Cluster-Based Approach to Analyzing Crash Injury Severity at Highway–Rail Grade Crossings

Yashu Kang and Aemal Khattak

The presence of unobserved heterogeneity in crash data can result in estimation of biased model parameters and incorrect inferences. The research presented in this paper investigated severity of crashes reported at highway–rail grade crossings by appropriately clustering the data, accounting for unobserved heterogeneity. A combination of data mining and statistical regression methods was used to cluster crash data into subsets and then to identify factors associated with crash injury severity levels. This research relied on highway–rail accident, incident, and crossing inventory databases for 2011 to 2015 obtained from FRA. Three clustering methods—*K*-means, traditional latent class cluster, and variational Bayesian latent class cluster—were considered, and the variational Bayesian latent class cluster method was chosen for partitioning the data set for model estimation. Unclustered data as well as the clustered subsets were used to estimate ordered logit models for crash injury severity. A comparison revealed that the cluster-based approach provided more relevant model parameters and identified factors relevant only to certain clusters of the data.

There are approximately 211,000 public and private highway–rail grade crossings (HRGCs) in the United States. Despite substantial crash frequency reduction at HRGCs (about 80% reduction between 1980 and 2014), severe injury outcomes remain an important issue at rail crossings (*1*). Crashes at HRGCs often result in serious consequences and can affect both highway and rail networks. For instance, HRGC crashes accounted for 32% of all rail-related fatalities, and 57.3% of HRGC crashes involved injuries or fatalities in 2013 (*2*).

Researchers have focused on HRGC crash frequency and explored contributing factors affecting crash occurrence (*3–5*). Research on crash severity is also available in the literature; the research presented here adds to the body of knowledge on injury severity of crashes reported at HRGCs.

Researchers have used a variety of modeling techniques that can be broadly divided into two sets. The first set of models includes discrete outcome models (e.g., logit, probit, and variations of multinomial logit models). These models do not explicitly consider the ordinal nature of injury severity commonly used in crash reporting (e.g., no injury, injury, and fatal injury). The second set of models, such as ordered logit (OL) and ordered probit, take into account the ordinal nature of injury severity. A limitation of this set of models is that exogenous variable impacts do not vary across alternatives (*6*). To statistically explore latent variables, data mining has been used in many scientific areas (*7, 8*). Among the data mining techniques, classification methods such as decision trees, kernel estimation, and neural networks and clustering techniques such as latent class and *K*-means are common.

This research examined the effectiveness of data clustering before modeling injury severity of train–vehicle crashes reported at HRGCs by using three clustering methods to appropriately identify factors associated with HRGC injury severity. The proposed approach combined the OL regression model with three clustering techniques: *K*-means, latent class, and Bayesian latent class. The advantage in clustering data into homogeneous subsets is that doing so addresses unobserved heterogeneity—an issue in statistical modeling that refers to the presence of unobserved relevant variables that are correlated with the observed variables in a model. The estimated parameters in a model, when unobserved heterogeneity is present, will be biased, and incorrect inferences could be drawn.

The organization of this paper is as follows. A review of published literature on HRGC crash injury severity follows the introduction. The methodology applied in this research is explained, followed by presentation of the data used in clustering and regression model estimation. Next is an examination of the modeling results. The paper concludes with a discussion of the results, including limitations of the presented research.

LITERATURE REVIEW

In this section, published literature covering HRGC crash injury severity and clustering data before injury severity modeling are discussed.

HRGC Crash Injury Severity

By investigating data obtained from FRA, Raub examined four specific warning device classes and compared safety effects by using univariate analysis (*9*). Raub reported that transforming crossbucks to stop signs created a false sense of improved safety with respect to crash frequency and injury severity but did not discuss details of injury severity. Using a generalized logit model, a study of HRGCs in Taiwan identified factors associated with crash injury severity that included the number of daily trains, the number of daily trucks, highway separation, an obstacle detection device, and approaching

Y. Kang, 330H Whittier Research Center, and A. Khattak, 330E Whittier Research Center, Department of Civil Engineering and Nebraska Transportation Center, College of Engineering, University of Nebraska–Lincoln, Lincoln, NE 68583-0851. Corresponding author: A. Khattak, khattak@unl.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2608, 2017, pp. 58–69.
http://dx.doi.org/10.3141/2608-07

crossing marks (*10*). The study did not include driver characteristics or environmental factors.

Eluru et al. analyzed the influence of various exogenous factors in an HRGC study with a focus on crash and crossing attributes (*11*). The issue of heterogeneity in the data set was addressed with a latent segmentation-based logit model. Their results highlighted risk segmentation related to the presence of active warning devices and the presence of permanent structures.

Researchers have looked at crash injury severity from another perspective by examining the impact of type of grade crossing control with an ordered probit model (*12*). Age of highway motor vehicle drivers, traffic volume, and weather conditions were identified as important factors. As well, driver gender and behavior were identified as influencing factors on the degree of injury (*13*).

Studies of pedestrian injury severity at HRGC crashes showed that higher train speed, female pedestrians, and commercial land use were associated with more severe injuries, whereas more crossing highway lanes and the presence of standard flashing light signals in clear weather decreased the likelihood of severe injuries (*14, 15*). In studies of the severity of train–motor vehicle crash injuries, factors related to more severe injuries included higher number of daily trains, adverse weather conditions, driver age over 60, and high train speed (*10–13, 16*).

In general, crash injury severity has not received as much attention as crash frequency, although this appears to be changing. There have been methodological advances and some variable database use (*11*). Detailed postcrash data are not available for the approach in Raub's work (*9*), but it is available for injury severity models conditioned on a crash having occurred (*17*). Among injury severity models, besides latent class segmentation applied in a few studies (*11, 14*), a simple ordinal response regression or multinomial regression has been the modus operandi (*12, 13, 16*).

Unobserved Heterogeneity in Injury Severity Modeling

In previous studies, it was revealed that some factors affecting crash frequency and severity could be not be observed or were not available, but they were correlated with some of the observed variables (*14, 17*). Referenced as unobserved heterogeneity, it can cause estimation of biased model parameters and lead to incorrect inferences (*17*). For example, a person's age is usually an explanatory variable in many injury models, while various underlying factors such as health status, driving age, and reaction time could be overlooked. Assuming age holds the same role across various populations on crash injury severity, potential bias and complications can creep into the modeling results.

Fan et al. compared the multinomial logit model and the OL model with respect to identifying key factors and prediction of crash severity at HRGCs (*18*). They reported that the multinomial logit model predicted crash severity outcomes better. Eluru et al. obtained homogeneous highway crash data by classifying possible factors that may account for highway motor vehicle crashes into six categories: vehicle characteristics, roadway characteristics, pedestrian characteristics, motorized vehicle driver characteristics, environmental factors, and crash characteristics (*19*). Similarly, the FRA database was used to categorize risk factors, and crossing characteristics were included (*11*).

Pertaining to the heterogeneity issue, researchers developed models based on traffic accident type (*20*). By estimating separate models, the difference in magnitude that risk factors have on injury outcomes could be described. As stated by Ulfarsson and Mannering, some factors may not be statistically significant for all accident types because of insufficient observations or the difference in magnitude (*21*).

The segmentation of crash data used to be done to study a specific problem or empirical decisions. Thus, heterogeneity cannot be precluded with traditional segmentation. To maintain a group of homogeneous observations within each segment, data mining techniques are usually used. Data mining pays more attention to the complexity of models and fits the idea of learning. Among various data mining techniques, some have been adopted for traffic safety research, such as artificial neural networks, classification, and regression trees (*22, 23*).

Clustering analysis is a frequently considered data mining approach along with the unsupervised learning technique. The practice of clustering forms subpopulations that consist of relatively homogeneous data. In a study by Kim and Yamashita, the *K*-means algorithm was applied to examine pedestrian-involved crashes in Hawaii (*24*). The authors compared hierarchical clustering techniques to *K*-means clustering and reported that both were useful tools for safety research. Prato et al. suggested that *K*-means as a descriptive technique is useful for classifying a large crash data set, although they also encountered problems regarding clear cutting among clusters (*25*). The clear-cut problem is discussed in the results section of this paper.

Latent class analysis has been used in traffic safety research. A latent class cluster (LCC) was used for a preliminary analysis, and then the full data set and each of the identified clusters were used to estimate a multinomial logit model (*26*). The authors suggested that clustered data yielded additional information compared with the full data set.

De Oña et al. applied an LCC along with Bayesian networks to investigate 3,229 crashes in Granada, Andalusia, Spain, between 2005 and 2008 (*27*). The crash database was segmented by means of latent class clustering, and the Bayesian network inference was established to obtain variables associated with fatalities or serious injuries. This research indicated that clustering analysis added value to subsequent injury analyses.

METHODOLOGY

Descriptive data mining was adopted in this study for data clustering to deal with unobserved heterogeneity within the data set. Ordered logistic regression modeling that took into account the ordinal nature of crash injury outcome was used for identifying the relationships between explanatory and response variable (crash injury severity). The effectiveness of this combination of data clustering and logistic regression is examined in this paper. Details of data clustering are described in the next section.

Clustering Analysis

Clustering analysis classifies data into groups to achieve homogeneity within each group and heterogeneity among different groups. As a category of unsupervised learning methods, various techniques are available, as described below.

K-*Means Clustering*

K-means clustering partitions data so that each observation belongs to a cluster with the nearest mean value with no hierarchical relations. A mapping of the interval (0, 1) can allow the distance calculation for the data set with a mixed type of variables (both continuous and categorical). This approach maximizes the similarity within each cluster and the dissimilarity between clusters by calculating distances between data set elements. Equivalently, K-means minimizes (*28*)

$$\sum_{k=1}^{K}\sum_{r=1}^{N}\sum_{i=1}^{P}(x_{\mathrm{rik}}-\bar{x}_{\mathrm{ik}})^2 \tag{1}$$

where x_{rik} denotes observation r in cluster k for variable i, and $\bar{x}_{\mathrm{ik}}$ indicates the mean of variable i in cluster k. If all variables are categorical, another k-modes algorithm is suggested instead. As with the determination of cluster numbers, there is no formal information criteria for K-means. One can calculate

$$W=\frac{\text{within sum of squares}}{\text{total sum of squares}} \tag{2}$$

and also refer to another statistical procedure, principal component analysis, to make sure the number of clusters makes sense. The W value is a measure of total variance in the training data set that is explained by the clustering. K-means attempts to minimize the within-group dispersion while maximizing between-group dispersion. In other words, a leveling of W values with increasing cluster numbers suggests that larger values of K are not needed.

Latent Class Clustering

LCC is used to deal with heterogeneity within a data set. Unlike the partitioning approach of K-means, LCC performs on the basis of a mixture model, allowing analysts to better understand probabilistic properties of the classifications. An in-depth explanation of LCC is available elsewhere (*29*). For this research, a package performing latent class analysis within a Bayesian framework was selected (*30*). Latent class analysis examines one or more unobserved categorical variables to identify the latent relationship between a set of known variables. Traditional LCC uses frequentist expectation–maximization to obtain posterior probability (probability an observation will be assigned to a subpopulation). This method is similar to classification methods allocating observations nonhierarchically, like K-means. The form of the latent model is

$$f(X_i|\theta)=\sum_{k=1}^{K}\pi_k\prod_{j=1}^{J}f_k(X_{\mathrm{ij}}|\theta_{\mathrm{jk}}) \tag{3}$$

where

$f_k(X_{\mathrm{ij}}|\theta_{\mathrm{jk}})$ = mixture probability density (*29*),
X_i = vector of observed variables from observation i,
k = number of clusters,
j = indicator variable, and
θ = cluster-specific parameter.

Traditional LCC based on the expectation–maximization algorithm may converge to a local maximum (rather than a global maximum), giving an inferior solution. Consequently, an alternative, the variational Bayesian–based LCC (VBLCC) was used in this research. VBLCC is an adaption of the LCC method that includes a Bayesian framework and can avoid local convergence and overfitting problems (*30*).

With respect to the model selection procedure, commonly addressed information criteria such as Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) can be used to select the optimal number of classes for data fitting. A higher value indicates a better fit to the data.

OL Models

Many researchers have used the OL model to model crash injury severity (*6, 19–20*). The structural model is of the form

$$y_i^*=\sum_{k=1}^{K}\beta_k X_{\mathrm{ki}}+\varepsilon_i \tag{4}$$

where

y_i^* = latent continuous variable mapped into the observed injury severity y_i,
β_k = vector of parameters,
ε_i = error term, and
X_{ki} = set of independent variables.

Response variable y_i in this research consists of three categories: (*a*) no injury, (*b*) injury, and (*c*) fatality. According to estimated model parameters, each observation or crash can be classified into one of the injury severity outcomes. The probability P of each observation with the injury level category is expressed as

$$P_k(i)=\frac{e^{\beta_k X_{\mathrm{ki}}}}{\sum_{k=1}e^{\beta_k X_{\mathrm{ki}}}} \tag{5}$$

The OL model was used in this research to account for the ordinal nature of the variable for crash injury severity. To summarize, the training data set (80% of the 5-year data) was divided into several clusters with each of the proposed clustering techniques. Then, OL regression models were estimated for each subpopulation as well as for the full training data. The remaining 20% data set was used for validation and comparison of the three clustering approaches.

DATA

Two FRA databases provided the data analyzed in this research; these were the 2011 to 2015 HRGC crash data and the HRGC inventory databases. These were merged together with the common crossing identification number variable. The HRGC crash database included information such as crash characteristics, highway user demographics, and environmental factors at the time of the crash. The HRGC inventory database contained details about the crossing features, such as daily train and highway traffic and types of warning devices.

The matched HRGC crash inventory data set contained 10,505 records and was classified into six categories according to available variables: crossing features, safety infrastructure, motorist characteristics, highway features, environmental factors, and crash characteristics. Observations with missing data were excluded to facilitate cluster formation, rendering an analysis data set of 7,606 observations. One indicator variable yielded very few population shares and was excluded from the analysis data set (rail yard land use—0.026%), as were records implying pedestrians and suicide attempts, resulting in an analysis data set consisting of train–vehicle crashes only. Table 1 presents a summary of the descriptive statistics for the analysis data

TABLE 1 Variable Frequencies for Analysis Data Set

Variable	Frequency	Percentage
Crash Characteristics		
Injury severity outcome		
No injury	4,727	62.11
Injury	2,252	29.61
Fatal injury	637	8.28
Crash type		
Rail equipment struck highway user	6,185	81.30
Rail equipment struck by highway user	1,421	18.70
Train speed		
<35 mph	4,270	57.90
35 to 50 mph	2,084	28.26
>50 mph	1,021	13.84
Vehicle speed		
<35 mph	7,189	95.93
35 to 50 mph	187	2.63
>50 mph	116	1.55
Number of locomotives		
No more than 2 locomotives	5,181	68.10
More than 2 locomotives	2,425	31.90
Motorist in vehicle or not		
Yes	6,298	82.79
No	1,308	17.21
Number of cars in train		
<30	3,295	43.31
30 to 50	908	11.93
>50	3,403	44.76
Train direction		
North	1,904	25.03
South	1,921	25.26
East	1,831	24.07
West	1,950	25.64
Action of highway user		
Went around gates	774	10.96
Went through gate	240	3.40
Stopped, then proceeded	470	6.65
Did not stop	2,925	41.44
Stopped on crossing	1,902	26.93
Went around or through barricade	13	0.18
Other	738	10.45
Crossing Features		
Land use		
Residential	1,649	21.69
Commercial	1,919	25.22
Industrial	1,495	19.66
Open space	2,010	26.42
Rail yard	2	0.026
Other	531	6.98
Crossing type		
Public crossing	6,391	84.03
Private crossing	1,215	15.97
Day through trains		
<15	2,446	32.15
16 to 50	1,702	22.37
51 to 85	1,175	15.47
>85	2,283	30.00
Night through trains		
<15	1,540	20.24
16 to 50	2,809	36.93
51 to 85	391	5.15
>85	2,866	37.67
Number of main tracks		
1	3,805	50.01
2	860	11.33
Other	2,941	38.66
Safety Infrastructure		
Warning device		
Flashing light	3,840	50.47
Wigwag	30	0.39
Stop sign	1,560	20.50
Crossbuck	5,302	69.69
Highway signal	268	3.52
Other device (including audible, watchman, and so on)	3,659	48.12
Pavement markings		
Yes	5,546	72.92
No	2,060	27.08
Highway User Characteristics		
Highway user age		
<20	662	8.70
20 to 30	1,607	21.12
30 to 40	1,360	17.88
40 to 50	1,355	17.82
50 to 60	1,332	17.52
>60	1,290	16.96
Highway user gender		
Male	5,651	74.28
Female	1,955	25.72
Highway user type		
Auto (including van)	3,835	50.38
Truck (including trailer and pickup)	3,113	40.92
Motorcycles and other motor vehicles	439	5.77
Bus	9	0.12
Other	210	2.76
Vehicle direction		
North	1,975	25.99
South	1,817	23.91
East	2,006	26.42
West	1,799	23.67
Highway Features		
Area type		
Urban	3,404	44.74
Rural	3,133	41.21
Other	1,069	14.05
Highway paved or not		
Yes	5,548	72.92
No	2,058	27.08
AADT		
<1,000	2,901	38.13
1,000 to 5,000	1,422	18.72
5,000 to 10,000	874	11.49
>10,000	2,409	31.66
Highway classification		
Interstate	112	1.47
Expressway	12	0.17
Arterial	1,543	20.28
Collector	1,551	20.40
Local	4,388	57.68
Environmental Factors		
Visibility		
Dawn	523	6.89
Day	4,533	59.60
Dusk	608	7.99
Dark	1,942	25.53
Crossing illuminated or not		
Yes	1,850	24.34
No	5,756	75.66
Road condition		
Dry	5,782	76.00
Wet	692	9.10
Snow	306	4.02
Ice	118	1.55
Gravel	171	2.27
Other	537	7.06
Weather		
Clear	5,230	68.74
Cloudy	1,527	20.07
Rain	522	6.86
Fog	90	1.18
Sleet	17	0.22
Snow	220	2.92

NOTE: AADT = annual average daily traffic.

set. For model estimation and validation, the analysis data set was randomly divided into a training data set (6,084 observations) and a validation data set (1,522 observations), according to the 80/20 rule (Pareto principle).

RESULTS

Use of the *K*-means, the latent class, and the variational-Bayesian latent class techniques allowed unobserved heterogeneity within the crash data to be addressed; salient features of the data set described the cluster subpopulations. For example, if one cluster had 90% female motorists while other clusters had a balanced distribution over the gender variable, this subpopulation could be described as "female highway users."

Clustering Process

Determination of the appropriate number of clusters was necessary before partitioning of the training data set. For selection criteria, BIC and AIC were chosen for the basic LCC approach, and Equation 2 was used for *K*-means. AIC Monte Carlo (AICM) and BIC Monte Carlo (BICM), adaptations for the variational Bayesian framework, were used for VBLCC (*31*). For the specific package, BayesLCA in R, a higher value with respect to a criterion is considered a better fit to the data (*30*). Figure 1 indicates that for the LCC method, BIC increased rapidly until six clusters, then started leveling, whereas there was still a monotonic increase in the AIC. BIC is usually more reliable with respect to big data sets than AIC is (*32*). The drop at AICM and BICM may be related to the variational Bayesian approximation. For the *K*-means, as described previously, the *W* value significantly decreased between four and five clusters, indicating that at least five clusters were appropriate. In this research, six clusters were used.

The use of the three clustering methods showed that results from the basic LCC and VBLCC were almost identical, with approximate classification probabilities 25.57%, 18.02%, 12.97%, 13.86%, 15.34%, and 14.22%. An examination of each observation showed that the deviation between LCC and VBLCC classification results was approximately 1.19%. The small dissimilarity indicated no local convergence problem for this data set. However, for the *K*-means results, the classification probabilities were 24.67%, 19.03%, 22.23%, 22.71%, 10.00%, and 1.36%. Considering the dissimilar grouping results, interpreting the variable distribution among clusters is important. Table 2 presents a summary of salient variables and their distributions in each cluster by LCC. The predominant variables in each cluster included land use around HRGCs, highway classification, and highway user gender. However, the variables and distributions in the *K*-means clustering results disclosed relatively few prominent variables. The three major principal components were obtained by principal component analysis and hence are used here as a coordinate axis for visualizing distributions of the observations. Each color indicates a different cluster. A clear separation between colors indicates a good clustering, such as the plot in Figure 2*a*, whereas there is no clear separation between clusters in the plot in Figure 2*b*. A detailed summary of the *K*-means results is not reported here.

According to the information included in Table 2, for Cluster 1, identifying variables were urban highway (0% rural highway), male highway user (female 0%), and crossing equipped with safety infrastructure other than a stop sign (stop sign 4.0%). With respect to Cluster 2, a substantial proportion of highway user gender (99.8% female) and type (89.0% highway user: auto) denotes its classification, and similarly to Cluster 1, there are few stop signs but are other safety devices. Cluster 3 also had some significant features: local highway (93.6% local highway) located at open spaces (100% open space) in rural areas (100% rural highway), with crossings having crossbucks other than flashing lights and paved markings. Within Cluster 4, 98.4% of the elements were rural highways with crossbucks. Cluster 5 consisted of male highway (0% female) users at rural HRGCs with few stop signs. In addition to other features, Cluster 6 mostly consisted of private crossings (96.3% private crossings).

Figure 3 shows plots for the item probability parameters. Item probability denotes the prior probability of belonging to a group or subpopulation. Some plots indicated well-separated groups such as auto, but other plots showed indistinct separation, such as most groups in TypeXing (type of crossing, public or private).

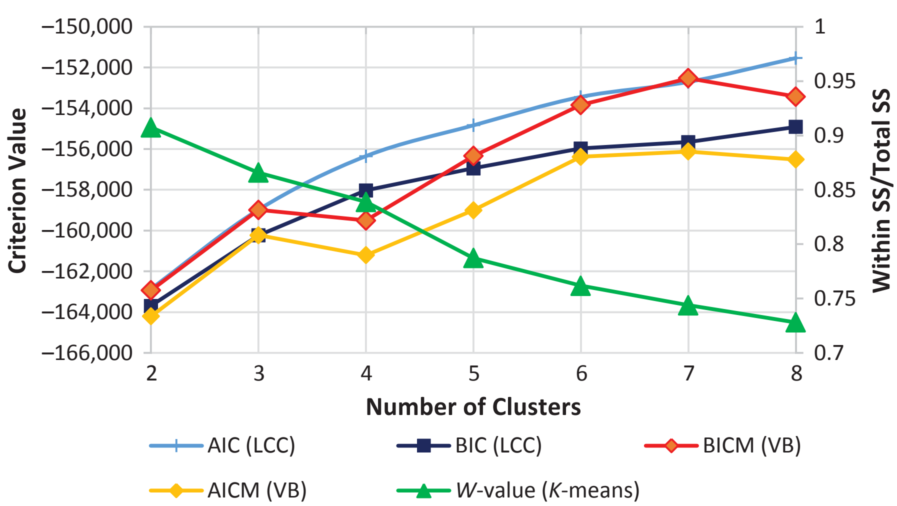


FIGURE 1 Information criteria and *W* value corresponding to number of clusters (SS = sum of squares).

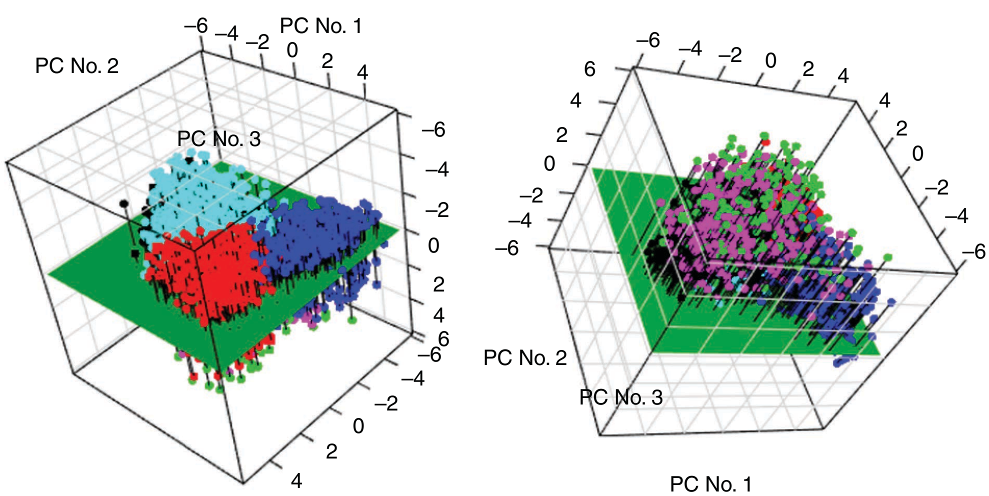
TABLE 2 Summary of Salient Variables and Distributions in Each Cluster, LCC

Variable	Full Data (%)	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)	Cluster 4 (%)	Cluster 5 (%)	Cluster 6 (%)
Private crossing	15.97	0.3	1.5	0.4	13.1	0.3	96.3
Land use							
Residential	21.69	20.7	27.2	0	51.8	25.0	23.6
Commercial	25.22	46.6	38.9	0	12.4	23.4	1.9
Open space	26.42	13.1	20.5	100	10.9	45.0	0.5
Rural highway	41.21	0	35	100	59.3	100	0.5
Highway classification							
Arterial highway	20.28	45.4	30.2	0.2	0.5	19.9	0
Collector highway	20.40	23.8	34.9	6.2	0.1	49.6	0.5
Local highway	57.68	29.7	34	93.6	98.4	30.4	0
Presence of safety infrastructure							
Stop sign	20.50	4.0	4.6	29.4	49.2	7.4	59.4
Crossbuck	69.69	62.3	65.7	96.6	91.6	66.6	48.3
Flashing	50.47	88.3	79.6	6.3	6.6	84.2	6.2
Paved markings	72.92	76.2	69.6	4.7	20.4	72.2	14.8
Other device	48.12	74.9	71.4	20.9	18.7	70.0	9.7
Gender: female	25.72	0	99.8	24	25	0	20.8
Age: 60 or older	16.96	19.0	21.5	19.0	25.2	21.2	22.1
Highway user: automobile	50.38	55.6	89.0	37.0	44.9	41.8	37.2
Roadway condition							
Dry	76.00	83.1	75.3	79.2	79.2	78.9	76.2
Snow	4.02	1.8	4.4	4.3	4.0	3.3	1.6
Gravel	2.27	0	0	6.5	2.6	0	4.9
Proportion of full data set (training data)		25.57	18.02	12.97	13.86	15.34	14.22

VBLCC takes significantly less time to calculate the classification probabilities although their maximum a posteriori results are identical. The clustering procedure provided credence for data segmentation. Examination of the performance of various clustering approaches was undertaken after the regression models yielded empirical results. Since VBLCC and traditional LCC yield similar classification results, the considered model specifications included (*a*) a traditional OL model, (*b*) a *K*-means–based segmentation with OL, and (*c*) a VBLCC-based segmentation with OL.

Estimation Results

There were some rank-deficient problems related to multicollinearity during estimation of the *K*-means–based cluster model parameters, but dropping several parameters eventually yielded parameter estimates. Table 3 summarizes the model estimation results for various model specifications. A smaller AIC value in this case is preferred; AIC value is also related to data size. Each clustering approach appears to have its own advantage with respect to AIC.



(a) (b)

FIGURE 2 *K*-means clustering results in three-dimensional plots (PC = principal component).

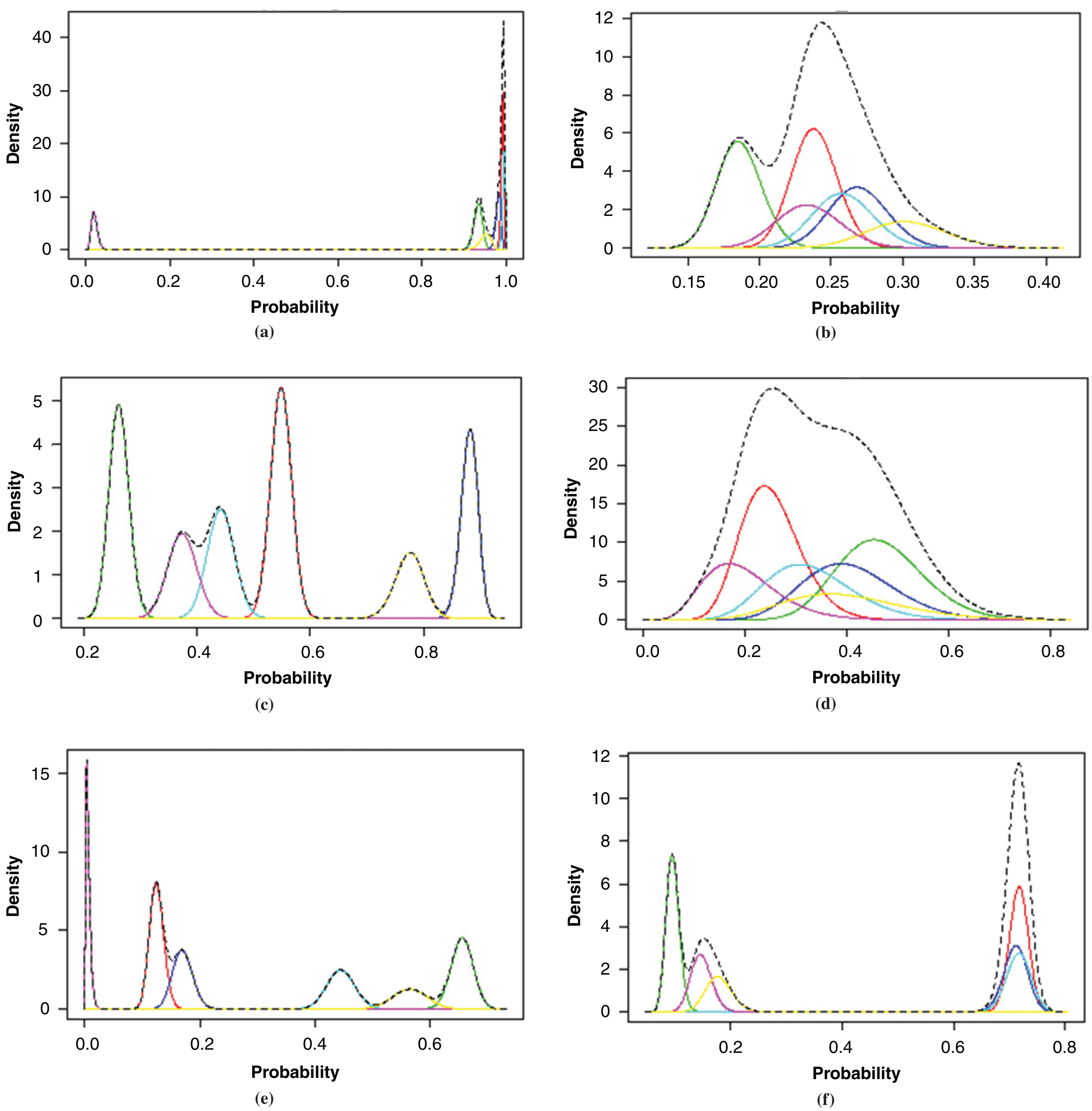


FIGURE 3 Posterior density plots for item probability parameters (LCC): (*a*) TypeXing, (*b*) landuseresidential, (*c*) auto, (*d*) roadsnow, (*e*) landuse_openspace, and (*f*) PaveMrkIDs.

TABLE 3 Model Estimation Results Using Different Approaches

	K-Means		VBLCC	
OL Model	AIC	Number of Observations	AIC	Number of Observations
Full data set	8,904.763	6,084	8,904.763	6,084
C1	1,562.294	1,031	2,107.712	1,556
C2	2,647.322	1,468	1,577.265	1,098
C3	394.676	306	1,494.443	789
C4	1,232.181	861	1,329.499	840
C5	2,596.939	1,937	1,362.941	936
C6	683.639	481	1,232.333	865

However, as with the model estimation process with the *K*-means segmentation, there were warnings related to rank-deficient issues. Again, considering the advantage of interpreting the subpopulation features, VBLCC proved a better approach than *K*-means.

Table 4 presents the OL model estimation results for the complete training data set ($n = 6{,}084$) as well as the OL model estimates for each VBLCC-based cluster (significant terms only based on a 10% significance level). The no-injury category was the base in all models, and therefore a positive coefficient denotes a higher probability of an injury or fatal injury.

Statistically significant independent variables in the OL model for the complete training data set included crash characteristics (hit by train, vehicle speed, train speed, and in vehicle, that is, a person inside instead of outside the vehicle). Other statistically significant variables were the number of rail cars in the train, highway features (urban area), crossing features (industrial land use), highway user characteristics (highway user gender or advanced age), safety infrastructure (flashing light), and environmental factors (dark and dusk visibility or roadway with presence of ice).

Individual OL models for the VBLCC-based clusters (C1, C2, . . . , C6) indicated more detailed findings compared with the OL estimates for the complete training data set model and identified additional independent variables that affected only certain clusters. For example, the variable "hit by train" was statistically significant in OL models estimated for C1, C2, C3, and C6 but not statistically significant in models for C4 and C5. An analyst relying on the results of the complete training data set would have missed the finding that "hit by train" is statistically relevant to C1, C2, C3, and C6 clusters.

Train speed and vehicle occupancy were two independent variable that were statistically significant in all the OL models. The estimated coefficient for train speed in C3 (0.005) was much smaller than estimated coefficients in the other cluster OLs as well as smaller than the estimated coefficient in the OL based on the complete training data set. The finding here is that train speed has more or less the same effect on injury severity across all HRGC train–vehicle crashes except those belonging to Cluster C3 (where the effect was smaller).

The estimated coefficients do not directly reflect the effect of independent variables on the three levels of injury severity. Therefore, Table 5 presents the calculated marginal effects.

Compared with the complete training data set model, some additional independent variables were identified that were statistically significant for certain clustered data subsets. For example, in the environment category, wet road conditions increased severity for crashes belonging to the C5 cluster, but snow on the road decreased injury severity for crashes in the C6 cluster.

An itemized comparison of the model results between the complete training data set and models based on the clustered data subsets is tedious and is not discussed here. However, Tables 4 and 5 illustrate the benefit of using a cluster-based approach to investigate HRGC train–vehicle crash injury severity. These benefits include estimation of more relevant model parameters by virtue of clustering the larger data set into subsets and identification of factors that may be relevant to certain clustered data subsets missed in the larger data set analysis.

Finally, a comparison of model predictions with the validation data set (20% of the 7,606 observations) provided validation and a means with which to compare prediction accuracy. Predictions from the models based on data for Clusters 1 through 5 yielded results that were close to the validation data set and were comparable to those obtained with the complete training data set model. Predictions using the Cluster 6 model, however, showed a 48.9% accuracy; this cluster mainly consisted of private crossings. Nonetheless, the results of this research indicate the credibility of the clustering regression approach, especially for public crossings.

SUMMARY AND DISCUSSION

The relationships between HRGC train–vehicle crash injury severity and a host of factors were investigated with HRGC crash data for 2011 to 2105, keeping in view the issue of unobserved heterogeneity. The training data set was partitioned into clusters through three techniques, with the objective of homogeneity within each cluster and heterogeneity among clusters. LCC and VBLCC yielded better results than the *K*-means method. However, there was little difference between LCC and VBLCC results. Crash injury severity was modeled with the OL model, and modeling results based on the complete training data set and the clustered subsets were compared. Results showed that clustering the data set into the clusters was useful for identifying factors contributing to injury severity. Factors that were consistently associated with HRGC train–vehicle crash injury severity in all the models included train speed and vehicle occupancy. Greater train speed and occupancy of the motor vehicle were associated with more severe crash injuries.

From a methodological point of view, the results of this research provide credence to a VBLCC- (or LCC-) based clustered data analysis approach for analysis of train–vehicle crash injury severity at HRGCs. Benefits of this approach include estimation of more relevant model parameters and identification of factors relevant only to certain clusters.

Greater train speed was associated with higher injury severity in train–vehicle crashes at HRGCs. However, the practical use of this finding to improve safety at HRGCs is somewhat questionable because of issues related to decreasing train speeds beyond the current speed limits. As such, the emphasis should be on ensuring that trains are not going faster than the set speed limits, which may yield some safety benefits.

Removal of observations with missing data from the matched data set to obtain the analysis data set is a limitation of this research. Although this approach ensured the formation of appropriate clusters, it is possible that modeling results will be different if somehow the analysis included both observations with missing data and observations with nonmissing data. Future work should investigate the effects of such a limitation on the results. As well, future research could focus on comparative analyses of alternative model structures (e.g., ordered probit versus multinomial logit).

TABLE 4 OL Model Parameter Estimates for Complete Training Data Set and Each VBLCC-Based Cluster

	Complete Training Data Set		C1		C2		C3		C4		C5		C6	
Variable	Coefficient	*P*-Value	Coefficient	*P*-Value	Coefficient	*P*-Value	Coefficient	*P*-Value	Coefficient	*P*-Value	Coefficient	*P*-Value	Coefficient	*P*-Value
Crash														
Hit by train	0.360	.000	0.438	.005	0.649	.000	0.861	.000					0.806	.022
Vehicle speed	0.027	.000	0.041	.000	0.034	.000	0.006	.006	0.019	.040	0.025	.000		
Train speed	0.037	.000	0.036	.000	0.038	.000	0.005	.000	0.046	.000	0.040	.000	0.038	.000
In vehicle	3.149	.000	2.905	.000	3.429	.000	0.042	.000	2.840	.000	3.021	.000	5.052	.000
No. of cars	0.003	.001			0.005	.013			0.005	.033			0.070	.042
No. of locomotives					−0.187	.017								
Crossing or highway														
Day through train							0.004	.069	0.008	.007				
Night through train							0.003	.044						
AADT													−0.000	.056
Land use industrial	−0.250	.571												
Urban area	−0.734	.093											−18.08	.014
Collector													2.165	.091
Flashing	0.022	.020	0.511	.007	0.441	.038					0.566	.030		
Stop sign					−0.612	.042			0.286	.055				
Crossbuck							1.018	.052						
Highway user														
Age 60 or older	0.509	.000	0.737	.000			0.473	.066	1.110	.000				
Age 50 to 60	0.217	.031												
Male	−0.332	.000					−0.491	.014						
Auto							−15.72	.066						
Motorcycle							−15.88	.057						
Truck							−16.04	.048					−0.754	.001
Other user					15.86	.024	−15.84	.065	1.734	.009	1.491	.002		
Environment														
Visibility dusk	0.023	.041									0.712	.018		
Visibility dark	0.197	.030									0.507	.012		
Lights					0.266	.096								
Road ice	−0.942	.000	−14.24	.020	−1.586	.022	−0.991	.045					−1.625	.031
Road wet											0.974	.023		
Road snow													−1.209	.075
Weather snow			−2.227	.082							14.03	.075		
Weather fog			−2.618	.083					−17.77	.002				
Weather cloudy											14.86	.099		

TABLE 5 Marginal Effects for Model Based on VBLCC Clusters

Variable	Full Data Set	C1	C2	C3	C4	C5	C6
No Injury							
Crash							
Hit by train	−7.6	−7.8	−12.9	−20.5	−3.2	3.4	−11.2
Vehicle speed	−0.6	−0.8	−0.7	−0.4	−0.5	−0.5	−0.2
Train speed	−0.8	−0.7	−0.8	−0.6	−1.1	−0.8	−0.6
In vehicle	−41.7	−33.4	−48.1	−48.8	−37.9	−37.1	−32.8
Number of cars	−0.1	0.0	−0.1	0.0	−0.1	0.0	−0.1
Number of locomotives	−0.1	−0.8	3.9	−1.7	1.3	−0.4	−1.2
Crossing or highway							
Day through train	0.0	0.0	0.1	0.2	−0.2	0.0	0.0
Night through train	0.0	0.0	0.0	−0.2	0.0	0.0	−0.1
AADT	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Land use industrial	5.3	8.9	5.4		−2.5	−4.5	2.0
Urban area	15.9		18.0		−0.4		25.0
Collector	−10.5	−10.4	−11.2	−31.6	37.1	−99.9	−48.9
Flashing	−4.8	−8.6	−9.0	−14.1	5.5	−9.8	−10.8
Stop sign	−2.8	1.7	11.7	1.9	−7.4	−7.6	−0.8
Crossbuck	−0.9	0.4	0.7	−25.3	5.3	0.8	−1.8
Highway user							
Age 60 or older	−11.8	−15.3	−3.2	−11.0	−26.2	−11.6	−2.1
Age 50 to 60	−5.0	−5.7	−7.4	−5.8	−9.5	−1.1	0.2
Male	7.5			12.2	6.5	−25.2	1.7
Auto	4.5	8.4	−67.2	99.5	−6.1	−4.9	
Motorcycle	3.9	6.0	−77.6	83.6		−10.2	4.6
Truck	14.0	16.7	−81.8	99.9	7.0		13.3
Other user	−24.6	−32.4	−72.1	57.4	−40.1	−34.5	−1.4
Environment							
Visibility dusk	−5.1	−0.6	2.6	−8.7	−8.4	−15.2	−9.1
Visibility dark	−4.3	−2.9	−1.0	−3.7	−6.2	−9.9	−0.1
Lights	−1.5	−3.4	−5.9	−3.5	3.2	2.2	6.2
Road ice	17.6	27.0	22.9	24.4	−5.1	12.2	16.8
Road wet	−0.1	3.5	−3.6	16.4	−5.3	−12.2	3.5
Road snow	4.9	−2.0	10.0	−1.0	−2.9	11.6	14.5
Weather snow	11.1	22.5	−81.0	21.8	25.5	−84.4	−82.6
Weather fog	16.7	23.3	−71.2	20.2	39.2	−80.2	−80.3
Weather cloudy	10.6	24.9	−97.5	19.0	25.6	−98.1	−99.1
Injury							
Crash							
Hit by train	6.2	6.7	11.2	14.3	2.7	−2.7	9.5
Vehicle speed	0.5	0.7	0.6	0.3	0.4	0.4	0.1
Train speed	0.7	0.6	0.7	0.4	0.9	0.6	0.5
In vehicle	34.5	29.0	41.5	37.2	33.2	29.9	27.9
Number of cars	0.0	0.0	0.1	0.0	0.1	0.0	0.1
Number of locomotives	0.1	0.7	−3.3	1.1	−1.1	0.3	1.0
Crossing or highway							
Day through train	0.0	0.0	−0.1	−0.1	0.1	0.0	0.0
Night through train	0.0	0.0	0.0	0.1	0.0	0.0	0.1
AADT	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Land use industrial	−4.3	−7.6	−4.7		2.1	3.6	−1.6
Urban area	−12.7		−15.2		0.3		−21.7
Collector	8.2	8.8	9.5	11.0	−32.9	0.6	31.8
Flashing	3.8	7.4	7.8	7.7	−4.6	7.9	8.9
Stop sign	2.2	−1.4	−10.3	−1.2	6.2	5.9	0.6
Crossbuck	0.7	−0.4	−0.6	18.5	−4.3	−0.6	1.5
Highway user							
Age 60 or older	9.1	12.7	2.8	6.4	20.2	9.1	1.8
Age 50 to 60	3.9	4.8	6.3	3.5	7.8	0.9	−0.1
Male	−5.9			−7.1	−5.3	21.3	−1.4
Auto	−3.6	−7.2	55.0	−3.8	5.0	3.9	
Motorcycle	−3.1	−5.1	−20.5	−47.0		7.9	−3.9
Truck	−11.2	−14.3	−16.7	−0.1	−5.8		−11.1
Other user	17.5	25.0	−25.4	−44.2	25.1	23.5	1.1

(continued on next page)

TABLE 5 *(continued)* **Marginal Effects for Model Based on VBLCC Clusters**

Variable	Full Data Set	C1	C2	C3	C4	C5	C6
Environment							
Visibility dusk	4.0	0.5	−2.3	5.1	6.8	11.7	7.5
Visibility dark	3.4	2.5	0.9	2.3	5.1	7.8	0.0
Lights	1.2	2.9	5.1	2.1	−2.6	−1.8	−5.2
Road ice	−14.6	−23.9	−20.4	−17.9	4.2	−10.0	−14.5
Road wet	0.1	−3.0	3.1	−11.5	4.4	9.5	−3.0
Road snow	−3.9	1.7	−8.8	0.7	2.4	−9.5	−12.4
Weather snow	−9.1	−19.9	−17.5	−15.8	−22.2	−13.4	−15.2
Weather fog	−13.9	−20.6	−26.1	−14.6	−34.6	−16.9	−17.2
Weather cloudy	−8.6	−21.6	−2.3	−13.0	−21.9	−1.6	−0.8
Fatal Injury							
Crash							
Hit by train	1.5	1.1	1.6	6.1	0.5	−0.7	1.6
Vehicle speed	0.1	0.1	0.1	0.2	0.1	0.1	0.0
Train speed	0.2	0.1	0.1	0.2	0.2	0.2	0.1
In vehicle	7.3	4.5	6.5	11.6	4.7	7.2	4.9
Number of cars	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Number of locomotives	0.0	0.1	−0.5	0.6	−0.2	0.1	0.2
Crossing or highway							
Day through train	0.0	0.0	0.0	−0.1	0.0	0.0	0.0
Night through train	0.0	0.0	0.0	0.1	0.0	0.0	0.0
AADT	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Land use industrial	−1.0	−1.2	−0.7		0.4	0.9	−0.3
Urban area	−3.2		−2.8		0.1		−3.4
Collector	2.3	1.6	1.7	20.6	−4.2	99.3	17.1
Flashing	1.0	1.2	1.2	6.4	−0.9	1.8	2.0
Stop sign	0.6	−0.2	−1.4	−0.7	1.3	1.6	0.1
Crossbuck	0.2	−0.1	−0.1	6.8	−1.0	−0.2	0.3
Highway user							
Age 60 or older	2.7	2.6	0.5	4.6	6.0	2.6	0.3
Age 50 to 60	1.1	0.9	1.1	2.2	1.8	0.2	0.0
Male	−1.6			−5.1	−1.2	3.9	−0.3
Auto	−0.9	−1.3	12.2	−95.7	1.0	1.0	
Motorcycle	−0.8	−0.8	98.1	−36.6		2.3	−0.7
Truck	−2.8	−2.4	98.5	−99.8	−1.2		−2.3
Other user	7.1	7.4	97.4	−13.2	15.0	11.0	0.2
Environment							
Visibility dusk	1.1	0.1	−0.4	3.6	1.6	3.6	1.6
Visibility dark	0.9	0.4	0.1	1.4	1.1	2.1	0.0
Lights	0.3	0.5	0.8	1.3	−0.5	−0.4	−0.9
Road ice	−2.9	−3.1	−2.4	−6.5	0.9	−2.1	−2.3
Road wet	0.0	−0.5	0.5	−4.8	1.0	2.7	−0.6
Road snow	−0.9	0.3	−1.2	0.4	0.5	−2.1	−2.0
Weather snow	−2.0	−2.7	98.4	−6.0	−3.2	97.8	97.8
Weather fog	−2.8	−2.7	97.3	−5.6	−4.5	97.1	97.5
Weather cloudy	−2.0	−3.3	99.8	−5.9	−3.7	99.8	99.9

REFERENCES

1. *Highway–Rail Grade Crossing Safety.* Association of American Railroads, Washington, D.C., 2015.
2. Operation Lifesaver. Crossing Collisions and Casualties by Year. https://oli.org/about-us/news/collisions-casualties. Accessed June 20, 2016.
3. Oh, J., S. P. Washington, and D. Nam. Accident Prediction Model for Railway–Highway Interfaces. *Accident Analysis and Prevention,* Vol. 38, No. 2, 2006, pp. 346–356. https://doi.org/10.1016/j.aap.2005.10.004.
4. Saccomanno, F., L. Fu, and L. Miranda-Moreno. Risk-Based Model for Identifying Highway–Rail Grade Crossing Blackspots. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1862, 2004, pp. 127–135. https://dx.doi.org/10.3141/1862-15.
5. Austin, R. D., and J. L. Carson. An Alternative Accident Prediction Model for Highway–Rail Interfaces. *Accident Analysis and Prevention,* Vol. 34, No. 1, 2002, pp. 31–42. https://doi.org/10.1016/S0001-4575(00)00100-7.
6. Yasmin, S., N. Eluru, C. R. Bhat, and R. Tay. A Latent Segmentation Based Generalized Ordered Logit Model to Examine Factors Influencing Driver Injury Severity. *Analytic Methods in Accident Research,* Vol. 1, 2014, pp. 23–38. https://doi.org/10.1016/j.amar.2013.10.002.
7. Felzenszwalb, P., D. McAllester, and D. Ramanan. A Discriminatively Trained, Multiscale, Deformable Part Model. Presented at Conference on Computer Vision and Pattern Recognition, IEEE, 2008.
8. Skrondal, A., and S. Rabe-Hesketh. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* CRC Press, Boca Raton, Fla., 2004. https://doi.org/10.1201/9780203489437.
9. Raub, R. A. Examination of Highway–Rail Grade Crossing Collisions over 10 Years in Seven Midwestern States. *ITE Journal,* Vol. 76, No. 4, 2006, p. 16.
10. Hu, S., C. Li, and C. Lee. Investigation of Key Factors for Accident Severity at Railroad Grade Crossings by Using a Logit Model. *Safety Science,* Vol. 48, No. 2, 2010, pp. 186–194. https://doi.org/10.1016/j.ssci.2009.07.010.
11. Eluru, N., M. Bagheri, L. F. Miranda-Moreno, and L. Fu. A Latent Class Modeling Approach for Identifying Vehicle Driver Injury Severity Factors at Highway–Railway Crossings. *Accident Analysis and Prevention,* Vol. 47, 2012, pp. 119–127. https://doi.org/10.1016/j.aap.2012.01.027.

12. Hao, W., and J. Daniel. Motor Vehicle Driver Injury Severity Study Under Various Traffic Control at Highway–Rail Grade Crossings in the United States. *Journal of Safety Research,* Vol. 51, 2014, pp. 41–48. https://doi.org/10.1016/j.jsr.2014.08.002.
13. Russo, B. Examination of Factors Affecting Frequency and Severity of Crashes at Rail–Grade Crossings. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
14. Zhao, S., A. Iranitalab, and A. Khattak. Investigation of Pedestrian Injury Severity in Crashes at Highway–Rail Grade Crossings Using Latent Class Analysis. Presented at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.
15. Zhao, S., and A. Khattak. Motor Vehicle Drivers' Injuries in Train–Motor Vehicle Crashes. *Accident Analysis and Prevention,* Vol. 74, 2015, pp. 162–168. https://doi.org/10.1016/j.aap.2014.10.022.
16. Fan, W., M.R. Kane, and E. Haile. Analyzing Severity of Vehicle Crashes at Highway–Rail Grade Crossings: Multinomial Logit Modeling. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
17. Mannering, F.L., and C.R. Bhat. Analytic Methods in Accident Research: Methodological Frontier and Future Directions. *Analytic Methods in Accident Research,* Vol. 1, 2014, pp. 1–22. https://doi.org/10.1016/j.amar.2013.09.001.
18. Fan, W., L. Gong, E.M. Washing, M. Yu, and E. Haile. Key Factors Contributing to Crash Severity at Highway–Rail Grade Crossings. *Journal of Modern Transportation,* 2016, https://doi.org/10.1007/s40534-016-0110-x, pp. 1–12.
19. Eluru, N., C.R. Bhat, and D.A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention,* Vol. 40, No. 3, 2008, pp. 1033–1054. https://doi.org/10.1016/j.aap.2007.11.010.
20. Kockelman, K.M., and Y. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis and Prevention,* Vol. 34, No. 3, 2002, pp. 313–321. https://doi.org/10.1016/S0001-4575(01)00028-8.
21. Ulfarsson, G.F., and F.L. Mannering. Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents. *Accident Analysis and Prevention,* Vol. 36, No. 2, 2004, pp. 135–147. https://doi.org/10.1016/S0001-4575(02)00135-5.
22. Abdelwahab, H., and M. Abdel-Aty. Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1784, 2002, pp. 115–125. https://dx.doi.org/10.3141/1784-15.
23. Chang, L., and H. Wang. Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques. *Accident Analysis and Prevention,* Vol. 38, No. 5, 2006, pp. 1019–1027. https://doi.org/10.1016/j.aap.2006.04.009.
24. Kim, K., and E.Y. Yamashita. Using a *K*-Means Clustering Algorithm to Examine Patterns of Pedestrian Involved Crashes in Honolulu, Hawaii. *Journal of Advanced Transportation,* Vol. 41, No. 1, 2007, pp. 69–89. https://doi.org/10.1002/atr.5670410106.
25. Prato, C.G., S. Bekhor, A. Galtzur, D. Mahalel, and J.N. Prashker. Exploring the Potential of Data Mining Techniques for the Analysis of Accident Patterns. In *Proceedings of the 12th World Conference on Transport Research,* Lisbon, Portugal, 2010.
26. Depaire, B., G. Wets, and K. Vanhoof. Traffic Accident Segmentation by Means of Latent Class Clustering. *Accident Analysis and Prevention,* Vol. 40, No. 4, 2008, pp. 1257–1266. https://doi.org/10.1016/j.aap.2008.01.007.
27. de Oña, J., G. López, R. Mujalli, and F.J. Calvo. Analysis of Traffic Accidents on Rural Highways Using Latent Class Clustering and Bayesian Networks. *Accident Analysis and Prevention,* Vol. 51, 2013, pp. 1–10. https://doi.org/10.1016/j.aap.2012.10.016.
28. Hartigan, J.A., and M.A. Wong. Algorithm AS 136: A *K*-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C, Applied Statistics,* Vol. 28, No. 1, 1979, pp. 100–108.
29. Vermunt, J.K., and J. Magidson. Latent Class Cluster Analysis. *Applied Latent Class Analysis,* Vol. 11, Cambridge University Press, Cambridge, United Kingdom, 2001, pp. 89-106.
30. White, A., and T.B. Murphy. BayesLCA: An R Package for Bayesian Latent Class Analysis. *Journal of Statistical Software,* Vol. 61, No. 1, 2014, pp. 1–28.
31. Raftery, A.E., M.A. Newton, J.M. Satagopan, and P.N. Krivitsky. *Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity.* In Bayesian Statistics, Vol. 8, Oxford University Press, Oxford, United Kingdom, 2007, pp. 1–45.
32. Mohamed, M.G., N. Saunier, L.F. Miranda-Moreno, and S.V. Ukkusuri. A Clustering Regression Approach: A Comprehensive Injury Severity Analysis of Pedestrian–Vehicle Crashes in New York, U.S., and Montreal, Canada. *Safety Science,* Vol. 54, 2013, pp. 27–37. https://doi.org/10.1016/j.ssci.2012.11.001.

The Standing Committee on Highway–Rail Grade Crossings peer-reviewed this paper.