Check for updates

Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
2022, Vol. 2677(7) 1–21
© National Academy of Sciences:
Transportation Research Board 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/03611981221150927
journals.sagepub.com/home/trr
SAGE

A Clustering Regression Approach to Explore the Heterogeneous Effects of Risk Factors Associated with Teen Driver Crash Severity

Md Mahmud Hossain[1], Huaguo Zhou[1], and Xiaoduan Sun[2]

Abstract

Teen drivers remain one of the most long-standing traffic safety concerns, as they continue to be overrepresented in the fatal and injury crash statistics. No previous studies have explored significant variations in the effect of risk attributes contributing to teen driver collisions in multiple crash circumstances by degrees of crash severity. Therefore, substantial exploration of teen driver crashes is expected to facilitate the strategic employment of countermeasures effectively. This study aimed to analyze teen driver crashes to investigate the heterogeneous effect of contributing factors on crash severity outcomes. Three years (2017 to 2019) of police-investigated crash information was used for the state of Alabama. This research first applied latent class clustering to minimize the heterogeneity in the extracted dataset by dividing the data into meaningful clusters (subgroups of the whole data). Then, multinomial logit models were constructed to illustrate the significant risk factors influencing the severity outcomes in different crash scenarios. Marginal effects were computed to understand the impacts of variable categories better. The findings suggested that the significance and estimated impact of variables varied within clusters and between crash severity levels in the same cluster. The results of latent class segmented submodels represented real-world crash patterns demonstrating the cumulative effect of variable attributes. Such contextual understandings of underlying risk factors could help to strengthen existing teen driver educational interventions. In addition, the study outcomes could assist practitioners and policy makers in developing safety improvement strategies to reduce the causalities associated with teen driver crashes in distinct circumstances.

Keywords

teen driver, crash severity, logistic regression, crash data analysis, latent class clustering

In the United States, all 50 states and the District of Columbia have enacted some form of graduated licensing system (GDL) for novice teenagers since 2006 to mitigate the impact of driving inexperience and inadequate skills. A typical GDL program starts with supervised driving in the learner phase and then permits restricted unsupervised driving until acquiring full driving privileges. The primary purpose of this program is to train and educate teenagers before initiating independent driving in hazardous circumstances (e.g., nighttime driving, driving with passengers). However, teen motorists remain one of the most long-standing traffic safety concerns. According to the Centers for Disease Control and Prevention, the risk of collision is still substantially higher for teenagers aged 16 to 19 years compared with any other age group (*1*). In addition, this teen driver cohort represented nearly two times more fatal crashes per vehicle miles traveled (VMT) than drivers aged 20 or older in 2019 (*2*). Teen drivers usually have a higher propensity for risky driving behaviors that can result from cognitive processes, including overestimating and underestimating driving

[1]Department of Civil and Environmental Engineering, Auburn University, Auburn, AL
[2]Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA

Corresponding Author:
Md Mahmud Hossain, mahmud@auburn.edu

skills, poor hazard perception, deficits in attention, and impulsivity (*3*).

Researchers heavily rely on police-reported crash data to explain why, where, and how teen driver crashes occurred. However, owing to inherent heterogeneity in the dataset, the contribution of risk factors may vary with crash circumstances (*4*). This study aimed to analyze teen driver crashes to explore the heterogeneous effect of contributing factors on crash severity outcomes. A comprehensive dataset has been developed using 3 years (2017 to 2019) of relevant crash information in Alabama. Alabama enacted the GDL program on October 1, 2002, with an initial aim to reduce the number of traffic deaths and injuries associated with teen drivers. In addition to GDL, the jurisdiction offers several driver education programs for teenagers, sponsored by the Children's Hospital of Alabama, VOICES for Alabama Children, SafeKids of Alabama, Children's First, and Think First (*5*). Despite these efforts, Alabama continues to show high fatal crashes of teen drivers per 100,000 VMT. From 2015 to 2019, the annual fatal crash rates were consistently more than 40% higher than the national average. Therefore, substantial exploration of teen driver crashes is expected to facilitate the strategic employment of countermeasures effectively. This research first applied latent class clustering (LCC) to divide the whole crash data into homogeneous groups. Then multinomial logit (MNL) models were utilized to identify the statistically significant variables for the entire crash data along with the subgroups by crash severity levels.

Literature Review

Contributing Factors of Teen Driver Crashes

Age and driving experience contribute differently to adolescent crash risk. For example, teen drivers aged 18 to 19 are more exposed to risky driving dispositions than those aged 16 to 17 (*6*, *7*). Conversely, newly licensed teenagers exhibit the highest crash rates in the first month after licensure, sharply declining through the next half-year, and steadily reducing as they spend more hours behind the wheel (*8*). Male teenagers have higher risk-taking tendencies (*9*, *10*), which could explain their twofold higher fatal crash rate than their counterparts. Female teenagers are less likely to be involved in injury collisions (*11*), but the related crashes are more likely to be rear-end (*12*, *13*). Because of the minimum legal drinking age (MLDA) and zero-tolerance laws, teenagers are less exposed to intoxicated driving than adults (*14*). Teen drivers with a blood alcohol concentration (BAC) of 0 mg/dL have a similar crash likelihood in respect of young adults with a BAC of 0.05 to 0.079 mg/dL (*15*). Therefore, adolescents consistently display a high risk of fatal and severe collisions at any BAC level (*16*). Although researchers have emphasized the severe outcomes from not wearing seatbelts while driving, a significant proportion teens still drive without being buckled up (*17*). The actual number is expected to be higher as unrestrained driving is less likely to be reported in minor injury crashes (*18*). Williams and Shabanova aimed to investigate the frequency of seatbelt usage among teenagers in multiple circumstances, and lower percentages were observed in nighttime driving, under intoxication, and with peer passengers (*19*). Adolescents often overestimate their vehicle control skills, which increases their participation in numerous secondary tasks during driving (*20*). Electronic devices such as cellphones are highly associated with any combination of visual, manual, and cognitive distractions. Repeatedly talking/manipulating cellphones can cause/amplify other risky driving maneuvers, for example, speeding and disregarding traffic signs and signals (*21*). In relation to carrying passengers, several teen driver studies have documented a positive correlation between the number of passengers in the vehicle before the crash incident and corresponding crash severity levels (*22*, *23*). However, further investigation is needed to come to conclusions about the interactions between collisions involving teenage drivers and the associated occupant characteristics (e.g., passenger age and gender). In relation to driver violation, failure to yield, careless operation, and following vehicles too closely are strongly linked with teen driver crashes (*24*). Several studies have reported teen drivers' gradual transition through the driving offenses with experience of more hours in the vehicle, for example, a high likelihood of speeding and intoxicated driving as progress toward independent driving expertise (*25*, *26*). The significance of vehicle choice in teen driving safety has also been reported in multiple studies, for example, aggressive driving habits while operating SUVs and pickup trucks that induce fatal consequences (*27*, *28*).

For teenagers, nighttime driving is often associated with distraction, alcohol consumption, and traffic offenses (*29*, *30*). In relation to total crash counts, teen driver crashes are prevalent during the peak hours, including 6 to 9 a.m. and 3 to 6 p.m. (*28*). The chance of fatal crashes is prominent on weekends, especially during Friday and Saturday midnight hours (*10*). One possible explanation could be more frequent involvement in high-risk conditions (e.g., impairment, unrestrained driving) on weekends, which increases the associated crash risk disproportionately more than that of adults (*31*). On freeways, the probability of teen drivers being associated with fatal and injury crashes gradually evolves in the order of road lighting: daylight, dark with streetlights, and dark without streetlights (*32*). However, Dissanayake and Amarasingha reported teenagers' lower risk of incidents during dark conditions than adult drivers aged 25 to 64 (*33*). One potential reason could be their lower exposure to nighttime driving because of the restrictions in the GDL program. In adverse weather conditions, teen drivers' proneness for being involved in injury collisions is strongly associated with failure to implement vehicle adjustment and inefficiency in braking (*34*, *35*), which may result in skid, drag, or run off road incidents (*28*).

In relation to area setting, multiple studies inferred teen drivers' elevated severe crash likelihood in rural areas (*36–38*). The estimated crash rates of teen drivers aged below 16 were higher on rural roads than teenagers aged 16 to 18 (*18*). Rural roads usually have unique hazards such as low visibility, higher posted speed limit, gravel road surface, and unsignalized intersections (*18*, *38*). Coupled with limited driving experience and a lack of safe driving habits, these driving challenges impose a greater safety threat (*37*). Duddu et al. found that teen drivers are less likely to be injured on interstates than on other highway types (*28*). More crashes while negotiating a curve than going straight are attributed to teenagers (*12*, *39*). Furthermore, researchers have outlined the misjudgments of adolescents at intersections resulting from limited driving skills in complex traffic environments (*40*). In fully access-controlled roads, sudden speed variations at control locations increase teen drivers' injury risk (*28*). With regard to fatal and severe injury crashes, teenagers are often found to be involved in intoxicated driving and driving without seatbelts on roadways with posted speed limits of $\geqslant 50$ mph (*41*). For teenagers, single-vehicle crashes are more likely to produce fatal and severe injuries (*41*). The majority of single-vehicle crashes involve fixed objects (*42*) and occur on roadways with high posted speed limits (*43*). According to McDonald et al., the most frequent crash scenarios among teen drivers are 1) going straight, the front vehicle stopped, rear-end, 2) approaching for left turns at intersections, and 3) negotiating curves/going straight, off the right edge of the road, and right roadside departure (*12*). These prevalent collision types are highly associated with hazardous driving behaviors such as alcohol intoxication, speeding, cellphone distraction, and traffic control violation (*20*, *23*, *44*). Rear-end and broadside collisions are prevalent on urban roads, whereas single-vehicle and rear-end crashes are more frequent in rural areas (*18*).

Application of LCC and Study Objectives

Researchers have conducted several studies to evaluate the impact of explanatory factors on teen driver crash severity using discrete choice models (*18*, *28*, *32*, *42*). However, some latent conditions might remain in the crash datasets, which result from multiple impact factors that cannot be directly observed or were not reported during the investigation (*45*). Because of this heterogeneity, the effect of a crash risk factor may vary under different conditions; for example, intoxicated driving may be influential for crashes on specific roadway segments, not for all roadway types. Such a condition might be combined with multiple crash attributes, for example, rural intersections in "dark-lighted" conditions, male drivers on roadways with high posted speed limits, and so on. Therefore, neglecting the significance of potential unobserved heterogeneity in crash data analysis could generate inaccurate model outcomes and conclusions (*46*). One way to overcome this heterogeneity issue is to classify the crash data into different homogeneous subgroups and apply conventional parametric models in each cluster separately to determine the significant factors for specific crash surroundings. However, the data segmentation method based on research needs and exogenous variables (e.g., driver age groups, collision type, crash location, posted speed limit, and so on) may not ensure homogeneous clusters (*47*). In recent years, LCC has been utilized in a substantial body of road-safety literature (*45*, *48*) to preprocess the crash data distribution to maximize heterogeneities between the clusters and homogeneity of factors within the subgroup.

Following the literature, no previous studies have explored significant variations in the effect of risk attributes contributing to teen driver collisions in multiple crash circumstances by degrees of crash severity. The objectives of this research were (1) to examine the application of LCC in minimizing the heterogeneity of the crash dataset involving teen drivers in Alabama and (2) to determine the factors that significantly influence the severity outcomes of identified crash scenarios using MNL models. In traffic safety research, MNL is the most widely utilized discrete choice modeling technique for severity analysis (*40*, *49*, *50*). According to Iranitalab and Khattak, LCC can improve the severity prediction performance of MNL model (*51*). Our crash severity analysis primarily follows the approaches of Depaire et al. (*47*) and Sun et al. (*48*). Depaire et al. used LCC to segment Belgium crash data into seven different classes and applied MNL models to the entire dataset and each identified cluster (*47*). Sun et al. adopted a similar integrated approach to identify significant risk factors in pedestrian severity outcomes (*48*). The present study evaluated and compared the effects of different crash contributing factors for the whole teen driver crash dataset as well as individual latent clusters. In addition, the marginal effects were estimated to illustrate the model results on the probability scale. The research outcomes could guide policy makers and safety practitioners to develop effective crash prevention strategies to improve teen driving safety.

Methodology

Latent Class Clustering

Clustering can handle the heterogeneity issue of datasets by splitting the data into multiple homogeneous subsets or clusters (*47*, *48*). This unsupervised machine learning approach maximizes the homogeneity between in-cluster features and dissimilarity among intercluster elements. LCC is a probability-based clustering approach that has exhibited several benefits over conventional clustering approaches such as *k*-means clustering, hierarchical clustering (*52*). A few key advantages are (1) it is not mandated to specify the number of clusters beforehand, and (2) it is appropriate for different variable classes (e.g., continuous, nominal, categorical) without any standardization process (*53*). LCC determines categorical latent variables by assessing class membership likelihoods and conditional probabilities that variables belong to specific class memberships (*54*). During estimation, all observed variables are assumed to be statistically independent, conditional on the values of latent variables, which is regarded as "local independence" (*55*).

Consider a dataset of T crashes with a latent variable, L, and a set of observed variables, $X_1 \ldots\ldots X_i$. These variables create a latent class model with U classes. Assume that each observed value includes a certain number of attributes, X_j, with C_j levels where $j = 1, 2, \ldots, i$. If the latent class is specified by the index, u, where $u = 1, 2, \ldots, U$, the basic form of latent class cluster can be expressed as

$$P_{X_j} = \sum_{u=1}^{U} P_{L_u} P_{X_j/L_u} \tag{1}$$

where

P is the likelihood of attaining response variable X_j,

P_{L_u} is the probability that a particular crash observation belongs to cluster u, and

P_{X_j/L_u} is the conditional likelihood that a particular crash has a response pattern X_j considering its membership in the u class of latent variable L.

As the observed variables are mutually exclusive with each latent class (*56*), Equation 1 can be presented as

$$P_{X_j} = \sum_{u=1}^{U} P_{L_u} \prod_{j=1}^{i} P_{X_j/L_u} \tag{2}$$

The model parameters are estimated using the maximum likelihood method (*56*). After model assessment, the Bayes rule can describe the likelihood of belonging to a latent class, also referred as the posterior membership probability.

$$P_{L_u/X_j} = \frac{P_{L_u} P_{X_j/L_u}}{P_{X_j}} \tag{3}$$

The Akaike information criterion (AIC), Bayesian information criterion (BIC), Consistent Akaike information criterion (CAIC), and entropy-based measures (EMs) have been widely used to determine the optimum number of clusters (*57*). AIC, BIC, and CAIC are penalized-likelihood information criteria that indicate the error between the calculated and actual likelihood function (*58*); therefore, lower values denote the better model. On the contrary, EM is a classification-based approach that exhibits data quality by specifying the uncertainty involved in the outcomes (*59*). The value ranges from 0 to 1, and a value close to 1 indicates a better clustering. Few researchers indicate that BIC is more reliable than AIC and CAIC in selecting the optimal number of clusters (*58*). In practice, the minimum value of BIC cannot be obtained when analyzing a large data sample. Therefore, some studies have suggested using the percentage reduction of BIC to determine the appropriate number of clusters (*48*, *60*). In this regard, a strategy that satisfies all four criteria could be a better approach (*54*).

$$AIC = -2\log\left(\frac{SSE}{t}\right) + 2w \tag{4}$$

$$BIC = -2\log\left(\frac{SSE}{t}\right) + \log(t)w \tag{5}$$

$$CAIC = -2\log\left(\frac{SSE}{t}\right) + [\log(t) + 1]w \tag{6}$$

$$EM = 1 - \frac{\sum_{j=1}^{t}\sum_{u=1}^{U} p_{ju}\ln\left(p_{ju}\right)}{t ln\left(\frac{1}{U}\right)} \tag{7}$$

where

SSE = sum of squares error,

w = number of parameters to be estimated,

t = number of observations, and

p_{ju} = posterior probability that a crash observation, j, belongs to latent class u.

Multinomial Logit Model

An MNL model indicates the likelihood of distinct outcomes based on the same observed covariates. This discrete choice model estimates a series of binary logit models considering one attribute of dependent variables as a reference category. In this study, the dependent variables were crash severity levels. Previously, several traffic safety studies have adopted an MNL approach to determine whether each category of crash contributing factor is statistically significant for different severity outcomes (*49*, *50*, *61*).

A propensity function, S_{ik}, is defined to quantify the propensity of crash i toward severity level k (*24*).

$$S_{ki} = \beta_k X_{ki} + \varepsilon_{ki} \tag{8}$$

where

β_k represents a vector of estimated parameters for crash severity category k,

X_{ki} denotes a vector of explanatory variables influencing the crash severity for i at severity level k, and

ε_{ki} is the random error term for unobservable impacts.

$$P_{ik} = \frac{\exp(\beta_k X_{ki})}{\sum_{\forall k} \exp(\beta_k X_{ki})} \tag{9}$$

where P_{ik} is the likelihood of crash i experiencing crash severity category k. The log likelihood function is as follows (*62*):

$$LL(\theta) = \sum_{n=1}^{N} \sum_{k} d_{ik} ln\left(\frac{\exp(\beta_k X_{ki})}{\sum_{\forall k} \exp(\beta_k X_{ki})}\right) \tag{10}$$

where N is the number of crash observations, and d_{ik} is the indicator variable (1 if individual crash i belongs to crash severity level k, and 0 otherwise).

Marginal effects are estimated to determine the effect of a one-unit increase in an explanatory variable (predictor) on severity-outcome probabilities (*63*). Compared with odds ratios, this method of parameter estimation is simpler to interpret and less affected by the extreme frequency of a particular category belonging to a covariate (*64*). The marginal effect of the t^{th} predictor variable is calculated as (*65*)

$$ME_{X_{ikt}}^{P_{ik}} = P_{ik}(X_{ikt} = 1) - P_{ik}(X_{ikt} = 0) \tag{11}$$

The presence and absence of the t^{th} predictor are denoted by 1 and 0, respectively, while all other explanatory variables remain constant. The marginal effects average is calculated for each variable attribute across all crash observations.

Data

Data Preprocessing

In this study, 3 years (2017 to 2019) of police-investigated teen driver crash data were extracted from the critical analysis reporting environment (CARE) system for the entire state of Alabama. It is noteworthy that CARE has been developed by the Center for Advanced Public Safety at the University of Alabama, and the database serves as a primary source of historical crash information for research in Alabama (*11*). Each crash record includes numerous risk factors covering driver, road, environment, vehicle, and crash characteristics. The potential contributing traits were selected based on previous studies and the availability of variables in the original database. In Alabama, crash severity has been recorded in the KABCO injury classification scale from 2009 (K: fatal, A: incapacitating injury, B: nonincapacitating injury, C: complaint or possible injury, and O: no injury). In the sorted teen driver crash data, the distribution of 57,098 unique crashes by maximum severity outcomes was K: 202 (0.35%), A: 1,679 (2.94%), B: 4,891 (8.56%), C: 5,766 (10.10%), and O: 44,560 (78.05%). In addition, the distribution of crash involvement by driver age was 15: 775 (1.36%), 16: 12,336 (21.60%), 17: 13,181 (23.08%), 18: 15,372 (26.92%), and 19: 15,434 (27.04%).

The categorization of selected variables was performed based on earlier related studies and engineering judgment. According to Alabama's GDL structure, the learner permit and probationary driving stage start from age 15 and 16, respectively. Therefore, "driver age" was classified into two categories (15 to 16 years and 17 to 19 years) to underline the teenagers in the restricted stages of driving. "Day of the week" was categorized following the National Highway Traffic Safety Administration classification (weekday: Monday 6:00 a.m. to Friday 5:59 p.m. and weekend: Friday 6:00 p.m. to Monday 5:59 a.m.) (*66*). In this study, "crash severity" was the dependent variable for the MNL model, and the attributes of this variable were regrouped as severe injury (fatal or incapacitating injury), minor injury (nonincapacitating or possible injury), and no injury, in parallel with previous studies (*67*, *68*). The final dataset contained 57,098 crashes, with 19 variables consisting of 62 categories.

Data Description

Table 1 shows the descriptive statistics of the explanatory variables by crash severity levels along with total crashes. Teenagers aged 15 to 16 years, either learners or intermediate license holders, were involved in 22.96% of crashes. Supervised driving restrictions limit novice teen drivers' exposure to critical circumstances in their early stages of driving (*69*). In the final dataset, male teen drivers had more frequent crashes than females (54.34% versus 45.66%). In addition, they were involved in 59.54% of severe injury crashes, slightly higher than the proportion of total crashes (54.34%). Earlier studies reported male teenagers' tendency toward traffic law violations and risky driving dispositions (*10*). Impairment and no seatbelt usage were found to be influential in severe crashes, reflecting the findings of previous studies (*24*, *41*). Compared with overall crashes, carrying multiple passengers seemed render teenagers vulnerable to incidents, owing to their overrepresentation in severe injury (12.23% versus 8.34%) and minor injury crashes (10.75% versus 8.34%). Reviewing the findings of 15 articles, Ouimet et al. concluded that the risk of young driver crashes comprising a high degree of severity

Table 1. Overview of Final Dataset by Crash Severity Levels

Variable	Category	Total crashes (57,098) (%)	Crashes by severity: Severe injury (1,881) (%)	Minor injury (10,657) (%)	No injury (44,560) (%)
Age	15–16 years	22.96	23.29	22.42	23.08
	17–19 years	77.04	76.71	77.58	76.92
Gender	Female	45.66	40.46	47.79	45.37
	Male	54.34	59.54	52.21	54.63
Impairment (alcohol/drug)	No	98.73	94.21	98.34	99.01
	Yes	1.27	5.79	1.66	0.99
No restraint usage	No	96.34	76.61	93.23	97.92
	Yes	3.66	23.39	6.77	2.08
Distraction	No	53.49	53.69	53.34	53.52
	Yes	14.13	13.18	14.16	14.16
	Unknown	32.38	33.12	32.50	32.32
Number of passengers	No	62.51	61.56	57.63	63.72
	One	29.15	26.21	31.61	28.68
	Multiple	8.34	12.23	10.75	7.60
Vehicle type	Passenger car	59.68	55.13	60.46	59.69
	Van/SUV	22.96	21.00	21.55	23.38
	Other	17.35	23.87	17.98	16.93
Crash time	6–11:59 a.m.	23.38	25.78	22.95	23.56
	Midday–5:59 p.m.	49.43	40.83	46.51	50.49
	6–11:59 p.m.	22.91	21.58	25.26	22.23
	Midnight–5:59 a.m.	4.28	11.80	5.27	3.72
Day of the week	Weekday	72.51	63.21	70.28	73.44
	Weekend	27.49	36.79	29.72	26.56
Lighting condition	Daylight	72.00	62.52	68.42	73.26
	Dark-lighted	14.01	10.15	15.24	13.88
	Dark not-lighted	9.37	23.02	12.08	8.15
	Other	4.61	4.31	4.26	4.71
Weather	Clear	63.74	62.20	64.56	63.61
	Cloudy	18.29	19.88	18.79	18.11
	Rain	14.22	12.97	13.05	14.55
	Other	3.75	4.94	3.60	3.73
Area setting	Rural	25.17	59.70	30.10	22.54
	Urban	74.83	40.30	69.90	77.46
Highway class	U.S. highway	13.19	13.88	13.52	13.08
	Interstate	7.30	6.38	6.64	7.50
	State	19.64	25.31	20.90	19.10
	County	19.13	35.94	21.72	17.80
	Other	40.74	18.50	37.21	42.52
Intersection	No	41.96	59.81	44.72	40.55
	Yes	58.04	40.19	55.28	59.45
Posted speed limit	30–35 mph	21.91	16.53	21.47	22.24
	40–45 mph	35.08	38.17	36.72	34.56
	50–55 mph	17.24	26.10	19.53	16.32
	$\geqslant$ 60 mph	7.69	11.16	7.75	7.52
	Other	18.09	8.03	14.54	19.36
Alignment	Straight level	70.26	51.11	67.08	71.84
	Straight grade	16.46	20.94	16.21	16.32
	Curve level	6.22	11.99	7.52	2.66
	Curve grade	7.06	15.95	9.19	6.17
Primary contributing factor	DUI/aggressive operation	2.45	9.52	3.08	2.00
	Disregard signs/signals	4.14	5.20	6.24	3.60
	Distraction	8.34	6.54	8.32	8.42
	Failure to yield	17.93	18.55	23.21	16.64
	Improper action/turning	27.39	8.24	17.51	30.57
	Speeding	8.80	20.52	11.94	7.55
	Other	30.94	31.43	29.70	31.22

(continued)

Table 1. (continued)

Variable	Category	Total crashes (57,098) (%)	Crashes by severity Severe injury (1,881) (%)	Minor injury (10,657) (%)	No injury (44,560) (%)
Collision type	Single-vehicle	20.84	48.12	28.86	17.73
	Rear-end	43.19	13.02	32.19	47.09
	Side-impact	17.81	18.46	23.19	16.46
	Angle	7.98	8.83	8.95	7.71
	Head-on	1.89	6.38	3.11	1.41
	Other	8.28	5.19	3.71	9.59

Note: SUV = sport utility vehicle; DUI = driving under the influence.

increased up to 2.92 times while driving with two or more passengers (*22*).

In relation to crash time, about 49.43% of crashes occurred during afternoon hours, consistent with the findings of Hossain et al. (*41*). However, teen drivers showed a slightly higher share of severe injury crashes in late-night hours (11.80%). Teenagers were more prevalent in severe and minor injury crashes on weekends (36.79% and 27.49%, respectively). In multiple studies, nighttime- and weekend driving have been highlighted as increasing the odds of severe injury crashes (*10*, *42*). "Dark not-lighted" conditions were found to be responsible for a high percentage of severe injury collisions (23.02%). The percentage of severe injury crashes was slightly higher in cloudy weather compared with total crashes (19.88% versus 18.29%). Adequate road visibility in dark conditions is a key road-safety parameter, specifically for inexperienced drivers. In relation to highway class, state and county highways accounted for 25.31% and 35.94% of severe injury collisions, respectively. Roadways with a posted speed limit of more than 40 mph appeared to be problematic, as teenagers were reported for severe injury crashes on such roads in high proportions. A similar distribution has been noted in previous studies (*24*, *41*). Teen drivers were more frequently involved in severe injury crashes on roadways with curves and vertical gradients. Following the primary contributing factor distribution, overrepresented attributes for severe and minor injury crashes, respectively, were driving under the influence (DUI)/aggressive operation (9.52% and 3.08%), disregarding signs/signals (5.20% and 6.24%), failure to yield (18.55% and 23.21%), and speeding (20.52% and 11.94%). Moving violations, that is, disregarding signs/signals, failure to yield, and speeding are more prevalent among teen drivers, which could significantly contribute to the increased risk of collisions (*25*). Single-vehicle and head-on crashes were found to result in severe injuries (48.12% and 6.38%, respectively) compared with other severity outcomes, as also reported in the research by Peek-Asa et al. (*18*).

Results and Discussions

Number of Cluster Selection

Several studies have utilized entropy value and penalized-likelihood information criteria including BIC, AIC, and CAIC to determine the optimal number of classes. Figure 1 shows the variations in BIC, AIC, CAIC, and entropy values while testing the entire dataset for Clusters 2 to 10. It was evident that the values of all the information criteria declined as the number of classes increased. In this study, changes in the BIC, AIC, and CAIC values by the number of clusters were also calculated in percentages (Table 2). From five classes onwards, the percentage reduction in all information criteria dropped to less than 1%. Moreover, the entropy value reached a maximum value of 0.892 at Cluster 5 (Figure 1). Therefore, the results indicated that dividing the dataset into five classes represented good separation, which could be set as an optimal number of clusters for the teen driver crash dataset. Previous relevant studies have noted a similar approach for class number selection (*48*, *54*).

Cluster Description

Table 3 illustrates the variables that describe the five clusters characterizing teen driver crash patterns. The attributes were selected because of their skewed distribution with respect to the entire dataset (*45*, *47*, *48*). The considerably larger percentages of categories for each subdataset in respect of the overall crashes are presented in bold for easier identification. While defining the clusters, classes may have overlapping features, meaning one variable category may present in multiple subgroups. Following the percentage distribution of Table 3, the characteristics of Cluster 4 were significantly overrepresented in relation to the number of passengers, area setting, presence of intersection, and collision type. About 78.66% of crashes occurred with no passengers, 73.43% in rural areas, 78.79% on segments, and 88.99% were

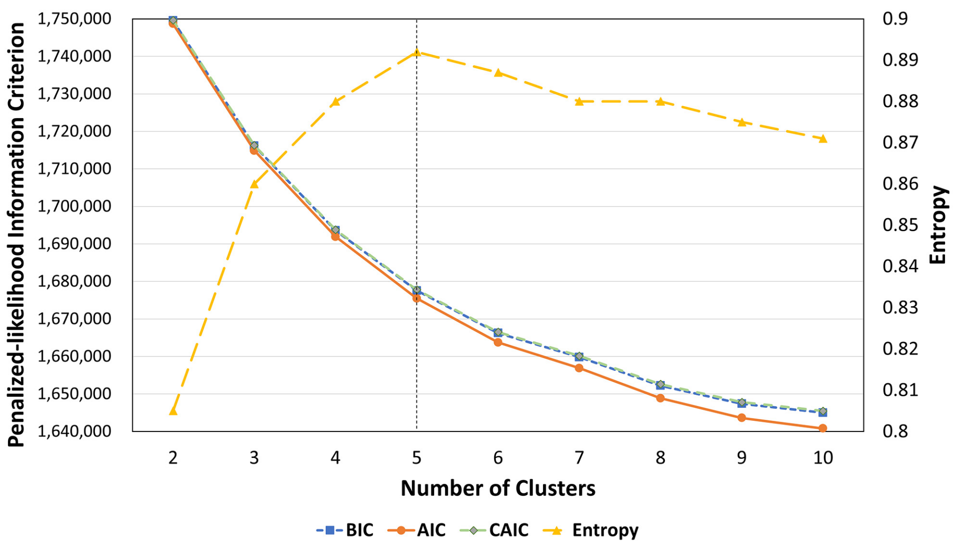

Figure 1. BIC, AIC, CAIC, and entropy values for different numbers of clusters.
Note: AIC = Akaike information criterion; BIC = Bayesian information criterion; CAIC = Consistent Akaike information criterion.

Table 2. Percentage of Reductions in Information Criteria by Number of Clusters

Number of clusters	BIC	AIC	CAIC
2	na	na	na
3	1.909	1.934	1.906
4	1.314	1.339	1.311
5	0.946	0.972	0.943
6	0.675	0.701	0.672
7	0.385	0.411	0.382
8	0.458	0.484	0.455
9	0.293	0.319	0.290
10	0.145	0.171	0.142

Note: AIC = Akaike information criterion; BIC = Bayesian information criterion; CAIC = Consistent Akaike information criterion; na = not applicable.

single-vehicle collisions. Therefore, Cluster 4 was denoted as a teen driver crash scenario in which single-vehicle crashes occurred on rural segments while driving alone. The definitions of five identified clusters are provided in the following:

- Cluster 1: rear-end collisions on urban roads (ReUr);
- Cluster 2: rear-end collisions during afternoon hours (ReAh);
- Cluster 3: side-impact collisions with failure to yield at intersections with straight-level alignment (SiFyInSl);
- Cluster 4: single-vehicle crashes on rural segments while driving alone (SvRuSgNp); and
- Cluster 5: weekend crashes during evening to midnight hours in dark-lighted conditions (WndEmhDl).

MNL Results

The results of MNL models for the entire data along with all five clusters are shown in Tables 4 to 9. The analyses were performed considering a cut-off significance level of 0.05. Only the standard errors and coefficient values of statistically significant attributes are presented. In discrete modeling, the signs of estimated coefficients may not accurately convey the influence of the covariates on crash severity outcomes (*45*, *54*). Therefore, the marginal effects of all significant categories were calculated to explain their impact on the severities of teen driver crashes. It should be noted that the results of categories assigned as “other” and “unknown” were excluded owing to the predictive uncertainty related to these crashes, which is difficult to describe. Changes in the variable impacts across different classes are addressed in detail in the following subsections.

Table 3. Distributions of Featured Variables Based on the Optimal Number of Clusters

Percentage of whole dataset		Whole data	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
		100	26.84	23.70	16.94	16.59	15.93
Variable	Attribute						
Number of passengers	None	62.51	64.89	54.50	51.31	**88.21**	54.90
Crash time	6–11:59 p.m.	22.91	10.71	6.06	7.17	31.36	**88.04**
	12–5:59 p.m.	49.43	55.54	**64.40**	55.09	31.81	0.62
Day of the week	Weekend	27.49	17.42	18.36	20.22	40.43	**62.85**
Area setting	Rural	25.17	0.45	36.85	27.37	**73.43**	19.24
	Urban	74.83	**99.55**	63.15	72.63	26.57	80.76
Intersection	No	41.96	30.28	64.21	21.34	**78.79**	38.26
	Yes	58.04	69.72	35.79	**78.66**	21.21	61.74
Lighting condition	Dark-lighted	14.01	2.58	3.34	3.86	5.32	**70.23**
Alignment	Straight level	68.17	70.44	70.76	**82.93**	33.63	75.59
Primary contributing factor	Failure to yield	17.93	1.75	1.57	**78.30**	1.01	23.34
Crash type	Rear-end	43.19	**70.08**	**77.90**	0.34	3.38	34.23
	Side-impact	17.81	4.86	3.02	**69.64**	1.83	22.53
	Single-vehicle	20.84	8.54	5.41	0.19	**88.99**	15.32

Driver and Vehicle Characteristics. Compared with adult teenagers aged 17 to 19 years, novice teen drivers aged 15 to 16 years were less likely to be involved in minor injury crashes in the whole data, Cluster 2 (ReAh), and Cluster 3 (SiFyInSl) with marginal effects of −0.0109, −0.0163, and −0.0187, respectively. However, the Cluster 4 (SvRuSgNp) results indicated that teenagers in the learner/probationary stages have an increased likelihood of minor injury collisions (+ 0.0182). In unsupervised driving conditions, the unique design challenges of rural roadways coupled with driving inexperience could contribute to the evaluated risk of collisions (*37*). Except for Cluster 1 (ReUr), female teenagers were more prone to crashes with lower severity levels, consistent with the findings of Duddu et al. (*28*). However, female teen drivers in Cluster 4 (SvRuSgNp) were found to be significantly associated with a high probability of severe crashes (+ 0.0022) compared with their counterparts. Adanu et al. documented the speeding tendencies of young female drivers on rural roads that resulted in fatal single-vehicle crashes (*11*). In this research outcomes, heterogeneity was observed in the impact of the "impairment" variable. For example, among teenagers, driving under the influence of alcohol/drugs was only found to contribute to the increased risk of severe injury crashes (+ 0.084) in Cluster 3 (SiFyInSl). While intoxicated, drivers take longer to react to critical events (*70*), therefore, leading to a high likelihood of failing to yield at intersections. Unrestrained driving seems to be an influential factor for teen drivers, as such risk-taking behavior increased the likelihood of severe and minor injury crashes in all clusters. Motorists not wearing seatbelts are more likely to be ejected during impact, which increases the risk of fatalities and injuries (*71*).

Between the classes, driving without being buckled up was more evident in Cluster 4 (SvRuSgNp), increasing the exposure to severe and minor injury crashes by + 0.2079 and −0.1591, respectively. As shown in the results, the "distraction" variable was not found to be significantly associated with injury crashes. Police officers often encounter difficulty specifying distracted driving in the investigated crashes, therefore, such related crashes are seriously underreported (*21*). According to the marginal effects, the risk magnitude of teen driver crashes increased with the number of passengers. For instance, compared with solo driving, driving with one and multiple passengers in Cluster 3 (SiFyInSl) significantly elevated the probability of severe injury crashes, with marginal effects of + 0.0148 and + 0.0352, respectively. A similar trend was also perceptible in relation to minor injury crashes, increasing the likelihood by + 0.0640 and + 0.0914, respectively. Along with peer influence (*10*), passengers can aggravate risky driving maneuvers among teenagers, including cellphone usage during driving (*21*) and speeding (*72*). Different vehicle classes have numerous effects on teen driver crash severity. Passenger cars were more likely to be associated with minor injury crashes in Cluster 1 (ReUr), Cluster 3 (SiFyInSl), and Cluster 5 (WndEmhDl). On the contrary, in Cluster 4 (SvRuSgNp), vans/SUVs had a positive impact on both severe and minor injury crashes, with marginal effects of + 0.0136 and + 0.0338, respectively. One previous study in North Carolina also reported the high risk of severe injuries and deaths in single-vehicle crashes on rural roads when driving vans/SUVs (*73*).

Table 4. MNL Results for Whole Data

Variable categories	Coefficient		Standard error		Marginal effect		
	Severe injury	Minor injury	Severe injury	Minor injury	Severe injury	Minor injury	No injury
(Intercept)	−5.311[c]	−2.184[c]	0.139	0.05	na	na	na
Age (ref: 17–19 years)							
15–16 years	na	−0.074[b]	na	0.027	0.0017	−0.0109	0.0092
Gender (ref: male)							
Female	na	0.180[c]	na	0.024	−0.0002	0.0248	−0.0269
Number of passengers (ref: no)							
One	0.510[c]	0.381[c]	0.064	0.026	0.0112	0.0508	−0.062
Multiple	1.049[c]	0.633[c]	0.084	0.039	0.0290	0.0856	−0.1145
No restraint usage (ref: no)							
Yes	2.036[c]	1.031[c]	0.073	0.054	0.0864	0.143	−0.2294
Lighting condition (ref: daylight)							
Dark-lighted	na	0.189[c]	na	0.046	−0.0017	0.0281	−0.0263
Dark not-lighted	na	0.181[c]	na	0.052	0.0020	0.0256	−0.0276
Weather (ref: clear)							
Rain	−0.465[c]	−0.302[c]	0.082	0.036	−0.0097	−0.0377	0.0474
Area setting (ref: urban)							
Rural	0.839[c]	0.102[b]	0.072	0.034	0.0245	0.0070	−0.0314
Highway class (ref: state)							
Interstate	−0.783[c]	−0.205[b]	0.145	0.068	−0.0189	−0.0222	0.0412
County	−0.311[c]	−0.111[b]	0.088	0.042	−0.0086	−0.0128	0.0214
Intersection (ref: no)							
Yes	−0.238[c]	−0.095[c]	0.059	0.025	−0.0059	−0.0116	0.0175
Posted speed limit (ref: 40–45 mph)							
30–35 mph	−0.427[c]	−0.174[c]	0.075	0.031	−0.0099	−0.0215	0.0315
50–55 mph	0.256[b]	0.145[c]	0.079	0.035	0.0069	0.0198	−0.0267
⩾ 60 mph	0.478[c]	na	0.128	na	0.0044	0.0081	−0.0246
Alignment (ref: straight level)							
Straight grade	0.323[c]	na	0.068	na	0.0175	0.0049	−0.0139
Curve level	0.234[a]	na	0.092	na	0.0059	0.0107	−0.0165
Curve grade	0.323[c]	0.171[c]	0.085	0.046	0.0078	0.0225	−0.0303
Primary contributing factor (ref: improper action/turning)							
DUI/aggressive operation	1.363[c]	0.437[c]	0.151	0.085	0.0376	0.0476	−0.0852
Disregarding signs/signals	1.152[c]	0.962[c]	0.162	0.067	0.0213	0.1418	−0.1631
Distraction	0.629[c]	0.375[c]	0.13	0.046	0.0115	0.0464	−0.0579
Failure to yield	0.691[c]	0.614[c]	0.119	0.047	0.0111	0.0837	−0.0948
Speeding	1.153[c]	0.576[c]	0.117	0.050	0.0266	0.0724	−0.099
Crash type (ref: rear-end)							
Angle	1.402[c]	0.287[c]	0.12	0.051	0.0318	0.0310	−0.0628
Head-on	2.663[c]	1.042[c]	0.134	0.075	0.0927	0.1387	−0.2314
Side-impact	1.406[c]	0.411[c]	0.111	0.045	0.0306	0.0498	−0.0804
Single-vehicle	1.513[c]	0.765[c]	0.141	0.061	0.0305	0.1099	−0.1404
Model parameters							
Log likelihood				−32,042.81			
AIC				64,277.61			

Note: MNL = multinomial logit; DUI = driving under the influence; AIC = Akaike information criterion; na = not applicable.
[a]$p < 0.05$; [b]$p < 0.01$; [c]$p < 0.001$.

Environmental Characteristics. Clusters 2 (ReAh) and 5 (WndEmhDl) contained the "crash time" variable attributes. In the rest of the classes, none of the crash hour intervals significantly influenced injury crashes with reference to no injury crashes. Compared with weekdays, teen driving on the weekends increased their exposure to severe injury crashes (+ 0.0196) in Cluster 4 (SvRuSgNp). Single-vehicle crashes were more frequent in rural areas (*42*), and multiple pieces of evidence suggest that risky driving behavior is more prevalent among teenagers during weekends (*41*, *74*). The collective impact of these understandings can contribute to crashes with a high degree of severity. The MNL results for the whole data indicated that dark-lighted conditions were significant, increasing the likelihood of minor injury collisions (+ 0.0281). However, dark

Table 5. MNL Results for Cluster 1 (ReUr)

Variable categories	Coefficient		Standard error		Marginal effect		
	Severe injury	Minor injury	Severe injury	Minor injury	Severe injury	Minor injury	No injury
(Intercept)	−6.093[c]	−2.118[c]	0.558	0.115	na	na	na
Number of passengers (ref: no)							
One	na	0.322[c]	na	0.054	0.0011	0.0363	−0.0375
Multiple	0.735[a]	0.561[c]	0.305	0.089	0.0059	0.0677	−0.0736
No restraint usage (ref: no)							
Yes	2.135[c]	0.839[c]	0.276	0.135	0.0373	0.1101	−0.1474
Vehicle type (ref: passenger car)							
Van/SUV	na	−0.118[a]	na	0.059	0.0025	−0.0133	0.0108
Lighting condition (ref: daylight)							
Dark not-lighted	na	0.801[c]	na	0.289	−0.0083	0.1176	−0.1093
Intersection (ref: no)							
Yes	−0.467[a]	−0.225[c]	0.192	0.052	−0.0035	−0.0251	0.0286
Posted speed limit (ref: 40–45 mph)							
50–55 mph	na	0.233[a]	na	0.107	−0.0045	0.0311	−0.0266
Alignment (ref: straight level)							
Curve level	0.955[b]	na	0.319	na	0.0101	0.0208	−0.0308
Curve grade	na	0.340[b]	na	0.116	0.0010	0.0421	−0.0421
Primary contributing factor (ref: improper action/turning)							
DUI/aggressive operation	2.265[c]	0.473[a]	0.392	0.202	0.0293	0.0460	−0.0753
Distraction	0.950[b]	0.555[c]	0.345	0.083	0.0053	0.0612	−0.0665
Failure to yield	na	0.401[a]	na	0.202	−0.0041	0.0432	−0.0391
Speeding	1.525[c]	1.022[c]	0.397	0.108	0.0107	0.1321	−0.1428
Model parameters							
Log likelihood				−6,379.32			
AIC				12,922.64			

Note: MNL = multinomial logit; ReUr = rear-end collisions on urban roads; SUV = sport utility vehicle; DUI = driving under the influence; AIC = Akaike information criterion; na = not applicable.
[a]$p < 0.05$; [b]$p < 0.01$; [c]$p < 0.001$.

conditions with no streetlighting had a significant impact on the minor injury crashes of teen drivers in Clusters 1 (ReUr) and 2 (ReAh), with marginal effects of + 0.1176 and + 0.0371, respectively. Darkness increases the closing speed of following vehicles by shortening the available reaction distance, resulting in an increased likelihood of rear-end crashes (*75*). The impact of cloudy weather on minor injury crashes varied according to class. For example, a marginal effect of −0.0237 was found for Cluster 4 (SvRuSgNp), whereas Cluster 5 (WndEmhDl) presented a marginal effect of + 0.0396. Such findings imply that cloudy weather conditions might be more influential in the presence of specific attributes related to environmental characteristics. Compared with clear weather, teen drivers were significantly less likely to be involved in minor injury crashes during rainy conditions in all classes except for Cluster 1, which means rainy weather had a protective effect on crash risk. One possible explanation is that teen drivers control their vehicles with greater caution while driving in wet conditions (*76*) or may be less likely to drive in bad weather.

Road Characteristics. Compared with urban settings, rural roads increased the severe crash likelihood in Cluster 2 (ReAh), Cluster 3 (SiFyInSl), and Cluster 5 (WndEmhDl) with marginal effects of + 0.0245, + 0.048, and + 0.0197, respectively. A similar finding in relation to rurality has been observed in previous teen driver safety studies (*18*, *37*). State highways seem to be problematic for teenagers, as all other highway classes are less likely to be associated with injury crashes. State highways are often recognized for crashes involving serious injuries resulting from heavy vehicular movements and high traffic volumes (*77*). At intersections, teen drivers showed a decreased probability of injury crashes in the data as a whole, in Cluster 1 (ReUr), Cluster 2 (ReAh), and Cluster 5 (WndEmhDl). Teenagers may become more aware of the complexities associated with intersections, and therefore more likely to adopt defensive driving in these locations. Heterogeneity in the impacts of posted speed limits was evident within each cluster. For example, compared with the posted speed limit of 40 to 45 mph, teen motorists in Cluster 5 (WndEmhDl) traveling on

Table 6. MNL Results for Cluster 2 (ReAh)

Variable categories	Coefficient		Standard error		Marginal effect		
	Severe injury	Minor injury	Severe injury	Minor injury	Severe injury	Minor injury	No injury
(Intercept)	−5.062[c]	−2.071[c]	0.250	0.088	na	na	na
Age (ref: 17–19 years)							
15–16 years	na	−0.138[a]	na	0.062	−0.0018	−0.0163	0.0181
Gender (ref: male)							
Female	na	0.171[b]	na	0.053	−0.0004	0.0212	−0.0217
No restraint usage (ref: no)							
Yes	1.719[c]	0.857[c]	0.239	0.158	0.0573	0.1179	−0.1752
Number of passengers (ref: no)							
One	na	0.251[c]	na	0.053	0.0024	0.0302	−0.0326
Multiple	0.620[c]	0.575[c]	0.185	0.077	0.0111	0.0754	−0.0865
Weather (ref: clear)							
Rain	na	−0.261[c]	na	0.079	−0.0057	−0.0297	0.0355
Lighting condition (ref: daylight)							
Dark not-lighted	na	0.269[a]	na	0.136	−0.0035	0.0371	−0.0336
Area setting (ref: urban)							
Rural	1.204[c]	0.154[b]	0.141	0.057	0.0245	0.0146	−0.039
Highway class (ref: state)							
Interstate	−0.941[c]	−0.281[b]	0.249	0.102	−0.0163	−0.0319	0.0482
U.S. highway	−0.324[a]	−0.140[a]	0.159	0.062	−0.0069	−0.0168	0.0236
County	−0.449[a]	−0.316[c]	0.216	0.083	−0.0087	−0.0372	0.0459
Intersection (ref: no)							
Yes	na	−0.121[a]	na	0.050	−0.0028	−0.0144	0.0172
Posted speed limit (ref: 40–45 mph)							
50–55 mph	0.509[b]	0.161[b]	0.179	0.061	0.0085	0.0184	−0.0269
⩾ 60 mph	1.121[c]	0.253[a]	0.254	0.105	0.0256	0.0271	−0.0527
Alignment (ref: straight level)							
Curve grade	0.647[a]	na	0.276	na	0.0148	0.0149	−0.0297
Primary contributing factor (ref: improper action/turning)							
DUI/aggressive operation	2.594[c]	0.861[b]	0.428	0.331	0.1064	0.0934	−0.1998
Distraction	0.570[b]	0.414[c]	0.181	0.07	0.0083	0.0499	−0.0582
Failure to yield	na	0.427[a]	na	0.202	−0.0090	0.0556	−0.0466
Speeding	na	0.739[c]	na	0.121	0.0443	0.0907	−0.1349
Model parameters							
Log likelihood				−6,755.62			
AIC				13,663.25			

Note: MNL = multinomial logit; ReAh = rear-end collisions during afternoon hours; DUI = driving under the influence; AIC = Akaike information criterion; na = not applicable.
[a]$p < 0.05$; [b]$p < 0.01$; [c]$p < 0.001$.

roadways with a posted speed limit of ⩾ 60 mph were more likely to be involved in severe and minor injury crashes, with marginal effects of + 0.0239 and + 0.078, respectively. Utilizing crash information from Nebraska, Dhungana and Qu inferred that speeding behavior was highly associated with teen drivers, high posted speed limits, weekends, and nighttime driving (*78*). According to the MNL outcomes, the risk of crashes with a high severity level increased with increasing posted speed limits. For instance, the exposure to severe injury crashes in Cluster 2 (ReAh) was estimated to be + 0.0085 for the posted speed limit of 50 to 55 mph, and + 0.0256 for the posted speed limit of ⩾ 60 mph. In high posted speed limits, drivers exhibit more significant speed deviations between lanes and among the segments, which elevates the probability of crash occurrence (*79*). In Cluster 4 (SvRuSgNp), roads with vertical grades significantly increased the probability of severe and minor injury crashes. Vertical alignments may also have a significant impact on rural road safety (*80*) owing to the difficulty of maintaining vehicle control (specifically for large vehicles), inadequate sight distance, and so forth (*81*). In Cluster 1 (ReUr), curved roads with no vertical alignment were found to contribute to the increase in severe injury crashes (+ 0.0101), whereas curved roads with grades significantly affected minor injury crashes (+ 0.0421). Teen drivers in Cluster 5 (WndEmhDl) were more prone to severe injury crashes (+ 0.0104) on straight roads with grades.

Table 7. MNL Results for Cluster 3 (SiFyInSl)

Variable categories	Coefficient		Standard error		Marginal effect		
	Severe injury	Minor injury	Severe injury	Minor injury	Severe injury	Minor injury	No injury
(Intercept)	−3.483[c]	−1.230[c]	0.202	0.086	na	na	na
Age (ref: 17–19 y)							
15–16 y	na	−0.108[a]	na	0.055	−0.0008	−0.0187	0.0195
Gender (ref: male)							
Female	na	0.123[a]	na	0.051	−0.0049	0.0203	−0.0252
Impairment (ref: no)							
Yes	1.522[a]	na	0.680	na	0.0841	0.0333	−0.1174
No restraint usage (ref: no)							
Yes	1.853[c]	0.852[c]	0.221	0.153	0.0937	0.1326	−0.2263
Number of passengers (ref: no)							
One	0.595[c]	0.387[c]	0.126	0.052	0.0148	0.0640	−0.0788
Multiple	1.105[c]	0.562[c]	0.171	0.081	0.0352	0.0914	−0.1266
Vehicle type (ref: passenger car)							
Van/SUV	na	−0.132[a]	na	0.061	0.0031	−0.0239	0.0208
Weather (ref: clear)							
Rain	na	−0.338[b]	na	0.101	0.0027	−0.0569	0.0542
Area setting (ref: urban)							
Rural	1.188[c]	0.207[b]	0.149	0.08	0.0480	0.0225	−0.0706
Highway class (ref: state)							
County	−0.849[c]	−0.331[c]	0.192	0.092	−0.0230	−0.0490	0.0720
Posted speed limit (ref: 40–45 mph)							
30–35 mph	−0.603[c]	−0.147[a]	0.168	0.064	−0.0169	−0.0215	0.0384
50–55 mph	0.468[b]	na	0.157	na	0.0199	0.0156	−0.0355
Model parameters							
Log likelihood				−6,464.42			
AIC				13,052.84			

Note: MNL = multinomial logit; SiFyInSl = side-impact collisions with failure to yield at intersections with straight level; SUV = sport utility vehicle; AIC = Akaike information criterion; na = not applicable.
[a]$p < 0.05$; [b]$p < 0.01$; [c]$p < 0.001$.

Crash Characteristics. In relation to primary contributing factors, DUI/aggressive driving were found to be significant in Cluster 1 (ReUr), Cluster 2 (ReAh), and Cluster 5 (WndEmhDl) for both severe and minor injury crashes. In Cluster 4 (SvRuSgNp), such teenage maneuvers significantly affected the prevalence severe injury crashes (+ 0.0593), as also reported in the research by Peek-Asa et al. (*18*). Compared with improper action/turning, disregarding traffic signs/signals added up to + 0.0219 and + 0.1464 likelihood of severe and minor injury crashes in Cluster 5 (WndEmhDl). Several studies have identified that traffic signal violations are prevalent on weekends and at night (*82*, *83*). Distracted driving significantly contributed to injury crashes in Clusters 1 (ReUr) and 2 (ReAh). While using cellphones, teen drivers take their attention away from the road for longer durations, therefore, having a high risk of rear-end collisions (*44*). Failure to yield seemed to render teenagers vulnerable in Cluster 5 (WndEmhDl), increasing the probability of severe injury crashes, with a marginal effect of + 0.0092. Speeding as a contributing behavior appeared to be influential in all classes other than Cluster 3 (SiFyInSl). One possible explanation could be that the lower posted speed limit at signalized intersections reduced the probability of injury crashes (*84*). Compared with rear-end collisions in Cluster 5 (WndEmhDl), head-on and single-vehicle crashes were the top two collision types that were most likely to result in severe injuries (+ 2.471 and + 1.712, respectively). Head-on and single-vehicle collisions often produce more severe injuries than rear-end crashes (*85*).

Key Findings

Figures 2 and 3 highlight the heterogeneous effects of variable categories on severe and minor injury crashes for the entire data and different clusters (i.e., subgroups of the whole dataset). Only statistically significant variables and their marginal effects are presented to compare the results of each submodel. Few contributing factors that were found to have negligible impacts in the entire dataset significantly affected teen driver crash severity in specific clusters and vice versa. This confirmed the heterogeneity within the crash dataset, which indicated that

Table 8. MNL Results for Cluster 4 (SvRuSgNp)

Variable categories	Coefficient		Standard error		Marginal effect		
	Severe injury	Minor injury	Severe injury	Minor injury	Severe injury	Minor injury	No injury
(Intercept)	−2.999[c]	−1.349[c]	0.328	0.196	na	na	na
Age (ref: 17–19 y)							
15–16 y	na	0.123[a]	na	0.060	0.0092	0.0182	−0.0274
Gender (ref: male)							
Female	0.351[c]	0.404[c]	0.085	0.054	0.0149	0.0661	−0.0811
No restraint usage (ref: no)							
Yes	2.295[c]	1.339[c]	0.097	0.083	0.2079	0.1591	−0.3669
Vehicle type (ref: passenger car)							
Van/SUV	0.256[b]	0.220[b]	0.101	0.065	0.0136	0.0338	−0.0474
Day of the week (ref: weekday)							
Weekend	0.262[b]	na	0.080	na	0.0196	−0.0027	−0.0169
Weather (ref: clear)							
Cloudy	na	−0.148[a]	na	0.065	−0.0078	−0.0237	0.0315
Rain	−0.696[c]	−0.422[c]	0.114	0.068	−0.0381	−0.0594	0.0975
Highway class (ref: state)							
Interstate	−0.946[c]	na	0.212	na	−0.0546	−0.0157	0.0703
Posted speed limit (ref: 40–45 mph)							
30–35 mph	−0.376[b]	−0.191[b]	0.112	0.072	−0.0218	−0.0253	0.0470
Alignment (ref: straight level)							
Straight grade	0.393[c]	0.213[b]	0.112	0.073	0.0238	0.0282	−0.0520
Curve grade	0.347[b]	0.201[b]	0.107	0.068	0.0203	0.0274	−0.0477
Primary contributing factor (ref: improper action/turning)							
DUI/aggressive operation	0.916[b]	na	0.305	na	0.0593	0.0158	−0.0751
Speeding	0.881[b]	0.333[a]	0.272	0.155	0.0520	0.0391	−0.0911
Model parameter							
Log likelihood				−7,464.77			
AIC				15,085.54			

Note: MNL = multinomial logit; SvRuSgNp = single-vehicle crashes on rural segments while driving alone; SUV = sport utility vehicle; DUI = driving under the influence; AIC = Akaike information criterion; na = not applicable.
[a]$p < 0.05$; [b]$p < 0.01$; [c]$p < 0.001$.

distinct explanatory variables influenced the severity of different types of teen driver crashes. In addition, the MNL outcomes of different clusters represented meaningful crash patterns demonstrating the cumulative effect of the variable attributes. The critical findings of this study include the following:

- Owing to the high likelihood of single-vehicle crashes with injuries, solo driving on rural road segments made teen drivers vulnerable (1) who were female, (2) in the learner/intermediate phases of driving, and (3) operating vans/SUVs.
- At intersections, side-impact crashes were more likely to be severe when intoxicated teen drivers failed to yield. On the contrary, weekend and vertical alignment were associated with severe single-vehicle crashes on rural road segments with no passengers.
- For teenagers, no restraint usage, carrying passengers, and driving in rural areas significantly increased the risk of severe- and minor injury crashes. In addition, they were more likely to be involved in rear-end crashes (1) while distracted, (2) driving in dark not-lighted conditions, and (3) operating vehicles on curved roads.
- Although rainy weather conditions showed a significant decrease in the severity of injuries, cloudy weather conditions had a significant impact on minor injury crashes when teenagers drove on weekends from evening to midnight hours in dark-lighted conditions.
- In the afternoon hours, rear-end crashes involving teen drivers were more likely to be severe on roadways with posted speed limits of more than 50 mph. In addition, the combined effect of weekend and dark conditions was significantly associated with disregarding traffic signs/signals and posted speed limits of $\geqslant$ 60 mph, which led to severe injury crashes.
- Speeding as a primary contributing factor was found to be significant except in relation to side-impact collisions with a failure to yield at intersections. State highways might therefore have a significant impact on teen driver crashes with injuries.

Table 9. MNL Results for Cluster 5 (WndEmhDl)

Variable categories	Coefficient		Standard error		Marginal effect		
	Severe injury	Minor injury	Severe injury	Minor injury	Severe injury	Minor injury	No injury
(Intercept)	−3.367[b]	−2.156[a]	0.267	0.124	na	na	na
Gender (ref: male)							
Female	na	0.120[a]	na	0.057	−0.0039	0.0191	−0.0151
No restraint usage (ref: no)							
Yes	1.907[c]	0.696[c]	0.219	0.145	0.0766	0.0889	−0.1655
Number of passengers (ref: no)							
One	0.562[b]	0.430[c]	0.165	0.062	0.0097	0.0608	−0.0705
Multiple	1.092[c]	0.592[c]	0.194	0.085	0.0252	0.0829	−0.1081
Vehicle type (ref: passenger car)							
Van/SUV	na	−0.147[a]	na	0.069	−0.0044	−0.0201	0.0245
Weather (ref: clear)							
Cloudy	na	0.247[b]	na	0.077	−0.0028	0.0396	−0.0368
Rain	na	−0.232[b]	na	0.086	−0.0052	−0.0308	0.0358
Area setting (ref: urban)							
Rural	0.663[b]	na	0.218	na	0.0197	−0.0121	−0.0075
Intersection (ref: no)							
Yes	−0.350[a]	na	0.154	na	−0.0084	−0.0046	0.0131
Posted speed limit (ref: 40–45 mph)							
30–35 mph	na	−0.171[a]	na	0.074	−0.0065	−0.024	0.0305
50–55 mph	na	0.296[c]	na	0.085	0.0071	0.0473	−0.0543
⩾ 60 mph	0.878[a]	0.498[b]	0.363	0.177	0.0239	0.078	−0.1020
Alignment (ref: straight level)							
Straight grade	0.397[a]	na	0.189	na	0.0104	0.0003	−0.0107
Primary contributing factor (ref: improper action/turning)							
DUI/aggressive operation	1.374[c]	0.512[b]	0.397	0.174	0.0307	0.0619	−0.0925
Disregarding signs/signals	1.285[c]	0.986[c]	0.331	0.132	0.0219	0.1464	−0.1683
Distraction	na	0.351[b]	na	0.127	0.0001	0.0459	−0.0461
Failure to yield	0.736[b]	0.762[c]	0.276	0.101	0.0092	0.1096	−0.1188
Speeding	1.226[b]	0.384[a]	0.390	0.157	0.0264	0.0441	−0.0703
Crash type (ref: rear-end)							
Angle	1.344[c]	na	0.298	na	0.0222	0.0089	−0.0311
Head-on	2.471[c]	0.797[c]	0.320	0.144	0.0618	0.1125	−0.1743
Side-impact	1.592[c]	0.370[c]	0.273	0.096	0.0277	0.0485	−0.0762
Single-vehicle	1.712[c]	0.639[c]	0.273	0.094	0.0289	0.0947	−0.1236
Model parameter							
Log likelihood				−5,104.93			
AIC				10,373.86			

Note: MNL = multinomial logit; WndEmhDl = weekend crashes during evening to midnight hours in dark-lighted conditions; SUV = sport utility vehicle; DUI = driving under the influence; AIC = Akaike information criterion; na = not applicable.
[a]$p < 0.05$; [b]$p < 0.01$; [c]$p < 0.001$.

Conclusions

Study Contribution and Practical Implementations

Teen drivers are vulnerable road users, and the numbers of casualties related to teen driver crashes have significant social and economic influences. This study utilized 3 years (2017 to 2019) of police-investigated crash information from Alabama to explore the effects of driver, vehicle, environmental, road, and crash characteristics on the severity of teen driver collisions. LCC was implemented to minimize the heterogeneity in the extracted dataset by dividing the entire data into meaningful classes. Finally, MNL models were constructed to illustrate the significant contributing factors influencing the severity of outcomes in different teen driver crash scenarios. Marginal effects were computed to better understand the impacts of variable categories. The traditional practice of employing discrete choice modeling to investigate heterogeneous crash data is unable to detect the hidden affiliations among injury severity levels and crash contributing factors, which can be vital in developing effective countermeasures and policies for improving teen driving safety. The findings of this cluster-based regression approach

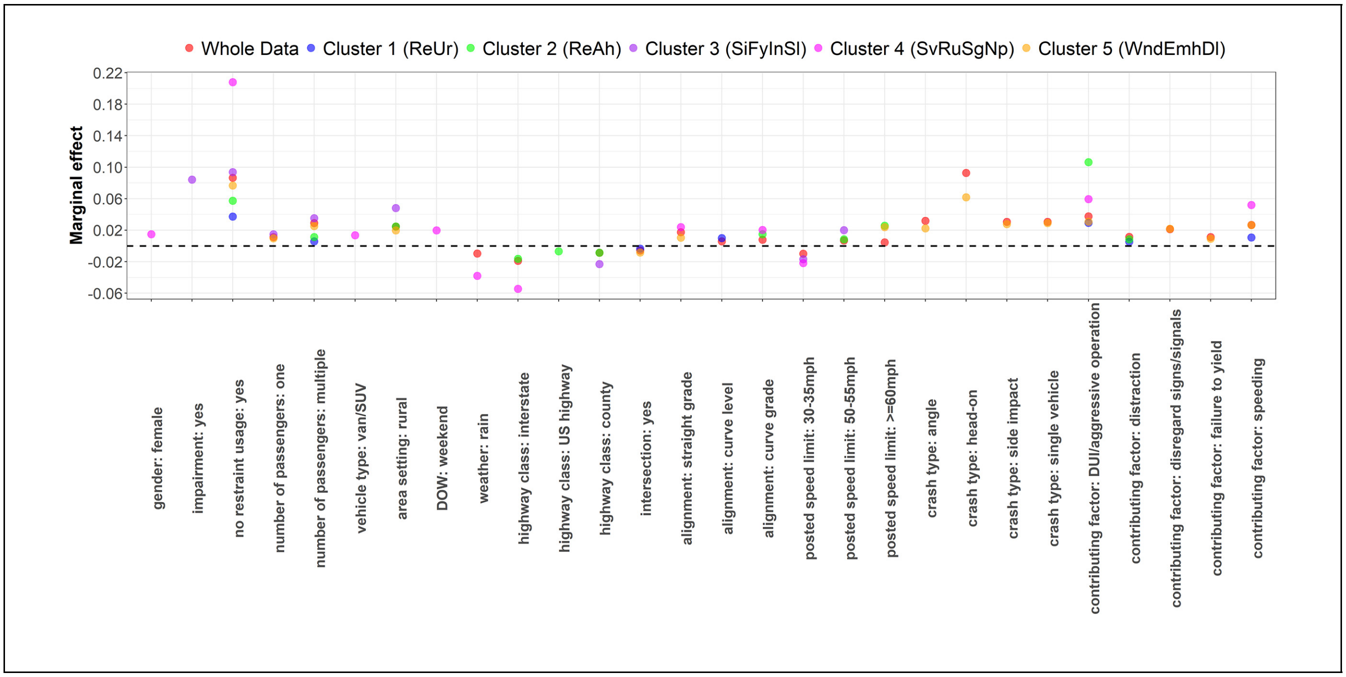

Figure 2. Marginal effects of significant predictors for severe injury crashes.
Note: ReUr = Cluster 1: rear-end collisions on urban roads; ReAh = Cluster 2: rear-end collisions during afternoon hours; SiFyInSl = Cluster 3: side-impact collisions with failure to yield at intersections with straight-level alignment; SvRuSgNp = Cluster 4: single-vehicle crashes on rural segments while driving alone; WndEmhDl = Cluster 5: weekend crashes during evening to midnight hours in dark-lighted conditions.

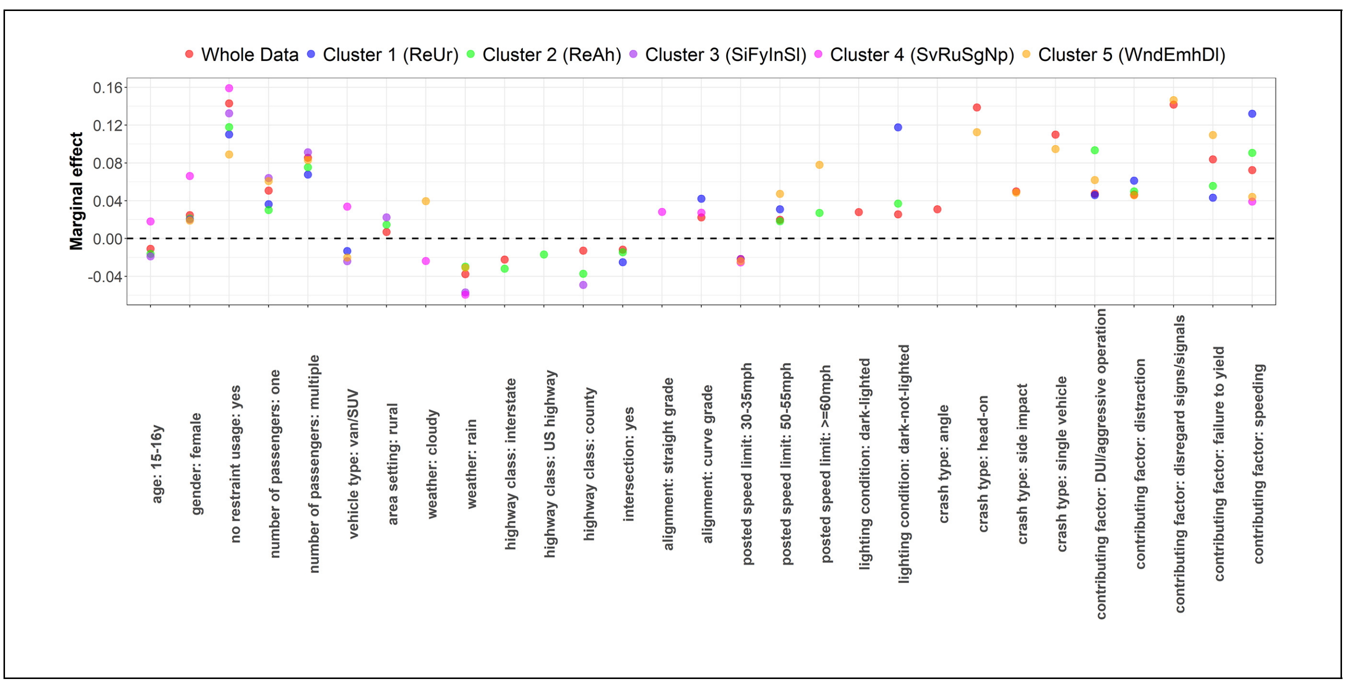

Figure 3. Marginal effects of significant predictors for minor injury crashes.
Note: ReUr = Cluster 1: rear-end collisions on urban roads; ReAh = Cluster 2: rear-end collisions during afternoon hours; SiFyInSl = Cluster 3: side-impact collisions with failure to yield at intersections with straight-level alignment; SvRuSgNp = Cluster 4: single-vehicle crashes on rural segments while driving alone; WndEmhDl = Cluster 5: weekend crashes during evening to midnight hours in dark-lighted conditions.

indicated that the significance and estimated impact of variables differed within clusters and between crash severity outcomes in the same cluster. This justifies the superiority of LCC in enhancing homogeneity in the dataset and heterogeneity within the separated subgroups. In addition, applying MNL in the systematically segmented datasets showed the grouping of attributes contributing to teen driver crashes. From an application perspective, the study outcomes could help practitioners and policy makers to develop different safety improvement strategies to reduce the injuries and deaths associated with teen driver crashes in distinct circumstances.

In this study, the results of latent class segmented submodels indicated real-world crash patterns of teen drivers, for example, in relation to severe side-impact crashes of intoxicated teen drivers while failing to yield at intersections, single-vehicle crashes on rural roads while driving alone, severe crashes on weekends under dark conditions while disregarding traffic signs/signals. Such contextual understandings of the underlying risk factors could be included in training materials to strengthen the existing educational interventions (e.g., TeenSMART, peer education programs) (*41*). Modern driving simulators could reproduce an extensive range of road, environment, and traffic conditions (e.g., numerous road lighting conditions, weather conditions, free-flow traffic, congested traffic conditions, interactions with other road users). The scenarios could also be adjusted for different driver groups stratified by age, gender, and driving experience. Researchers have conducted simulator experiments by mimicking on-road driving scenarios to evaluate the driving performance of teen drivers in hazardous circumstances (*86*). The association of contributing factors identified in the current study could be helpful in developing simulator scenarios for assessing teen drivers' safe driving needs.

This study has provided insights into the factors that could be controlled to ensure safe interactions between teen drivers and other road users. The results indicated that solo driving rendered teenagers in the learner/probationary stages on rural road segments vulnerable. According to Alabama's GDL law, supervised driving is mandatory for learner permit holders. Furthermore, intermediate license holders must be occupied by a parent, legal guardian, or licensed driver aged 21 or older between midnight and 6 a.m. Therefore, further investigation is required to confirm whether the GDL regulations were complied with or not. In this study, teenagers tended to violate the primary seatbelt law along with MLDA and zero-tolerance regulations. Periodical and intensive enforcement campaigns could be implemented to prevent such offenses (*30*). For teenagers, crashes in rural areas and on curved roads were more likely to result in severe injuries. Road-safety assessments with respect to geometric features and functional elements could be prioritized in hotspot locations. Following the MNL results, teen drivers' safety could be improved by increasing road visibility in dark conditions. In recent years, transportation planners have prioritized modern technologies to improve existing road illumination. FHWA suggests installing LEDs and adaptive lighting systems (*87*). Crash avoidance technologies such as side-view assist, lane departure warning, vehicle stability control, forward collision warning, and adaptive headlights have been found to be effective in reducing several types of collision (e.g., rear-end, head-on, single-vehicle, side-swipe) (*88*). A successful intervention design may require a mix of environmental, technological, educational, and enforcement efforts to improve teen driving safety.

Study Limitations

This study had some limitations that can be addressed in follow-up studies. Ordinal features of the crash severity level were not considered. Rather than considering fixed effects, a mixed effect ordered model could be applied in future studies. A similar study approach could be used to compare teen driver crash characteristics with other age groups. Further investigations could be conducted, including site-specific-, occupant-, and situational characteristics, to better comprehend the effect and transferability of risk factors. Driver attitudes and behavior vary with geographic locations (*89*), and such variations might be more understandable for teenagers in relation to state-based GDL policies. This study only considered Alabama crash data, therefore, it may not be possible to generalize the findings to other states and jurisdictions until further investigations are performed. Underreporting issues in relation to driver distraction information may have influenced the outcomes of this research. Other data sources, such as Strategic Highway Safety Program 2 Naturalistic Driving Study, could be used in future studies to minimize the specified constraint. Further research could be undertaken to recognize the crash mechanisms (e.g., sequence of events) of teen drivers in the identified crash scenarios by utilizing narratives and collision diagrams of crash reports. Real-world observations combined with laboratory experiments could be more strategic in targeting specific countermeasures.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: M.M. Hossain, H. Zhou; data collection: M.M. Hossain; analysis and interpretation of results: M.M. Hossain; draft manuscript preparation: M.M. Hossain, H. Zhou, X. Sun. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Md Mahmud Hossain https://orcid.org/0000-0002-2737-6951
Huaguo Zhou https://orcid.org/0000-0001-5210-036X
Xiaoduan Sun https://orcid.org/0000-0001-7282-1340

References

1. Centers for Disease Control and Prevention (CDC). National Center for Injury Prevention and Control (NCIPC). WISQARS (Web-Based Injury Statistics Query and Reporting System). https://www.cdc.gov/injury/wisqars/index.html.
2. Insurance Institute for Highway Safety (IIHS). Fatality Facts 2020: Teenagers. U.S. Department of Transportation, Washington, D.C., 2022. https://www.iihs.org/topics/fatality-statistics/detail/teenagers.
3. Romer, D. Adolescent Risk Taking, Impulsivity, and Brain Development: Implications for Prevention. *Developmental Psychobiology*, Vol. 52, No. 3, 2010, pp. 263–276. https://doi.org/10.1002/dev.20442.
4. Das, S., A. Dutta, X. Kong, and X. Sun. Hit and Run Crashes: Knowledge Extraction From Bicycle Involved Crashes Using First and Frugal Tree. *International Journal of Transportation Science and Technology*, Vol. 8, No. 2, 2019, pp. 146–160. https://doi.org/10.1016/j.ijtst.2018.11.001.
5. Monroe, K., E. Irons, M. Crew, J. Norris, M. Nichols, and W. D. King. Trends in Alabama Teen Driving Death and Injury. *Journal of Trauma and Acute Care Surgery*, Vol. 77, No. 3 Supplement 1, 2014, pp. S51–S54. https://doi.org/10.1097/TA.0000000000000399.
6. Mayhew, D. R., and H. M. Simpson. *New to the Road. Young Drivers and Novice Drivers: Similar Problems and Solutions?* HS-041 039. Traffic Injury Research Foundation (TIRF), Ottowa, Ontario, Canada, 1990.
7. McCartt, A. T., D. R. Mayhew, K. A. Braitman, S. A. Ferguson, and H. M. Simpson. Effects of Age and Experience on Young Driver Crashes: Review of Recent Literature. *Traffic Injury Prevention*, Vol. 10, No. 3, 2009, pp. 209–219. https://doi.org/10.1080/15389580802677807.
8. Mayhew, D. R., H. M. Simpson, and A. Pak. Changes in Collision Rates Among Novice Drivers During the First Months of Driving. *Accident Analysis & Prevention*, Vol. 35, No. 5, 2003, pp. 683–691. https://doi.org/10.1016/S0001-4575(02)00047-7.
9. Byrnes, J. P., D. C. Miller, and W. D. Schafer. Gender Differences in Risk Taking: A Meta-Analysis. *Psychological Bulletin*, Vol. 125, No. 3, 1999, pp. 367–383. https://doi.org/10.1037/0033-2909.125.3.367.
10. Shope, J. T., and C. R. Bingham. Teen Driving. Motor-Vehicle Crashes and Factors That Contribute. *American Journal of Preventive Medicine*, Vol. 35, No. 3 Supplement, 2008, pp. S261–S271. https://doi.org/10.1016/j.amepre.2008.06.022.
11. Adanu, E., P. Penmetsa, S. Jones, and R. Smith. Gendered Analysis of Fatal Crashes Among Young Drivers in Alabama, USA. *Safety*, Vol. 4, No. 3, 2018, p. 29. https://doi.org/10.3390/safety4030029.
12. McDonald, C. C., A. E. Curry, V. Kandadai, M. S. Sommers, and F. K. Winston. Comparison of Teen and Adult Driver Crash Scenarios in a Nationally Representative Sample of Serious Crashes. *Accident Analysis & Prevention*, Vol. 72, 2014, pp. 302–308. https://doi.org/10.1016/j.aap.2014.07.016.
13. Bingham, C. R., and J. P. Ehsani. The Relative Odds of Involvement in Seven Crash Configurations by Driver Age and Sex. *Journal of Adolescent Health*, Vol. 51, No. 5, 2012, pp. 484–490. https://doi.org/10.1016/j.jadohealth.2012.02.012.
14. Hossain, M. M., M. A. Rahman, X. Sun, and E. Mitran. Investigating Underage Alcohol-Intoxicated Driver Crash Patterns in Louisiana. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2675: 769–782. https://doi.org/10.1177/03611981211019742.
15. Mayhew, D. R., A. C. Donelson, D. J. Beirness, and H. M. Simpson. Youth, Alcohol and Relative Risk of Crash Involvement. *Accident Analysis & Prevention*, Vol. 18, No. 4, 1986, pp. 273–287. https://doi.org/10.1016/0001-4575(86)90042-4.
16. Voas, R. B., P. Torres, E. Romano, and J. H. Lacey. Alcohol-Related Risk of Driver Fatalities: An Update Using 2007 Data. *Journal of Studies on Alcohol and Drugs*, Vol. 73, No. 3, 2012, pp. 341–350. https://doi.org/10.15288/jsad.2012.73.341.
17. Shults, R. A., T. M. Haegerich, G. Bhat, and X. Zhang. Teens and Seat Belt Use: What Makes Them Click? *Journal of Safety Research*, Vol. 57, 2016, pp. 19–25. https://doi.org/10.1016/j.jsr.2016.03.003.
18. Peek-Asa, C., C. Britton, T. Young, M. Pawlovich, and S. Falb. Teenage Driver Crash Incidence and Factors Influencing Crash Injury by Rurality. *Journal of Safety Research*, Vol. 41, No. 6, 2010, pp. 487–492. https://doi.org/10.1016/j.jsr.2010.10.002.
19. Williams, A. F., and V. I. Shabanova. Situational Factors in Seat Belt Use by Teenage Drivers and Passengers. *Traffic Injury Prevention*, Vol. 3, No. 3, 2002, pp. 201–204. https://doi.org/10.1080/15389580213650.
20. Neyens, D. M., and L. N. Boyle. The Effect of Distractions on the Crash Types of Teenage Drivers. *Accident Analysis & Prevention*, Vol. 39, No. 1, 2007, pp. 206–212. https://doi.org/10.1016/j.aap.2006.07.004.
21. Hossain, M. M., H. Zhou, S. Das, X. Sun, and A. Hossain. Young Drivers and Cellphone Distraction: Pattern Recognition From Fatal Crashes. *Journal of Transportation Safety & Security*, 2022, pp. 1–26. https://doi.org/10.1080/19439962.2022.2048763.
22. Ouimet, M. C., A. K. Pradhan, A. Brooks-Russell, J. P. Ehsani, D. Berbiche, and B. G. Simons-Morton. Young

Drivers and their Passengers: A Systematic Review of Epidemiological Studies on Crash Risk. *Journal of Adolescent Health*, Vol. 57, No. 1, 2015, pp. S24–S35.e6. https://doi.org/10.1016/j.jadohealth.2015.03.010.
23. Weiss, H. B., S. Kaplan, and C. G. Prato. Analysis of Factors Associated With Injury Severity in Crashes Involving Young New Zealand Drivers. *Accident Analysis & Prevention*, Vol. 65, 2014, pp. 142–155. https://doi.org/10.1016/j.aap.2013.12.020.
24. Rahman, M. A., M. M. Hossain, E. Mitran, and X. Sun. Understanding the Contributing Factors to Young Driver Crashes: A Comparison of Crash Profiles of Three Age Groups. *Transportation Engineering*, Vol. 5, 2021, p. 100076. https://doi.org/10.1016/j.treng.2021.100076.
25. Chapman, E. A., S. V. Masten, and K. K. Browning. Crash and Traffic Violation Rates Before and After Licensure for Novice California Drivers Subject to Different Driver Licensing Requirements. *Journal of Safety Research*, Vol. 50, 2014, pp. 125–138. https://doi.org/10.1016/j.jsr.2014.05.005.
26. McCartt, A. T., V. I. Shabanova, and W. A. Leaf. Driving Experience, Crashes and Traffic Citations of Teenage Beginning Drivers. *Accident Analysis & Prevention*, Vol. 35, No. 3, 2003, pp. 311–320. https://doi.org/10.1016/S0001-4575(02)00006-4.
27. Paleti, R., N. Eluru, and C. R. Bhat. Examining the Influence of Aggressive Driving Behavior on Driver Injury Severity in Traffic Crashes. *Accident Analysis & Prevention*, Vol. 42, No. 6, 2010, pp. 1839–1854. https://doi.org/10.1016/j.aap.2010.05.005.
28. Duddu, V. R., V. M. Kukkapalli, and S. S. Pulugurtha. Crash Risk Factors Associated With Injury Severity of Teen Drivers. *IATSS Research*, Vol. 43, No. 1, 2019, pp. 37–43. https://doi.org/10.1016/j.iatssr.2018.08.003.
29. Adanu, E. K., A. Hainen, and S. Jones. Latent Class Analysis of Factors That Influence Weekday and Weekend Single-Vehicle Crash Severities. *Accident Analysis & Prevention*, Vol. 113, 2018, pp. 187–192. https://doi.org/10.1016/j.aap.2018.01.035.
30. Alderman, E. M., B. D. Johnston, C. Breuner, L. K. Grubb, M. Powers, K. Upadhya, S. Wallace, B. D. Hoffman, K. Quinlan, P. Agran, S. Denny, M. Hirsh, L. Lee, K. Monroe, J. Schaechter, M. Tenenbein, and M. R. Zonfrillo. The Teen Driver. *Pediatrics*, Vol. 142, No. 4, 2018. https://doi.org/10.1542/peds.2018-2163.
31. Bingham, C. R., J. T. Shope, J. E. Parow, and T. E. Raghunathan. Crash Types: Markers of Increased Risk of Alcohol-Involved Crashes Among Teen Drivers. *Journal of Studies on Alcohol and Drugs*, Vol. 70, No. 4, 2009, pp. 528–535. https://doi.org/10.15288/jsad.2009.70.528.
32. Wang, J., and G. Zhang. Modeling and Examining the Teenage and Adult Freeway Crash Risks and Injury Severities in Washington State. *Journal of Transportation Safety and Security*, Vol. 3, No. 3, 2011, pp. 207–221. https://doi.org/10.1080/19439962.2011.591518.
33. Dissanayake, S., and N. Amarasingha. Safety Experience of Young Drivers in Comparison to Experienced Drivers. In *Proc., 2nd Transportation and Development Institute Congress, T&DI Congress 2014: Planes, Trains, and Automobiles* (A. Varma and G. D. Gosling, eds.), Orlando, FL, American Society of Civil Engineers, June 8–11, 2014, pp. 344–355. https://doi.org/10.1061/9780784413586.033.
34. Mueller, A. S., and L. M. Trick. Driving in Fog: The Effects of Driving Experience and Visibility on Speed Compensation and Hazard Avoidance. *Accident Analysis & Prevention*, Vol. 48, 2012, pp. 472–479. https://doi.org/10.1016/j.aap.2012.03.003.
35. Mueller, A. S., and J. B. Cicchino. Teen Driver Crashes Potentially Preventable by Crash Avoidance Features and Teen-Driver-Specific Safety Technologies. *Journal of Safety Research*, Vol. 81, 2022, pp. 305–312. https://doi.org/10.1016/j.jsr.2022.03.007.
36. Vachal, K., R. Faculty, and D. Malchose. What can we Learn About North Dakota's Youngest Drivers From their Crashes? *Accident Analysis & Prevention*, Vol. 41, No. 3, 2009, pp. 617–623. https://doi.org/10.1016/j.aap.2009.02.014.
37. Hamann, C., M. Price, and C. Peek-Asa. Characteristics of Crashes and Injuries Among 14 and 15 Year Old Drivers, by Rurality. *Journal of Safety Research*, Vol. 73, 2020, pp. 111–118. https://doi.org/10.1016/j.jsr.2020.02.019.
38. Lin, C., D. Wu, H. Liu, X. Xia, and N. Bhattarai. Factor Identification and Prediction for Teen Driver Crash Severity Using Machine Learning: A Case Study. *Applied Science*, Vol. 10, No. 5, 2020, p. 1675. https://doi.org/10.3390/app10051675.
39. Kim, W., A. M. Svancara, and T. Kelley-Baker. Understanding the Impact of Road Design Characteristic on Teen Driver's Fatality. *Traffic Injury Prevention*, Vol. 21, No. 5, 2020, pp. 313–318. https://doi.org/10.1080/15389588.2020.1753038.
40. Wu, Q., G. Zhang, Y. Ci, L. Wu, R. A. Tarefder, and A. "Dely" Alcántara. Exploratory Multinomial Logit Model–Based Driver Injury Severity Analyses for Teenage and Adult Drivers in Intersection-Related Crashes. *Traffic Injury Prevention*, Vol. 17, No. 4, 2016, pp. 413–422. https://doi.org/10.1080/15389588.2015.1100722.
41. Hossain, M. M., X. Sun, E. Mitran, and M. A. Rahman. Investigating Fatal and Injury Crash Patterns of Teen Drivers With Unsupervised Learning Algorithms. *IATSS Research*, Vol. 45, No. 4, 2021, pp. 561–573. https://doi.org/10.1016/j.iatssr.2021.07.002.
42. Shults, R. A., G. Bergen, T. J. Smith, L. Cook, J. Kindelberger, and B. West. Characteristics of Single Vehicle Crashes With a Teen Driver in South Carolina, 2005–2008. *Accident Analysis & Prevention*, Vol. 122, 2019, pp. 325–331. https://doi.org/10.1016/j.aap.2017.08.002.
43. Carney, C., D. Mcgehee, K. Harland, M. Weiss, and M. Raby. *Using Naturalistic Driving Data to Assess the Prevalence of Environmental Factors and Driver Behaviors in Teen Driver Crashes*. *Report*. AAA Foundation for Traffic Safety, Washington DC, 2015, pp. 3–4.
44. Carney, C., K. K. Harland, and D. V. McGehee. Examining Teen Driver Crashes and the Prevalence of Distraction: Recent Trends, 2007–2015. *Journal of Safety Research*, Vol. 64, 2018, pp. 21–27. https://doi.org/10.1016/j.jsr.2017.12.014.
45. Lin, Z., and W. (David) Fan. Exploring Bicyclist Injury Severity in Bicycle-Vehicle Crashes Using Latent Class

Clustering Analysis and Partial Proportional Odds Models. *Journal of Safety Research*, Vol. 76, 2021, pp. 101–117. https://doi.org/10.1016/j.jsr.2020.11.012.
46. 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.
47. Depaire, B., G. Wets, and K. Vanhoof. Traffic Accident Segmentation by Means of Latent Class Clustering. *Accident Analysis & Prevention*, Vol. 40, No. 4, 2008, pp. 1257–1266. https://doi.org/10.1016/j.aap.2008.01.007.
48. Sun, M., X. Sun, and D. Shan. Pedestrian Crash Analysis With Latent Class Clustering Method. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 50–57.
49. Chen, Z., and W. (David) Fan. A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity in North Carolina. *International Journal of Transportation Science and Technology*, Vol. 8, No. 1, 2019, pp. 43–52. https://doi.org/10.1016/j.ijtst.2018.10.001.
50. Abrari Vajari, M., K. Aghabayk, M. Sadeghian, and N. Shiwakoti. A Multinomial Logit Model of Motorcycle Crash Severity at Australian Intersections. *Journal of Safety Research*, Vol. 73, 2020, pp. 17–24. https://doi.org/10.1016/j.jsr.2020.02.008.
51. Iranitalab, A., and A. Khattak. Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 27–36. https://doi.org/10.1016/j.aap.2017.08.008.
52. DiStefano, C.Cluster Analysis and Latent Class Clustering Techniques. In *Handbook of Developmental Research Methods* (B. Laursen, T. D. Little, and N. A. Card, eds.), The Guilford Press, New York, NY, 2012, pp. 645–666.
53. Sun, Z., Y. Xing, J. Wang, X. Gu, H. Lu, and Y. Chen. Exploring Injury Severity of Bicycle-Motor Vehicle Crashes: A Two-Stage Approach Integrating Latent Class Analysis and Random Parameter Logit Model. *Journal of Transportation Safety and Security*, Vol. 14, No. 11, 2022, pp. 1838–1864. https://doi.org/10.1080/19439962.2021.1971814.
54. Song, L., and W. Fan. Combined Latent Class and Partial Proportional Odds Model Approach to Exploring the Heterogeneities in Truck-Involved Severities at Cross and T-Intersections. *Accident Analysis & Prevention*, Vol. 144, 2020, p. 105638. https://doi.org/10.1016/j.aap.2020.105638.
55. Hennig, C., and T. F. Liao. Comparing Latent Class and Dissimilarity Based Clustering for Mixed Type Variables With Application to Social Stratification. *Journal of the Royal Statistical Society*, Vol. 62, No. 3, 2013, pp. 309–369.
56. Vermunt, J. K., and J. Magidson. Latent Class Cluster Analysis. In *Applied Latent Class Analysis* (J. A. Hagenaars, and A. L. McCutcheon, eds.), Cambridge University Press, New York, NY, 2002, pp. 89–106.
57. Liu, P., and W. Fan. 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*, Vol. 135, 2020, p. 105388. https://doi.org/10.1016/j.aap.2019.105388.
58. Dziak, J. J., D. L. Coffman, S. T. Lanza, R. Li, and L. S. Jermiin. Sensitivity and Specificity of Information Criteria. *Briefings in Bioinformatics*, Vol. 21, No. 2, 2020, pp. 553–565. https://doi.org/10.1093/bib/bbz016.
59. Chen, Y., K. Wu, X. Chen, C. Tang, and Q. Zhu. An Entropy-Based Uncertainty Measurement Approach in Neighborhood Systems. *Information Sciences*, Vol. 279, 2014, pp. 239–250. https://doi.org/10.1016/j.ins.2014.03.117.
60. Sivasankaran, S. K., and V. Balasubramanian. Exploring the Severity of Bicycle – Vehicle Crashes Using Latent Class Clustering Approach in India. *Journal of Safety Research*, Vol. 72, 2020, pp. 127–138. https://doi.org/10.1016/j.jsr.2019.12.012.
61. Wahab, L., and H. Jiang. A Multinomial Logit Analysis of Factors Associated With Severity of Motorcycle Crashes in Ghana. *Traffic Injury Prevention*, Vol. 20, No. 5, 2019, pp. 521–527. https://doi.org/10.1080/15389588.2019.1616699.
62. Ye, F., and D. Lord. Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2241: 51–58.
63. Islam, M., P. Hosseini, and M. Jalayer. An Analysis of Single-Vehicle Truck Crashes on Rural Curved Segments Accounting for Unobserved Heterogeneity. *Journal of Safety Research*, Vol. 80, 2022, pp. 148–159. https://doi.org/10.1016/j.jsr.2021.11.011.
64. Bogard, M. Marginal Effects vs Odds Ratios. *Econometrics Sense*, March 11, 2016, p. 1.
65. Washington, S., M. G. Karlaftis, F. Mannering, and P. Anastasopoulos. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall/CRC, New York, NY, 2020.
66. Stewart, T. *Overview of Motor Vehicle Crashes in 2020*. DOT HS 813 266. National Highway Traffic Safety Administration, Washington, D.C., 2022.
67. Islam, S., S. L. Jones, and D. Dye. Comprehensive Analysis of Single- and Multi-Vehicle Large Truck at-Fault Crashes on Rural and Urban Roadways in Alabama. *Accident Analysis & Prevention*, Vol. 67, 2014, pp. 148–158. https://doi.org/10.1016/j.aap.2014.02.014.
68. Adanu, E. K., W. Agyemang, R. Islam, and S. Jones. A Comprehensive Analysis of Factors That Influence Interstate Highway Crash Severity in Alabama. *Journal of Transportation Safety & Security*, Vol. 14, No. 9, 2022, pp. 1552–1576. https://doi.org/10.1080/19439962.2021.1949414.
69. Sivak, M. *Evaluation of Changes in Teenage Driver Exposure*. HLDI, 2013.
70. Yadav, A. K., and N. R. Velaga. Modelling the Relationship Between Different Blood Alcohol Concentrations and Reaction Time of Young and Mature Drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 64, 2019, pp. 227–245. https://doi.org/10.1016/j.trf.2019.05.011.
71. Viano, D. C., and C. S. Parenteau. Ejection and Severe Injury Risks by Crash Type and Belt Use With a Focus on Rear Impacts. *Traffic Injury Prevention*, Vol. 11, No. 1, 2010, pp. 79–86. https://doi.org/10.1080/15389580903479186.

72. Simons-Morton, B. G., M. C. Ouimet, R. Chen, S. G. Klauer, S. E. Lee, J. Wang, and T. A. Dingus. Peer Influence Predicts Speeding Prevalence Among Teenage Drivers. *Journal of Safety Research*, Vol. 43, No. 5–6, 2012, pp. 397–403.
73. Gong, L., and W. (David) Fan. Modeling Single-Vehicle Run-Off-Road Crash Severity in Rural Areas: Accounting for Unobserved Heterogeneity and Age Difference. *Accident Analysis & Prevention*, Vol. 101, 2017, pp. 124–134. https://doi.org/10.1016/j.aap.2017.02.014.
74. Lombardi, D. A., W. J. Horrey, and T. K. Courtney. Age-Related Differences in Fatal Intersection Crashes in the United States. *Accident Analysis & Prevention*, Vol. 99, 2017, pp. 20–29. https://doi.org/10.1016/j.aap.2016.10.030.
75. Anarkooli, A. J., and M. H. Hosseinlou. Analysis of the Injury Severity of Crashes by Considering Different Lighting Conditions on Two-Lane Rural Roads. *Journal of Safety Research*, Vol. 56, 2016, pp. 57–65. https://doi.org/10.1016/j.jsr.2015.12.003.
76. Eluru, N., and C. R. Bhat. A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity. *Accident Analysis & Prevention*, Vol. 39, No. 5, 2007, pp. 1037–1049. https://doi.org/10.1016/j.aap.2007.02.001.
77. Bham, G. H., and U. R. R. Manepalli; Mack-Blackwell Rural Transportation Center. *Identification and Analysis of High Crash Segments on Interstate, US, and State Highway Systems of Arkansas*. Final Project Report. Arkansas Highway and Transportation Department, Fayettefille, 2009, p. 68.
78. Dhungana, P., and M. Qu. The Risks of Driving on Roadways With 50 Miles per Hour Posted Speed Limit. *Journal of Safety Research*, Vol. 36, No. 5, 2005, pp. 501–504. https://doi.org/10.1016/j.jsr.2005.10.013.
79. Yu, R., and M. Abdel-Aty. Investigating the Different Characteristics of Weekday and Weekend Crashes. *Journal of Safety Research*, Vol. 46, 2013, pp. 91–97. https://doi.org/10.1016/j.jsr.2013.05.002.
80. Choi, J., S. Kim, T. Y. Heo, and J. Lee. Safety Effects of Highway Terrain Types in Vehicle Crash Model of Major Rural Roads. *KSCE Journal of Civil Engineering*, Vol. 15, No. 2, 2011, pp. 405–412. https://doi.org/10.1007/s12205-011-1124-x.
81. Tang, Z., S. Chen, J. Cheng, S. A. Ghahari, and S. Labi. Highway Design and Safety Consequences: A Case Study of Interstate Highway Vertical Grades. *Journal of Advanced Transportation*, Vol. 2018, 2018. https://doi.org/10.1155/2018/1492614.
82. Kulanthayan, S., W. K. Phang, and K. S. Hayati. Traffic Light Violation Among Motorists in Malaysia. *IATSS Research*, Vol. 31, No. 2, 2007, pp. 67–73. https://doi.org/10.1016/s0386-1112(14)60224-7.
83. Zhang, G., Y. Tan, and R. C. Jou. Factors Influencing Traffic Signal Violations by Car Drivers, Cyclists, and Pedestrians: A Case Study From Guangdong, China. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 42, 2016, pp. 205–216. https://doi.org/10.1016/j.trf.2016.08.001.
84. Sharma, A., L. Rilett, Z. M. Wu, and S. Wang. Speed Limit Recommendation in Vicinity of Signalized High-Speed Intersection. Nebraska Department of Transportation Research Report. Nebraska LTAP at DigitalCommons@University of Nebraska – Lincoln, 2012, p. 37.
85. Farmer, C. M. Reliability of Police-Reported Information for Determining Crash and Injury Severity. *Traffic Injury Prevention*, Vol. 4, No. 1, 2003, pp. 38–44. https://doi.org/10.1080/15389580309855.
86. Ābele, L., S. Haustein, M. Møller, and L. M. Martinussen. Consistency Between Subjectively and Objectively Measured Hazard Perception Skills Among Young Male Drivers. *Accident Analysis & Prevention*, Vol. 118, 2018, pp. 214–220. https://doi.org/10.1016/j.aap.2018.02.022.
87. Gibbons, R., J. Cheung, and P. Lutkevich. *The Future of Roadway Lighting*. Federal Highway Administration, Washington, D.C., 2015.
88. Jermakian, J. S. Crash Avoidance Potential of Four Passenger Vehicle Technologies. *Accident Analysis & Prevention*, Vol. 43, No. 3, 2011, pp. 732–740. https://doi.org/10.1016/j.aap.2010.10.020.
89. Eiksund, S. A Geographical Perspective on Driving Attitudes and Behaviour Among Young Adults in Urban and Rural Norway. *Safety Science*, Vol. 47, No. 4, 2009, pp. 529–536. https://doi.org/10.1016/j.ssci.2008.07.034.