Journal of Safety Research 76 (2021) 101–117

Contents lists available at ScienceDirect

Journal of Safety Research

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

Exploring bicyclist injury severity in bicycle-vehicle crashes using latent class clustering analysis and partial proportional odds models

Zijing Lin [a], Wei (David) Fan [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] 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 3261, 9201 University City Boulevard, Charlotte, NC 28223-0001, United States

A R T I C L E I N F O

Article history:
Received 11 June 2020
Received in revised form 17 July 2020
Accepted 23 November 2020
Available online 13 December 2020

Keywords:
Bicycle-involved crashes
Injury severity
Latent class analysis
Partial proportional odds models

A B S T R A C T

Introduction: Bicyclists are more vulnerable compared to other road users. Therefore, it is critical to investigate the contributing factors to bicyclist injury severity to help provide better biking environment and improve biking safety. According to the data provided by National Highway Traffic Safety Administration (NHTSA), a total of 8,028 bicyclists were killed in bicycle-vehicle crashes from 2007 to 2017. The number of fatal bicyclists had increased rapidly by approximately 11.70% during the past 10 years (NHTSA, 2019). *Methods:* This paper conducts a latent class clustering analysis based on the police reported bicycle-vehicle crash data collected from 2007 to 2014 in North Carolina to identify the heterogeneity inherent in the crash data. First, the most appropriate number of clusters is determined in which each cluster has been characterized by the distribution of the featured variables. Then, partial proportional odds models are developed for each cluster to further analyze the impacts on bicyclist injury severity for specific crash patterns. *Results:* Marginal effects are calculated and used to evaluate and interpret the effect of each significant explanatory variable. The model results reveal that variables could have different influence on the bicyclist injury severity between clusters, and that some variables only have significant impacts on particular clusters. *Conclusions:* The results clearly indicate that it is essential to conduct latent class clustering analysis to investigate the impact of explanatory variables on bicyclist injury severity considering unobserved or latent features. In addition, the latent class clustering is found to be able to provide more accurate and insightful information on the bicyclist injury severity analysis. *Practical Applications:* In order to improve biking safety, regulations need to be established to prevent drinking and lights need to be provided since alcohol and lighting condition are significant factors in severe injuries according to the modeling results.

1. Introduction

Compared to other transportation modes, cycling is considered to be an environmentally friendly and sustainable means of travelling since it can help relieve the congestion pressure (Behnood & Mannering, 2017), reduce energy consumption and emissions (Pucher et al., 2011), and provide potential benefits in terms of environment, health, and society (Rojas-Rueda et al., 2011; Xia et al., 2013; Kelly et al., 2014; Macmillan et al., 2014; Götschi et al., 2016). Therefore, city planners and policy makers have been continuously encouraging and promoting cycling, and improving the bicycle facilities in order to construct a bicycle-friendly city and provide better cycling environment (Nabors et al., 2012).

As cycling has become more popular among citizens, especially for recreation and short-distance commuting trips (Klassen et al., 2014), there are certain issues that need to be addressed. One of the most critical concerns is cycling safety, which is highly associated with the fact that bicyclists are more vulnerable in comparison to other road users (Vanparijs, et al., 2015; Nilsson et al., 2017). From 2007 to 2017, there were 8,028 bicyclists killed in bicycle-motor vehicle crashes. The number of fatalities has increased by approximately 11.70% for the past 10 years. And in 2017, 50,000 bicyclists were injured accounting for 1.82% of total injuries in traffic crashes (NHTSA, 2019). In addition, bicyclists are found to be fatally injured with high probability especially in the United States (Pucher & Dijkstra 2003). Based on this situation, it is essential to identify and analyze the contributing factors to

* Corresponding author.
E-mail addresses: zlin4@uncc.edu (Z. Lin), wfan7@uncc.edu (Wei (David) Fan).

https://doi.org/10.1016/j.jsr.2020.11.012
0022-4375/

bicyclist injury severity resulting from bicycle-vehicle crashes using reasonable and appropriate modeling methods.

This paper aims to investigate the potential factors that significantly affect bicyclist injury severity using latent class clustering analysis and the partial proportional odds (PPO) model. The police reported bicycle-involved crash data collected from 2007 to 2014 in North Carolina were used in this study. Information including bicyclist, driver, vehicle, crash, roadway, temporal, and environmental characteristics was recorded in the database. Based on the crash data, latent class clustering analysis is conducted first to separate the whole data into homogenous segmentations, and PPO models are developed within each cluster to model the bicyclist injury severity. The contribution of this paper is the application of latent class clustering analysis to reduce the heterogeneity revealed in the bicycle-motor vehicle crash data and the method of combining it with a PPO model to explore the significant contributing factors to each type of crashes.

The remainder of this paper is structured as follows. Section 2 reviews previous studies on the relevant research topics. Section 3 describes the bicycle-motor vehicle crash data collected for this research study and the explanatory variables considered in bicyclist injury severity analysis. Section 4 explains the methodology used for this study including latent class clustering analysis and partial proportional odds model. Section 5 discusses the model results in detail. Finally, Section 6 summarizes this paper with a conclusion and provides recommendations for future work.

2. Literature review

Many researchers have conducted studies on the impact of different explanatory variables on bicyclist injury severity utilizing discrete choice models (Kim et al., 2007; Eluru et al., 2008; Yan et al., 2011; Kröyer, 2015; Behnood & Mannering, 2017; Chen et al., 2017; Robartes & Chen, 2017). However, some underlying conditions might exist in traffic crashes due to the heterogeneity, which results from the unobserved impact factors that cannot be reported or revealed from the collected data (Valent et al., 2002; Ulfarsson & Mannering, 2004; Pai & Saleh, 2007). This issue makes it difficult to analyze and evaluate the effects of significant factors on bicyclist injury severity resulting from such traffic crashes. In addition, the model bias cannot be neglected, which might lead to inaccurate conclusions (Shaheed & Gkritza, 2014; Mannering & Bhat, 2014). To overcome this problem, researchers have applied the segmentation method to resolve the heterogeneity issue by concentrating on specific crashes including crashes that occurred at different locations (Moore et al., 2011; Rash-ha Wahi et al., 2018; Lin & Fan, 2019a, 2019b), certain types of crashes (Decker et al., 2016), and different age groups (Gong & Fan, 2017; Li & Fan, 2019). However, the segmentation method mentioned above is usually based on the research need or the experience from the previous studies, which may not be able to guarantee the homogenous groups of the data (Depaire et al., 2008). Therefore, cluster analysis has been leveraged to separate the whole data and identify homogenous crash segmentations. Recently, latent class clustering analysis has been utilized (Depaire et al., 2008; Yasmin et al., 2014a, 2014b; Liu & Fan, 2020; Li & Fan, 2019) as the data clustering techniques to preprocess the distribution of data. While analyzing the injury severity considering the heterogeneity within each cluster, researchers adopted different models such as binary logit model (Sasidharan et al., 2015; Sivasankaran & Balasubramanian, 2020), multinomial logit model (Depaire et al., 2008; Sun et al., 2019), mixed logit model (Liu & Fan, 2018), and ordered probit model (Mohamed et al., 2013). To examine the unobserved heterogeneity underlying in the crash data, other data mining techniques including k-means clustering (Mohamed et al., 2013), decision tree and Bayesian networks (Prati et al., 2017), and classification and regression tree (Kashani & Mohaymany, 2011) are also utilized.

To model bicyclist injury severity, there are two categories of model structures that are ordered framework and unordered framework (Eluru, 2013). The basic discrete choice model within ordered framework is ordered logit model, which is employed for outcomes in ordinal nature (Mooradian et al., 2013). Based on this model, the partial proportional odds model relaxes the proportional odds assumption, which allows for variable coefficients across different levels. In addition to ordered framework models, the multinomial logit model and mixed logit models within unordered framework are also employed by researchers to conduct bicyclist injury severity analysis. However, the unordered models might neglect the inherent ordinal nature of injury severity.

By reviewing the previous relevant research studies, it can be concluded that there is a need to develop an innovative and combined method, which sequentially conducts the latent class clustering analysis and develops the partial proportional odds models, to model bicyclist injury severity in bicycle-vehicle crashes. The necessity of applying latent class clustering is to uncover the unobserved or latent features within the crash data, and to classify the bicycle-vehicle crashes into optimal homogenous groups for further analysis of heterogeneity between categorical segmentations. To explore the impact of various explanatory variables on bicyclist injury severity within each cluster, the partial proportional odds model can be developed in order to consider both the ordinal nature of injury severity levels and the limitation of proportional odds assumption associated with standard ordered logit models.

3. Data

The data utilized in this research study are the police reported bicycle-involved crash data collected from 2007 to 2014 in North Carolina, which record information including bicyclist injury severity, driver and bicyclist demographics, vehicle type and traveling speed, crash types and locations, roadway and environmental characteristics, and crash time, etc. The potential explanatory variables considered in this paper are carefully selected based on numerous previous studies as well as the availability of this dataset.

After data cleaning by removing the missing, unknown, and obviously incorrect data, a total of 4,012 bicycle-involved crash data are kept for descriptive analysis and model development. The bicyclist injury severity is categorized into five levels, which are no injury, possible injury, evident injury, disabling injury, and fatal injury, accounting for 8.85%, 38.96%, 43.62%, 5.76%, 2.81% total crashes, respectively. The same categorization method of injury severity levels can be found in research studies conducted by Eluru et al. (2008), and Liu et al. (2020).

Table 1 presents a descriptive analysis of the crash dataset, which contains the detailed information on the distribution of bicyclist injury severity levels and specific crashes. In addition, the potential factors obtained from the dataset are classified into eight categories to describe crashes from different aspects. The percentages of each injury severity level with different characteristics are presented in the table. The explanatory variables of the models are coded as dummy variables (0–1), which are listed in Table 1.

According to the data presented in Table 1, some potential variables that might be associated with severe injuries can be discovered. Alcohol usage for both bicyclists and drivers is a critical factor impacting bicyclist injury severity. Comparing the different effects on bicyclists and drivers, it can be seen that the percentage of fatal injuries associated with drivers in alcohol usage (16.00%) is higher

Table 1
Descriptive statistics of bicycle-motor vehicle crashes severity outcomes and explanatory variables.

Variable	Categories	Injury Severity					
	Description	Total	No injury	Possible injury	Evident injury	Disabling injury	Fatal injury
Bicycle-vehicle crashes		4012	8.85%	38.96%	43.62%	5.76%	2.81%
Cyclist characteristics							
Gender	Male	3426	9.28%	38.35%	43.70%	5.78%	2.89%
	Female	586	6.31%	42.49%	43.17%	5.63%	2.39%
Age	<16	726	9.37%	38.57%	45.04%	6.34%	0.69%
	16–24	860	9.88%	39.65%	42.56%	6.16%	1.74%
	25–54	1916	8.77%	39.25%	43.42%	5.22%	3.34%
	55+	510	6.67%	37.25%	44.12%	6.27%	5.69%
Alcohol usage	Yes	280	8.57%	30.36%	43.21%	10.00%	7.86%
	No	3732	8.87%	39.60%	43.65%	5.44%	2.44%
Driver characteristics							
Gender	Male	2194	8.48%	37.42%	44.80%	5.70%	3.60%
	Female	1818	9.30%	40.81%	42.19%	5.83%	1.87%
Age	<25	824	7.89%	37.74%	44.54%	5.83%	4.00%
	25–59	2418	9.47%	37.80%	44.13%	6.00%	2.61%
	60+	770	7.92%	43.90%	41.04%	4.94%	2.21%
Alcohol usage	Yes	75	4.00%	28.00%	38.67%	13.33%	16.00%
	No	3937	8.94%	39.17%	43.71%	5.61%	2.57%
Vehicle characteristics							
Veh_Type	Passenger Car	3015	9.59%	39.97%	42.82%	5.37%	2.26%
	Pickup	586	4.61%	38.05%	44.20%	8.19%	4.95%
	Van	221	4.98%	37.10%	50.23%	5.43%	2.26%
	Bus	25	24.00%	32.00%	36.00%	0.00%	8.00%
	Single Unit Truck	46	4.35%	21.74%	58.70%	6.52%	8.70%
	Motorcycle	26	15.38%	19.23%	61.54%	0.00%	3.85%
	Others	93	17.20%	32.26%	39.78%	6.45%	4.30%
Veh Speed	<20 mph	2052	10.87%	46.64%	40.11%	2.00%	0.39%
	20–30 mph	534	8.43%	33.71%	51.12%	5.62%	1.12%
	30–40 mph	650	6.62%	34.31%	46.92%	8.31%	3.85%
	40–50 mph	516	5.62%	26.94%	47.29%	13.18%	6.98%
	50–60 mph	251	5.58%	24.30%	41.04%	15.14%	13.94%
	60 + mph	9	11.11%	33.33%	22.22%	0.00%	33.33%
Crash characteristics							
Bike Direction	With traffic	2968	7.99%	35.14%	46.93%	6.70%	3.23%
	Facing traffic	1044	11.30%	49.81%	34.20%	3.07%	1.63%
Crash Type	Motorist overtaking bicyclist	839	5.96%	32.66%	45.77%	8.70%	6.91%
	Backing vehicle	34	23.53%	38.24%	38.24%	0.00%	0.00%
	Bicyclist failed to yield	610	8.69%	35.90%	43.77%	9.67%	1.97%
	Bicyclist Turn/Merge	367	10.08%	33.79%	43.60%	8.72%	3.81%
	Bicyclist overtaking motorist	55	16.36%	32.73%	49.09%	0.00%	1.82%
	Head-On	121	4.96%	32.23%	43.80%	9.92%	9.09%
	Motorist Failed to Yield	816	9.68%	52.94%	35.54%	1.47%	0.37%
	Motorist Turn/Merge	674	7.42%	36.05%	52.82%	3.12%	0.59%
	Crossing Paths	211	14.69%	46.45%	34.60%	3.79%	0.47%
	Parallel Paths	69	14.49%	30.43%	44.93%	7.25%	2.90%
	Other Crash Types	216	10.19%	37.96%	44.44%	4.17%	3.24%
Speeding	Yes	90	6.67%	22.22%	45.56%	12.22%	13.33%
	No	3922	8.90%	39.34%	43.57%	5.61%	2.58%
Rural/Urban	Urban	2700	9.78%	41.67%	42.59%	4.19%	1.78%
	Rural	1312	6.94%	33.38%	45.73%	8.99%	4.95%
Crash Location	Intersection	2329	9.32%	42.12%	42.81%	4.68%	1.07%
	Non-intersection	1683	8.20%	34.58%	44.74%	7.25%	5.23%
Roadway characteristics							
Road Geometry	Curve	230	5.65%	29.13%	52.17%	6.52%	6.52%
	Straight	3782	9.04%	39.56%	43.10%	5.71%	2.59%
Road Type	One-way	143	14.69%	46.15%	32.17%	4.90%	2.10%
	Two-way	3869	8.63%	38.69%	44.04%	5.79%	2.84%
Divided Road	Yes	801	9.11%	39.83%	43.07%	4.24%	3.75%
	No	3211	8.78%	38.74%	43.76%	6.14%	2.58%
Road Condition	wet, water, ice, snow, mud	323	8.05%	38.39%	44.89%	7.43%	1.24%
	dry	3689	8.92%	39.01%	43.51%	5.61%	2.95%
Traffic Control	Yes	2502	7.99%	41.49%	43.01%	5.44%	2.08%
	No	1510	10.26%	34.77%	44.64%	6.29%	4.04%
No. of Lanes	1	61	8.20%	40.98%	44.26%	6.56%	0.00%
	2	2395	8.89%	37.33%	44.51%	6.43%	2.84%
	3	316	10.13%	43.35%	41.14%	2.85%	2.53%
	4	611	8.84%	39.44%	42.88%	5.56%	3.27%
	5	394	7.11%	43.91%	40.61%	5.33%	3.05%
	6	127	8.66%	42.52%	42.52%	3.15%	3.15%
	7	38	7.89%	36.84%	50.00%	5.26%	0.00%

(*continued on next page*)

Table 1 (*continued*)

Variable	Categories	Injury Severity					
	Description	Total	No injury	Possible injury	Evident injury	Disabling injury	Fatal injury
	8	40	12.50%	35.00%	47.50%	2.50%	2.50%
	9+	30	13.33%	36.67%	43.33%	6.67%	0.00%
Land characteristics							
Work Zone	Yes	17	11.76%	29.41%	47.06%	11.76%	0.00%
	No	3995	8.84%	39.00%	43.60%	5.73%	2.83%
Temporal Characteristics							
Crash Time	0:00–5:59	137	7.30%	32.85%	39.42%	12.41%	8.03%
	6:00–9:59	547	7.31%	38.76%	46.98%	4.57%	2.38%
	10:00–14:59	1066	9.76%	42.78%	40.99%	4.13%	2.35%
	15:00–17:59	1093	9.15%	39.89%	43.37%	5.58%	2.01%
	18:00–23:59	1169	8.64%	35.41%	45.17%	7.19%	3.59%
Environmental Characteristics							
Weather	Clear	3326	8.81%	38.82%	44.11%	5.53%	2.74%
	Cloudy	511	8.61%	40.51%	40.90%	6.26%	3.72%
	Fog, smog, smoke	9	11.11%	33.33%	22.22%	0.00%	33.33%
	Rain	161	9.94%	36.02%	44.72%	9.32%	0.00%
	Snow	5	20.00%	80.00%	0.00%	0.00%	0.00%
Light Condition	Daylight	2970	8.99%	40.27%	43.80%	5.02%	1.92%
	Dusk or Dawn	176	6.82%	38.64%	44.32%	9.09%	1.14%
	Dark - lighted roadway	442	9.50%	37.33%	44.12%	6.79%	2.26%
	Dark - roadway not lighted	424	8.02%	31.60%	41.51%	8.49%	10.38%

than that for bicyclists (7.86%). In addition, vehicle speed is another variable that is related to bicyclist injury severity. Based on the bicyclist injury severity for different vehicle speeds shown in Table 1, the percentage of fatal injuries corresponding to each vehicle speed range increases with a higher vehicle speed. Similarly, speeding has a strong relationship with severe bicyclist injuries, as the percentage of fatal injuries resulting from speeding is 13.33%. Furthermore, environmental characteristics are the external causes to bicyclist injury severity. From the data presented in Table 1, the percentages of fatal injuries under adverse weather and dark not lighted roadway conditions are high.

4. Methodology

4.1. Latent class clustering (LCC)

Latent class clustering is a probability based cluster analysis approach (Depaire et al., 2008; Collins & Lanza, 2010), which has been widely used recently for traffic crash data segmentation in order to identify optimal homogenous groups. It is assumed that the whole crash dataset is divided into exclusive latent classes with each cluster being classified by an unobserved or latent categorical variable, which maximizes the homogeneity within each class and the heterogeneity between classes (Lanza & Rhoades, 2013; Sasidharan et al., 2015). Therefore, to uncover the unobserved and latent features underlying the bicycle-vehicle crash data, latent class clustering analysis is conducted. By applying this innovative method, the impact of various factors within the identified latent clusters on bicyclist injury severity can be investigated without omitting the potential heterogeneity between each segmentation. This method can be employed to conduct similar studies on injury severities of other road users. Compared to other clustering methods, latent class clustering does not require the number of clusters to be predetermined (Sun et al., 2019; Depaire et al., 2008). In addition, it allows different types of variables (e.g., numerical, categorical variables) (Sasidharan et al., 2015; Sun et al., 2019), and accepts different statistical criteria to determine the number of clusters (Sasidharan et al., 2015).

Let us consider a crash data sample where C latent clusters are assumed to be estimated based on K categorical items. Let $c = 1, 2, \ldots, C$ denote the latent class membership, and Y_i (=$Y_{i1}, \ldots, Y_{ik}$) represent crash i's responses to K categorical items in which Y_i is a categorical variable with possible values being $1, \ldots, r_k$. Let $I(y_k = r_k)$ be an indicator factor that equals to 1 when y_k equals to r_k, and 0 otherwise. Then, the probability function describing the response of crash is shown as follows:

$$P(Y_i = y) = \sum_{c=1}^{C} \gamma_c \prod_{k=1}^{K} \prod_{r_k=1}^{R_k} \rho_{k,r_k|c}^{I(y_k=r_k)} \tag{1}$$

where γ_c represents the probability of latent class membership for cluster c, and ρ denotes the item-response probability conditional on latent class membership. In this study, the latent class clustering analysis is conducted using the LCA procedure installed in SAS 9.4, which is developed by the Penn State Methodology Center (Lanza et al., 2007).

To determine the appropriate number of clusters, different number of clusters need to be tested by trying multiple models. Information criteria including Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), Consistent Akaike Information Criterion (CAIC), and entropy-based measures can be utilized to select the optimal number of clusters. The minimal values of AIC, BIC, CAIC indicate the best number of clusters. In addition, some researchers believe that BIC is better compared to AIC and CAIC for identifying the number of clusters (Biernacki & Govaert, 1999). However, it is also found that increasing the number of clusters might not always result in a minimum value, especially for large samples (e.g., traffic crash data) (Bijmolt et al., 2004). Therefore, the percentage reduction in BIC between tested models can be utilized (Sasidharan et al, 2015). As for the entropy measure, it is a weighted average of the posterior membership probability of an individual ranging from 0 to 1. Larger entropy values are associated with better latent class segmentation and 0.9 is suggested as a satisfied entropy value (McLachlan & Peel, 2000). In this paper, AIC, BIC, CAIC, and entropy measures are used for identifying the appropriate number of clusters.

4.2. Partial proportional odds model (PPO)

The partial proportional odds model is developed based on the ordered logit model. In the ordered logit model, the proportional odds (PO) assumption is subjected. It can be interpreted that the estimated parameters are restricted to be the same across all the alternatives. However, this assumption is unrealistic. To relax the assumption, the PPO model is developed.

The explanatory variables associated with each bicycle-vehicle crash are categorized into two groups. One contains parameters satisfying the PO assumption, which is presented as vector X_i; the other includes parameters that violate the PO assumption, which is shown as vector Z_i. The variables that violate the PO assumption are able to affect the response variables differently, while others remaining fixed parameters have the same effect across different levels. Thus, the PPO model with logit function is presented as follows (Peterson & Harrell, 1990):

$$P(Y_i \geq j) = \frac{\exp\left[\theta_j - \left(X_i'\beta_j + Z_i'\gamma_j\right)\right]}{1 + \exp\left[\theta_j - \left(X_i'\beta_j + Z_i'\gamma_j\right)\right]} \tag{2}$$

where j denotes the level of bicyclist injury severity and Y_i represents the crash injury resulting from bicycle-motor vehicle crash i, β and γ represent the coefficients that will be estimated, and θ_j demonstrates the threshold for jth cumulative logit.

To examine whether or not the explanatory variables violate the PO assumption, the Wald Chi-square tests are utilized during the model development (Wang & Abdel-Aty, 2008; Sasidharan & Menéndez, 2014). This procedure helps divide the explanatory variables into two groups that belong to either vector X_i or vector Z_i.

Before developing the PPO models, ordered logit models are built to help select the explanatory variables that will be considered later for the PPO models, and all parameters are assumed to violate the PO assumption as the base for PPO model development for each cluster. SAS 9.4 is used to conduct the PPO model estimation procedure. Since the sign of the estimated coefficient may not always explain the effect of explanatory variables on the bicyclist injury severity (Wooldridge, 2010; Washington et al., 2010), the marginal effects are applied for the PPO model result interpretation.

4.3. Marginal effect

To examine the impact of significant variables included in the partial proportional odds models on the likelihood of bicyclist injury severity, the marginal effect of each significant variable is calculated. Since all the variables in this research study are dummy-coded, it is not appropriate to apply the marginal effect equation for continuous variables (Yu et al., 2019). Therefore, the marginal effects of all the explanatory variables for each bicyclist i and severity level j are expressed as follows:

$$E_{X_{ijk}}^{P_{ij}} = P_{ij}(X_{ijk} = 1) - P_{ij}(X_{ijk} = 0) \tag{3}$$

where $E_{X_{ijk}}^{P_{ij}}$ represents the marginal effects of the k-th dummy variable X_{ijk}, $P_{ij}(X_{ijk} = 1)$ and $P_{ij}(X_{ijk} = 0)$ denote the probability when dummy variable X_{ijk} equals to 1 and 0, respectively.

The marginal effect of each significant explanatory variable for different bicyclist injury severity levels is calculated when the k-th dummy variable X_{ijk} varies from 0 to 1 based on the corresponding probabilities. Then, the average of marginal effects is computed for each parameter over all observations.

5. Results and discussions

5.1. Latent class clustering results

Several models are examined using the bicycle involved crash data including all the explanatory variables to identify the optimal number of clusters by testing from cluster 1 to cluster 10. Entropy value and information criteria including AIC, BIC, and CAIC are utilized to determine the most appropriate number of clusters and the variation of these values are shown in Fig. 1. Based on the values of AIC, BIC, and CAIC presented in this figure, all information criteria decrease with the increase of the number of clusters. However, no minimum value is reached for all the tested models. Hence, the reduction percentage of less than 1% in all three information criteria is applied to select the number of clusters for this study following the research conducted by Chang et al. (2019). From seven clusters afterwards, little improvement (<1%) of AIC, BIC, and CAIC is shown and the entropy has reached a high value of 0.97, which indicates a good separation between the clusters. Therefore, the bicycle-motor vehicle crashes are classified into seven clusters for further partial proportional odds analysis.

Following the research study conducted by Depaire et al. (2008), the final seven-cluster models can be described by the skewed feature distributions of each variable in each cluster, which can be found in Table 2. To characterize each cluster as a specific crash pattern, some featured variables are identified based on the variable distribution across clusters, and the variables with significantly different percentages are set to be bold in Table 2.

For cluster 1, 50.29% of the crashes occurred with estimated vehicle speed ranging from 20 to 30 mph. In addition, 55.32% of the crashes occurred when bicyclists failed to yield. Therefore, cluster 1 can be referred to as "crashes occurred with vehicle speed ranging from 20 to 30 mph when the bicyclist failed to yield." All the crashes in cluster 2 are caused by drivers aged from 25 to 59 with low estimated vehicle speed, which is less than 20 mph. One can define this cluster as "crashes caused by drivers from 25 to 59 years old with less than 20 mph vehicle speed." For cluster 3, 86.34% of the crashes occurred at non-intersection locations, and 61.8% occurred on dark roadway without light. Cluster 3 can be described as "crashes occurred on not lighted non-intersection roadways in dark." In cluster 4, only one variable is found to have significantly different distribution (63.81%) from that in other clusters, which is dark – lighted roadway. Hence, cluster 4 is named as "crashes occurred on lighted roadways in dark." Cluster 5 and cluster 6 overlap with cluster 3 on the same crash location, which is non-intersection location, but differ from it by the skewed feature of vehicle type (pickup) and lighting condition (daylight) respectively. It is noted that 65.43% of the crashes in cluster 5 are caused by pickups, and all the crashes in cluster 6 are occurred in daylight. Therefore, cluster 5 and cluster 6 can be referred to as "crashes caused by pickups at non-intersection locations" and "crashes occurred at non-intersection locations in daylight," respectively. Cluster 7 shares an overlapped variable (vehicle speed less than 20 mph) with cluster 2. All the crashes occurred with the estimated vehicle speed less than 20 mph, but none of them are caused by drivers aged from 25 to 59. Instead, 50.4% of the crashes are caused by old drivers (more than 60 years old). So, cluster 7 is described as "crashes caused by drivers elder than 60 years old with less than 20 mph vehicle speed." Finally, Table 3 summarizes the definition, key variables and the distribution of each cluster.

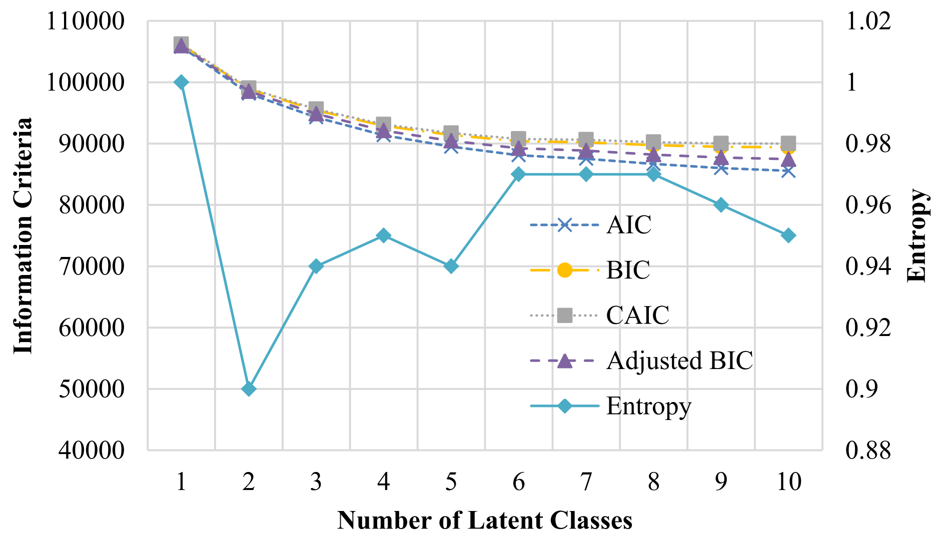

Fig. 1. Variation of entropy and information criteria for different number of clusters.

Table 2
Distribution of variables describing each cluster (bold).

Variable	Description	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Bicycle-vehicle crashes		12.89%	24.08%	8.02%	13.81%	11.46%	14.21%	15.53%
Driver Age	25–59	61.9%	**100%**	62.11%	58.12%	63.7%	55.61%	0%
	60+	16.63%	0%	14.29%	16.06%	24.13%	21.75%	**50.4%**
Veh_Type	Pickup	12.19%	11.59%	0%	9.03%	**65.43%**	0%	9.63%
Veh Speed	<20 mph	0%	**100%**	4.35%	71.12%	5%	5.61%	**100%**
	20–30 mph	**50.29%**	0%	11.18%	12.82%	11.74%	19.82%	0%
Crash Type	Bicyclist failed to yield	**55.32%**	13.15%	1.86%	19.49%	0%	0.18%	13.16%
Crash Location	Non-intersection	2.13%	27.95%	**86.34%**	13.54%	**91.74%**	**84.56%**	23.27%
Light Condition	Daylight	99.81%	100%	0%	0%	64.13%	**100%**	100%
	Dark - lighted roadway	0%	0%	20.81%	**63.18%**	5.43%	0%	0%
	Dark - roadway not lighted	0%	0%	**61.8%**	19.31%	25.65%	0%	0%

Table 3
Definition and distribution of each cluster.

Cluster	Key Variables	Description	Percentage
Cluster 1	Vehicle speed 20–30 mph Bicyclist failed to yield	Crashes occurred with vehicle speed from 20 to 30 mph when the bicyclist failed to yield	12.89%
Cluster 2	Driver age 25–59 Vehicle speed less than 20 mph	Crashes caused by drivers from 25 to 59 years old with less than 20 mph vehicle speed	24.08%
Cluster 3	Crash location non-intersection Dark - roadway not lighted	Crashes occurred on not lighted non-intersection roadways in dark	8.02%
Cluster 4	Dark - lighted roadway	Crashes occurred on lighted roadways in dark	13.81%
Cluster 5	Vehicle type pickup Crash location non-intersection	Crashes caused by pickups at non-intersection locations	11.46%
Cluster 6	Crash location non-intersection Daylight	Crashes occurred at non-intersection locations in daylight	14.21%
Cluster 7	Driver age over 60 Vehicle speed less than 20 mph	Crashes caused by drivers elder than 60 years old with less than 20 mph vehicle speed	15.53%

5.2. Partial proportional odds model results

Based on the LCC results, PPO models are developed for each cluster. However, for cluster 4, cluster 6, and cluster 7, all the variables are found to pass the Wald Chi-square tests for the PO assumption, which make these three PPO models collapse into ordered logit models. To compare the restricted model developed based on the whole data and the sub-models developed based on latent class clustering, a PPO model is developed using the whole dataset, and a likelihood ratio test is conducted. According to the model estimation results, the log likelihood value at convergence for the restricted PPO model is −4530.08, while the sum of the log likelihood values at convergence for all seven sub-models is −4411.21. The value of χ^2 test statistics is 237.75 with 48 degrees of freedom, which indicates a better fitness for seven sub-models. Therefore, the PPO/ORL model results for each cluster and the whole data are presented in Tables 4a–4h.

As is mentioned in Section 4.2., the sign of the estimated parameters may not accurately reveal the effect of bicyclist injury severity, marginal effects need to be calculated for the interpretation of the variable impacts. The marginal effects of significant variables in each model for each cluster and the whole data are presented in Tables 5a–5h.

Comparing the significant factors for the whole dataset and separate clusters, three critical findings can be discussed. First, differences can be found between the significant explanatory variables for the whole dataset and the seven clusters. Some variables that do not significantly affect the bicyclist injury severity in the whole dataset are found to have significant impacts in specific clusters

Table 4a
PPO model for cluster 1 in bicyclist-vehicle crashes.

Cluster 1		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−3.0494	0.0010**	−1.1260	0.1405	0.0559	0.9474	0.5546	0.5538
Vehicle characteristics											
Veh Speed	20–30 mph			−3.1930	0.0025**	−2.6773	<0.0001**	−0.7646	0.2706	0.7874	0.3418
	30–40 mph			−5.4050	0.0004**	−2.5565	<0.0001**	−1.1038	0.1131	0.2602	0.7511
	40–50 mph			−4.0777	0.0120**	−1.3928	0.0352**	−0.3270	0.6606	2.2216	0.0828*
Crash characteristics											
Bike Direction	Facing traffic			1.5879	0.0837*	−0.3658	0.3682	−0.5812	0.0119**	−0.0941	0.8110
Crash Type	Bicyclist Turn/Merge	1.0534	0.0458**								
	Bicyclist overtaking motorist	−2.8403	0.0325**								
	Motorist Failed to Yield	−0.5971	0.0500*								
Roadway characteristics											
Road Geometry	Curve			4.1579	0.0011**	0.4645	0.4458	0.6454	0.2220	13.4334	0.9851
Road Type	Two-way	1.0816	0.0170**								
Model Performance Results											
Log Likelihood with Constant Only		−619.40									
Log Likelihood at Convergence		−577.81									
AIC		1213.63									

* Level of significance >90%.
** Level of significance >95%.

Table 4b
PPO model for cluster 2 in bicyclist-vehicle crashes.

Cluster 2		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−18.8777	0.9821	−4.2560	<0.0001**	−0.6332	0.1332	2.8783	<0.0001**
Cyclist characteristics											
Gender	Male			11.7300	0.9889	−0.1434	0.8007	−0.2588	0.1502	−1.6871	0.0003**
Vehicle characteristics											
Veh_Type	Passenger Car	0.8759	0.0089**								
	Pickup	1.1659	0.0020**								
	Van	0.9009	0.0284**								
	Single Unit Truck	2.1506	0.0052**								
	Motorcycle	2.4802	0.0436**								
Crash characteristics											
Bike Direction	Facing traffic			−13.0156	0.9877	−1.1769	0.0370**	−0.7558	<0.0001**	−0.2122	0.3103
Crash Type	Motorist Turn/Merge	0.3634	0.0309**								
Rural/Urban	Urban	−0.4694	0.0080**								
Roadway characteristics											
Road Geometry	Curve	0.7416	0.0384**								
Road Condition	wet, water, ice, snow, mud	0.8671	0.0028**								
Traffic Control	Yes	0.3975	0.0042**								
Temporal Characteristics											
Crash Time	0:00–5:59	1.9987	0.0377**								
	6:00–9:59	−0.3212	0.0635*								
Environmental Characteristics											
Weather	Fog, smog, smoke, rain, snow	−0.4284	0.0279**								
Model Performance Results											
Log Likelihood with Constant Only		−1029.52									
Log Likelihood at Convergence		−971.68									
AIC		1993.37									

* Level of significance >90%.
** Level of significance >95%.

(i.e., types of bicycle-vehicle crashes). This finding confirms the necessity and importance of conducting the latent class clustering analysis, which can provide the insight into the differences between various types of crashes and reveal the heterogeneity within the data. Second, the significant explanatory variables vary across the seven clusters, indicating that different types of crashes are affected by distinctive factors. Third, significant variables in the PPO model that are developed based on the whole dataset can still be found in the sub-models, which provides clear evidence that the sub-models can interpret the effects of bicyclist injury severity more exhaustively.

Differences of the impacts of significant variables identified across seven clusters will be discussed in detail in the following sections.

Table 4c
PPO model for cluster 3 in bicyclist-vehicle crashes.

Cluster 3		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−3.1950	<0.0001**	−1.9005	<0.0001**	0.3748	0.0197**	2.5665	<0.0001**
Driver characteristics											
Age	<25	0.7157	0.0039**								
Alcohol usage	Yes	1.2941	0.0104**								
Vehicle characteristics											
Veh Speed	50–60 mph			1.5335	0.0004**	1.0300	0.0026**	0.1141	0.7237	−0.1120	0.8345
Crash characteristics											
Bike Direction	Facing traffic			−1.3387	0.1420	0.4606	0.3477	−0.8320	0.0499**	−1.2660	0.0239**
Speeding	Yes	1.4298	0.0154**								
Roadway characteristics											
Road Condition	wet, water, ice, snow, mud	−0.8877	0.0296**								
Temporal Characteristics											
Crash Time	15:00–17:59			1.5150	0.1110	−1.0948	0.0808*	−0.6792	0.0783*	−1.0940	0.0366**
Environmental Characteristics											
Weather	Fog, smog, smoke, rain, snow	0.9195	0.0053**								
Model Performance Results											
Log Likelihood with Constant Only		−454.59									
Log Likelihood at Convergence		−971.68									
AIC		879.97									

* Level of significance >90%.
** Level of significance >95%.

Table 4d
ORL model for cluster 4 in bicyclist-vehicle crashes.

Cluster 4		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−3.3756	<0.0001**	−1.6385	<0.0001**	1.2704	<0.0001**	3.6950	<0.0001**
Cyclist characteristics											
Age	<16	−0.4630	0.0710*								
Vehicle characteristics											
Veh Speed	<20 mph	−1.3416	<0.0001**								
	20–30 mph	−0.8078	0.0106**								
Crash characteristics											
Crash Type	Bicyclist overtaking motorist	−1.4275	0.0723*								
	Motorist Failed to Yield	−0.6251	0.0008**								
	Parallel Paths	−1.7609	0.0029**								
Speeding	Yes	−1.4821	0.0455**								
Model Performance Results											
Log Likelihood with Constant Only		−637.81									
Log Likelihood at Convergence		−607.89									
AIC		1237.78									

* Level of significance >90%.
** Level of significance >95%.

5.2.1. Bicyclist characteristics

Bicyclist characteristic factors that have significant impacts on bicyclist injury severity include the gender of bicyclists, bicyclist age, and bicyclists under the influence of alcohol. Differences of impact variables exist between each cluster. For example, male bicyclists in cluster 2 are found to be more likely to have fatal injury or no injury with marginal effects being 0.0012 and 0.1079, respectively.

Young bicyclists (<16) have a lower probability of suffering from severe injuries (fatal injury, disabling injury, and evident injury) in cluster 4 (marginal effects −0.0042, −0.0167, and −0.0822) and cluster 5 (marginal effects −0.06, −0.0512, and −0.114). Similarly, bicyclists (16–24 and 25–54) in cluster 5 are less likely to sustain severe injury, including fatal injury, disabling injury, and evident injury (marginal effects −0.0476, −0.038, and −0.0697; and marginal effects −0.0748, −0.054, and −0.053). This result is in line with a previous research study conducted by Kaplan et al. (2014).

In cluster 7, bicyclists under the influence of alcohol have a higher likelihood to be severely injured (fatal injury, disabling injury, and evident injury). It can be noted that, although the estimated driving speed is relatively low, the injury severity level can be high. This result is probably valid because drinking alcohol might decrease the reaction speed of a bicyclist to an incident and therefore have a negative impact on his/her physical condition. In addition, the influence of elderly drivers cannot be neglected. Therefore, regulations can be made to prevent bicyclists from drinking alcohol while riding a bicycle, so that biking safety might be improved.

Table 4e
PPO model for cluster 5 in bicyclist-vehicle crashes.

Cluster 5		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−0.8829	0.0416**	−0.0170	0.9678	2.4620	<0.0001**	5.0348	<0.0001**
Cyclist characteristics											
Age	<16	−1.0458	0.0022**								
	16–24	−0.7379	0.0214**								
	25–54	−0.9294	0.0002**								
Driver characteristics											
Age	25–59	0.4041	0.0348**								
Vehicle characteristics											
Veh_Type	Single Unit Truck	0.8254	0.0665*								
Veh Speed	<20 mph	−1.1364	0.0076**								
	20–30 mph	−0.9011	0.0021**								
	30–40 mph	−0.5539	0.0112**								
Crash characteristics											
Speeding	Yes			1.6194	0.0046**	1.2554	0.0160**	0.9392	0.1619	−1.1144	0.1742
Roadway characteristics											
No. of Lanes	<=4	−0.7549	0.0247**								
Environmental Characteristics											
Light Condition	Daylight	−0.3785	0.0611*								
	Dark - lighted roadway	−0.8676	0.0487**								
Model Performance Results											
Log Likelihood with Constant Only		−584.12									
Log Likelihood at Convergence		−552.88									
AIC		1143.76									

* Level of significance >90%.
** Level of significance >95%.

Table 4f
ORL model for cluster 6 in bicyclist-vehicle crashes.

Cluster 6		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−0.4500	0.7383	0.8233	0.5412	3.2915	0.0151**	5.4747	<0.0001**
Vehicle characteristics											
Veh Speed	30–40 mph	0.4648	0.0262**								
	40–50 mph	0.7711	0.0004**								
	50–60 mph	1.0369	0.0002**								
Crash characteristics											
Crash Type	Bicyclist Turn/Merge	−0.4803	0.0074**								
	Head-On	1.2955	0.0005**								
Roadway characteristics											
Road Type	Two-way	−2.8927	0.0266**								
Traffic Control	Yes	0.3239	0.0514*								
No. of Lanes	<=4	−0.5629	0.0661*								
Environmental Characteristics											
Weather	Fog, smog, smoke, rain, snow	−0.3908	0.0882*								
Model Performance Results											
Log Likelihood with Constant Only		−718.46									
Log Likelihood at Convergence		−692.91									
AIC		1411.81									

* Level of significance >90%.
** Level of significance >95%.

5.2.2. Driver characteristics

Young drivers might increase the probability of severe injuries (fatal injury, disabling injury, and evident injury) suffered by bicyclists in crashes that occurred on not lighted, non-intersection locations in dark with marginal effects being 0.0673, 0.0513, and 0.027. Similar results can be found in cluster 5, where mid-aged drivers (25–59) are more likely to cause severe injuries including fatal, disabling, and evident injury in crashes involving pickups at non-intersection locations.

Unsurprisingly, drivers under the influence of alcohol can probably increase the likelihood of getting severe injuries such as fatal and disabling injuries (marginal effects 0.1501 and 0.0976) for crashes occurred on not lighted non-intersection locations in dark. This result is in line with previous research in (Kim et al., 2007; Moore et al., 2011). Policies can be implemented to prohibit drivers from drinking alcohol to avoid severe injuries to bicyclists in bicycle-vehicle crashes.

5.2.3. Vehicle characteristics

Several types of vehicles are found to have significant impacts on bicyclist injury severity, especially in cluster 2 and cluster 5, which are crashes caused by drivers from 25 to 59 years old with

Table 4g
ORL model for cluster 7 in bicyclist-vehicle crashes.

Cluster 7		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−4.6780	<0.0001**	−3.6478	<0.0001**	0.2588	0.1795	3.2116	<0.0001**
Cyclist characteristics											
Alcohol usage	Yes	2.0006	0.0115**								
Crash characteristics											
Bike Direction	Facing traffic	−0.3601	0.0435**								
Crash Type	Motorist Turn/Merge	0.5508	0.0044**								
	Crossing Paths	−1.4198	<0.0001**								
	Parallel Paths	−2.5955	0.0241**								
Roadway characteristics											
Road Geometry	Curve	0.8632	0.0394**								
Divided Road	Yes	−0.4608	0.0164**								
Temporal Characteristics											
Crash Time	10:00–14:59	−0.6423	0.0017**								
	15:00–17:59	−0.5938	0.0050**								
Model Performance Results											
Log Likelihood with Constant Only		−626.19									
Log Likelihood at Convergence		−589.05									
AIC		1204.11									

*Level of significance >90%.
** Level of significance >95%.

Table 4h
PPO model for the whole data in bicyclist-vehicle crashes.

Whole Data		All Levels		Fatal Injury		Disabling Injury		Evident Injury		Possible Injury	
Variable		Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.	Coef.	Sig.
Intercept				−3.3425	<0.0001**	−2.3343	<0.0001**	0.0064	0.9756	2.2166	<0.0001**
Cyclist characteristics											
Age	<16			−1.6649	0.0007**	−0.5829	0.0025**	−0.2271	0.0585*	−0.3235	0.0640*
	16–24			−0.9676	0.0011**	−0.2544	0.1355	−0.1984	0.0797*	−0.3341	0.0358**
	25–54	−0.2531	0.0085**								
Alcohol usage	Yes			0.5943	0.0192**	0.5355	0.0031**	0.2172	0.1061	−0.1600	0.4877
Driver characteristics											
Age	<25	0.2511	0.0092**								
	25–59	0.1332	0.0909*								
Alcohol usage	Yes			1.5997	<0.0001**	1.1884	<0.0001**	0.4384	0.0922*	0.6569	0.2690
Vehicle characteristics											
Veh Speed	<20 mph			−2.1048	<0.0001**	−1.5920	<0.0001**	−0.5245	<0.0001**	−0.3628	0.0196**
	20–30 mph			−1.0894	0.0107**	−0.8127	<0.0001**	−0.0567	0.6307	−0.1684	0.4075
	40–50 mph	0.3912	0.0005**								
	50–60 mph			1.1516	<0.0001**	0.9503	<0.0001**	0.4712	0.0034**	0.1895	0.5353
Veh_Type	Pickup			0.3012	0.1733	0.2610	0.0761*	0.1139	0.2343	0.7111	0.0006**
	Van	0.2453	0.0599*								
	Single Unit Truck	0.7171	0.0136**								
Crash characteristics											
Bike Direction	Facing traffic			−0.1770	0.5107	−0.4532	0.0092**	−0.4762	<0.0001**	−0.1654	0.2090
Speeding	Yes	0.3182	0.1275								
Crash Type	Bicyclist Failed to Yield			0.2613	0.4250	0.8624	<0.0001**	0.3064	0.0018**	0.1970	0.2342
	Head-On	0.7386	<0.0001**								
	Motorist Turn/Merge			−0.6942	0.1939	−0.1762	0.4417	0.4386	<0.0001**	0.3913	0.0210**
Roadway characteristics											
Road Geometry	Curve	0.4310	0.0012**								
Road Type	Two-way	0.2931	0.0772*								
Temporal Characteristics											
Crash Time	10:00–14:59	−0.1334	0.0535*								
Model Performance Results											
Log Likelihood with Constant Only		−4848.98									
Log Likelihood at Convergence		−4530.08									
AIC		9178.16									

* Level of significance >90%.
** Level of significance >95%.

less than 20 mph vehicle speed, and crashes caused by pickups at non-intersection locations. To be specific, passenger cars, pickups, vans, single unit trucks, and motorcycles are types of vehicles that might increase the probability of severe injuries (fatal, disabling, and evident injuries). Similar results can be found in Gårder (1994), Gårder et al. (1998), and Stone and Broughton (2003).

Table 5a
Average marginal effects for cluster 1.

Variable		Cluster1				
		F	D	E	P	N
Vehicle characteristics						
Veh Speed	20–30 mph	−0.1230	−0.1726	0.1349	0.2164	−0.0558
	30–40 mph	−0.1100	−0.1214	−0.0145	0.2641	−0.0182
	40–50 mph	−0.0470	−0.0463	0.0197	0.1546	−0.0811
Crash characteristics						
Bike Direction	Facing traffic	0.0385	−0.0687	−0.1042	0.1277	0.0067
Crash Type	Bicyclist Turn/Merge	0.0273	0.1042	0.0682	−0.1498	−0.0499
	Bicyclist overtaking motorist	−0.0228	−0.0786	−0.4150	0.0600	0.4564
	Motorist Failed to Yield	−0.0099	−0.0348	−0.0933	0.0878	0.0502
Roadway characteristics						
Road Geometry	Curve	0.2642	−0.2161	0.0842	−0.0501	−0.0822
Road Type	Two-way	0.0150	0.0530	0.1821	−0.1404	−0.1096

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

Table 5b
Average marginal effects for cluster 2.

Variable		Cluster2				
		F	D	E	P	N
Cyclist characteristics						
Gender	Male	0.0012	−0.0045	−0.0551	−0.0496	0.1079
Vehicle characteristics						
Veh_Type	Passenger Car	0.0008	0.0161	0.1621	−0.0732	−0.1058
	Pickup	0.0021	0.0380	0.2177	−0.1711	−0.0867
	Van	0.0014	0.0274	0.1614	−0.1212	−0.0690
	Single Unit Truck	0.0073	0.1203	0.2974	−0.3209	−0.1041
	Motorcycle	0.0105	0.1610	0.2910	−0.3550	−0.1075
Crash characteristics						
Bike Direction	Facing traffic	−0.0017	−0.0197	−0.1499	0.1499	0.0215
Crash Type	Motorist Turn/Merge	0.0004	0.0080	0.0741	−0.0486	−0.0339
Rural/Urban	Urban	−0.0006	−0.0111	−0.0954	0.0652	0.0418
Roadway characteristics						
Road Geometry	Curve	0.0011	0.0209	0.1472	−0.1113	−0.0578
Road Condition	wet, water, ice, snow, mud	0.0013	0.0253	0.1701	−0.1298	−0.0669
Traffic Control	Yes	0.0004	0.0076	0.0794	−0.0454	−0.0420
Temporal Characteristics						
Crash Time	0:00–5:59	0.0061	0.1049	0.2934	−0.3040	−0.1005
	6:00–9:59	−0.0003	−0.0060	−0.0638	0.0353	0.0348
Environmental Characteristics						
Weather	Fog, smog, smoke, rain, snow	−0.0004	−0.0079	−0.0842	0.0449	0.0476

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

Another significant impact factor of vehicle characteristic is the estimated vehicle speed, which is highly associated with the bicyclist injury severity. Low vehicle speed (less than 30 mph) is less likely to result in severe injuries including fatal, disabling, and evident injuries in cluster 4 and cluster 5 (i.e., crashes occurred on lighted roadways in dark and crashes caused by pickups at non-intersection locations, respectively). Interestingly, the estimated vehicle speed from 30 to 40 mph has different effects for cluster 5 and cluster 6. In cluster 5, this range of vehicle speed is less likely to cause severe injuries for crashes caused by pickups at non-intersection locations, while in cluster 6, the vehicle speed might increase the probability of severe injuries for crashes occurred at non-intersection locations in daylight. For vehicle speed from 40 to 60 mph, it is clear that the likelihood of suffering from severe injuries is increased for crashes occurring at non-intersection locations in daylight. It can be seen that even in daylight, driving fast may increase the probability of severe injuries of bicyclists resulting from crashes occurred at non-intersection locations. Therefore, speed limit is critical for providing safer cycling environment. When determining the speed limit for roadways that are popular among bicyclists, one needs to carefully consider avoiding setting the speed limit over 40 mph based on the model estimation results since high speed may increase the likelihood of severe injuries of bicyclists.

Table 5c
Average marginal effects for cluster 3.

Variable		Cluster3				
		F	D	E	P	N
Driver characteristics						
Age	<25	0.0673	0.0513	0.0270	−0.1040	−0.0415
Alcohol usage	Yes	0.1501	0.0976	−0.0241	−0.1675	−0.0560
Vehicle characteristics						
Veh Speed	50–60 mph	0.1708	0.0098	−0.1565	−0.0320	0.0079
Crash characteristics						
Bike Direction	Facing traffic	−0.0785	0.1541	−0.2622	0.0588	0.1277
Speeding	Yes	0.1744	0.1046	−0.0416	−0.1792	−0.0582
Roadway characteristics						
Road Condition	wet, water, ice, snow, mud	−0.0606	−0.0526	−0.0814	0.1155	0.0791
Temporal Characteristics						
Crash Time	15:00–17:59	0.1785	−0.3097	−0.0211	0.0497	0.1026
Environmental Characteristics						
Weather	Fog, smog, smoke, rain, snow	0.0890	0.0663	0.0247	−0.1273	−0.0527

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

Table 5d
Average marginal effects for cluster 4.

Variable		Cluster4				
		F	D	E	P	N
Cyclist characteristics						
Age	< 16	−0.0042	−0.0167	−0.0822	0.0563	0.0468
Vehicle characteristics						
Veh Speed	<20 mph	−0.0178	−0.0688	−0.2189	0.2107	0.0948
	20–30 mph	−0.0070	−0.0277	−0.1360	0.0798	0.0910
Crash characteristics						
Crash Type	Bicyclist overtaking motorist	−0.0082	−0.0344	−0.2374	0.0777	0.2023
	Motorist Failed to Yield	−0.0056	−0.0227	−0.1153	0.0840	0.0596
	Parallel Paths	−0.0092	−0.0385	−0.2786	0.0572	0.2692
Speeding	Yes	−0.0085	−0.0355	−0.2430	0.0745	0.2125

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

5.2.4. Crash characteristics

Biking direction is a significant impact factor that affects bicyclist injury severity. Since bicyclists will have the ability to prepare while biking facing traffic, the probability of suffering from severe injury severities is decreased, which is consistent with the model results revealed in cluster 2, cluster 3, and cluster 7 (i.e., crashes caused by mid-aged drivers with low speed, crashes occurred on not lighted non-intersection locations in dark, and crashes caused by elderly drivers with low speed). However, in cluster 1, bicyclists are more likely to sustain fatal injury (marginal effect 0.0385), which is related to the specific crash pattern (crashes occurred with vehicle speed from 20 to 30 mph when bicyclists failed to yield). This result indicates that severe injuries might occur due to the fault of bicyclists in bicycle-vehicle crashes. This is consistent with a research study conducted by Kim et al. (2007). Considering this model results, it is necessary to specify the right of way when there is a conflict between drivers and bicyclists. A yield sign will help remind drivers of yielding to bicyclists to reduce the probability of a collision. Furthermore, bicyclists need to wear reflective materials so as to be clearly seen by other road users.

Different crash types will affect bicyclist injury severity distinctively, and different effects may exist for specific crash patterns. In cluster 1, more severe injuries might be suffered by bicyclists when bicyclists turn or merge, while in cluster 6, the opposite results can be concluded that bicyclists are less likely to be severely injured in crashes occurred at non-intersection locations in daylight. That is probably associated with the crash locations (non-intersection) and the lighting condition (daylight). When bicyclists overtake motorist, the likelihood of bicyclists getting severe injuries (including fatal, disabling, and evident injuries) is low for both cluster 1 (marginal effects −0.0228, −0.0786, and −0.415) and cluster 4 (marginal effects −0.0082, −0.0344, and −0.2374), which is probably related to the driving speed of a driver. Similar results can be found when motorists failed to yield in cluster 1 and cluster 4. In contrast, head on crashes and motorists turning or merging might increase the likelihood of suffering severe injuries for crashes occurred at non-intersection locations in daylight, and crashes caused by mid-aged drivers (25–59) and elderly drivers (60+) with low speed (20 mph), respectively, based on the marginal effects shown in Table 5. For crashes occurred at crossing paths and par-

Table 5e
Average marginal effects for cluster 5.

Variable		Cluster5				
		F	D	E	P	N
Cyclist characteristics						
Age	< 16	−0.0600	−0.0512	−0.1140	0.1612	0.0640
	16–24	−0.0476	−0.0380	−0.0697	0.1150	0.0403
	25–54	−0.0748	−0.0540	−0.0530	0.1420	0.0398
Driver characteristics						
Age	25–59	0.0294	0.0230	0.0311	−0.0656	−0.0179
Vehicle characteristics						
Veh_Type	Single Unit Truck	0.0825	0.0523	0.0112	−0.1206	−0.0254
Veh Speed	<20 mph	−0.0590	−0.0522	−0.1394	0.1754	0.0751
	20–30 mph	−0.0527	−0.0453	−0.0984	0.1447	0.0516
	30–40 mph	−0.0379	−0.0306	−0.0477	0.0896	0.0266
Crash characteristics						
Speeding	Yes	0.2066	0.0162	−0.0606	−0.2360	0.0738
Roadway characteristics						
No. of Lanes	<=4	−0.0721	−0.0471	−0.0181	0.1125	0.0247
Environmental Characteristics						
Light Condition	Daylight	−0.0298	−0.0224	−0.0234	0.0604	0.0153
	Dark - lighted roadway	−0.0498	−0.0422	−0.0968	0.1372	0.0517

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

Table 5f
Average marginal effects for cluster 6.

Variable		Cluster6				
		F	D	E	P	N
Vehicle characteristics						
Veh Speed	30–40 mph	0.0193	0.0326	0.0500	−0.0704	−0.0316
	40–50 mph	0.0340	0.0561	0.0784	−0.1201	−0.0483
	50–60 mph	0.0556	0.0837	0.0728	−0.1574	−0.0547
Crash characteristics						
Crash Type	Bicyclist Turn/Merge	−0.0162	−0.0301	−0.0644	0.0738	0.0369
	Head-On	0.0814	0.1146	0.0508	−0.1892	−0.0577
Roadway characteristics						
Road Type	Two-way	−0.3554	−0.2037	0.1903	0.2946	0.0742
Traffic Control	Yes	0.0125	0.0220	0.0386	−0.0508	−0.0223
No. of Lanes	<=4	−0.0262	−0.0428	−0.0516	0.0876	0.0330
Environmental Characteristics						
Weather	Fog, smog, smoke, rain, snow	−0.0129	−0.0239	−0.0532	0.0591	0.0310

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

allel paths, bicyclists are more likely to have possible injury or no injury.

Speeding can affect bicyclist injury severity differently in cluster 3–5. For crashes that occurred on not lighted non-intersection locations in dark (cluster 3), and crashes caused by pickups at non-intersections (cluster 5), speeding could increase the probability of fatal and disabling injuries, while for crashes that occurred on lighted roadways in dark (cluster 4), speeding could only be more likely to cause possible injury or even no injury. Comparing cluster 3 and cluster 4, it can be concluded that lighting condition could be a critical impact factor to the severe injury severities. Therefore, it is important to provide better light condition to reduce the likelihood of bicyclists suffering from severe injuries. Street lights are recommended to be built to ensure good light condition for bicyclists.

For crashes that occurred in urban areas, bicyclists are less likely to suffer severe injuries including fatal, disabling, and evident injuries (marginal effects −0.0006, −0.0111, and −0.0954) in cluster 2 (i.e., crashes caused by mid-aged drivers with low vehicle speed). This is reasonable since low speed usually has negative impacts on severe injuries. In addition, roadways in urban areas have better access control and road conditions.

5.2.5. Roadway characteristics

Curved roadway has a positive impact on severe injury severities, especially for cluster 1, cluster 2, and cluster 7, which indicates a higher probability of severe injuries for particular crash patterns including crashes occurred with vehicle speed from 20 to 30 mph when the bicyclists failed to yield and crashes caused by mid-aged drivers (25+) with low vehicle speed (<20 mph).

Table 5g
Average marginal effects for cluster 7.

Variable		Cluster7				
		F	D	E	P	N
Cyclist characteristics						
Alcohol usage	Yes	0.0366	0.0552	0.3050	−0.3259	−0.0708
Crash characteristics						
Bike Direction	Facing traffic	−0.0021	−0.0036	−0.0749	0.0542	0.0264
Crash Type	Motorist Turn/Merge	0.0038	0.0065	0.1159	−0.0911	−0.0350
	Crossing Paths	−0.0050	−0.0087	−0.2565	0.1117	0.1585
	Parallel Paths	−0.0059	−0.0104	−0.3518	−0.0408	0.4089
Roadway characteristics						
Road Geometry	Curve	0.0082	0.0136	0.1720	−0.1482	−0.0456
Divided Road	Yes	−0.0026	−0.0044	−0.0942	0.0651	0.0360
Temporal Characteristics						
Crash Time	10:00–14:59	−0.0039	−0.0066	−0.1313	0.0937	0.0480
	15:00–17:59	−0.0036	−0.0061	−0.1191	0.0828	0.0460

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

Table 5h
Average marginal effects for the whole dataset.

Variable		Whole Data				
		F	D	E	P	N
Cyclist characteristics						
Age	<16	−0.0265	−0.0098	−0.0162	0.0246	0.0279
	16–24	−0.0193	0.0024	−0.0290	0.0173	0.0286
	25–54	−0.0065	−0.0112	−0.0407	0.0381	0.0202
Alcohol usage	Yes	0.0175	0.0256	0.0069	−0.0634	0.0134
Driver characteristics						
Age	<25	0.0066	0.0118	0.0394	−0.0390	−0.0188
	25–59	0.0033	0.0059	0.0216	−0.0201	−0.0107
Alcohol usage	Yes	0.0714	0.0464	−0.0184	−0.0590	−0.0404
Vehicle characteristics						
Veh Speed	<20 mph	−0.0365	−0.0606	−0.0276	0.0964	0.0282
	20–30 mph	−0.0197	−0.0278	0.0345	−0.0009	0.0140
	40–50 mph	0.0104	0.0189	0.0609	−0.0629	−0.0273
	50–60 mph	0.0390	0.0461	0.0223	−0.0934	−0.0140
Veh_Type	Pickup	0.0080	0.0112	0.0071	0.0191	−0.0454
	Van	0.0067	0.0117	0.0379	−0.0385	−0.0178
	Single Unit Truck	0.0234	0.0390	0.0961	−0.1155	−0.0430
Crash characteristics						
Bike Direction	Facing traffic	−0.0042	−0.0245	−0.0834	0.0986	0.0135
Speeding	Yes	0.0089	0.0156	0.0483	−0.0505	−0.0223
Crash Type	Bicyclist Failed to Yield	0.0071	0.0652	−0.0020	−0.0555	−0.0148
	Head-On	0.0240	0.0402	0.0992	−0.1190	−0.0443
	Motorist Turn/Merge	−0.0137	0.0020	0.1119	−0.0721	−0.0280
Roadway characteristics						
Road Geometry	Curve	0.0124	0.0215	0.0643	−0.0690	−0.0292
Road Type	Two-way	0.0065	0.0121	0.0492	−0.0420	−0.0258
Temporal Characteristics						
Crash Time	10:00–14:59	−0.0032	−0.0059	−0.0218	0.0201	0.0108

N - No Injury.
P - Possible Injury.
E - Evident Injury.
D - Disabling Injury.
F - Fatal Injury.

Different effects are identified for the impact of two-way road on bicyclist injury severities. To be specific, this factor could increase the likelihood of severe injuries including fatal, disabling, and evident injuries in cluster 1, while decreasing the probability of fatal and disabling injuries in cluster 6. This result indicates the necessity of conducting latent class clustering analysis to reveal and interpret the variation of the effect of variables across different clusters.

Furthermore, divided roadway is found to be less likely to result in severe injuries (fatal, disabling, and evident injuries) in cluster 7 with marginal effect being −0.0026, −0.0044, and −0.0942, respectively. Therefore, in order to improve cycling safety, constructing divided roadways is recommended. Bad road conditions could result in different bicyclist injury outcomes in cluster 2 and cluster 3. Bicyclists might be more likely to be severely injured in cluster 2, while being less likely to have severe injuries in cluster 3, which indicates the heterogeneity between clusters. Some similarities can be found in cluster 2 and cluster 6, where traffic control has a positive impact on the high level of bicyclist injury severity. This is probably because traffic control is always related to the intersection areas, which may be a reason for severe injuries outcomes. For the impact of the number of lanes on bicyclist injury severity, a smaller number of lanes has a negative effect on severe injuries, which means that severe bicyclist injuries could be less likely to occur on fewer lanes, especially in cluster 5 and cluster 6.

5.2.6. Temporal characteristics

Different times of day could have various impacts on bicyclist injury severity according to the marginal effects in Table 5. Biking during midnight (0:00–5:59) may increase the likelihood of suffering severe injury severities including fatal, disabling, and evident injuries especially in cluster 2. Also, in cluster 2, different effects can be found during 6:00 to 9:59 in the morning, when bicyclists are less likely to be severely injured. Similar trend can be seen in cluster 7, where bicyclists are at a lower risk to get severe injuries during midday from 10:00 to 14:59. However, for crashes occurred during 15:00 to 17:59, bicyclists are more likely to be fatally injured in cluster 3, while being less likely to have fatal injury in cluster 7, which shows the heterogeneity features between different clusters. This result might correspond to the characterization of cluster 7, which is described as crashes caused by elderly drivers (60+) with low vehicle speed (<20 mph). Summarizing the results regarding the various impact of temporal characteristics on bicyclist injury severity for different clusters, several findings can be concluded. First, temporal characteristics are to a great extent associated with the light condition. The effects of the time periods reflect a part of the influence of light condition in another aspect. Second, the traffic volume varies during different time periods, which is related to the bicyclist injury severities.

5.2.7. Environmental characteristics

Adverse weather conditions are found to have a negative impact on bicyclist injury severity in cluster 2 and cluster 6, while it could increase the probability of severe injuries in cluster 3 (marginal effects 0.089, 0.663, and 0.0247). Since cluster 3 is characterized as crashes occurring on not lighted non-intersection locations in dark, the high probability of severe bicyclist injuries might result from the dark lighting condition. It can be inferred that adverse weather conditions are not the determining factors to severe injuries.

Lighting condition is a significant impact factor to bicyclist injury severity especially for cluster 5. Compared to dark, not lighted roadways, daylight and lighted roadways could decrease the likelihood of severe bicyclist injuries in cluster 5. Maintaining clear sight is essential to enhance cycling safety. As previously mentioned, building appropriate street lights and wearing reflective materials may decrease the risk of collisions.

6. Conclusions and recommendations

This study aims to investigate the differences of the effects of impact factors on bicyclist injury severity existing in various crash patterns. Since bicyclists are more vulnerable compared to other road users, it is essential to evaluate the variables contributing to severe bicyclist injuries. Based on the police reported data collected from 2007 to 2014 in North Carolina, latent class clustering analysis and the subsequent partial proportional odds models and/or ordered logit models are developed. Seven clusters are identified, and four PPO models and three ORL models are built to observe and interpret the impacts on bicyclist injury severity for certain patterns of crashes. It is tested that the sub-models have better goodness of fit compared to the single model developed with the whole dataset, which is consistent with the results of other research studies conducted by Depaire et al. (2009) and Sun et al. (2019).

To better analyze and interpret the bicyclist injury outcomes, marginal effects are calculated to reveal the accurate effect of each significant variable. Results show that some variables only have significant influence in specific clusters, and the impact of the same variable for various clusters can be different, which indicate that latent class clustering provides as a more accurate and insightful method to explore the information on the impact of contributing factors for further analysis of bicyclist injury severity and the improvement of bicyclist safety.

In addition, the findings of the model results provide some more useful information and guidance to help mitigate severe bicyclist injuries and improve biking safety. To be specific, since bicyclists under the influence of alcohol are found to be more likely to suffer severe injuries, it is critical to establish relative regulations to prevent bicyclists from drinking alcohol while riding a bicycle. Similar enforcement should be enhanced for drivers to inhibit driving under the influence of alcohol. Furthermore, the lighting condition is a significant contributing factor to severe injuries. Therefore, constructing lights in the areas with high bicyclist activities might help improve biking safety.

However, there are still some limitations in this research study. It should be noted that it is possible to model the latent class and the bicyclist injury severity simultaneously, for example, by combining latent class and multinomial logit models that could be more effectively compared to the sequential approach adopted in this study (Yasmin et al., 2014a, 2014b). In addition, although latent class clustering tries to maximize the homogeneity within clusters, the unobserved heterogeneity might be neglected when developing PPO/ORL models. Therefore, in the future study, the authors will try to apply mixed effect PPO model (Eluru & Yasmin, 2015) for the analysis of bicyclist injury severity.

Declaration of Competing Interest

The authors 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

Behnood, A., & Mannering, F. (2017). Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research, 16*, 35–47.

Biernacki, C., & Govaert, G. (1999). Choosing models in model-based clustering and discriminant analysis. *Journal of Statistical Computation and Simulation, 64*(1), 49–71.

Bijmolt, T. H., Paas, L. J., & Vermunt, J. K. (2004). Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing, 21*, 323–340.

Chang, F., Xu, P., Zhou, H., Chan, A. H., & Huang, H. (2019). Investigating injury severities of motorcycle riders: A two-step method integrating latent class cluster analysis and random parameters logit model. *Accident Analysis & Prevention, 131*, 316–326.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. John Wiley & Sons.

Decker, S., Otte, D., Cruz, D. L., Müller, C. W., Omar, M., Krettek, C., & Brand, S. (2016). Injury severity of pedestrians, bicyclists and motorcyclists resulting from crashes with reversing cars. *Accident Analysis & Prevention, 94*, 46–51.

Depaire, B., Wets, G., & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. *Accident Analysis & Prevention, 40*(4), 1257–1266.

Eluru, N., & Yasmin, S. (2015). A note on generalized ordered outcome models. *Analytic Methods in Accident Research, 8*, 1–6.

Eluru, N. (2013). Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. *Accident Analysis & Prevention, 55*, 1–11.

Eluru, N., Bhat, C. R., & Hensher, D. A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis & Prevention, 40*(3), 1033–1054.

Gårder, P. (1994). Bicycle accidents in Maine: An analysis. *Transportation Research Record*. 34-34.

Gårder, P., Leden, L., & Pulkkinen, U. (1998). Measuring the safety effect of raised bicycle crossings using a new research methodology. *Transportation Research Record, 1636*(1), 64–70.

Gong, L., & Fan, W. D. (2017). Modeling single-vehicle run-off-road crash severity in rural areas: Accounting for unobserved heterogeneity and age difference. *Accident Analysis & Prevention, 101*, 124–134.

Götschi, T., Garrard, J., & Giles-Corti, B. (2016). Cycling as a part of daily life: A review of health perspectives. *Transport Reviews, 36*(1), 45–71.

Kaplan, S., Vavatsoulas, K., & Prato, C. G. (2014). Aggravating and mitigating factors associated with cyclist injury severity in Denmark. *Journal of Safety Research, 50*, 75–82.

Kashani, A. T., & Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety Science, 49*(10), 1314–1320.

Kelly, P., Kahlmeier, S., Götschi, T., Orsini, N., Richards, J., Roberts, N., ... Foster, C. (2014). Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship. *International Journal of Behavioral Nutrition and Physical Activity, 11*(1), 132.

Kim, J. K., Kim, S., Ulfarsson, G. F., & Porrello, L. A. (2007). Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis & Prevention, 39*(2), 238–251.

Klassen, J., El-Basyouny, K., & Islam, M. T. (2014). Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A City of Edmonton case study. *Safety Science, 62*, 295–304.

Kröyer, H. R. (2015). The relation between speed environment, age and injury outcome for bicyclists struck by a motorized vehicle–a comparison with pedestrians. *Accident Analysis & Prevention, 76*, 57–63.

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

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.

Li, Y., & Fan, W. (2019). 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.

Lin, Z., & Fan, W. (2019a). Cyclist injury severity analysis with mixed-logit models at intersections and nonintersection locations. *Journal of Transportation Safety & Security*, 1–23.

Lin, Z., & Fan, W. (2019b). Modeling bicyclist injury severity in bicycle–motor vehicle crashes that occurred in urban and rural areas: A mixed logit analysis. *Canadian Journal of Civil Engineering, 46*(10), 924–933.

Liu, J., Khattak, A. J., Li, X., Nie, Q., & Ling, Z. (2020). Bicyclist injury severity in traffic crashes: A spatial approach for geo-referenced crash data to uncover non-stationary correlates. *Journal of Safety Research*.

Liu, P., & Fan, W. (2018). Modeling head-on crash severity on NCDOT freeways: A mixed logit model approach. *Canadian Journal of Civil Engineering, 46*(6), 322–328.

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

Macmillan, A., Connor, J., Witten, K., Kearns, R., Rees, D., & Woodward, A. (2014). The societal costs and benefits of commuter bicycling: Simulating the effects of specific policies using system dynamics modeling. *Environmental Health Perspectives, 122*(4), 335–344.

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.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

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.

Mooradian, J., Ivan, J. N., Ravishanker, N., & Hu, S. (2013). Analysis of driver and passenger crash injury severity using partial proportional odds models. *Accident Analysis & Prevention, 58*, 53–58.

Moore, D. N., Schneider Iv, W. H., Savolainen, P. T., & Farzaneh, M. (2011). Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and nonintersection locations. *Accident Analysis & Prevention, 43*(3), 621–630.

Nabors, D., Goughnour, E., Thomas, L., DeSantis, W., & Sawyer, M. (2012). *Bicycle road safety audit guidelines and prompt lists (No. FHWA-SA-12-018)*. United States: Federal Highway Administration. Office of Safety.

NHTSA (2019). *Traffic Safety Facts 2017: A compilation of motor vehicle crash data from the fatality analysis reporting system and the General Estimates System*. Washington DC: U. S Department of Transportation.

Nilsson, P., Stigson, H., Ohlin, M., & Strandroth, J. (2017). Modelling the effect on injuries and fatalities when changing mode of transport from car to bicycle. *Accident Analysis & Prevention, 100*, 30–36.

Pai, C. W., & Saleh, W. (2007). An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK. *Safety Science, 45*(8), 832–847.

Peterson, B., & Harrell, F. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 39*(2), 205–217.

Prati, G., Pietrantoni, L., & Fraboni, F. (2017). Using data mining techniques to predict the severity of bicycle crashes. *Accident Analysis & Prevention, 101*, 44–54.

Pucher, J., & Dijkstra, L. (2003). Promoting safe walking and cycling to improve public health: Lessons from the Netherlands and Germany. *American Journal of Public Health, 93*(9), 1509–1516.

Pucher, J., Buehler, R., & Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A: Policy and Practice, 45*(6), 451–475.

Rash-ha Wahi, R., Haworth, N., Debnath, A. K., & King, M. (2018). Influence of type of traffic control on injury severity in bicycle–motor vehicle crashes at intersections. *Transportation Research Record, 2672*(38), 199–209.

Robartes, E., & Chen, T. D. (2017). The effect of crash characteristics on cyclist injuries: An analysis of Virginia automobile-bicycle crash data. *Accident Analysis & Prevention, 104*, 165–173.

Rojas-Rueda, D., de Nazelle, A., Tainio, M., & Nieuwenhuijsen, M. J. (2011). The health risks and benefits of cycling in urban environments compared with car use: Health impact assessment study. *BMJ, 343* d4521.

Sasidharan, L., & Menéndez, M. (2014). Partial proportional odds model - An alternate choice for analyzing pedestrian crash injury severities. *Accident Analysis & Prevention, 72*, 330–340.

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.

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

Sivasankaran, S. K., & Balasubramanian, V. (2020). Exploring the severity of bicycle–Vehicle crashes using latent class clustering approach in India. *Journal of Safety Research, 72*, 127–138.

Stone, M., & Broughton, J. (2003). Getting off your bike: Cycling accidents in Great Britain in 1990–1999. *Accident Analysis & Prevention, 35*(4), 549–556.

Sun, M., Sun, X., & Shan, D. (2019). Pedestrian crash analysis with latent class clustering method. *Accident Analysis & Prevention, 124*, 50–57.

Ulfarsson, G. F., & Mannering, F. L. (2004). Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. *Accident Analysis & Prevention, 36*(2), 135–147.

Valent, F., Schiava, F., Savonitto, C., Gallo, T., Brusaferro, S., & Barbone, F. (2002). Risk factors for fatal road traffic accidents in Udine, Italy. *Accident Analysis & Prevention, 34*(1), 71–84.

Vanparijs, J., Panis, L. I., Meeusen, R., & de Geus, B. (2015). Exposure measurement in bicycle safety analysis: A review of the literature. *Accident Analysis & Prevention, 84*, 9–19.

Wang, X., & Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis & Prevention, 40*(5), 1674–1682.

Washington, S. P., Karlaftis, M. G., & Mannering, F. (2010). *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Xia, T., Zhang, Y., Crabb, S., & Shah, P. (2013). Cobenefits of replacing car trips with alternative transportation: a review of evidence and methodological issues. Journal of Environmental and Public Health, 2013.

Yan, X., Ma, M., Huang, H., Abdel-Aty, M., & Wu, C. (2011). Motor vehicle-bicycle crashes in Beijing: Irregular maneuvers, crash patterns, and injury severity. *Accident Analysis & Prevention, 43*(5), 1751–1758.

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

Yasmin, S., Eluru, N., Bhat, C. R., & Tay, R. (2014b). A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. *Analytic Methods in Accident Research, 1*, 23–38.

Yu, H., Li, Z., Zhang, G., & Liu, P. (2019). A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. *Analytic Methods in Accident Research, 24* 100110.

Zijing Lin is a Ph.D. student in the Infrastructure and Environmental Systems (INES) program at the University of North Carolina at Charlotte, Charlotte, NC, U.S.A. She is with the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE). Her research topics include transportation safety data analysis, and non-motorized transportation.

Wei (David) Fan currently serves as an associate professor in the Department of Civil and Environmental Engineering at The University of North Carolina at Charlotte (UNC Charlotte). He is the Director of the USDOT Tier I University Transportation Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE). Dr. Fan holds a Ph.D. (May 2004) in Civil Engineering – Transportation from the University of Texas at Austin (Hook 'em Horns!). He was a Senior Analytical Optimization Software Developer for the R&D Department at SAS Institute Inc. located in Cary, North Carolina from June 2004 – August 2006. Dr. Fan's primary research interests include operations research, transportation network modeling & optimization, multimodal transportation systems planning and operations, traffic operations and simulation, transportation safety and crash analysis, intelligent transportation systems (ITS), and computer software development. Dr. Fan serves as an associate editor of IEEE Transactions – Intelligent Transportation Systems while also serving as a member of three other transportation journal editorial boards. Dr. Fan also serves as a member on the National Science Foundation (NSF) review panels, two National Cooperative Highway Research Program (NCHRP) panels, one Transit Cooperative Research Program (TCRP) panel, and three American Society of Civil Engineers (ASCE) Committees (on the Connected & Autonomous Vehicles Impacts, Public Transport, and Rail Transportation), as well as two Transportation Research Board (TRB) committees. He has been working as an active ad hoc reviewer for more than 20 prestigious journals as well as the national Highway Capacity Manual (HCM) (2010 version). He was selected by the students as the Best Civil Engineering Professor of the Year in 2007 at The University of Texas at Tyler and was a proud ASCE Excellence in Civil Engineering Education (ExCEEd) Teaching Fellow in 2007.

Dr. Fan is involved in many sponsored projects (over 14 million dollars in funding), having been principal or co-principal investigator on many research studies for the U. S. Department of Transportation (USDOT), Federal Highway Administration (FHWA), National Cooperative Highway Research Program (NCHRP), SHRP2 Education Connection, Texas Department of Transportation (TxDOT) and North Carolina Department of Transportation (NCDOT). He had published more than 100 journal articles, proceeding papers and technical reports on transportation network modeling, multimodal transportation systems planning and operations, traffic operations and simulations, traffic safety and crash analysis, and operations research applications in transportation. He is a registered professional engineer in Texas.