006–2015	Louisiana	14,236	(Sun et al., 2019)
Latent class clustering and binary logit	whole and each cluster	2009–2012	Switzerland	9,659	(Sasidharan et al., 2015)
Latent class with ordered probit method, k-means with MNL	whole and each cluster	2002–2006 (NYC), 2003–2006 (M)	New York City, Montreal	5,820	(Mohamed et al., 2013)
Latent class clustering and PPO	each cluster	2007–2014	North Carolina	10,875	(Li & Fan, 2019a)
Latent-class logit and mixed logit models.	period (pre-recession, recession, and post-recession)	2005–2012	Chicago	19,895	(Behnood & Mannering, 2016)
Considering spatial and temporal patterns					
Bernoulli model and logistic regression	spatial clusters	2000–2007	Georgia	7,763	(Dai, 2012)
Kernel density estimation analysis and MNL	Spatiotemporal patterns	2001–2013	Tunisia	1,922	(Ouni & Belloumi, 2018)
Geographically and temporally weighted ordinal logistic regression	–	2007–2014	North Carolina	13,854	(Liu, Hainen, Li, Nie, & Nambisan, 2019)

$$Z_S = \begin{cases} \frac{S-1}{SD(S)}, S > 0 \\ 0, S = 0 \\ \frac{S+1}{SD(S)}, S < 0 \end{cases} \quad (4)$$

When $T \geq 10$, statistic S follows the normal distribution approximately. $SD(S)$ denotes the stand error of the S. For a given confidence level α, if $|Z_S| \geq |Z_{S,1-\alpha/2}|$, then the null hypothesis is rejected. Also, $Z_S > 0$ and $Z_S < 0$ indicate the uptrend and downtrend in bin values.

3.1.2. Average nearest neighbor

The average nearest neighbor analysis is used to provide a reasonable reference for the distance interval of the grid. If the average distance of the data is less than the average distance for a hypothetical random distribution, the distribution of the data being analyzed is considered clustered (Ebdon, 1985). The average nearest neighbor ratio (ANN) can be expressed as:

$$ANN = \frac{\overline{D_O}}{\overline{D_E}} = \frac{\sum_{i=1}^{n} d_i/n}{0.5/\sqrt{n/A}} = \frac{2\sum_{i=1}^{n} d_i}{\sqrt{nA}} \quad (5)$$

$$Z_{ANN} = \frac{\overline{D_O} - \overline{D_E}}{SD} \quad (6)$$

where $\overline{D_O}$ represents the observed average distance between each data point and its nearest neighbor. $\overline{D_E}$ means the expected average distance for the randomly distributed data points. d_i denotes the distance between data i and its nearest neighboring data. n corresponds to the total number of data, and A is the area of a minimum enclosing rectangle around all data points. If ANN is less than 1, the point pattern exhibits clustering. If ANN equals 1, there has no trend. If ANN is greater than 1, the trend is dispersion.

3.1.3. Spatial autocorrelation test

The Global Moran's I is a spatial autocorrelation test and can be used to evaluate the clustered, dispersed, or random spatial pattern in observations (Moran, 1948). This paper utilizes the I index to provide a reasonable reference for the neighborhood distance of bins used in the spatial–temporal analysis.

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \times C_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \times D(x_i)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \times (x_i - \overline{x})(x - \overline{x})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \times \frac{1}{n}\sum_{1}^{n}(x_i - \overline{x})^2} \quad (7)$$

$$Z_I = \frac{I - E(I)}{\sqrt{D(I)}} \quad (8)$$

where x_i is the attribute value of j spatial location/grid. $w_{ij} = 1$ if the jth grid is within the spatial neighborhood distance of the ith grid and 0 otherwise. C_{ij} denotes the attribute similarity matrix. $E(I) = -1/(n-1)$. $D(I) = E\left(I^2\right) - E(I)^2$. If I is positive and close to 1, it denotes the incremental spatial autocorrelation (clustered pattern) between neighborhoods; if I is equal to 0, it means a random pattern of features; and if I is less than 0, it represents a dispersed pattern of features.

3.2. Latent class random parameters model

3.2.1. Latent class clustering

The latent class clustering (LCC) is a statistical model-based approach that can classify the dataset into homogenous subsets by maximizing the heterogeneity between classes (Lanza, Collins, Lemmon, & Schafer, 2007). It is assumed that the LCC segments the whole dataset with J discrete category variables into M classes. The probability of response Y can be calculated as:

$$P(Y_i = j) = \sum_{m=1}^{M} \gamma_m \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{k,r_j|m}^{I(y_j=r_j)} \tag{9}$$

where each observation i contains J categorical variables, Y_i denotes the response of the observation i for J category, and Y_i = 1, 2, ...r_j. γ_m is the membership probability for latent class cluster m (m = 1, 2, ..., M). $\rho_{k,r_j|m}^{I(y_j=r_j)}$ represents the item-response probability that observation i has response r_j being conditioned on latent class membership m. ρ means 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.

To determine an appropriate number of classes, four commonly used criteria including Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criteria (BIC) and Entropy-based Measures (EM) are utilized (Song & Fan, 2020). Smaller values of the AIC, BIC, and CAIC indicate a better clustering result. Meanwhile, the EM indicates the information quality of the cluster and closing to 1 means a better clustering result (McLachlan & Peel, 2004).

3.2.2. Random parameter logit model

Following the LCC, random parameter logit model (or mixed logit model) is developed to further explore the unobserved heterogeneity for each segmented crash data. The utility function is defined as a linear function for individual i with severity level j:

$$U_{ij} = \beta_i X_{ij} + \xi_{ij} + \varepsilon_{ij} \tag{10}$$

where X_{ij} is a vector of independent variables, β_i denotes the corresponding parameter. ξ_{ij} represents the error component that has a general distribution correlated among severity levels and heteroscedastic for individuals, and ε_{ij} is the random term with independently and identically distributed Gumbel distribution over severity levels and individuals (McFadden & Train, 2000).

The probability of individual i for injury severity j is the integral of the condition choice probability $P_i(j|\xi_{ij})$ over the distribution of ξ_{ij}.

$$P_i(j|\xi_{ij}) = \frac{exp(\beta_i X_{ij} + \xi_{ij})}{\sum_{j=1}^{J} exp(\beta_i X_{ij} + \xi_{ij})} \tag{11}$$

Since the $P_i(j)$ does not always have a closed-form solution, 200 Halton draws are used in the simulation-based maximum likelihood method for parameter estimation. All the random parameters are assumed to be normally distributed since the parameters could be positive and negative. Marginal effects are used to illustrate the impact of the explanatory variable in the changing values of severity probability outcomes (Derr, 2013).

4. Data description

The data used in this paper are obtained from the North Carolina Department of Transportation (NCDOT), which include 33,707 pedestrian-vehicle crash observations in North Carolina from 2007 to 2018. The whole spatiotemporal analysis process uses a 5% significant level. The distance interval for the temporal trend analysis is set as 382 m, which is obtained by the average nearest neighbor test (ANN ratio: 0.286; Z-score: −250.845; P-value <0.0001). The inverse neighborhood distance is set as 8,000 m, which is determined by the empirical results of the spatial autocorrelation test and the total number of hotspots detected. As shown in Fig. 1, the Moran's Index at 8,000 m is 0.36, which denotes a clustering pattern within the neighborhood distance, and the z-score reaches 322. Meanwhile, the total number of hotspots reaches 5,810 with a change rate less than 5% after 8,000 m. A total number of 17,013 pedestrian-vehicle crashes at hotspots with upward annual trend are detected and is shown in Fig. 2.

To further model the contributing factors to pedestrian injury severity at hotspots with an upward trend, 13,303 pedestrian injury observations are filtered after selecting the pedestrian with the highest injury severity in single vehicle involved crashes and deleting observations with missing variables. The pedestrian severity is classified into three levels (i.e., fatal/incapacitating injury (F/I), non-incapacitating injury (NI), and no/possible injury (N/P)) by considering both the severity features and crash frequency. As shown in Appendix Table A1, explanatory variables are classified into the human, vehicle, crash, locality and roadway, environment and time, and traffic control categories.

5. Results and discussions

5.1. Latent class clustering results

The LCC is implemented to maximize the heterogeneity between the datasets. As shown in Fig. 3. The values of AIC, CAIC, and BIC all decrease with the increase of the class numbers, and the rate of change is less than 3% after four classes. Meanwhile, the entropy value for the 4-class model reaches a local maximum of 0.91, which is close to 1 and denotes a good segmentation of the data. Hence, this paper uses the LCC to segment the crash data into four latent classes.

All explanatory variables in Appendix Table A1 are utilized in the LCC analysis. Table 2 only shows the featured variables having a proportion that is significantly different from other latent classes, while other variables are not shown in Table 3 since the proportions of them are comparatively small and less descriptive. The combination of these featured variables is utilized as a latent variable/label to describe each class. For example, according to the feature variables in class 1, about 87.78% of the crashes happened in rural areas, 49.53% occurred in state secondary routes, 56.59% happened in dark without roadway lights, and 44.79% are set with double yellow lines, no passing zone sign. Hence, class 1 could be labeled as a condition of rural, state secondary route, dark without roadway lights, and double yellow line, no passing zone sign control. Similarly, class 2 can be specified as a circumstance of urban, public vehicular area, daylight, and without traffic control. Class 3 can be described as a scenario of urban, local street and driveway, daylight, and no traffic control. Class 4 can be defined as a situation of urban, dark with lighted roadway, and no traffic control.

5.2. Random parameter logit model results

After obtaining the LCC results, the heterogeneity within the crash is further investigated with four random parameter logit models. To obtain significant variables in each latent class, all explanatory variables are first utilized as the inputs in the random parameter logit model. The Chi-square test is applied as the selection criterion for both significant fixed variables and random parameter variables at a 5% significance level. Final variable coefficient estimation results are shown in Appendix Tables A2–A5.

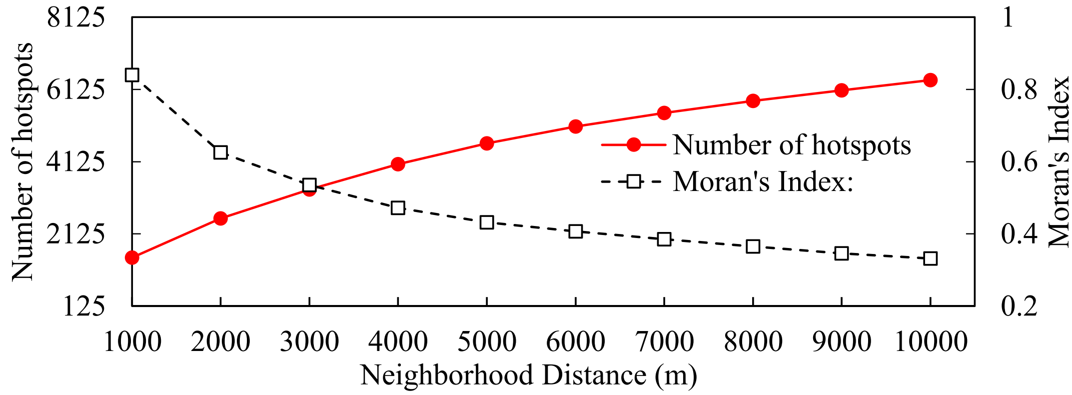

Fig. 1. Number of hotspots and the Moran's Index for different neighborhood distance.

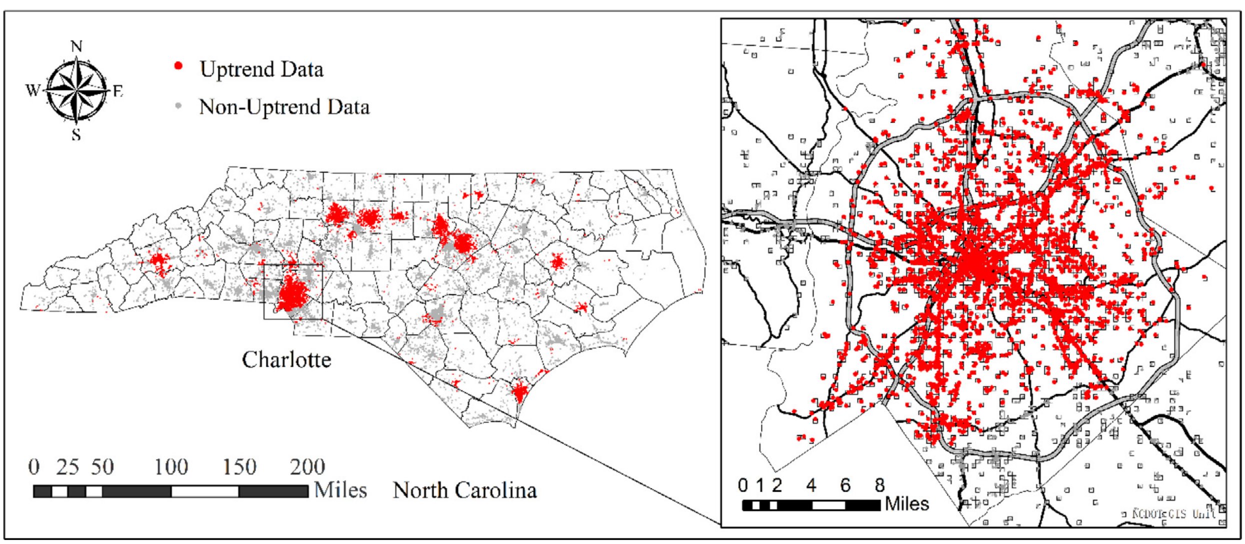

Fig. 2. Spatiotemporal trend analysis result for hotspots with upward trend in North Carolina.

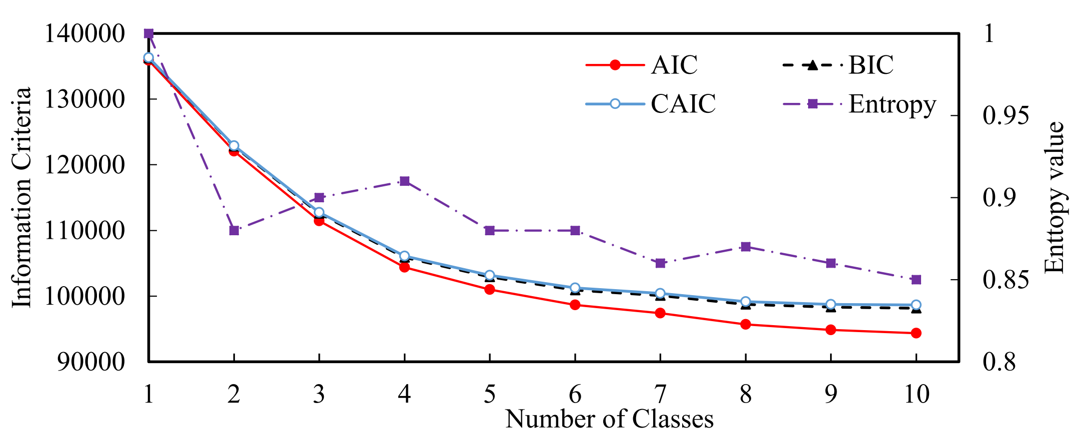

Fig. 3. Latent class results of AIC, BIC CAIC, and Entropy value for different class numbers.

Table 2
Distributions of featured variables (bold) and statistics for each latent class.

Variable	No.	Description	Class 1	Class 2	Class 3	Class 4
Locality	2	Urban	12.22%	**91.92%**	**97.64%**	**97.98%**
RdClass	4	State Secondary Route	**49.53%**	0%	0.05%	0.31%
	5	Local Street, Driveway	6.55%	9.97%	**93.95%**	**87.09%**
	6	Public Vehicular Area	1.16%	**89.89%**	1.14%	3.33%
LightCond	1	Daylight	37.05%	**77.62%**	**91.87%**	2.59%
	3	Dark – Lighted Roadway	3.06%	17.42%	2.28%	**70.86%**
	4	Dark – Roadway Not Lighted	**56.59%**	1.21%	0.22%	22.49%
Control	1	No Control Present	48.25%	**95.26%**	50.62%	66.68%
	4	Double Yellow Line, No Passing Zone	**44.79%**	0.15%	1.4%	1.7%

Table 3
Marginal effects of explanatory variables in class 1 and class 2.

Variable	Description	Class 1			Class 2		
	Severity level	F/I[a]	NI[b]	N/P[c]	F/I[a]	NI[b]	N/P[c]
PedAge2	24 < PedAge ≤ 54 (base: <24)				−0.019	−0.048	0.067
PedAge4	PedAge ≥ 65	0.231	−0.136	−0.095	−0.007	0.096	−0.090
DrvrVehTyp2	Middle (base: small)				0.022	0.029	−0.051
DrvrVehTyp3	Heavy	0.235	−0.139	−0.096	0.111	0.070	−0.181
AmbulanceR2	No ambulance rescue (base: yes)	−0.197	−0.185	0.382	−0.043	−0.143	0.186
CrashGrp2	Crossing roadway with vehicle not turning (base: walking along roadway)	0.160	−0.096	−0.064			
CrashGrp4	Off roadway				−0.010	0.142	−0.132
CrashGrp6	Dash/dart-out	0.221	−0.130	−0.090			
CrashGrp7	Backing vehicle	0.137	−0.325	0.188	−0.011	0.152	−0.142
CrashGrp10	Other/unusual circumstances				0.014	0.259	−0.272
Locality2	Urban (base: rural)				−0.025	0.009	0.016
Development2	Commercial (base: residential)				−0.015	−0.069	0.084
RdGrad2	Grade (base: level)	−0.040	0.091	−0.052			
RdClass2	Interstate (base: US route)	0.158	−0.094	−0.064			
RdClass5	Local street, driveway				0.038	0.054	−0.092
RdConfig3	Two-way, divided (base: one-way, not divided)	0.143	−0.004	−0.139			
LightCond4	Dark – roadway not lighted (base: daylight)	0.111	−0.067	−0.044			
Hour6	0:00–5:59 (base 6:00–9:59)				0.032	0.124	−0.156

Note:
[a] F/I – Fatal/ Incapacitating injury.
[b] NI – Non-incapacitating injury.
[c] N/P – No/Possible injury.

5.3. Marginal effects

The marginal effect results of the explanatory variables at a 5% significance level are shown in Tables 3 and 4. Variations of such impacts at the different severity levels and different latent classes are also detected. It is also noted that even though some variables have comparatively small proportions in the latent class subsets, they are still found to be a significant factor in the final results. The following subsections provide specific analyses and comparison for the impacts of factors on F/I and NI severity across different latent classes.

5.3.1. Pedestrian characteristics

Age and alcohol involvement are identified as significant variables. Compared to the young pedestrians (age <24), the results in classes1, 3, and 4 indicate an increasing tendency in which the probability of pedestrian suffering F/I and NI injury would both increase with the increase of the age stage. For example, the probability of pedestrians being F/I in class 4 increases from 0.035 to 0.086 and 0.175 with the increase of age stage. A similar result could be found in (Yasmin, Eluru, & Ukkusuri, 2014). However, class 2 shows heterogeneity results as the middle-age pedestrians (age within 24 to 54) and the elder pedestrians (age >65) could decrease the F/I injury by −0.019 to −0.007, respectively. Such heterogeneity within the demographical variables (age or gender) among different latent classes can also be supported by some previous studies (Abay, 2013; Aziz, Ukkusuri, & Hasan, 2013). Besides, for pedestrians under the influence of the alcohol, the possibility of pedestrians being F/I injured increases by 0.079 and 0.071 in classes 3 and 4, respectively. This result is in line with (Abay, 2013; Sasidharan, Wu, & Menendez, 2015), and more specific enforcements to limit/protect intoxicated pedestrians are needed.

5.3.2. Vehicle characteristics

This study shows that pedestrians are more vulnerable with the increase of the vehicle weight during pedestrian-vehicle crashes. For example, compared to small vehicles, the probability of pedestrians suffering F/I injury involving the middle and heavy vehicles in class 2 increases from 0.022 to 0.111, respectively.

5.3.3. Crash characteristics

Compared to ambulance rescue situations, situations without an ambulance rescue are found to have less F/I and NI injuries in all classes. This could be possibly explained by the situation under which people might not call the ambulance when the pedestrian has no/possible injury. For the hit and run situation, heterogeneous results show that there are a −0.024 probability decrease and a 0.033 probability increase for the F/I injury in classes 3 and 4,

Table 4
Marginal effects of the explanatory variables in class 3 and class 4.

Variable	Description	Class 3			Class 4		
	Severity level	F/I[a]	NI[b]	N/P[c]	F/I[a]	NI[b]	N/P[c]
PedAge2	24 < Pedage ≤ 54 (base: <24)				0.035	−0.021	−0.014
PedAge3	55 < Pedage ≤ 64	0.033	−0.015	−0.017	0.086	−0.050	−0.035
PedAge4	Pedage ≥ 65	0.082	−0.038	−0.044	0.175	−0.102	−0.073
PedAlcFlag2	PedAlcFlag = 'yes' (base: no)	0.079	0.083	−0.161	0.071	0.024	−0.095
DrvrVehTyp2	Middle (base: small)	0.034	−0.016	−0.018			
DrvrVehTyp3	Heavy				0.160	−0.093	−0.067
HitRun2	Hit and run (base: no)	−0.024	−0.072	0.096	0.033	−0.102	0.069
AmbulanceR2	No ambulance rescue (base: yes)	−0.066	−0.248	0.315	−0.081	−0.177	0.258
CrashGrp2	Crossing roadway with vehicle not turning (base: walking along roadway)				0.089	−0.011	−0.078
CrashGrp3	Crossing roadway with vehicle turning	−0.078	−0.063	0.142	−0.105	−0.067	0.172
CrashGrp5	Pedestrian in roadway				0.109	−0.064	−0.045
CrashGrp6	Dash/dart-out	−0.030	0.204	−0.174	0.092	0.016	−0.108
CrashGrp7	Backing vehicle				0.057	−0.177	0.119
CrashGrp8	Multiple threat/trapped	−0.022	0.150	−0.128			
CrashGrp9	Bus related vehicle	−0.030	0.206	−0.176	0.212	−0.123	−0.089
Locality2	Urban (base: rural)	0.023	−0.153	0.131			
Development2	Commercial (base: residential)				0.043	−0.026	−0.018
Development4	Institutional				0.046	−0.141	0.095
RdCurve2	Curve (base: straight)	0.046	−0.021	−0.024	0.104	−0.061	−0.043
RdGrad2	Grade (base: level)	0.023	−0.011	−0.012	0.076	−0.045	−0.031
RdClass2	Interstate (base: US route)	0.046	−0.272	0.226			
RdClass5	Local street, driveway	−0.031	−0.130	0.161	−0.106	0.063	0.044
RdClass6	Public vehicular area	−0.068	−0.249	0.317	−0.089	−0.100	0.189
RdConfig2	Two-way, not divided (base: one-way, not divided)				0.024	−0.074	0.050
RdConfig3	Two-way, divided	0.020	−0.010	−0.011	0.098	−0.057	−0.041
LightCond3	Dark – lighted roadway (base: daylight)	0.197	−0.091	−0.106			
LightCond4	Dark – roadway not lighted				0.061	−0.036	−0.025
Weather2	Cloudy (base: clear)				0.073	−0.043	−0.030
Hour2	10:00–14:59 (base 6:00–9:59)	−0.029	0.014	0.015			
Hour3	15:00–17:59	−0.018	0.009	0.010	−0.096	0.058	0.038
Hour5	21:00–23:59				−0.017	0.052	−0.035
Hour6	0:00–5:59				0.053	0.039	−0.092
TraffCntrl2	Signs (base: no control)	−0.016	−0.074	0.091	0.028	−0.086	0.059
TraffCntrl4	Double yellow line, no passing zone	0.084	−0.039	−0.045	0.093	−0.055	−0.039

Note:
[a] F/I – Fatal/ Incapacitating injury.
[b] NI – Non-incapacitating injury.
[c] N/P – No/Possible injury.

respectively. Meanwhile, both two classes show a probability decrease for NI injury and a probability increase for N/P injury.

For crash type factors, the situation when the pedestrian is walking along the roadway is set as the base. Pedestrians crossing the roadway with the vehicle not turning would result in a 0.16 and 0.089 probability increase for the F/I injury in classes 1 and 4, respectively. In comparison, results show a probability decrease of −0.078 and −0.105 for the F/I injury when crossing a roadway with the turning vehicle in classes 3 and 4, respectively. One possible reason for explaining such a difference might be that the speed of the vehicles is much lower when turning than when vehicles are traveling straight. Also, the situation when vehicles are off the roadway decreases the probability with −0.01 for F/I injury in class 2. This result is also in line with Kim, Ulfarsson, Shankar, and Mannering (2010) as the driver would reduce the speed when driving off the roadway. For the situation when the pedestrian is in the roadway, a 0.109 probability increase for the F/I injury is found in class 4. A similar conclusion was also drawn in (Mohamed, Saunier, Miranda-Moreno, & Ukkusuri, 2013). In multiple threat/trapped situation, a −0.022 probability decrease for the F/I injury and a 0.15 probability increase for the NI injury are identified.

Heterogeneities also exist in variables of dash/dash-out, backing vehicle, and bus-related cases across different latent classes. In the dash/dart-out case, heterogeneity results are found in the probability of F/I injury. While, the probability of N/P injury decreases in all classes, which indicate that a severe injury outcome would occur under the dash/dart-out situation, and the results are in accord with Sun et al. (2019). Also, for the backing vehicle case, there is a 0.137 and 0.057 probability increase for the F/I injury in classes 1 and 4, respectively, while case 2 shows a −0.011 probability decrease for the F/I injury. For the bus-related case, heterogeneity results are also observed in the F/I injury, while the probability decrease for the N/P injury indicates an increase in the severe outcome in bus-related crashes. All these heterogeneities indicate a need to analyze the influences of these factors under the specific scenario, which again shows the superiority of using latent class random parameter models.

5.3.4. Locality and roadway characteristics

Compared to the rural area, crashes that occurred in the urban area also show heterogeneous results for the F/I injury. Similar conclusions on such heterogeneity were also made in Li and Fan (2019a). Debates on whether pedestrian-vehicle crashes happened in the urban area are safer than those in the rural area could be found in past studies. Some scholars concluded that rural is more dangerous because of the higher speed of vehicles and lack of medical resource (Sasidharan et al., 2015; Ulak et al., 2017), while others argued that urban has more complex traffic conditions with sufficiently high speed for fatality (Sun et al., 2019). These two explanations could well illustrate the heterogeneity in severe injuries that occurred in complex urban areas. In regard to land development, heterogeneous results could also be found in commercial land. Both commercial and institutional land show about a 0.04 probability increase for F/I injury in class 4, and class 4 denotes a latent class of urban local street without traffic control.

Roadway alignment, class level, and settlement are three major significant factors within the category of roadway features detected in this study. Compared to the straight-road, the curve-road shows a 0.046 and a 0.104 probability increase for the F/I injury in classes 3 and 4, respectively. Results on the grade-road show that the probability of pedestrians being F/I injured is increased by 0.023 and 0.076 compared to the level-road in classes 3 and 4, respectively. These locations are accident-prone areas as the driver has bad sight condition and the vehicle is difficult to control. Similar results could be referred to Sasidharan et al. (2015). Compared to the U.S. route, results on the interstate roads indicates a 0.158 and 0.046 probability increase for the F/I injury in classes 1 and 3. Also, the public vehicular area shows a −0.068 and −0.089 probability decrease for the F/I injury in classes 3 and 4. The reason for this might be that the public vehicular area (e.g., parking lot) has much lower traveling speeds than the U.S. route (Li & Fan, 2019b). Heterogeneities are also found in local streets and driveways. Classes 3 and 4 decrease the probability of pedestrians being F/I injured by −0.031 and −0.106, respectively, while class 2 shows a 0.038 increase in the F/I injury.

5.3.5. Environment and time characteristics

Compared to the daylight environment, both with/without lighting in the dark environment increase the probability of pedestrians being F/I injured in classes 1, 2, and 3 by 0.111, 0.197, and 0.061, respectively. The significant possibility decrease of the F/I injury requires a better lighting facility in these hotspots and this finding is in accordance with Yasmin et al. (2014). In comparison with the clear weather, the cloudy weather situation increases the probability of pedestrians being F/I injured by 0.073. A similar result could be referred to Aziz et al. (2013), and one possible reason for this is the decrease of sight in cloudy condition.

Though previous research has already pointed out the positive correlation between the vehicle-vehicle crash injury severity level with the peak hour (Mohamed et al., 2013), pedestrian-vehicle crash frequency in this study does not show a significant difference between the peak and non-peak hours as vehicle-vehicle crash does. Hence, this paper categorizes the crash time into six different periods mainly according to different features of the light condition, the frequency of the total crashes, and frequency of F/I injuries. Compared to the "morning" period (6:00–9:59), the "early morning" (0:00–5:59) shows a 0.053 probability increase for the F/I injury in class 4. Similar conclusions are showed in Haleem, Alluri, and Gan (2015). The "noon" (10:00–14:59), "afternoon" (15:00–17:59), and "early night" (21:00–23:59) in class 3 and class 4 all show a probability decrease of pedestrians being F/I injured. Furthermore, comparing the periods within class 3, results show that the afternoon hour has a higher probability of pedestrians being F/I injured than the noon hour. Result in class 4 indicates that pedestrian-vehicle crashes during the early night hour have a higher probability of being F/I injured for pedestrians compared to the afternoon hour. The increase of the F/I injury in the early morning might be caused by the combined impacts of dark environment, high speed, and fatigue of the driver in the early morning.

5.3.6. Traffic control characteristics

Compared to the situation of no traffic control, heterogeneity is found under the traffic sign control situation. Results on traffic sign control indicate a 0.016 probability decrease for the F/I injury in class 3, and a 0.028 probability increase for the F/I injury in class 4. The downward tendency of the probability of the NI injury in classes 3 and 4 is also detected. There are debates on the heterogeneous effects of traffic sign control on the safety of pedestrians. Kim et al. (2010) observed heterogeneity for traffic sign control in pedestrian fatalities, and the correlation between pedestrian age and traffic sign was detected. Possible explanations for such a difference in the effect of this factor could be concluded as follows: (a) the mitigatory outcome might result from the warning function of the traffic signs; and (b) the deteriorative result might be the consequence of the dangerous and complex environment where the traffic signs were installed.

6. Conclusions

This study explores factors of pedestrian-injury severity in pedestrian-vehicle crashes at hotspots with an upward trend considering the heterogeneity within and between the datasets. Twelve years of the police-reported pedestrian-vehicle crash data from 2007 to 2018 in North Carolina are used. Spatiotemporal trend analysis combined with the average nearest neighbor analysis and the spatial autocorrelation test are implemented to test the spatial clustering pattern and the temporal tendency of the crashes. The latent class clustering and four random parameter logit models are implemented to further investigate the heterogeneity within each class. Marginal effects are further calculated for better interpreting the impacts of categorical variables on the severity levels.

The random parameter variables detected across observations and the heterogeneous results between the subgroups indicate the superiority of combining the latent class clustering with random parameter logit models. Significant impacts of pedestrian behaviors, such as dash/dart-out and crossing or staying in the roadway, also require more attention to improve the transportation facilities to provide better protection for pedestrians. Meanwhile, there is a need to strengthen law enforcement and education to prohibit playing in roadways, crossing divided roadways without permission, and drunk walking in/across the roadways. Also, more appropriate traffic control management, such as adjusting the signal phase to decrease the behavior of crossing with the red light, is needed for both drivers and pedestrians. Besides, the zebra crossing sign could be equipped with flashing lights to alert the driver when the pedestrian is crossing since early night hour (0:00–5:59) is found to be the most dangerous period for pedestrians. Furthermore, a patrol route considering hotspots with an upward trend could help to reduce the response time to reach crash locations.

This paper provides a framework for researchers and engineers to identify crash hotspots considering spatiotemporal patterns and explore contribution factors to crashes considering unobserved heterogeneity. However, the temporal fluctuations may still exist in different time scales and may be caused by different factors such as the global recession (Behnood & Mannering, 2016). Further studies are still needed to investigate the heterogeneities within the time–space scale, spatial and temporal correlations of the factors, and the temporal fluctuation and instability of the crash data.

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).

References

Abay, K. A. (2013). Examining pedestrian-injury severity using alternative disaggregate models. *Research in Transportation Economics, 43*(1), 123–136. https://doi.org/10.1016/j.retrec.2012.12.002.

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention, 41*(3), 359–364. https://doi.org/10.1016/j.aap.2008.12.014.

Aziz, H. M. A., Ukkusuri, S. V., & Hasan, S. (2013). Exploring the determinants of pedestrian–vehicle crash severity in New York City. *Accident Analysis & Prevention, 50*, 1298–1309. https://doi.org/10.1016/j.aap.2012.09.034.

Behnood, A., & Mannering, F. L. (2016). An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. *Analytic Methods in Accident Research, 12*, 1–17. https://doi.org/10.1016/j.amar.2016.07.002.

Blazquez, C. A., Picarte, B., Calderón, J. F., & Losada, F. (2018). Spatial autocorrelation analysis of cargo trucks on highway crashes in Chile. *Accident Analysis and Prevention, 120*(August), 195–210. https://doi.org/10.1016/j.aap.2018.08.022.

Chen, Z., & Fan, W. (. (2019). A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. *International Journal of Transportation Science and Technology, 8*(1), 43–52. https://doi.org/10.1016/j.ijtst.2018.10.001.

Dai, D. (2012). Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment. *Journal of Transport Geography, 24*, 206–214. https://doi.org/10.1016/j.jtrangeo.2012.02.005.

Derr, B. (2013). Ordinal response modeling with the LOGISTIC procedure. *SAS Global Forum*, 1–20.

Ebdon, D. (1985). *Statistics in geography*. Blackwell.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Gudes, O., Varhol, R., Sun, Q. (., & Meuleners, L. (2017). Investigating articulated heavy-vehicle crashes in western Australia using a spatial approach. *Accident Analysis & Prevention, 106*, 243–253. https://doi.org/10.1016/j.aap.2017.05.026.

Haleem, K., Alluri, P., & Gan, A. (2015). Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accident Analysis & Prevention, 81*, 14–23. https://doi.org/10.1016/j.aap.2015.04.025.

Hu, L., Wu, X., Huang, J., Peng, Y., & Liu, W. (2020). Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China. *Safety Science, 127*, 104710. https://doi.org/10.1016/j.ssci.2020.104710.

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

Kendall, M. G., Gibbons, J. D. (1990). Rank correlation methods (fifth ed.).

Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis & Prevention, 42*(6), 1751–1758. https://doi.org/10.1016/j.aap.2010.04.016.

Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 671–694. https://doi.org/10.1080/10705510701575602.

Li, D., Ranjitkar, P., Zhao, Y., Yi, H., & Rashidi, S. (2017). Analyzing pedestrian crash injury severity under different weather conditions. *Traffic Injury Prevention, 18*(4), 427–430. https://doi.org/10.1080/15389588.2016.1207762.

Li, Y., & Fan, W. (David) (2019a). Modelling severity of pedestrian-injury in pedestrian-vehicle crashes with latent class clustering and partial proportional odds model: A case study of North Carolina. *Accident Analysis and Prevention, 131*, 284–296. https://doi.org/10.1016/j.aap.2019.07.008.

Li, Y., & Fan, W. (David) (2019b). Pedestrian injury severities in pedestrian-vehicle crashes and the partial proportional odds logit model: Accounting for age difference. *Transportation Research Record, 2673*(5), 731–746. https://doi.org/10.1177/0361198119842828.

Liu, J., Hainen, A., Li, X., Nie, Q., & Nambisan, S. (2019). Pedestrian injury severity in motor vehicle crashes: An integrated spatio-temporal modeling approach. *Accident Analysis & Prevention, 132*, 105272. https://doi.org/10.1016/j.aap.2019.105272.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica, 13*(3), 245. https://doi.org/10.2307/1907187.

Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research, 17*, 1–13. https://doi.org/10.1016/j.amar.2017.10.002.

Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research, 1*, 1–22. https://doi.org/10.1016/j.amar.2013.09.001.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*(5), 447–470. https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1.

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. *Safety Science, 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. *Institute of Transportation Engineers. ITE Journal, 89*(10), 25–33.

Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2020). Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. *International Journal of Transportation Science and Technology*, 1–16. https://doi.org/10.1016/j.ijtst.2020.01.001.

Moran, P. A. P. (1948). The interpretation of statistical maps. In Journal of the Royal Statistical Society. Series B (Methodological) (Vol. 10, pp. 243–251). Wiley Royal Statistical Society. https://doi.org/10.2307/2983777.

Mujalli, R. O., Garach, L., López, G., & Al-Rousan, T. (2019). Evaluation of injury severity for pedestrian-vehicle crashes in Jordan using extracted rules. *Journal of Transportation Engineering Part A: Systems, 145*(7), 1–13. https://doi.org/10.1061/JTEPBS.0000244.

NHTSA. (2019). Traffic safety facts, 2017 data: Pedestrians. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812681.

Ouni, F., & Belloumi, M. (2018). Spatio-temporal pattern of vulnerable road user's collisions hot spots and related risk factors for injury severity in Tunisia. *Transportation Research Part F: Traffic Psychology and Behaviour, 56*, 477–495. https://doi.org/10.1016/j.trf.2018.05.003.

Plug, C., Xia, J., & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis & Prevention, 43*(6), 1937–1946. https://doi.org/10.1016/j.aap.2011.05.007.

Sasidharan, L., Wu, K. F., & Menendez, M. (2015). Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. *Accident Analysis & Prevention, 85*, 219–228. https://doi.org/10.1016/j.aap.2015.09.020.

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. *Accident Analysis & Prevention, 144*, 105638. https://doi.org/10.1016/j.aap.2020.105638.

Songchitruksa, P., & Zeng, X. (2010). Getis-ord spatial statistics to identify hot spots by using incident management data. *Transportation Research Record, 2165*, 42–51. https://doi.org/10.3141/2165-05.

Sun, M., Sun, X., & Shan, D. (2019). Pedestrian crash analysis with latent class clustering method. *Accident Analysis & Prevention, 124*, 50–57. https://doi.org/10.1016/j.aap.2018.12.016.

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. *Transportation Research Record, 2635*(1), 90–97. https://doi.org/10.3141/2635-11.

Wang, C. W., & Chan, C. L. W. (2016). Estimated trends and patterns of road traffic fatalities in China, 2002–2012. *Traffic Injury Prevention, 17*(2), 164–169. https://doi.org/10.1080/15389588.2015.1045501.

Yalcin, G., & Sebnem Duzgun, H. S. (2015). Spatial analysis of two-wheeled vehicles traffic crashes: Osmaniye in Turkey. *KSCE Journal of Civil Engineering, 19*(7), 2225–2232. https://doi.org/10.1007/s12205-015-0661-0.

Yasmin, S., Eluru, N., & Ukkusuri, S. V. (2014). Alternative ordered response frameworks for examining pedestrian injury severity in New York City. *Journal of Transportation Safety & Security, 6*(4), 275–300. https://doi.org/10.1080/19439962.2013.839590.

Li Song received the M.S. degree in transportation engineering from the Harbin Institute of Technology. He is currently a Ph.D. Research Assistant of the USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE) at the Department of Civil and Environmental Engineering in University of North Carolina at Charlotte. His research interests include traffic safety and traffic control.

Wei (David) Fan received the Ph.D. degree in transportation engineering from the University of Texas at Austin. He is currently the Director of the USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE) and the Professor at the Department of Civil and Environmental Engineering in University of North Carolina at Charlotte. His research interests include big data analytics for transportation, connected and autonomous vehicles, multimodal transportation and shared mobility; traffic system operation and control, and transportation system analysis and network modeling.

Yang Li received the M.S. degree in transportation engineering from the Clemson University. He is currently a Ph.D. Research Assistant in the USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE) at the Department of Civil and Environmental Engineering in University of North Carolina at Charlotte.His research interests include traffic safety and data analysis.

Peijie Wu received the B.S. degree in transportation engineering from Chongqing Jiaotong University. She is currently a Ph.D. student at the School of Transportation Science and Engineering in Harbin Institute of Technology. Her research interests include traffic safety and data analysis.

Appendix A

Table A1
Statistics of explanatory variables for pedestrian-vehicle crashes at hotspots with an upward trend.

Variable	Description		Total	F/I[a]	NI[b]	N/P[c]
	Number of observations		13303	1415(10.64%)	5046(37.93%)	6842(51.43%)
Pedestrian Characteristics						
PedAge	**PedAge ≤ 24**	1	4305	395(9.18%)	1788(41.53%)	2122(49.29%)
	24 < PedAge ≤ 54	2	6459	680(10.53%)	2330(36.07%)	3449(53.4%)
	55 < PedAge ≤ 64	3	1299	164(12.63%)	449(34.57%)	686(52.81%)
	PedAge ≥ 65	4	1240	176(14.19%)	479(38.63%)	585(47.18%)
PedSex	**Male**	1	7593	972(12.8%)	2978(39.22%)	3643(47.98%)
	Female	2	5710	443(7.76%)	2068(36.22%)	3199(56.02%)
PedAlcFlag	**PedAlcFlag = ’No’**	1	11867	1026(8.65%)	4416(37.21%)	6425(54.14%)
	PedAlcFlag = ’Yes’	2	1436	389(27.09%)	630(43.87%)	417(29.04%)
Vehicle Type						
DrvrVehTyp	**Small**	1	7998	754(9.43%)	3045(38.07%)	4199(52.5%)
	Middle	2	4940	593(12%)	1865(37.75%)	2482(50.24%)
	Heavy	3	365	68(18.63%)	136(37.26%)	161(44.11%)
Crash Characteristics						
AmbulanceR	**Ambulance Rescue**	1	9880	1287(13.03%)	4225(42.76%)	4368(44.21%)
	No Ambulance Rescue	2	3423	128(3.74%)	821(23.98%)	2474(72.28%)
HitRun	**No Hit and Run**	1	11653	1279(10.98%)	4494(38.57%)	5880(50.46%)
	Hit and Run	2	1650	136(8.24%)	552(33.45%)	962(58.3%)
CrashGrp	**Walking Along Roadway**	1	845	117(13.85%)	346(40.95%)	382(45.21%)
	Crossing Roadway with Vehicle Not Turning	2	2692	502(18.65%)	1079(40.08%)	1111(41.27%)
	Crossing Roadway with Vehicle Turning	3	2092	60(2.87%)	718(34.32%)	1314(62.81%)
	Off Roadway	4	1632	59(3.62%)	486(29.78%)	1087(66.61%)
	Pedestrian in Roadway	5	696	135(19.4%)	256(36.78%)	305(43.82%)
	Dash/Dart-Out	6	1182	175(14.81%)	632(53.47%)	375(31.73%)
	Backing Vehicle	7	1325	54(4.08%)	342(25.81%)	929(70.11%)
	Multiple Threat/Trapped	8	214	15(7.01%)	113(52.8%)	86(40.19%)
	Bus related Vehicle	9	134	12(8.96%)	64(47.76%)	58(43.28%)
	Other/Unusual Circumstances	10	2491	286(11.48%)	1010(40.55%)	1195(47.97%)
Locality and Roadway Characteristics						
Locality	**Rural**	1	1282	276(21.53%)	485(37.83%)	521(40.64%)
	Urban	2	12021	1139(9.48%)	4561(37.94%)	6321(52.58%)
Development	**Residential**	1	4584	484(10.56%)	1897(41.38%)	2203(48.06%)
	Commercial	2	7705	768(9.97%)	2769(35.94%)	4168(54.09%)
	Industrial	3	76	3(3.95%)	37(48.68%)	36(47.37%)
	Institutional	4	485	27(5.57%)	172(35.46%)	286(58.97%)
	Farms, Woods, Pastures	5	453	133(29.36%)	171(37.75%)	149(32.89%)
RdCurve	**Straight**	1	12770	1306(10.23%)	4846(37.95%)	6618(51.82%)
	Curve	2	533	109(20.45%)	200(37.52%)	224(42.03%)
RdGrad	Level	1	10996	1055(9.59%)	4116(37.43%)	5825(52.97%)
	Grade	2	1718	277(16.12%)	685(39.87%)	756(44%)
	Hillcrest	3	502	66(13.15%)	203(40.44%)	233(46.41%)
	Bottom	4	87	17(19.54%)	42(48.28%)	28(32.18%)
RdClass	**US Route**	1	415	128(30.84%)	170(40.96%)	117(28.19%)
	Interstate	2	241	103(42.74%)	72(29.88%)	66(27.39%)
	State Route	3	329	78(23.71%)	138(41.95%)	113(34.35%)
	State Secondary Route	4	465	100(21.51%)	196(42.15%)	169(36.34%)
	Local Street, Driveway	5	8479	882(10.4%)	3521(41.53%)	4076(48.07%)
	Public Vehicular Area	6	3374	124(3.68%)	949(28.13%)	2301(68.2%)

(*continued on next page*)

Appendix A (*continued*)

Variable	Description		Total	F/I[a]	NI[b]	N/P[c]
RdConfig	**One-Way, Not Divided**	1	1275	69(5.41%)	421(33.02%)	785(61.57%)
	Two-Way, Not Divided	2	9119	820(8.99%)	3395(37.23%)	4904(53.78%)
	Two-Way, Divided	3	2909	526(18.08%)	1230(42.28%)	1153(39.64%)
Environment and Temporal Characteristics						
LightCond	**Daylight**	1	7977	519(6.51%)	2923(36.64%)	4535(56.85%)
	Dawn/Dusk Light	2	604	62(10.26%)	207(34.27%)	335(55.46%)
	Dark – Lighted Roadway	3	3336	456(13.67%)	1390(41.67%)	1490(44.66%)
	Dark – Roadway Not Lighted	4	1386	378(27.27%)	526(37.95%)	482(34.78%)
Weather	**Clear**	1	10242	1075(10.5%)	3909(38.17%)	5258(51.34%)
	Cloudy	2	1800	204(11.33%)	643(35.72%)	953(52.94%)
	Rain	3	1141	125(10.96%)	440(38.56%)	576(50.48%)
	Snow, Sleet, Hail, Freezing Rain/Drizzle	4	80	4(5%)	38(47.5%)	38(47.5%)
	Fog, Smog, Smoke	5	40	7(17.5%)	16(40%)	17(42.5%)
Hour	**6:00–9:59**	1	1889	174(9.21%)	709(37.53%)	1006(53.26%)
	10:00–14:59	2	3252	199(6.12%)	1107(34.04%)	1946(59.84%)
	15:00–17:59	3	2805	200(7.13%)	1064(37.93%)	1541(54.94%)
	18:00–20:59	4	2668	330(12.37%)	1036(38.83%)	1302(48.8%)
	21:00–23:59	5	1598	269(16.83%)	671(41.99%)	658(41.18%)
	0:00–5:59	6	1091	243(22.27%)	459(42.07%)	389(35.66%)
Traffic Control Type						
TraffCntrl	**No Control Present**	1	8876	996(11.22%)	3351(37.75%)	4529(51.03%)
	Signs	2	1126	73(6.48%)	367(32.59%)	686(60.92%)
	Signal	3	2613	220(8.42%)	1056(40.41%)	1337(51.17%)
	Double Yellow Line, No Passing Zone	4	548	119(21.72%)	230(41.97%)	199(36.31%)
	Human Control	5	140	7(5%)	42(30%)	91(65%)

Note: variables in bold and numbered with 1 are set as the base for the explanatory variables.
[a] F/I – Fatal/Incapacitating injury. [b]NI – Non-incapacitating injury. [c] N/P – No/Possible injury.

Table A2
Random parameter logit model's significant variable coefficients for class 1.

Variable	Description	F/I[a]		NI[b]	
		Coef.	t value	Coef.	t value
Intercept		−1.0814	−5.05	0.3277	3.11
PedAge4	PedAge $\geq$ 65	1.2747	3.68		
PedAlcFlag2	PedAlcFlag = 'Yes'	0.4824	1.92		
DrvrVehTyp3	Heavy	1.2946	2.97		
HitRun2	Hit and Run	−0.8608	−1.89		
AmbulanceR2	No ambulance rescue	−2.2402	−5.24	−1.5076	−6.97
CrashGrp2	Crossing roadway with vehicle not turning	0.9042	3.32		
CrashGrp6	Dash/Dart-Out	1.2366	3.46		
CrashGrp7	Backing vehicle			−2.2548	−2
Locality2	Urban	−0.6561	−1.8		
RdGrad2	Grade			0.4057	2.11
RdClass2	Interstate	0.8898	2.41		
RdConfig3	Two-Way, Divided	1.1638	3.57	0.5615	2.52
LightCond4	Dark – Roadway Not Lighted	0.7111	2.85		
Std. dev.		1.2284	1.81		

Note: Number of observations: 901. Log-likelihood at convergence: −862.41. Log-likelihood (constant only): −989.85.
[a] F/I – Fatal/Incapacitating injury. [b] NI – Non-incapacitating injury.

Table A3
Random parameter logit model's significant variable coefficients for class 2.

Variable	Description	F/I[a]		NI[b]	
		Coef.	t value	Coef.	t value
Intercept		−1.9696	−5.98	−1.395	−4.94
PedAge2	24 < PedAge ≤ 54	−0.6698	−3.34	−0.309	−3.09
PedAge4	PedAge ≥ 65			0.4966	3.84
DrvrVehTyp2	Middle	0.681	3.41	0.2114	2.34
DrvrVehTyp3	Heavy	1.9061	4.97	0.6345	2.61
AmbulanceR2	No Ambulance Rescue	−1.9317	−6.46	−0.9216	−8.56
CrashGrp4	Off Roadway			0.7939	2.35
Std. dev.				1.3123	2.3
CrashGrp7	Backing Vehicle			0.8114	3
CrashGrp10	Other/Unusual Circumstances	0.9622	4.6	1.3561	4.97
Locality2	Urban	−0.5765	−2.05		
Development2	Commercial	−0.5448	−2.63	−0.4106	−4.09
RdClass5	Local Street, Driveway	0.9522	3.09	0.3767	2.32
Weather2	Cloudy	−0.5732	−1.79		
Hour6	0:00–5:59	1.0055	2.92	0.7169	3.92
TraffCntrl4	Double Yellow Line, No Passing Zone	2.463	1.95		

Note: Number of observations: 3517. Log-likelihood at convergence: −2432. Log-likelihood (constant only): −3864.

Table A4
Random parameter logit model's significant variable coefficients for class 3.

Variable	Description	F/I[a]		NI[b]	
		Coef.	t value	Coef.	t value
Intercept		−0.7485	−3.1	1.4015	4.23
PedAge3	55 < PedAge ≤ 64	0.4373	2.48		
PedAge4	PedAge ≥ 65	0.9448	5.5		
PedAlcFlag2	PedAlcFlag = 'Yes'	1.2401	5.04	0.6698	2.3
DrvrVehTyp2	Middle	0.4965	4.23		
HitRun2	Hit and Run	−0.6081	−2.67	−0.4504	−2.4
AmbulanceR2	No Ambulance Rescue	−1.8857	−8.92	−1.5342	−5
CrashGrp3	Crossing Roadway with Vehicle Turning	−1.7686	−9.54	−0.5091	−3.74
CrashGrp6	Dash/Dart-Out			0.9431	3.69
CrashGrp8	Multiple Threat/Trapped			0.7118	2.47
CrashGrp9	Bus related Vehicle			0.973	2.3
Locality2	Urban			−0.7269	−2.86
Std. dev.				2.1824	2.86
Development5	Farms, Woods, Pastures	3.7492	3.47	2.2967	1.71
RdCurve2	Curve	0.5734	2.5		
RdGrad2	Grade	0.3173	2.17		
RdClass2	Interstate			−1.9342	−2.84
RdClass5	Local Street, Driveway	−0.8247	−3.69	−0.7765	−2.96
RdClass6	Public Vehicular Area	−2.7034	−2.64	−1.875	−2.72
RdConfig3	Two-Way, Divided	0.2928	2.29		
LightCond2	Dawn/Dusk Light			−0.3717	−1.66
LightCond3	Dark – Lighted Roadway	1.728	4.33		
Weather4	Snow, Sleet, Hail, Freezing Rain/Drizzle			1.6402	1.92
Hour2	10:00–14:59	−0.4597	−3.21		
Hour3	15:00–17:59	−0.2804	−1.98		
TraffCntrl2	Signs	−0.4593	−2.24	−0.4451	−2.57
TraffCntrl4	Double Yellow Line, No Passing Zone	0.9243	2.56		

Note: Number of observations: 5255. Log-likelihood at convergence: −4358. Log-likelihood (constant only): −5773.

Table A5
Random parameter logit model's significant variable coefficients for class 4.

Variable	Description	F/I[a]		NI[b]	
		Coef.	t value	Coef.	t value
Intercept		−1.7074	−8.05	0.3974	3.75
PedAge2	24<PedAge $\leq$ 54	0.28	2.27		
PedAge3	55<PedAge $\leq$ 64	0.6125	3.44		
PedAge4	PedAge $\geq$ 65	1.1367	5.39		
PedAlcFlag2	PedAlcFlag = 'Yes'	0.7554	6.28	0.365	3.1
Std. dev.				1.3175	1.93
DrvrVehTyp3	Heavy	1.0348	2.8		
HitRun2	Hit and Run			−0.4551	−2.08
Std. dev.				1.5665	1.88
AmbulanceR2	No Ambulance Rescue	−1.3145	−8.11	−1.1283	−9.9
CrashGrp2	Crossing Roadway with Vehicle Not Turning	0.8225	5.94	0.2264	2.21
CrashGrp3	Crossing Roadway with Vehicle Turning	−1.4071	−5.24	−0.6261	−4.98
CrashGrp5	Pedestrian in Roadway	0.7605	4.16		
CrashGrp6	Dash/Dart-Out	0.9161	4.9	0.4135	2.88
CrashGrp7	Backing Vehicle			−0.843	−2.81
CrashGrp9	Bus related Vehicle	1.317	2.17		
Development2	Commercial	0.3504	3.08		
Development4	Institutional			−0.6541	−2.42
RdCurve2	Curve	0.717	3.13		
RdGrad2	Grade	0.5511	4.07		
RdClass5	Local Street, Driveway	−0.7426	−5.06		
RdClass6	Public Vehicular Area	−1.2944	−3.36	−0.7613	−2.59
RdConfig2	Two-Way, Not Divided			−0.3195	−3.71
RdConfig3	Two-Way, Divided	0.7398	6.45		
LightCond4	Dark – Roadway Not Lighted	0.4588	3.98		
Weather2	Cloudy	0.534	3.74		
Hour3	15:00–17:59	−0.9898	−3.06		
Hour5	21:00–23:59			0.2262	2.56
Hour6	0:00–5:59	0.6473	5.02	0.3936	3.34
TraffCntrl2	Signs			−0.3854	−2.59
TraffCntrl4	Double Yellow Line, No Passing Zone	0.6512	1.99		

Note: Number of observations: 3630. Log-likelihood at convergence: −3361. Log-likelihood (constant only): −3988.