TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–20

DOI: 10.1177/03611981241297976
journals.sagepub.com/home/trr
Sage

Investigating Factor Associations in Barrier Crashes through Cluster Correspondence Analysis

Rohit Chakraborty[1], Subasish Das[1], Mahmuda Sultana Mimi[1] and Boniphace Kutela[2]

Abstract

Roadside and median barriers have proven effective in preventing crashes; however, a significant number of crashes still occur that involve road barriers across United States. This study focuses on analyzing the factors related to barrier crashes across Texas. The dataset includes 63,475 crashes involving road barriers and covering six years of crash data (2017–2022). Using cluster correspondence analysis (CCA) data mining approach, the study identifies six different clusters of crashes. Each cluster identifies factors contributing to barrier-related crashes. Before implying the CCA, variable importance analysis is also performed to identify the significance of the variables. The analysis highlighted that dry road surface conditions and clear weather are highly associated with high-speed crashes. Driver distraction and absence of traffic control devices can attribute to the crashes on roads with lower speed limits. Moreover, along with other factors, adverse weather conditions are also found to be a contributing factor that can influence the crash frequency and type of crashes. The analysis also implies that non-complex crash type related to a single vehicle are highly correlated with barrier crashes. The study concludes by making several policy-making implications to assist the transportation planners in reducing the frequency and severity of barrier-related crashes.

Keywords

safety, safety performance and analysis, crash analysis, crash data, crash frequency, crash severity, data mining

Traffic barriers, such as guardrails, cable barriers, and concrete barriers, serve to protect vehicles from obstacles including trees, poles, and steep slopes. However, crashes with these barriers account for a significant proportion of road-related injuries and fatalities. Data from the Fatality Analysis Reporting System (FARS) and the National Highway Traffic Safety Administration (NHTSA) indicated that, in 2021, the United States witnessed 8,884 fatalities resulting from collisions with fixed objects, marking a 3% increase from 2020 (*1*). Of these fixed object crashes in 2021, collisions with trees and utility poles were notably prevalent, representing 45% and 11% of the total crash-related fatalities, respectively. Traffic barriers emerged as the third most common object involved in these incidents, contributing to 9% (823) of the deaths. There are three potential outcomes when a vehicle leaves the roadway and hits a median or roadside barrier: (i) vehicle containment, where the barrier successfully holds the vehicle, indicating a successful or non-crossover event; (ii) vehicle redirection, where the vehicle is redirected back onto the road after hitting the barrier, also indicating a successful or non-crossover event; and (iii) barrier penetration, where the vehicle breaks through, goes under, or vaults over the barrier, signifying a crossover event or a failure of the barrier (*2*).

However, the types of barriers and associated crashes are being extensively studied. According to a study, barriers such as concrete and cable guardrails increased the overall number of crashes as a result of collisions with

[1]Department of Civil Engineering, Texas State University, San Marcos, TX
[2]Texas A&M Transportation Institute, Bryan, TX

Corresponding Author:
Rohit Chakraborty, rohitchakraborty@txstate.edu

the barrier itself, although they are effective in reducing more severe cross-median crashes and rollovers (*3*). In addition, the presence of median barriers was associated with a higher number of barrier-related crashes. Another study found that light vehicles are more susceptible to severe crashes with rigid barriers (*4*). These insights underscore the significant role that transportation infrastructure plays in road safety and the critical need for ongoing evaluation and enhancement of traffic barrier systems. It was also found that crashes involving cable barriers tended to result in less severe consequences than crashes with other types of barriers and the distance between the median barrier and the roadway influenced the likelihood of sustaining severe injuries (*5*). Another study found that injury outcomes were associated with crash parameters such as collision speed, angle, and potential barrier penetration (*6*). Injury and fatal crashes can be reduced by increasing barrier offset (*5*, *7*), whereas more lanes and wider medians correlated with lower crash frequencies. These studies have evaluated various types of barriers, examining their potential for crashes, and exploring how enhancements in their characteristic factors can mitigate crash severities and prevent injuries. Nonetheless, there has been limited research into how other associated factors influence the severity of crashes involving barriers.

This study examines an often-overlooked aspect of road safety: barrier-related crashes. Despite the effectiveness of roadside and median barriers in averting crashes, a considerable number of incidents still occur nationwide. Focused specifically on Texas, this study meticulously analyzes factors contributing to such crashes. This study aims to bridge the current research gap by investigating the potential influencing factors through a unique data mining approach known as cluster correspondence analysis (CCA). Previously, researchers adopted several cluster analysis techniques to identify the relevant factors in crash safety analysis (*8–10*). Through systematic analysis, this study generates clusters that include meaningful grouping of crash observations. Using CCA, influential factors associated with barrier crashes were determined. The rest of the paper is structured in the following manner: the literature review section explores an in-depth examination of literature related to barrier crashes; the methodology section outlines the procedures undertaken to prepare the data for analysis, offers a comprehensive explanation of the CCA methodology employed in this research, and the results and discussions are explained in the respective section. Finally, this study concludes by summarizing the investigative findings and providing policy-making implications.

Literature Review

Most barrier-crash analysis has focused on reducing crash severity and evaluating the initiating factors. Several studies analyzed the factors that are associated with the crashes (*3*, *6*, *7*, *11*). One study examined wire rope safety barrier crashes leading to fatal and serious injuries, utilizing crash datasets and computer simulations to investigate contributing factors (*6*). The results indicated that injury outcomes are linked to crash parameters such as collision speed, angle, and potential barrier penetration. Another study analyzed factors influencing median-related crashes using statistical modeling with multi-year data (*7*). According to the study, significant factors included lane number, differential elevation, and cable barrier offset. In addition, higher traffic volume, curves, and greater elevation between lanes increased crash frequency. Several studies identified that increased barrier offset reduced injury and fatal crashes (*5*, *7*), while more lanes and wider medians correlate with lower crash frequencies. A study that analyzed motorcycle-to-barrier crash frequency on curved roads using police-reported data found that curve radius emerged as the strongest predictor, suggesting a criterion for countermeasure placement (*11*). It was also found that curves meeting the existing criterion of 820 ft or less increased crash frequency rate by 10 times. A study developed crash modification factors (CMFs) and estimated average crash costs for different road barrier scenarios like concrete barriers, W-beam guardrails, and high-tension cable barriers (*3*). The study analyzed how barriers affected crash frequency, type, and severity. Their results showed barriers increased total crashes attributed to collisions but effectively reduced cross-median crashes, rollovers, and collisions with hazards. Another study analyzed barrier-related crashes over seven years on major Alabama highways, examining various risk factors' impact on injury severity and barrier-hit outcomes (*2*). The results highlighted the association of barrier penetration, female drivers, and driver fatigue with higher injury probabilities. Furthermore, they found that longer barrier lengths and specific barrier types influenced barrier-hit outcomes. They suggested designing longer barrier run lengths that may reduce barrier penetration likelihood. A group of researchers analyzed five years of rural divided highway data, using a nested logit model to estimate median barrier-crash severity (*5*). They found that collisions with cable median barriers were associated with less-severe outcomes compared with concrete or guardrail barriers.

A study presented the development of a motorcycle barrier and evaluated crash performance to enhance road safety by redirecting motorcyclists during collisions (*12*). The results aided in addressing growing motorcycle safety concerns within road restraint systems. Another study analyzed fatal motorcycle collisions with roadside barriers, aiming to understand crash mechanics and inform crash test protocols including barrier and motorcycle types, crash postures, kinematics, speeds, collision angles, and energy dissipation (*13*). In a further study, a novel crash box system was explored for frontal vehicle crashes, incorporating energy absorption from a pre-inverted pipe to counter external inversion forces (*14*). A group of researchers evaluated roadside restraint systems' safety focusing on occupant injury metrics by analyzing 33 barriers (*15*). Their results suggested the need for improved testing protocols considering actual head and neck data to accurately assess occupant safety in barrier collisions.

Crespo et al. aimed to address differences between past and current Euro New Car Assessment Program (NCAP) protocols, proposing countermeasures including redesign of sill and B-pillars and optimization of door beams (*16*). Their results showed significant reductions in pelvic forces, offering a cost-effective strategy for enhancing occupant protection in mid-sized cars. A study assessed the performance of a central median wire rope barrier (WRB) and found that even with design compromises, a well-installed and maintained WRB system could effectively reduce crash severity on highways (*17*). Another study was conducted on full-scale crash tests involving barriers and vehicle types to evaluate their relative performance and assess capacity, deflection, energy absorption, and collision severity (*18*).

Several studies conducted crash severity analysis of different barriers using statistical models. One study used crash simulations to assess the structural crash probability and potential occupant injuries of an off-road utility vehicle (OUV) including body structure, front bumper, chassis, powertrain, fuel tank, seats, and container (*19*). Another study discussed electronic stability control (ESC), a vehicle safety system aimed at preventing loss-of-control crashes by keeping vehicles on the intended path (*20*). It examined the impact of ESC on fatal crashes involving roadside barriers, finding that ESC significantly reduced the likelihood of such crashes for both cars and light trucks/vans.

A study investigated the impact of traffic barrier geometric characteristics on crashes on non-interstate roads in Wyoming (*21*). Data from 137 mi of barriers, including height, side-slope rate, post-spacing, and lateral offset, were analyzed alongside crash reports from 2008 to 2017. Their results showed that specific barrier configurations were associated with varying crash frequencies and severity. Another study investigated the effect of different variables on the severity of crashes involving traffic barriers (*22*). Traffic barriers with a height between 28 and 31 in. were identified as safer, while end treatments located nearer to the traffic lane exhibited lower crash severity. A similar study investigated factors affecting crash severity across different traffic barrier and vehicle types through ordinal logistic regression (*4*). They found that guardrail segments had higher severity crashes with speeds over 55 mph, whereas cable barrier crashes were less-severe in high-speed areas. Light vehicles were more prone to severe crashes with rigid barriers, whereas cable and rigid barriers showed no significant difference in truck crash severity.

Rezapour and Ksaibati used structural equation modeling (SEM) to understand the complex relationships affecting traffic barrier crash severity, considering human, environmental, and road/traffic barrier characteristics (*23*). They found collision force as a key factor influencing crash severity. Another study investigated the optimization of roadside barrier heights to mitigate run-off-the-road crashes (ROR), which historically contributed to fatalities and severe injuries on highways (*24*). Their results indicated substantial benefits in crash reduction. The effectiveness of traffic barriers in mitigating ROR was examined in another study, acknowledging the necessity of considering crash frequency and severity (*25*). Results revealed significant contributions of factors such as rollover and driving under the influence on barrier crash costs, particularly at higher quantiles. A similar study aimed to reduce severe vehicle crashes by analyzing factors affecting traffic barrier crash severity using a Bayesian hierarchical model (*26*). Factors such as citation record, curve negotiation, non-compliance with speed limits, alcohol involvement, and emotional signs increased crash severity, while factors like turning before collision, younger age, and adverse weather reduced severity. In another study, Zulkiffli et al. (*27*) assessed crash severity on rural expressways depending on median barrier types. They focused on fatal and non-fatal crashes and analyzed data from seven expressways in Malaysia from 2009 to 2011. Their findings suggested that crashes involving w-beam guardrails were more likely to result in fatalities compared with other barrier types. Earlier research did not focus the diverse factors related to crashes involving roadside barriers or median barriers. To gain a deeper understanding of the underlying factors and patterns associated with barrier crashes, it is essential to examine the factors related to speed, weather conditions, vehicle stability, and intersection that may influence barrier crashes. This study aims to fill this research gap by exploring the potential factors that are associated with such crashes.

This study adopted CCA, which is a unique data mining approach to assess factors related to barrier crashes. This method has been widely used in studies for crash analysis. Das et al. used the CCA to investigate motor scooter-related fatal crashes (*28*). Another study adopted the CCA to study 3,381 crashes occurred during rainy weather in Louisiana (*29*). Similarly, CCA method was used to evaluate the associated factors contribute to vehicle encroachment related crashes in another study (*30*).

These studies have demonstrated the effectiveness of this method to identify crucial factors that lead to crashes and explain the underlying relationships. Therefore, this research employed CCA to investigate the influential attributes of barrier crashes and provide insights into the relationships among these attributes. In addition, CCA offers distinct advantages over traditional regression models in analyzing crashes. Unlike regression models, which assume linear relationships between variables, CCA is a data mining technique that can uncover complex patterns and relationships within the dataset. By identifying distinct clusters of crashes, CCA provides a clear understanding of the factors contributing to barrier-related incidents, allowing for more targeted interventions.

Methodology

The research methodology of this study is described in this section. Before implementing the CCA, data preparation, variable importance analysis, and exploratory data analysis were performed. The adopted methodology is described in the following sub-sections.

Data Preparation

The data set that has been used in this study includes 63,745 barrier crashes that occurred between 2017 and 2022 across Texas in the United States. Recent years of crash data (2017–2022) were specifically chosen for analysis to provide the most up-to-date and relevant information on barrier related crash trends and patterns. By focusing on this time frame, the study captures recent changes and developments in road safety, including any emerging trends or shifts in crash dynamics. Initially, the dataset includes sixteen variables that are related to barrier crashes. Based on this dataset, variable importance analysis was performed to finalize the variables that have a higher likelihood of influencing the crashes. Figure 1 illustrates the procedure that was followed in data preparation and execution of correspondence analysis on the prepared data in this research.

The COVID-19-related analysis was not conducted in this study and is recommended as a potential avenue for future research. Table 1 presents crash data from 2017 to 2022, showing a relatively stable number of crashes across the years, including during the pandemic (2020–2021). Despite possible reductions in traffic volume during the pandemic, no significant changes in crash numbers were observed. This suggests that the overall pattern of crashes, particularly barrier-related incidents, may not have been heavily affected by the pandemic. The color gradient in the table represents the frequency of crashes categorized by severity levels and years, emphasizing that the majority of fatal and injury crashes occurred in 2021 and 2022.

Variable Importance Analysis

The variable importance analysis is a critical step in understanding the factors that most influence crash severity outcomes. By ranking the variables based on their contributions, this analysis allows us to identify the factors that have the greatest effect on model accuracy and prediction. To perform this, this study utilized two R packages 'vip' and 'xgboost' (*31*, *32*). The current study used the XGBoost classification algorithm, which constructs an ensemble of decision trees in a sequential manner, where each subsequent tree attempts to correct the errors of its predecessors. The model optimizes a loss

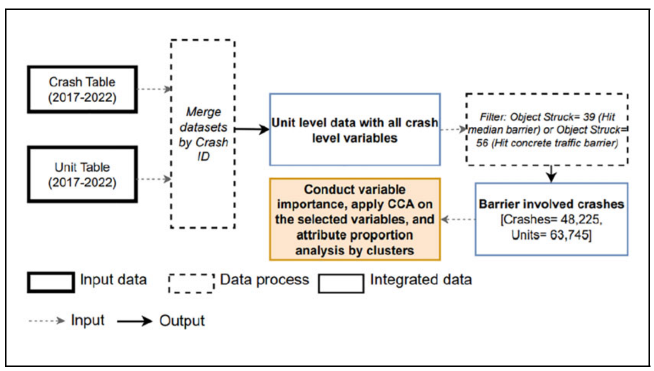

Figure 1. Flow chart of study design.

Table 1. Number of Crashes by Year and Crash Severity Type

Severity level	2017	2018	2019	2020	2021	2022	Total
Fatal	136	170	126	171	216	195	1,014
Incapacitating injury	428	357	350	411	518	542	2,606
Non-incapacitating	1,423	1,174	1,173	1,187	1,519	1,630	8,106
Possible injury	1,580	1,595	1,698	1,703	1,740	1,452	9,768
Not injured	6,793	6,444	6,698	6,994	7,969	7,353	42,251
Total	10,360	9,740	10,045	10,466	11,962	11,172	63,745

Note: Color online only.

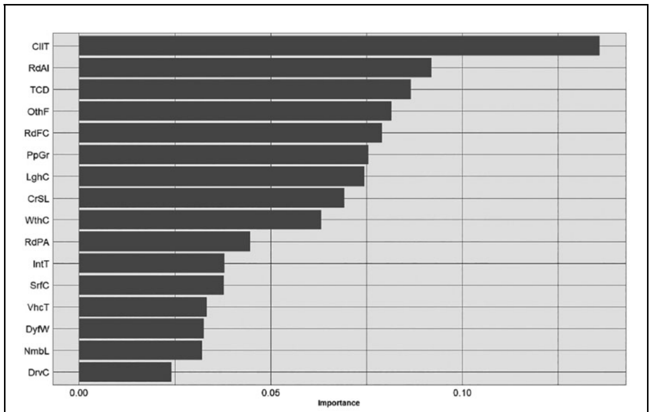

Figure 2. Variable importance plot.

function using gradient boosting, which enhances predictive performance by reducing bias and variance. Figure 2 illustrates the results of a variable importance analysis, revealing the relative influence of each predictor on the model performance. In this chart, the variable "ClIT," which represents "Collision With," emerges as the most influential variable with an importance score of approximately 0.136. This indicates a strong predictive ability of this variable within the model. Following "ClIT," "RdAl," or "Roadway Alignment," is the second most significant variable, with an importance score near 0.092. "TCD," which denotes "Traffic Control Device" also shows a notable level of importance at roughly 0.087. Other variables with a substantial impact include "OthF" (Other Contributing Factors) and "RdFC" (Roadway Functional Class), with importance scores of about 0.082 and 0.079, respectively. On the lower end of the importance value, importance scores of "VhcT" (Vehicle Type), "DyfW" (Day of the Week), "Nmbl" (Number of Lane), and "DrvC" (Driver Condition) keep descending from roughly 0.033 to 0.025. Although these variables have some impact, their lower scores indicate they are less pivotal in the predictive process compared with the others. Therefore, the variables with an importance score lower than 0.035 are removed in further analysis. Finally, the current study included the top 12 variables from the variable importance analysis chart.

Exploratory Data Analysis

Table 2 presents descriptive statistics of barrier crash data, highlighting the relationship between injury severity and a variety of contributing factors. The data show that straight-moving single motor vehicles (Str-OMV) predominantly feature in crashes where no injuries are reported, making up 77.1% of such incidents, in comparison with a lower percentage of 57.9% in fatal outcomes. This disparity is more evident with rear-end collisions involving two motor vehicles (REnd-TMV), which are more common in fatal crashes at 10.6% than in incidents where no injuries occur at 5.36%. It suggests that the type of crash plays a significant role in determining the severity of injuries. The analysis of population density in the crash locations reveals that rural areas have a disproportionate share of fatalities at 43.7%, whereas they are less represented in crashes that result in no injuries at 31.1%. In contrast, areas with over 250,000 people have a lower percentage of fatal outcomes at 30.6% compared with their share of crashes with no injuries at 34.4%, possibly indicating the impact of better safety measures in more densely populated areas.

Crashes on straight, level roads are common across all levels of injury severity, but there is a higher occurrence of these in incidents without injuries (59.2%) than in those with fatalities (50.6%). This suggests that crashes on roads with more complex alignments may have a greater chance of resulting in severe injuries. The speed limit appears to be a critical factor in crash severity. Higher speed limits, particularly ranging from 60 to 70 mph, correspond to most fatalities at 59.2% however, they are less frequently linked to crashes where no injuries are sustained (46.1%). This highlights the increased danger of fatal consequences when crashes occur at higher speed. Weather conditions during crashes also seem to play a role, with clear conditions present in most no-injury incidents (63.6%) and a slightly higher percentage in fatal outcomes (68.2%). However, rainy conditions account for a more substantial proportion of no-injury crashes (18.9%) compared with fatal ones (13.4%), indicating that rain contributes to crash occurrences but not necessarily to an increase in fatalities. Surface conditions reflect a similar pattern, with dry conditions most reported across all injury severities. However, wet conditions are observed more frequently in no-injury crashes (23.2%) compared with fatalities (18.6%), which suggests that although wet conditions are hazardous, they may not directly increase the risk of fatalities.

Considering intersection types, non-intersection crashes make up a larger proportion of no-injury crashes (89.4%) compared with fatal crashes (93.5%). This marks intersections as potential zones of higher risk for severe injuries. The p-values are less than 0.001 for these variables, which indicate that the differences observed in the data are statistically significant and not random. This analysis underscores the complexity of factors that lead to traffic crashes and the profound effect they have on the severity of the outcomes.

Cluster Correspondence Analysis

CCA is recognized as a developed method for analyzing categorical data. CCA investigates the connections

Table 2. Percentage Distribution of Key Attributes by Injury Severity

	Fatal ($N=1,014$)	Incapacitating injury ($N=2,606$)	Non-incapacitating ($N=8,106$)	Not injured ($N=42,251$)	Possible injury ($N=9,768$)	p.overall
Collision type (CllT)						0.000
RearEnd-Two motor vehicle (REnd-TMV)	107 (10.6%)	306 (11.7%)	881 (10.9%)	2,266 (5.36%)	1,129 (11.6%)	
Sideswipe-Two motor vehicle (Sswipe-TMV)	95 (9.37%)	193 (7.41%)	1,052 (13.0%)	4,053 (9.59%)	1,499 (15.3%)	
Straight-One motor vehicle (Str-OMV)	587 (57.9%)	1681 (64.5%)	5,227 (64.5%)	32,584 (77.1%)	6,075 (62.2%)	
Turning-One motor vehicle (Turn-OMV)	8 (0.79%)	28 (1.07%)	130 (1.60%)	1,623 (3.84%)	160 (1.64%)	
Other	217 (21.4%)	398 (15.3%)	816 (10.1%)	1725 (4.08%)	905 (9.26%)	
Population group (PpGr)						<0.001
250,000 +	310 (30.6%)	776 (29.8%)	2,887 (35.6%)	14,517 (34.4%)	4,016 (41.1%)	
100,000–249,999	107 (10.6%)	239 (9.17%)	1,044 (12.9%)	4,886 (11.6%)	1,340 (13.7%)	
25,000–49,999	28 (2.76%)	167 (6.41%)	460 (5.67%)	2,464 (5.83%)	541 (5.54%)	
Rural	443 (43.7%)	991 (38.0%)	2,376 (29.3%)	13,131 (31.1%)	2,316 (23.7%)	
Other	126 (12.4%)	433 (16.6%)	1,339 (16.5%)	7,253 (17.2%)	1,555 (15.9%)	
Road alignment (RdAl)						<0.001
Curve, grade	141 (13.9%)	298 (11.4%)	762 (9.40%)	4136 (9.79%)	781 (8.00%)	
Curve, level	156 (15.4%)	296 (11.4%)	842 (10.4%)	5,023 (11.9%)	994 (10.2%)	
Straight, grade	115 (11.3%)	293 (11.2%)	1034 (12.8%)	5,171 (12.2%)	1,231 (12.6%)	
Straight, level	513 (50.6%)	1555 (59.7%)	4,954 (61.1%)	25,009 (59.2%)	6052 (62.0%)	
Other	89 (8.78%)	164 (6.29%)	514 (6.34%)	2,912 (6.89%)	710 (7.27%)	
Other factors (OthF)						<0.001
Attention diverted	91 (8.97%)	429 (16.5%)	1,520 (18.8%)	7,372 (17.4%)	1,748 (17.9%)	
Lane change	48 (4.73%)	119 (4.57%)	700 (8.64%)	2428 (5.75%)	931 (9.53%)	
Lost control (LostCtrl)	78 (7.69%)	188 (7.21%)	644 (7.94%)	5,303 (12.6%)	835 (8.55%)	
Other	204 (20.1%)	544 (20.9%)	1,873 (23.1%)	9,660 (22.9%)	2,400 (24.6%)	
Not applicable	593 (58.5%)	1326 (50.9%)	3,369 (41.6%)	17,488 (41.4%)	3,854 (39.5%)	
Crash speed limit (CrSL)						<0.001
<30 mph	21 (2.07%)	160 (6.14%)	466 (5.75%)	3,253 (7.70%)	609 (6.23%)	
>75 mph	168 (16.6%)	394 (15.1%)	943 (11.6%)	5,069 (12.0%)	897 (9.18%)	
35–55 mph	225 (22.2%)	787 (30.2%)	2,511 (31.0%)	14,441 (34.2%)	3097 (31.7%)	
60–70 mph	600 (59.2%)	1,265 (48.5%)	4,186 (51.6%)	19,488 (46.1%)	5,165 (52.9%)	
Road functional class (RdFC)						<0.001
Rural interstate (RurInt)	273 (26.9%)	629 (24.1%)	1,896 (23.4%)	9,488 (22.5%)	2,276 (23.3%)	
Rural local (RurLoc)	289 (28.5%)	909 (34.9%)	2,643 (32.6%)	14,861 (35.2%)	2,904 (29.7%)	
Rural principal arterial (RurPrinArt)	139 (13.7%)	285 (10.9%)	995 (12.3%)	4,800 (11.4%)	1,119 (11.5%)	
Urban interstate (UrbInt)	142 (14.0%)	338 (13.0%)	1,104 (13.6%)	5,995 (14.2%)	1,714 (17.5%)	
Other	171 (16.9%)	445 (17.1%)	1,468 (18.1%)	7,107 (16.8%)	1,755 (18.0%)	
Lighting condition (LghC)						<0.001
Dark, lighted	243 (24.0%)	630 (24.2%)	2,041 (25.2%)	11,164 (26.4%)	2,305 (23.6%)	
Dark, not lighted	336 (33.1%)	626 (24.0%)	1508 (18.6%)	8,510 (20.1%)	1,508 (15.4%)	
Dawn	14 (1.38%)	33 (1.27%)	125 (1.54%)	607 (1.44%)	112 (1.15%)	
Daylight	394 (38.9%)	1266 (48.6%)	4289 (52.9%)	20,912 (49.5%)	5,644 (57.8%)	
Other	27 (2.66%)	51 (1.96%)	143 (1.76%)	1,058 (2.50%)	199 (2.04%)	

(continued)

Table 2. (continued)

	Fatal (*N* = 1,014)	Incapacitating injury (*N* = 2,606)	Non-incapacitating (*N* = 8,106)	Not injured (*N* = 42,251)	Possible injury (*N* = 9,768)	p.overall
Traffic control device (TCD)						<0.001
Center stripe/divider	151 (14.9%)	320 (12.3%)	903 (11.1%)	4,716 (11.2%)	995 (10.2%)	
Marked lanes	637 (62.8%)	1,610 (61.8%)	4,994 (61.6%)	24,920 (59.0%)	6,098 (62.4%)	
No passing zone	75 (7.40%)	116 (4.45%)	280 (3.45%)	1,431 (3.39%)	249 (2.55%)	
Other	78 (7.69%)	245 (9.40%)	762 (9.40%)	3,570 (8.45%)	889 (9.10%)	
None	73 (7.20%)	315 (12.1%)	1,167 (14.4%)	7,614 (18.0%)	1,537 (15.7%)	
Weather condition (WthC)						<0.001
Clear	692 (68.2%)	1,885 (72.3%)	5,602 (69.1%)	26,879 (63.6%)	6,523 (66.8%)	
Cloudy	162 (16.0%)	389 (14.9%)	1,259 (15.5%)	5,771 (13.7%)	1,538 (15.7%)	
Fog	11 (1.08%)	17 (0.65%)	69 (0.85%)	431 (1.02%)	93 (0.95%)	
Rain	136 (13.4%)	285 (10.9%)	1,054 (13.0%)	7,998 (18.9%)	1,480 (15.2%)	
Other	13 (1.28%)	30 (1.15%)	122 (1.51%)	1,172 (2.77%)	134 (1.37%)	
Road part (RdPA)						<0.001
Entrance/on ramp	25 (2.47%)	62 (2.38%)	241 (2.97%)	1569 (3.71%)	328 (3.36%)	
Exit/off ramp	25 (2.47%)	107 (4.11%)	424 (5.23%)	2,390 (5.66%)	469 (4.80%)	
Main/proper lane	896 (88.4%)	2,233 (85.7%)	6,601 (81.4%)	33,374 (79.0%)	7,866 (80.5%)	
Other	17 (1.68%)	28 (1.07%)	109 (1.34%)	717 (1.70%)	130 (1.33%)	
Unknown	51 (5.03%)	176 (6.75%)	731 (9.02%)	4,201 (9.94%)	975 (9.98%)	
Surface condition (SrfC)						<0.001
Dry	803 (79.2%)	2,144 (82.3%)	6483 (80.0%)	30,085 (71.2%)	7588 (77.7%)	
Ice	5 (0.49%)	27 (1.04%)	113 (1.39%)	1,178 (2.79%)	126 (1.29%)	
Standing water	4 (0.39%)	18 (0.69%)	90 (1.11%)	673 (1.59%)	131 (1.34%)	
Wet	189 (18.6%)	393 (15.1%)	1,382 (17.0%)	9,823 (23.2%)	1,866 (19.1%)	
Other	13 (1.28%)	24 (0.92%)	38 (0.47%)	492 (1.16%)	57 (0.58%)	
Intersection type (IntT)						<0.001
Driveway access	2 (0.20%)	23 (0.88%)	83 (1.02%)	415 (0.98%)	100 (1.02%)	
Intersection (Int)	37 (3.65%)	156 (5.99%)	448 (5.53%)	790 (1.87%)	512 (5.24%)	
Intersection related (IntRel)	27 (2.66%)	121 (4.64%)	445 (5.49%)	3,278 (7.76%)	545 (5.58%)	
Non-intersection (NonInt)	948 (93.5%)	2306 (88.5%)	7130 (88.0%)	37,768 (89.4%)	8,611 (88.2%)	

between categorical variables (*33*). This approach is designed for the examination of categorical data, with its main goal being the formation of significant clusters within the dataset based on a defined group of observable variables. Specifically, CCA focuses on identifying cluster allocations and scaling figures for categorical variables, aiming to enhance the variance between groups. The CCA method was first introduced by Van de Velden et al. (*34*). This method is created by joining correspondence analysis and K-means cluster analysis. This obtains a cluster allocation as well as optimal scaling values (i.e., coordinates) for the categories of p categorical variables maximizing variation between the groups. This was achieved by considering the cross-tabulation of cluster memberships by variable categories and determining both the cluster allocation and the category weights (i.e., the coordinates in reduced space) that simultaneously maximize between clusters and the between categories variances. The method produces simple visualizations that enable the interpretation of a standard biplot. The following discussion will explore the mathematical algorithm utilized in CCA.

Initially, a normal data matrix X that has n observations containing q number of categorical variables needs to be converted to a new matrix Z which is named as super indicator matrix. This matrix will be generated through a one-hot encoding process which transforms each categorical variable to a binary matrix, that is, $Z = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, ..., \boldsymbol{Z}_q]$, where Zj is an $n \times pj$ matrix of the encoded j-categorical variable with pj number of categories. The indicator matrix Z has the same n number of rows and $Q = \sum_{j=1}^{q} Pj$ number of columns. One can define Zk as a $n \times K$ binary matrix indicating the memberships of each observation into the K number of clusters. To consider the clusters' relationship with categorical variables, cross-tabulation of indicator matrix and membership matrix will be constructed as a $K \times pj$ matrix, that is, $F = \boldsymbol{Z}_K^{\mathrm{T}}\boldsymbol{Z}$. By applying CCA to contingency matrix F, scaling values corresponding to clusters and categories that maximize the inter-group variances will be optimized. In the optimal condition, the clusters separating the observations in a way that maximum variances from distributions over categorical variables and simultaneously distributions of categories in each variable will be obtained.

The CCA procedure starts by randomly distributing observations into clusters, forming an initial membership matrix Z_K and contingency matrix F. Following this, correspondence analysis is applied to matrix F to derive the category quantifications matrix B. Furthermore, the object coordinate matrix $\boldsymbol{Y} = \frac{1}{q}\left(\boldsymbol{I}_n - \frac{\boldsymbol{1}_n\boldsymbol{1}_n'}{n}\right)\boldsymbol{ZB}$ is calculated. Finally, k-means clustering is applied to Y and Zk is continuously being updated. The process is repeated until Zk values converge to a fixed number (*34*).

Table 3. Eigenvalue and Percentage of Variance by the First Ten Dimensions

Dim	Eigenvalue	% of variance	Cumulative % of variance
dim 1	0.2314	1.4894	1.4894
dim 2	0.2072	1.3333	2.8227
dim 3	0.1621	1.0431	3.8658
dim 4	0.1582	1.0184	4.8842
dim 5	0.1541	0.9916	5.8759
dim 6	0.1375	0.8850	6.7609
dim 7	0.1354	0.8714	7.6323
dim 8	0.1326	0.8534	8.4857
dim 9	0.1307	0.8414	9.3271
dim 10	0.1272	0.8188	10.1459

Note: Dim = dimension or axis.

The resulting Zk gives G which is the optimum cluster centroid matrix and category quantification matrix B. The coordinate matrix G, B is used to present a biplot of the clusters and categories. However, to facilitate the interpretation of the biplots, the two matrices will be scaled by a constant value of $\boldsymbol{\gamma} = \left(\frac{K}{Q}\cdot\frac{TrB^TB}{TrG^TG}\right)^{1/4}$. The new coordinate matrixes $\boldsymbol{G_s} = \boldsymbol{\gamma G}$, and $\boldsymbol{B_s} = \frac{1}{\gamma}\boldsymbol{B}$ have the same average squared deviation from the origin that will be used for biplot presentations of analysis (*34*).

Results and Discussions

This study applied CCA on the barrier crash data. In this study, the eigenvalues for the first 10 dimensions (see Table 3) are considered. Determination of the number of dimensions is the first step of this method. On analyzing the scree plot (see Figure 3), the current study considered the first two dimensions based on the cumulative variance explained, as it is a common practice to retain dimensions that together account for a significant portion of the total variance (*35–37*). In this analysis, the first two dimensions explain approximately 28.23% of the total variance, which is deemed sufficient for a robust interpretation (see Table 3). The scree plot illustrates the distribution of eigenvalues across the dimensions of the analysis, with the y-axis representing the eigenvalue (variance explained) and the x-axis showing the dimensions. Figure 3 illustrates a sharp decline in eigenvalues after the first two dimensions, indicating that dimensions 1 and 2 explain the most significant proportion of the variance in the barrier crash data. These two retained dimensions capture the key relationships within the barrier crash data. The first dimension (dim 1) primarily represents high-speed rural interstate crashes, as it is heavily

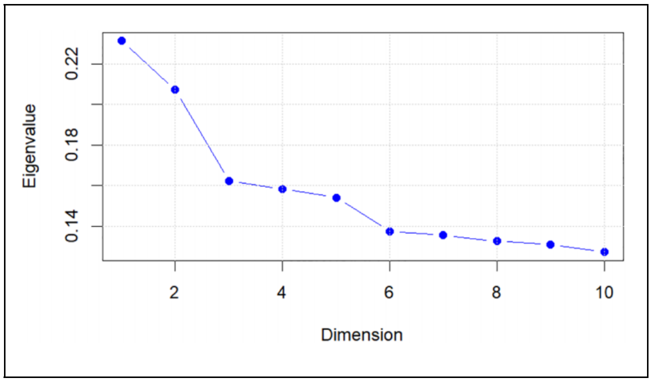

Figure 3. Scree plot on eigenvalues.

associated with variables such as road surface condition (e.g., dry surfaces) and crash speed limits (e.g., 60–70 mph). Crashes in this dimension typically occur under favorable weather conditions (e.g., clear weather), suggesting that road characteristics and speed limits play a dominant role in crash outcomes. The second dimension (dim 2), by contrast, is composed mainly of variables that reflect adverse weather conditions. This dimension highlights crashes occurring in the presence of rain, wet road surfaces, or standing water. These crashes, while less frequent in the overall dataset, represent a significant subset where environmental conditions are the primary contributing factors. The curve starts to flatten from dimension 3 onward, suggesting that additional dimensions contribute progressively smaller amounts of variance. This pattern, often referred to as the 'elbow,' suggests that retaining the first two or three dimensions would be most effective, as they capture the bulk of the information, while further dimensions add diminishing value. Thus, the scree plot justifies focusing on the first few dimensions in the analysis. In addition, this cutoff ensures that the dimensionality is kept manageable for interpretation while retaining the most important information.

The second step is to determine the optimum number of clusters. The determination of the optimal number of clusters is essential to ensure that the clusters are both interpretable and meaningful. A common approach to this is using centroids, which represent the central position of each cluster. The process of identifying the optimal number of clusters involves an iterative approach, where different numbers of clusters are tested through trial and error. Each iteration evaluates the ability of the clusters to capture significant patterns within the dataset. The effectiveness of each clustering solution is assessed based on its interpretability and relevance to the research question. This process continues until a small number of clusters is identified that not only maximizes the clarity of the results but also balances the trade-off between the complexity of the model and the insight it provides.

The CCA method was employed to investigate the influencing factors that are responsible for the barrier crashes. It was implemented using the 'clustrd' package in the R software is in this analysis (*34*). Table 4 provided outlines the results from a cluster analysis of vehicle crashes involving barriers, capturing a total of 63,745 incidents. The table presents six clusters from the barrier crash analysis, each described by its position (centroids) in two key dimensions (dim 1 and dim 2), the spread of data within each cluster (within-cluster sum of squares), and the number of crashes in each cluster (size). Larger clusters, like cluster 1 and 2, have smaller variation, meaning crashes in these groups are more similar, whereas smaller clusters, like cluster 6, have higher variation, showing more diverse crash patterns. The centroids indicate where each cluster lies relative to the others in the dataset.

The CCA scores illustrate the relative positions of categories along the dimensions. These scores are indicative of the similarity between categories, influencing their approximate placements. Figure 4 displays a scatter plot derived from vehicle crash data involving barriers. The plot is spread across two dimensions, labeled as dim 1 and dim 2, representing different attributes or factors analyzed in the crash data. The individual points on the plot represent specific crashes, and they are grouped around labeled points which indicate the centers of

Table 4. Centroids and Size of the Clusters

	Barrier crashes			
Cluster	Dim 1	Dim 2	Within cluster sum of squares	Size
Cluster 1	−0.0002	0.0023	0.0118	25,929
Cluster 2	−0.0014	0.0007	0.0105	15,012
Cluster 3	0.0028	−0.0014	0.0249	9,062
Cluster 4	0.0059	−0.0037	0.0170	6,327
Cluster 5	−0.0037	−0.0028	0.0178	4,826
Cluster 6	−0.0074	−0.0075	0.0143	2,589

Note: Dim = dimension or axis.

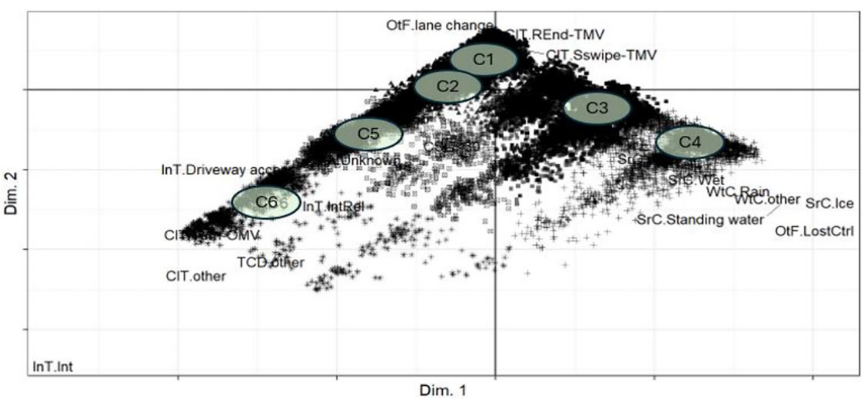

Figure 4. Clusters developed from vehicle crash data involving barriers. (Color online only)

clusters. These are the centroids mentioned earlier. The oval shapes in green are representing the centroid of each cluster.

The clusters form a broad, triangular distribution across the plot. This suggests there are varying degrees of similarity between crashes within each cluster, and there are some overlaps in characteristics between different clusters. The clusters are marked with terms for intersection type: driveway access (Int.T Driveway access) or collision type REnd-TMV, are referring to specific types of barrier-related crash scenarios such as crashes occurring near driveways or involving rear-end collisions. The labels farther from the center, for example, other factors: lane change (Otf.Lane change), surface condition: wet (Sfc.Wet), or weather condition: rain (WthC.Rain), represent more specific conditions or rarer events in the data. Factors such as wet road surfaces or rain are also contributing factors to crashes.

The CCA analysis generate six clusters from the dataset. It is crucial to understand the trends among each cluster for explaining the relationship between the contributing factors. For providing better explanations, cluster-level bar plots are subsequently developed to investigate into each cluster, offering a more comprehensive understanding of the attributes and their associations within the clusters.

For the interpretation of the clusters, it is crucial to identify the attributes that deviate most from the independence condition. The figures in Figures 5 to 7 are based on the top 20 attributes with the highest residuals. These residuals can be positive or negative. A positive residual signifies that the attribute has above average frequency within the clusters, whereas a negative residual indicates that the attribute has below average frequency. The vertical axis denotes the number of attributes, and the horizontal axis shows the deviation in the frequencies of the attributes from the average frequency.

Cluster 1 (C1)—High-Speed Rural Interstate Crashes

Figure 5*a* illustrates different attributes related to vehicle crashes in C1, which accounts for 40.7% of the data within the analysis. Each bar length indicates the frequency or weight of the attribute within this cluster relative to the center of the cluster. The center, at zero on the horizontal axis, is defined as the baseline. The figure includes the top 20 attributes that have strongest association within this cluster. The association can either be positive or negative based on the extension of the bars that represent different attributes. The attributes extending to the right, such as “CSL.60-70” (Crash Speed Limit: 60–70 mph), “SrfC.Dry” (Surface Condition: Dry), and “WthC.Clear” (Weather Condition: Clear), have a positive association with this cluster. These conditions have strongest association with the crashes within C1. Conversely, attributes to the left of the center, such as “CSL. 35-55” (Crash Speed Limit: 35–55 mph), “SrfC.Wet” (Surface Condition: Wet), and “WthC.Rain,” are negatively associated with this cluster. This means that these conditions are less frequent or less characteristic of crashes in Cluster 1 compared with the overall dataset.

Notably, attributes like “REnd-TMV” and “RFC.RurInt” (Road Functional Class: Rural Interstate) are among those with the strongest positive association with C1. This suggests that rear-end collisions on rural interstates with higher speed limits and dry, clear conditions are prevalent in this cluster. The analysis reflects the defining characteristics of this cluster and can help to develop targeted safety measures. For example, since high-speed rear-end collisions are frequent in dry, clear conditions, it can be beneficial to focus on awareness campaigns or enforcement of safe following distances on rural interstates. A higher speed

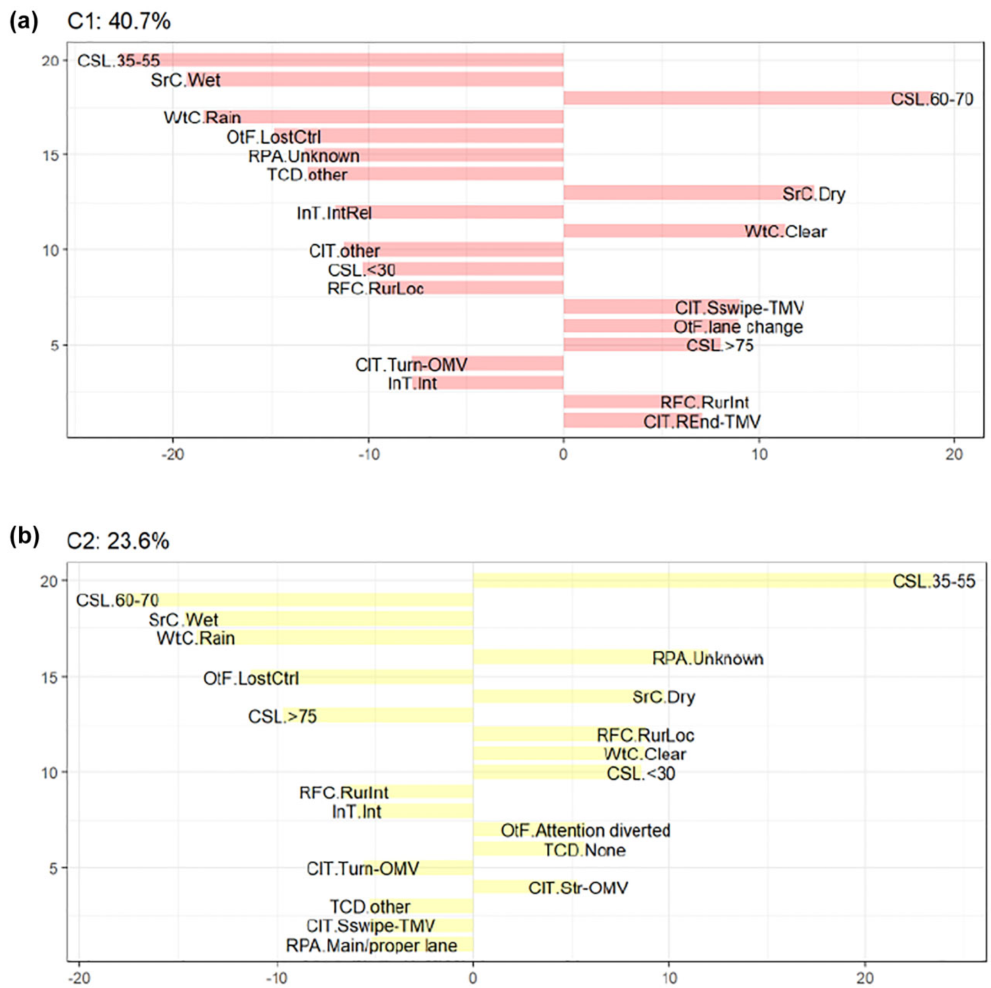

Figure 5. Cluster 1 (C1) and cluster 2 (C2): (*a*) cluster 1 high-speed rural interstate crashes and (*b*) cluster 2 low-speed rural local crashes.

limit is usually highly correlated with the increase in fatalities on rural interstates (*38*).

Cluster 2 (C2)—Low-Speed Rural Local Crashes

The C2 is represented in Figure 5*b*. The bar chart depicts the attributes associated with C2, which comprises 23.6% of the analyzed crash data. Attributes extending to the right in yellow, such as "CSL.35-55," "RFC.RurLoc" (Road Functional Class: Rural Local), and "WthC.Clear," have a positive association within this cluster. This indicates that crashes within this cluster are more likely to occur under these conditions. On the left side, the negative association is represented by bars extending in the opposite direction. Here the attributes "CSL.60-70," "SrfC.Wet," "WthC.Rain," and others indicate that these conditions are less associated with the crashes in cluster 2 relative to the overall data. This association can relate with the behavior of speeding depending on the weather conditions. Generally, people speed up during clear weather when the surface condition is dry. The combination of these factors can result in crashes. Whereas a rainy weather condition makes the surface wet, and it also forces the drivers to move at a lower speed which can effectively reduce the crashes and the severity.

The prominent positively associated attributes for C2, such as "ClT.Str-OMV" (Collision Type: Straight-One Motor Vehicle), "TCD.None" (Traffic Control Device: None), "OtF.Attention diverted" (Other Factors: Attention Diverted), and "CSL < 30" (Crash Speed Limit: < 30 mph), provide a significant outcome of the CCA. It highlighted that if the driver's attention is diverted on a road with a lower speed limit; there is a

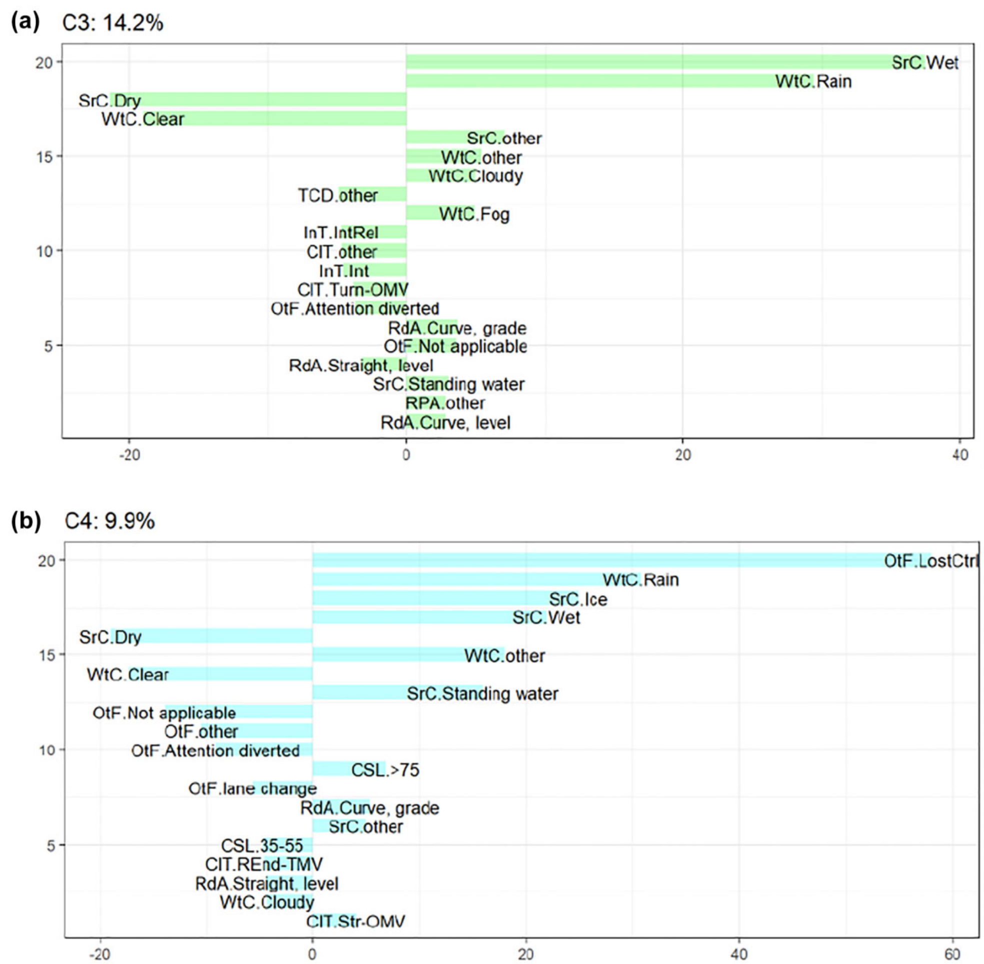

Figure 6. Cluster 3 (C3) and cluster 4 (C4): (*a*) cluster 3 adverse weather crashes and (*b*) cluster 4 adverse weather and loss of control crashes.

higher likelihood that it will result in a crash with the barrier. This can be highlighted as an example of driver distraction-related crashes which can be influenced by road design. Several previous studies have identified the effect of driver distraction on crashes and severity (*39*, *40*). Moreover, absence of traffic control devices, especially in specific road segments, can also increase crashes on the roadway (*41*).

Cluster 3 (C3)—Adverse Weather Crashes

Around 14% of barrier crash data are involved in C3 that is illustrated in Figure 6*a*. This bar chart showcases the variables associated with C3. On the right side, the attributes such as "SrfC.Wet," and "WthC.Rain" dominate positively. This indicates that wet surface conditions and rainy weather are significantly more-associated with the crashes in this cluster. Conversely, on the left side with negative association, attributes such as "SrfC.Dry" and "WthC.Clear" appear. These conditions are less prevalent in crashes within C3, suggesting that dry and clear weather are not as typical for crashes in this group as they might be in the overall dataset.

The clear distinction of wet and rainy conditions as key characteristics of crashes in C3 suggests a strong influence of adverse weather on these incidents. The other positively associated attributes such as "RdA.Curve, level" (Road Alignment: Curve, Level) and "SrC.Standing water" (Surface Condition: Standing Water) also denotes that standing water on the road at a curve can significantly increase the likelihood of crashes. Moreover, the weather condition attributes "WtC.Fog" (Weather Condition: Fog) and "WtC.Coludy" (Weather condition: Cloudy) also bolster this statement. Previously, few studies have found the correlation between adverse weather and traffic crashes. The adverse condition of weather can also affect driving behavior and results in crashes (*42*). Weather related factors such as windy or cloudy weather conditions can increase the likelihood of severe injury crashes (*43*).

Cluster 4 (C4)—Adverse Weather and Loss of Control Crashes

Figure 6*b* illustrates that C4 accounts for 9.9% of the crash data. To the right, the positively associated attributes with the longest bars are "WthC.Rain," "SrfC.Ice" (Surface Condition: Ice), and "SrfC.Wet." These suggest that C4 crashes are significantly associated with icy, wet conditions and rainy weather. To the left, the negatively associated attributes include "SrfC.Dry," "WthC.Clear," and "Otf.Not applicable" (Other Factors: Not applicable), indicating that these conditions are underrepresented in this cluster compared with the average.

"Otf.LostCtrl" (Other Factors: Lost control) shows the strongest positive deviation, significantly characterizing this cluster. Similar to previous studies that found a correlation between adverse weather and traffic crashes, few studies have also found a high association between loss of vehicle-control and traffic crashes. Vehicle stability is an important factor in reducing crashes (*44*). This points to a pattern where a loss of vehicle control, potentially because of adverse weather and slippery road conditions, is a defining feature of crashes in this cluster. Based on the defining attributes, C4 can be appropriately named "Adverse Weather and Loss of Control Crashes" highlighting the primary conditions under which these crashes occur and suggesting a focus on vehicle control and weather adaptation as key areas for intervention. This cluster has a similarity in weather conditions with C3. This also justifies the explanation of Figure 4 where the authors stated the overlapping of attributes between clusters, indicating the possibility of similarities in the attributes between clusters.

Cluster 5 (C5)—Turning Crashes at Intersections

Figure 7*a* represents C5 that accounts for 7.6% crashes of the dataset. The longitudinal bars extending to the right, in purple (color online only), have a positive association with this cluster. The most prominent positive attributes include "CIT.Turn-OMV" (Collision Type: Turning-One motor vehicle), "IntT.Driveway access" (Intersection Type: Driveway access), and "TCD.other" (Traffic Control Device: other). These attributes are significantly associated with the crashes in C5, suggesting that crashes involving vehicles turning at intersections and driveways are common within this cluster. Conversely, the negative association to the left includes "IntT.NonInt" (Intersection Type: Non-intersection), "CSL.60-70," and "CIT.Str-OMV" (Collision Type: Straight-One motor vehicle), indicating these conditions occur less frequently in crashes of this cluster.

The positive association with turning crashes at intersections and driveway access, and the negative association with non-intersection crashes, indicates that the crashes in C5 likely happen in more complex traffic environments where turning is involved. Turning-related crashes, particularly left-turn crashes, are highly associated with crashes that occur at an intersection even at signalized intersections (*45*). This could be important for urban planning and traffic management, as it may point to the need for improved signage, road design, or signal timing at intersections.

Cluster 6 (C6)—Diverse Intersection Crashes

C6 in Figure 7*b* includes 4.1% of the dataset. The attributes that are most positively associated with C6, extending to the right, are "IntT.Int" (Intersection Type: Intersection) and "CIT.other" (Collision Type: other), followed by "TCD.other" (Traffic Control Device: other). This indicates that crashes within this cluster frequently occur at intersections and involve a variety of collision types not specified in the other categories. Conversely, the attributes extending to the left are negatively associated with C6, meaning they are less representative of the crashes in this cluster compared with the centroid. These include "IntT.NonInt," "CIT.Str-OMV," and "TCD.Marked lanes" (Traffic Control Device: Marked lanes). Considering these associations, crashes in Cluster 6 is characterized by a diversity of unspecified factors occurring at intersections. Since "TCD.other" is also a significant factor, these crashes can involve uncommon or less typical traffic control devices, or situations where standard devices like marked lanes are not present. Studies have also justified that factors related to intersections such as type of intersection, posted speed limit, and visibility can affect type and severity of crashes at intersections (*46*).

Cluster-Based Comparative Analysis

In this section, cluster-based comparative statistics are explained. The descriptive statistics in Table 5 categorizes vehicle crash data that involve barriers in six distinct clusters and includes different distribution of attributes such as crash type, collision type, road conditions, and environmental factors. Most collisions are categorized as "Straight-One motor vehicle," especially prominent in C1- C5. This indicates that single-vehicle, non-complex crash scenarios are common. Crashes occur most commonly in clear weather conditions and on dry road surfaces across several clusters, with C3 and C4 as an

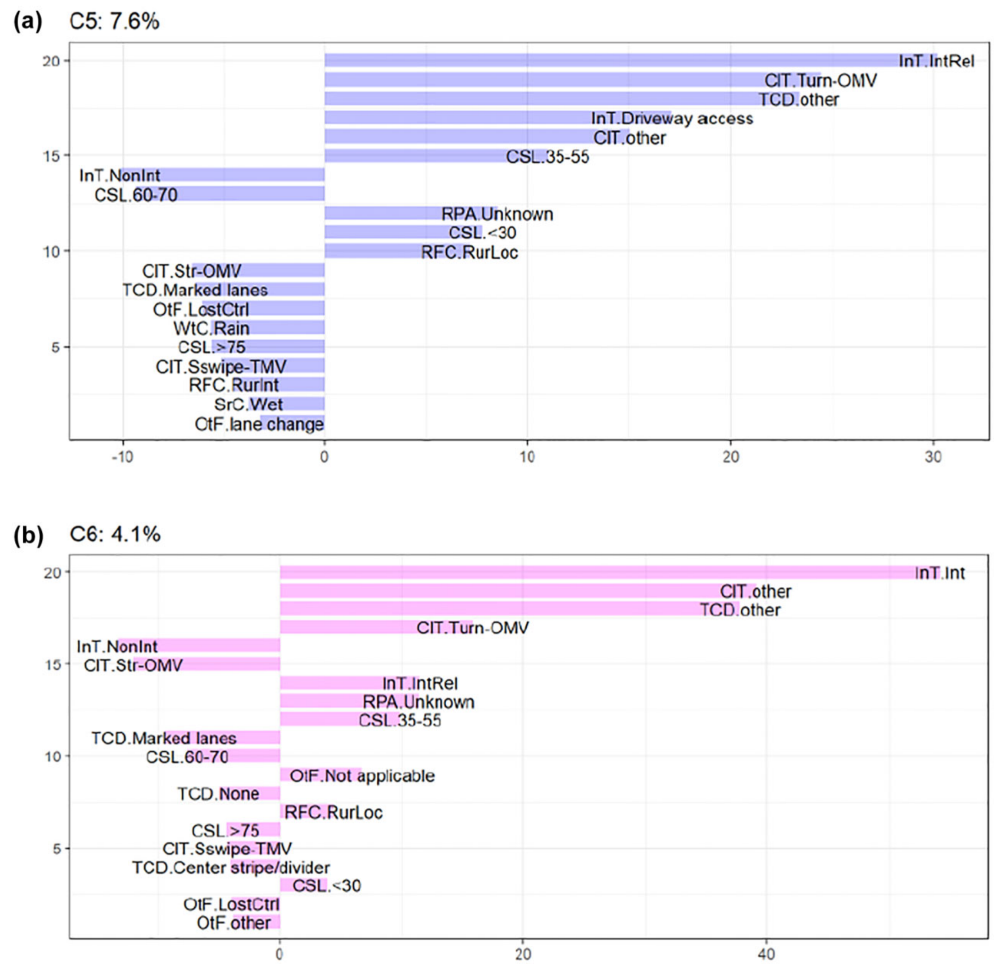

Figure 7. Cluster 5 (C5) and cluster 6 (C6): (*a*) cluster 5 turning crashes at intersections and (*b*) cluster 6 diverse intersection crashes.

exception where rain and wet surfaces predominate. This suggests a correlation with adverse weather conditions. Speed limits of 35–55 mph are frequently represented in C2, whereas higher speed limits of 60–70 mph are more characteristic of C1. Rural local roads have a higher occurrence in C2 and C5, which may reflect rural traffic risks. Daylight is the most common lighting condition across all clusters, underscoring that most crashes happen during the day. Moreover, straight and level road alignments are dominant across all clusters, which signify a high influence of these alignment types on crashes. The higher percentages of variables within each cluster are marked in darker shades of red.

C1 is characterized by a predominance of collisions with barriers involving a single motor vehicle (70.90%), occurring mostly on straight, level roads (66.30%). Notably, there is a significant representation of higher speed limits within this cluster, with 77.50% of crashes happening at a speed limit of 60–70 mph. This indicates the risk associated with higher speeds. The descriptive of this cluster verifies the explanation of C1 that was provided previously in this study. The cluster represents mostly the attributes that are strongly associated with barrier crashes that involve one motor vehicle. C2 includes 15,012 crashes, and similarly to C1, it has a high proportion of non-injury incidents (66.70%). However, the cluster is distinguished by a significant representation of crashes on rural local roads (49.00%) and most of them occurring within a lower speed limit range of 35–55 mph (72.80%). Compared with C1, this cluster refers a different risk profile, associated with rural road conditions.

C3 is dominated by weather, with 63.40% of crashes occurring during rain and 87.10% occurring on wet surfaces, the highest across all clusters. It also has the largest share of crashes where drivers were not injured (72.20%), which could indicate lower severity outcomes compared with other clusters. Like C1, most of the crashes within

Table 5. Cluster Based Comparative Analysis

	High-speed rural interstate crashes (cluster 1) (N = 25,929)	Low-speed rural local crashes (cluster 2) (N = 15,012)	Adverse weather crashes (cluster 3) (N = 9,062)	Adverse weather and loss of control crashes (cluster 4) (N = 6,327)	Turning crashes at intersections (cluster 5) (N = 4,826)	Diverse intersection crashes (cluster 6) (N = 2,589)
Crash type (CrST)						
Fatal	1.79%	1.82%	1.45%	1.00%	0.99%	1.39%
Incapacitating injury	4.47%	4.09%	2.96%	2.29%	5.49%	5.99%
Non-incapacitating	14.50%	12.60%	9.77%	8.74%	11.30%	18.10%
Not injured	62.40%	66.70%	72.20%	76.50%	68.90%	52.50%
Possible injury	16.90%	14.70%	13.70%	11.50%	13.40%	22.00%
Collision type (CllT)						
RearEnd-Two motor vehicle(REnd-TMV)	11.70%	5.00%	7.11%	1.71%	3.29%	0.15%
Sideswipe-Two motor vehicle (Sswipe-TMV)	17.40%	5.78%	10.80%	6.48%	2.01%	0.77%
Straight-One motor vehicle (Str-OMV)	70.90%	85.80%	79.70%	88.00%	43.40%	0.04%
Turning-One motor vehicle (Turn-OMV)	0.00%	0.17%	0.50%	1.14%	25.20%	22.80%
Other	0.00%	3.22%	1.90%	2.72%	26.10%	76.30%
Population group (PpGr)						
250,000+	37.50%	34.60%	39.20%	28.60%	30.90%	29.20%
100,000-249,999	11.90%	10.60%	12.60%	14.50%	11.60%	12.80%
25,000-49,999	4.71%	6.48%	6.12%	5.33%	7.83%	7.53%
Rural	33.20%	27.90%	25.80%	34.10%	26.80%	26.30%
Other	12.80%	20.50%	16.30%	17.60%	22.90%	24.10%
Road alignment (RdAl)						
Curve, grade	7.04%	9.99%	13.90%	17.20%	7.83%	2.74%
Curve, level	8.02%	16.00%	15.10%	12.60%	11.90%	3.71%
Straight, grade	12.60%	11.10%	12.20%	14.90%	11.90%	11.50%
Straight, level	66.30%	56.90%	50.10%	44.50%	61.80%	77.60%
other	6.04%	6.08%	8.76%	10.80%	6.57%	4.48%
Other factors (OthF)						
Attention diverted	19.90%	24.60%	11.60%	0.16%	18.00%	14.20%
lane change	11.80%	3.66%	5.39%	0.03%	2.28%	0.97%
Lost control (LostCtrl)	0.00%	0.00%	8.17%	98.50%	0.54%	1.82%
other	28.80%	22.80%	24.40%	0.35%	27.00%	10.10%
Not applicable	39.50%	49.00%	50.40%	0.92%	52.10%	72.90%
Crash speed limit (CrSL)						
<30 mph	0.91%	13.80%	6.92%	5.31%	17.80%	14.50%
>75 mph	17.90%	2.00%	11.20%	22.30%	1.76%	1.00%
35–55 mph	3.76%	72.80%	30.40%	20.70%	66.00%	73.50%
60–70 mph	77.50%	11.30%	51.50%	51.70%	14.40%	11.00%
Road functional class (RdFC)						
Rural interstate (RurInt)	30.50%	13.60%	23.20%	25.30%	11.70%	13.00%
Rural local (RurLoc)	21.30%	49.00%	30.70%	30.00%	55.60%	52.50%
Rural principal arterial (RurPrinArt)	14.80%	7.22%	11.70%	13.90%	5.78%	7.49%
Urban interstate (UrbInt)	18.30%	10.10%	17.20%	14.10%	8.23%	7.61%
other	15.00%	20.10%	17.20%	16.80%	18.70%	19.40%

(continued)

Table 5. Continued

	High-speed rural interstate crashes (cluster 1) (N = 25,929)	Low-speed rural local crashes (cluster 2) (N = 15,012)	Adverse weather crashes (cluster 3) (N = 9,062)	Adverse weather and loss of control crashes (cluster 4) (N = 6,327)	Turning crashes at intersections (cluster 5) (N = 4,826)	Diverse intersection crashes (cluster 6) (N = 2,589)
Lighting condition (LghC)						
Dark, lighted	26.90%	26.90%	25.90%	19.00%	26.10%	22.10%
Dark, not lighted	20.60%	21.90%	19.10%	17.80%	16.00%	8.42%
Dawn	1.42%	1.17%	1.57%	2.17%	0.81%	1.08%
Daylight	49.20%	47.70%	49.80%	58.50%	54.50%	67.10%
Other	1.86%	2.32%	3.64%	2.56%	2.53%	1.27%
Traffic control device (TCD)						
Center stripe/divider	12.30%	12.00%	10.60%	12.10%	6.86%	1.66%
Marked lanes	70.20%	55.90%	64.20%	63.30%	34.20%	8.19%
No passing zone	3.44%	4.32%	2.85%	3.89%	2.09%	0.19%
Other	0.00%	4.14%	3.16%	3.49%	44.50%	87.60%
None	14.10%	23.70%	19.20%	17.20%	12.40%	2.32%
Weather condition (WthC)						
Clear	85.70%	86.10%	6.96%	1.77%	77.40%	75.30%
Cloudy	13.80%	13.30%	21.40%	6.99%	15.70%	15.30%
Fog	0.39%	0.56%	2.81%	1.50%	1.20%	1.04%
Rain	0.03%	0.01%	63.40%	75.00%	5.10%	8.00%
Other	0.02%	0.03%	5.41%	14.70%	0.64%	0.39%
Road part (RdPA)						
Entrance/on ramp	2.07%	5.37%	4.13%	3.16%	5.06%	2.43%
Exit/off ramp	5.14%	6.24%	6.27%	4.17%	5.04%	2.74%
Main/proper lane	91.30%	66.40%	79.70%	85.70%	65.40%	59.40%
Other	1.16%	1.40%	2.90%	2.45%	1.20%	0.58%
Unknown	0.37%	20.60%	7.03%	4.47%	23.40%	34.90%
Surface condition (SrfC)						
Dry	98.60%	98.50%	4.04%	0.03%	86.40%	85.90%
Ice	0.00%	0.00%	2.43%	19.30%	0.17%	0.04%
Standing water	0.01%	0.00%	2.82%	10.10%	0.21%	0.39%
Wet	1.37%	1.35%	87.10%	67.40%	12.30%	13.40%
Other	0.05%	0.18%	3.63%	3.22%	0.91%	0.27%
Intersection type (IntT)						
Driveway access	0.00%	0.73%	0.07%	0.06%	9.78%	1.24%
Intersection (Int)	0.00%	0.00%	0.04%	0.02%	2.53%	70.10%
Intersection related (IntRel)	0.00%	6.39%	2.22%	3.08%	48.20%	28.40%
Non-intersection (NonInt)	100%	92.90%	97.70%	96.80%	39.50%	0.19%

Note: Color online only.

this cluster are associated with a higher speed limit of 60–70 mph (51.50%). C4 also portrays statistics highlighting the influence of adverse weather conditions on barrier crashes. In addition to that, C4 includes the driver behavior by addressing most crashes where the driver lost control (98.50%). This suggests a strong correlation with driver behavior and extreme road conditions. Most of these crashes took place on straight, level roads (44.50%) and in foggy or rainy weather, reinforcing the importance of addressing vehicle control in adverse weather interventions.

C5 is distinctive for the high percentage of intersection-related crashes (48.20%). This is significantly higher than in other clusters. This cluster also has a higher rate of crashes resulting in possible injuries (13.40%) and frequently occurs within the speed limit range of 35–55 mph (66.00%), pointing to potential issues with urban traffic environments and intersection design. C6, although the smallest with 2,589 incidents, has a remarkable 70.10% of crashes occurring at intersections. It also has a high rate of crashes involving vehicles in turning movements (Turn-OMV at 22.80%). Unlike other clusters, where straight-line movements are more common, this one is characterized by a diverse range of incident types, as shown by the high percentage of "other" collision types (76.30%).

Conclusions

This study offers a unique perspective on barrier crashes by delineating distinct crash clusters and their associated factors. Through the analysis of six clusters of vehicle crash data, it identifies distinct patterns with significant policy implications. For instance, findings reveal a prevalence of high-speed and rear-end crashes on rural interstates without injuries, suggesting potential safety overestimation in these areas. High-speed crashes on rural interstates, particularly rear-end collisions, indicate a need for enhanced speed enforcement, such as speed cameras and increased patrols. Educational campaigns targeting the risks of high-speed driving, even in clear conditions, could further improve safety. In rural areas, where crashes occur despite low speeds, improving road infrastructure, such as clearer signage and lane markings, alongside rural driving safety programs, can mitigate crash risks. Adverse weather conditions such as rain, wet, and icy surfaces also play a considerable role in crash occurrences, contributing to vehicle loss of control incidents. This indicates that while barriers are designed to prevent roadway departure (RwD) crashes, these environmental factors complicate their effectiveness. DOTs and policymakers should prioritize road design and maintenance strategies that account for weather-related risks, possibly enhancing barrier designs or implementing complementary measures like improved road surface treatments to mitigate crashes under such conditions. Advanced driver assistance systems (ADAS) and infrastructure improvements like better drainage and high-friction road surfaces can enhance safety in wet or icy conditions. Real-time weather updates and public advisories via message boards could also help reduce weather-related crashes.

Intersection-related crashes, prevalent in urban settings, underscore the need for smarter intersection design, such as intelligent traffic signals and dedicated turning lanes, to reduce complexity and improve visibility at critical points. Moreover, the diversity of intersection related crashes indicates the complexity of intersection safety. Policy implications here can include a review of intersection design standards and the adoption of roundabouts where feasible, which have been shown to significantly reduce the incidence and severity of intersection crashes. Safety cushions, improved barrier design, and adaptive traffic management systems could greatly reduce crash occurrences, ensuring that barriers not only prevent roadway departures but also minimize unintended crash risks. Policymakers are encouraged to focus on enhancing both infrastructure and driver awareness to make roadways safer for all users.

A key limitation of this study is the lack of detailed information concerning barrier design, barrier types, and the specific locations of barriers. This limitation arises because the available crash data does not include such specific countermeasure details for each roadway segment. However, this information is crucial, as different barrier designs and placements may influence crash outcomes in various ways. Without this data, it is difficult to assess the effectiveness of barriers in preventing crashes or to identify whether certain designs contribute to unintended crash risks. Future studies should focus on incorporating detailed information concerning barrier types, design specifications, and exact placements to better understand their impact on crash dynamics. This would enable a more comprehensive analysis of how barrier features interact with road and environmental conditions, helping transportation planners design more effective safety interventions. Such insights could guide improvements in barrier design and placement, ensuring they prevent crashes while minimizing unintended risks associated with barrier collisions.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Rohit Chakraborty, Subasish Das; data collection: Subasish Das, Boniphace Kutela; analysis and interpretation of results: Rohit Chakraborty; draft manuscript preparation: Rohit Chakraborty, Mahmuda Sultana Mimi. 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

Rohit Chakraborty https://orcid.org/0000-0002-1660-9764
Subasish Das https://orcid.org/0000-0002-1671-2753
Mahmuda Sultana Mimi https://orcid.org/0009-0007-8534-3633
Boniphace Kutela https://orcid.org/0000-0002-5450-1623

References

1. FARS, and Traffic Safety Administration. Crashes and All Victims. https://www-fars.nhtsa.dot.gov/States/StatesCrashesAndAllVictims.aspx. Accessed March 19, 2024.
2. Hosseinpour, M., and K. Haleem. Examining Crash Injury Severity and Barrier-Hit Outcomes from Cable Barriers and Strong-Post Guardrails on Alabama's Interstate Highways. *Journal of Safety Research*, Vol. 78, 2021, pp. 155–169. https://doi.org/10.1016/j.jsr.2021.06.009.
3. Zou, Y., and A. P. Tarko. Barrier-Relevant Crash Modification Factors and Average Costs of Crashes on Arterial Roads in Indiana. *Accident Analysis & Prevention*, Vol. 111, 2018, pp. 71–85. https://doi.org/10.1016/j.aap.2017.11.020.
4. Molan, A. M., M. Rezapour, and K. Ksaibati. Investigating the Relationship Between Crash Severity, Traffic Barrier Type, and Vehicle Type in Crashes Involving Traffic Barrier. *Journal of Traffic and Transportation Engineering (English Edition)*, Vol. 7, No. 1, 2020, pp. 125–136. https://doi.org/10.1016/j.jtte.2019.03.004.
5. Hu, W., and E. T. Donnell. Median Barrier Crash Severity: Some New Insights. *Accident Analysis & Prevention*, Vol. 42, No. 6, 2010, pp. 1697–1704.
6. Candappa, N. L. *An Investigation of Fatal and Serious Injury Wire Rope Safety Barrier Crashes Using Crash Data and Software Simulation*. 2021. https://doi.org/10.26180/13728343.
7. Chimba, D., D. Emaasit, S. Allen, B. Hurst, and M. Nelson. Factors Affecting Median Cable Barrier Crash Frequency: New Insights. *Journal of Transportation Safety & Security*, Vol. 6, No. 1, 2014, pp. 62–77. https://doi.org/10.1080/19439962.2013.812168.
8. Chang, F., P. Xu, H. Zhou, A. H. S. Chan, and H. Huang. Investigating Injury Severities of Motorcycle Riders: A Two-Step Method Integrating Latent Class Cluster Analysis and Random Parameters Logit Model. *Accident Analysis & Prevention*, Vol. 131, 2019, pp. 316–326. https://doi.org/10.1016/j.aap.2019.07.012.
9. Li, Z., C. Chen, Y. Ci, G. Zhang, Q. Wu, C. Liu, and Z. Qian. Examining Driver Injury Severity in Intersection-Related Crashes Using Cluster Analysis and Hierarchical Bayesian Models. *Accident Analysis & Prevention*, Vol. 120, 2018, pp. 139–151. https://doi.org/10.1016/j.aap.2018.08.009.
10. Nilsson, D., M. Lindman, T. Victor, and M. Dozza. Definition of Run-off-Road Crash Clusters-for Safety Benefit Estimation and Driver Assistance Development. *Accident; Analysis & Prevention*, Vol. 113, 2018, pp. 97–105. https://doi.org/10.1016/j.aap.2018.01.011.
11. Gabauer, D. J., and X. Li. Influence of Horizontally Curved Roadway Section Characteristics on Motorcycle-to-Barrier Crash Frequency. *Accident Analysis & Prevention*, Vol. 77, 2015, pp. 105–112. https://doi.org/10.1016/j.aap.2015.02.006.
12. Atahan, A. O., J. M. Hiekmann, J. Himpe, and J. Marra. Development of a Continuous Motorcycle Protection Barrier System Using Computer Simulation and Full-Scale Crash Testing. *Accident Analysis & Prevention*, Vol. 116, 2018, pp. 103–115. https://doi.org/10.1016/j.aap.2017.04.005.
13. Bambach, M. R., R. H. Grzebieta, and A. S. McIntosh. The Crash Mechanics of Fatal Motorcycle–Barrier Collisions in Australasia. *Journal of Transportation Safety & Security*, Vol. 5, No. 1, 2013, pp. 66–77. https://doi.org/10.1080/19439962.2012.700389.
14. Cha, J. G., B. Chul Hon, and B. K. Han, and National Highway Traffic Safety Administration. Experimental Study for the Possibility of Using of External Inversion Crash Box on Sloped Crash Barrier. *Proc., 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Seoul, South Korea, 2013.
15. Chell, J., C. E. Brandani, S. Fraschetti, J. Chakraverty, and V. Camomilla. Limitations of the European Barrier Crash Testing Regulation Relating to Occupant Safety. *Accident Analysis & Prevention*, Vol. 133, 2019, p. 105239. https://doi.org/10.1016/j.aap.2019.07.015.
16. Crespo, S., D. Perez-Rapela, J. Román-Marín, F. Martin-Vázquez, J. Luzón-Narro, and C. Arregui-Dalmases, and International Research Council on Biomechanics of Injury (IRCOBI). Structural Optimisation in Vehicle Development for the Current Euro NCAP Side Crash Protocol: How to Minimise the Structural Changes due to the Current Barrier Stiffness and Geometry. Presented at the IRCOBI Conference Proceedings, Malaga, Spain, 2016.
17. Crowther, S., and R. Swears. Longswamp to Rangiriri Wire Rope Barrier: Increased Crash Numbers but Improved Road Safety. *Proc., ARRB Conference*, Melbourne, Australia, 2010.
18. Hammonds, B. R., and R. Troutbeck. Crash Test Outcomes for Three Generic Barrier Types. *Proc., ARRB Conference*, Perth, Australia, 2012.
19. Hu, H., W.-J. Lu, and Z. Lu. Impact Crash Analyses of an Off-Road Utility Vehicle—Part II: Simulation of Frontal Pole, Pole Side, Rear Barrier and Rollover Impact Crashes. *International Journal of Crashworthiness*, Vol. 17, No. 2,

2012, pp. 163–172. https://doi.org/10.1080/13588265.2011.632858.

20. Johnson, N. S., and H. C. Gabler. Reduction in Fatal Longitudinal Barrier Crash Rate due to Electronic Stability Control. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2521: 79–85.
21. Molan, A. M., M. Moomen, and K. Ksaibati. The Impact of Traffic Barrier Geometric Features on Crash Frequency and Injury Severity of Non-Interstate Highways. *Journal of Safety Research*, Vol. 75, 2020, pp. 155–165. https://doi.org/10.1016/j.jsr.2020.09.005.
22. Molan, A. M., M. Rezapour, and K. Ksaibati. Modeling Traffic Barriers Crash Severity by Considering the Effect of Traffic Barrier Dimensions. *Journal of Modern Transportation*, Vol. 27, No. 2, 2019, pp. 141–151. https://doi.org/10.1007/s40534-019-0186-1.
23. Rezapour, M., and K. Ksaibati. Application of Multi-Group Structural Equation Modelling for Investigation of Traffic Barrier Crash Severity. *International Journal of Injury Control and Safety Promotion*, Vol. 27, No. 2, 2020, pp. 232–242. https://doi.org/10.1080/17457300.2020.1734943.
24. Rezapour, M., and K. Ksaibati. Application of Machine Learning Technique for Optimizing Roadside Design to Decrease Barrier Crash Costs, a Quantile Regression Model Approach. *Journal of Safety Research*, Vol. 78, 2021, pp. 19–27. https://doi.org/10.1016/j.jsr.2021.06.002.
25. Rezapour, M., K. Ksaibati, and M. Moomen. Application of Quantile Mixed Model for Modeling Traffic Barrier Crash Cost. *Accident Analysis & Prevention*, Vol. 148, 2020, p. 105795. https://doi.org/10.1016/j.aap.2020.105795.
26. Rezapour, M., S. S. Wulff, and K. Ksaibati. Bayesian Hierarchical Modelling of Traffic Barrier Crash Severity. *International Journal of Injury Control and Safety Promotion*, Vol. 28, No. 1, 2020, pp. 94–102. https://doi.org/10.1080/17457300.2020.1849312.
27. Zulkiffli, N. S., A. Roslan, and N. Z. Harun. *Effect of Median Barrier Types on Crash Severity at Rural Expressway*. Publication 9789675967924. Malaysian Institute of Road Safety Research, Kuala Lumpur, 2016, 53p.
28. Das, S., M. M. Hossain, M. A. Rahman, X. Kong, X. Sun, and G. M. A. Mamun. Understanding Patterns of Moped and Seated Motor Scooter (50 Cc or Less) Involved Fatal Crashes Using Cluster Correspondence Analysis. *Transportmetrica A Transport Science*, Vol. 19, No. 2, 2023, p. 2029613. https://doi.org/10.1080/23249935.2022.2029613.
29. Ashifur Rahman, M., S. Das, and X. Sun. Using Cluster Correspondence Analysis to Explore Rainy Weather Crashes in Louisiana. *Transportation Research Record: Journal of the Transportation Research Board*, 2022. 2676: 159–173.
30. Das, S., M. Tabesh, B. Dadashova, and C. Dobrovolny. Diagnosis of Encroachment-Related Work-Zone Crashes by Applying Pattern Recognition. *Transportation Research Record: Journal of the Transportation Research Board*, 2023. 2677: 222–236.
31. Greenwell, B. M., and B. C. Boehmke. Variable Importance Plots—An Introduction to the Vip Package. *The R Journal*, Vol. 12, No. 1, 2020, pp. 343–366. https://doi.org/10.32614/RJ-2020-013.
32. Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, et al. Xgboost: Extreme Gradient Boosting. 2014, p. 1.7.8.1. https://doi.org/10.32614/CRAN.package.xgboost.
33. Sourial, N., C. Wolfson, B. Zhu, J. Quail, J. Fletcher, S. Karunananthan, K. Bandeen-Roche, F. Béland, and H. Bergman. Correspondence Analysis Is a Useful Tool to Uncover the Relationships Among Categorical Variables. *Journal of Clinical Epidemiology*, Vol. 63, No. 6, 2010, pp. 638–646. https://doi.org/10.1016/j.jclinepi.2009.08.008.
34. Van De Velden, M., A. I. D'Enza, and F. Palumbo. Cluster Correspondence Analysis. *Psychometrika*, Vol. 82, No. 1, 2017, pp. 158–185. https://doi.org/10.1007/s11336-016-9514-0.
35. Greenacre, M. J. Correspondence Analysis. *WIREs Computational Statistics*, Vol. 2, No. 5, 2010, pp. 613–619. https://doi.org/10.1002/wics.114.
36. Zhou, Y., X. Zhang, J. Wang, and Y. Gong. Research on Speaker Feature Dimension Reduction Based on CCA and PCA. Presented at the 2010 International Conference on Wireless Communications & Signal Processing (WCSP), Suzhou, China, October 21–23, 2010.
37. Hsu, L. L., and A. C. Culhane. Correspondence Analysis for Dimension Reduction, Batch Integration, and Visualization of Single-Cell RNA-Seq Data. *Scientific Reports*, Vol. 13, No. 1, 2023, p. 1197. https://doi.org/10.1038/s41598-022-26434-1.
38. Warner, J., H. Chawla, C. Zhou, and P. T. Savolainen. Analysis of Rural Interstate Fatality Rates in Consideration of Recent Increases in Maximum Statutory Speed Limits. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 48–58.
39. Jazayeri, A., J. R. B. Martinez, H. S. Loeb, and C. C. Yang. The Impact of Driver Distraction and Secondary Tasks With and Without Other Co-Occurring Driving Behaviors on the Level of Road Traffic Crashes. *Accident Analysis & Prevention*, Vol. 153, 2021, p. 106010. https://doi.org/10.1016/j.aap.2021.106010.
40. Hasan, A. S., M. Jalayer, E. Heitmann, and J. Weiss. Distracted Driving Crashes: A Review on Data Collection, Analysis, and Crash Prevention Methods. *Transportation Research Record: Journal of the Transportation Research Board*, 2022. 2676: 423–434.
41. Bryden, J. E. Work Zone Crashes Involving Traffic Control Devices, Safety Features, and Work Vehicles and Equipment. *Transportation Research Record: Journal of the Transportation Research Board*, 2007. 2015: 64–70.
42. Pisano, P. A., L. C. Goodwin, and M. A. Rossetti. U.S. Highway Crashes in Adverse Road Weather Conditions. *Proc., 88th American Meteorological Society Annual Meeting*, New Orleans, LA, 2008.
43. Pathivada, B. K., A. Banerjee, and K. Haleem. Impact of Real-Time Weather Conditions on Crash Injury Severity in

Kentucky Using the Correlated Random Parameters Logit Model with Heterogeneity in Means. *Accident Analysis & Prevention*, Vol. 196, 2024, p. 107453. https://doi.org/10.1016/j.aap.2023.107453.

44. Bahouth, G. Real World Crash Evaluation of Vehicle Stability Control (VSC) Technology. *Annual Proceedings / Association for the Advancement of Automotive Medicine*, Vol. 49, 2005, pp. 19–34.
45. Wang, X., and M. Abdel-Aty. Modeling Left-Turn Crash Occurrence at Signalized Intersections by Conflicting Patterns. *Accident Analysis & Prevention*, Vol. 40, No. 1, 2008, pp. 76–88. https://doi.org/10.1016/j.aap.2007.04.006.
46. Galloway, A. J., M. Bareiss, T. Hasegawa, R. Sherony, and L. E. Riexinger. Evaluation of Intersection Crashes Using Naturalistic Driving Data Through the Lens of Future I-ADAS. *Traffic Injury Prevention*, Vol. 24, No. 7, 2023, pp. 577–582. https://doi.org/10.1080/15389588.2023.2237621.