Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Key factors affecting motorcycle-barrier crash severity: an innovative cluster-regression technique

Reuben Tamakloe, Subasish Das, Emmanuel Kofi Adanu & Dongjoo Park

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2023.2230310

Key factors affecting motorcycle-barrier crash severity: an innovative cluster-regression technique

Reuben Tamakloe [a], Subasish Das [b], Emmanuel Kofi Adanu [c] and Dongjoo Park [d]

[a]Department of Transportation Engineering, University of Seoul, Seoul, South Korea; [b]Texas State University, San Marcos, TX, USA; [c]Alabama Transportation Institute, The University of Alabama System, Tuscaloosa, AL, USA; [d]Department of Transportation Engineering & Department of Urban Big Data Convergence, University of Seoul, Seoul, South Korea

ABSTRACT

Highway motorcycle-barrier crashes are uncommon but are associated with severe ramifications. Little has been done to understand the factors related to these crashes, making it difficult to establish appropriate mitigation policies. This study identifies homogeneous groups of motorcycle-barrier crashes on highways and investigates cluster-specific key factor associations and the determinants of injury severity. Cluster Correspondence Analysis was employed to discover latent clusters and cluster-specific key factor associations using motorcycle-barrier crashes from Massachusetts. Further, an ordered probit regression technique was employed to investigate the effect of factors on injury severity outcomes at the cluster level. Three highway access control type-related clusters were identified. While seniors, collectors, intersections/roundabouts, daylight, and summer were associated with no/partial access-controlled segment crashes, interstates, ramps, medians, dark-lighted roads, and winter correlated with full access-controlled segment crashes. Factors influencing fatalities differed for each cluster. From the insightful findings, targeted countermeasures geared at improving motorcycle safety are suggested.

ARTICLE HISTORY
Received 7 September 2022
Accepted 9 May 2023

KEYWORDS
Motorcycle; cluster correspondence analysis; ordered probit regression; access control; injury severity

Introduction

Background

Motorcycle usage is increasing globally due to the growing delivery market size and its usefulness in terms of low operation cost and easy manoeuvrability (Chen, Chou, and Hung 2019; Trinh, Sano, and Hatoyama 2021; Tamakloe, Das, et al. 2022). Despite their numerous advantages, they are associated with high fatalities in terms of traffic crashes, making them a global health concern. In 2020, over 5,500 motorcyclists in the US were killed in crashes, representing 14% of all traffic fatalities. According to statistics from the National Highway Traffic Safety Administration, this represents the highest number of motorcycle fatalities ever recorded since 1975. Although not a frequent crash type, motorcycles crashing with

CONTACT Reuben Tamakloe drtamakloe@uos.ac.kr

fixed roadside objects, particularly barriers and guardrails, have been identified to be more correlated with fatalities in the US (Venkatraman et al. 2021). Motorcycle crashes with other vehicles have been well studied in the literature as it is the most common crash type. However, there are numerous questions regarding the factors influencing motorcycle-barrier crashes since it is less studied.

The effect of roadway barriers on motorcycle safety has been a major concern for transport experts. Barriers prevent cars from running off the road, crashing into dangerous objects, or falling into water bodies. However, they do little regarding the protection of motorcycles as they were primarily designed for protecting vehicles such as cars. Unlike occupants of these vehicles, whose bodies are protected on impact, motorcyclists may end up being ejected from the bike after hitting a barrier or may be wounded after hitting the sharp parts of the barrier. In the traffic safety literature, barriers are noted to contribute significantly to both motorcycle crash risk and fatalities (Vlahogianni, Yannis, and Golias 2012). In the US, while 25% of motorcycle collisions with fixed objects, including barriers, resulted in fatalities, passenger car- and light truck-fixed object crashes resulted in only 18% and 14% of fatalities, respectively (Venkatraman et al. 2021). In Massachusetts, motorcycle fatalities increased by 11.4% in 2020 (44 fatalities in 2019 and 49 fatalities in 2020). Motorcycle-barrier crashes resulted in the third-highest fatalities in Massachusetts from 2016 to 2020. Due to the high number of fatalities associated with this crash type, it is important to explore factors associated with them and to develop targeted countermeasures to mitigate them.

As a crash contributing factor, curve radius has been found to be the key factor influencing motorcycle-barrier crashes, and researchers suggested that increasing horizontal curve radius could reduce these crashes (Gabauer and Li 2015; Xin et al. 2017). Concerning crash severity, researchers have associated motorcycle crash fatalities with factors such as rider ejections, barrier type, precrash speed, impact angles, crash postures, motorcyclist kinematics, older riders, darkness, and non-level roadway profiles (Bambach, Grzebieta, and McIntosh 2013; Daniello, Cristino, and Gabler 2013; Gabauer and Li 2015; Li, Park, and Lambert 2018). The individual effect of a range of factors influencing motorcycle safety in terms of barrier crashes has been established. Although a study has explored patterns in motorcycle crashes in general (Das 2021), the patterns and associations of risk factors influencing motorcycle-barrier crashes of homogeneous clusters with similar characteristics are unknown. Besides, the risk factors affecting injury severities of motorcycle-barrier crashes clustered into similar groups have not been explored.

Against this backdrop, this study aims to bridge the literature gap by applying a robust cluster-regression approach to first identify latent clusters in a motorcycle-barrier crash dataset and discover the patterns of attributes most likely to impact motorcycle-barrier crashes in each cluster. The study then proceeds to investigate the individual influence of risk factors on crash injury severity outcomes for each cluster. Identifying the latent patterns of factors associated with motorcycle-barrier crashes and the impact of factors on injury severity outcomes for each homogeneous cluster will help policymakers take more targeted actions to mitigate the crashes. Motorcycle safety can be improved by controlling all or some of the variables identified in the key factor chains influencing motorcycle-barrier crashes and the factors affecting the severity outcomes of these crashes.

A brief review and critique of prior studies

Several studies have been conducted to understand the influence of risk factors on the crash risk, frequency, or severity of motorcycle crashes. Due to the aim of this research, this review focuses on briefly discussing the literature on motorcycle-barrier crashes followed by a review of the literature on applying the cluster-regression technique in the safety field.

Motorcycle-barrier crashes

Many studies on motorcycle safety have been conducted to understand the factors affecting the frequency and severity of collisions involving motorcycles. In the area of severity, studies identified factors such as inattentiveness, good road surfaces, high-speed limits, alcohol involvement, horizontal curves, heavy vehicle involvement, dry pavement, darkness, and fixed objects that increased the likelihood of fatal/severe injuries (Tamakloe, Das, et al. 2022; Farid and Ksaibati 2021; Agyemang, Adanu, and Jones 2021; Waseem, Ahmed, and Saeed 2019). Some researchers noted a close association of motorcycle crashes with fatalities particularly when they occur around or hit barriers or roadside objects. For example, it was revealed that the risk of sustaining fatal injuries increases significantly when a motorcycle collides with roadside objects (Daniello and Gabler 2011b). An in-depth study conducted in New Jersey to explore rider trajectory and injury outcomes in motorcycle-barrier crashes discovered that the location of the crash has a significant effect on the rider's trajectory after the crash. The two critical locations identified to be highly correlated with post-impact trajectory are the entrance and the exit of ramps. Most riders were thrown into the barrier when they collided on exit ramps. The authors further noted that the likelihood of observing serious injuries increased by 2.91 and 4.73 times when the motorcycle strikes a barrier or ends up being thrown into the barrier after impact, respectively (Daniello, Cristino, and Gabler 2013).

Upon analysing crash data from Australia and New Zealand, researchers identified that collisions on clear days, weekends, daylight, dry roads, over-speeding, and alcohol involvement are key features in fatal motorcycle-barrier collisions (Jama et al. 2011). Other studies demonstrated that injury severity was highly correlated with speed prior to the crash. Moreover, in fatal motorcycle-barrier crashes, the thorax region was more likely to receive maximum injury (Bambach, Grzebieta, and McIntosh 2012; Grzebieta, Bambach, and McIntosh 2013). Further, the impact trajectory angle and the kinetic energy dissipated onto the barrier by the motorcyclist were key crash mechanics identified to be associated with motorcycle-barrier crashes (Bambach, Grzebieta, and McIntosh 2013).

To gain insights for the development of safer barrier treatments for motorcycles, researchers primarily focused on identifying risk factors influencing crash frequency outcomes after a motorcycle-barrier crash occurs. Recently, a motorcycle-barrier crash frequency study was conducted to explore the influence of roadway design features on crash frequency and to propose appropriate locations for placing motorcycle-friendly barriers. Based on the statistical analysis conducted, it was determined that features such as long curves, isolated curves, and smaller curve radii increase motorcycle-barrier crash risk (Gabauer and Li 2015). In other literature, curved road sections (Jama et al. 2011), smaller radii, and grades above 3% (Gabauer 2016) were associated with a higher proportion of motorcycle-barrier crashes. Further, vertical grades, high access density, small curve radius, narrow lanes, rough pavement, and high AADT and truck volume (Abdul Manan, Jonsson, and Várhelyi 2013; Schneider, Savolainen, and Moore 2010; Xin et al. 2017) have a large impact on motorcycle crash risk. Although motorcyclists are at a remarkably high risk of collisions with barriers (Bambach, Grzebieta, and McIntosh 2012; Mehrara Molan, Rezapour, and Ksaibati 2020; Venkatraman et al. 2021), and these kinds of crashes have been noted to be associated with high severity outcomes, research on motorcycle-barrier crash safety has not received much attention.

Cluster-regression analyses

Regarding research methodologies for exploring motorcycle-barrier crash datasets, safety experts mainly employed statistical regression-based methods to analyse the whole data without focusing on investigating the factors influencing critical groups in the data (Bambach, Grzebieta, and McIntosh 2012; Daniello and Gabler 2011a; Gabauer and Li 2015). While regression models constitute a powerful tool with a sound structure for investigating the individual influence of factors on dependent variables, it is also imperative to consider other methods that can help discover key factors influencing specific groups for the development of targeted countermeasures for addressing critical safety concerns.

It has been established that the effect of risk factors on crash outcomes is mostly unstable. Researchers first segment the crash data into unique groups to account for this instability before running different regression models on each dataset (Mannering 2018; Tamakloe et al. 2021). Recent studies in the safety field using data segmentation are gaining popularity as it helps researchers discover various risk factors influencing crashes in each homogeneous data segment with similar characteristics. For example, some studies manually segmented their data and employed statistical techniques to perform these analyses, which led to the discovery of newer insights for dealing with issues pertaining to each group identified (Fountas et al. 2020; Obaid et al. 2022; Tamakloe et al. 2021). In other experiments, researchers first employed robust machine learning techniques to automatically segment data into homogeneous subgroups. They then conducted further analyses on each subgroup with similar crash observations (Mohamed et al. 2013a; Z. Li et al. 2018b), which led to identifying associations between variables specific to each cluster and the quantitative impact of individual risk factors on injury severity outcomes to provide more targeted countermeasures for dealing with the safety problems.

In the literature, it has been identified that exploring crash data after segmenting it into different groups with homogeneous features using a cluster-regression approach often leads to identifying more interesting insights than employing only a traditional regression model (Liu, Lin, and Fan 2021). A study recently segmented a crash dataset into clusters using latent class analysis (LCA) and employed a binary regression model to analyse the data. The analysis identified interesting findings for each cluster that could not have been identified without data segmentation (Sivasankaran and Balasubramanian 2020). Similar results were obtained by combining LCA and mixed logit models (Liu, Lin, and Fan 2021; Mohamed et al. 2013b). LCA is a probabilistic modelling tool which allows for clustering based on some pre-estimated regression model – that is to say, the latent classes depend on the probabilities for the latent classes, and the analyst is expected to specify the number of latent classes based on some statistical measures. The LCA operates based on the premise that the observed distribution of the variables arises from a finite hidden mixture of underlying distributions (Sinha, Calfee, and Delucchi 2021). As the validity of the results depends on the model's assumptions, the research always has to consider them. Violations of these assumptions can lead to biased and unreliable estimates of the model's parameters, making the conclusions drawn from the analysis incorrect or misleading. Like LCA, cluster analyses employ algorithms to partition populations into homogeneous clusters that share common attributes. Various clustering algorithms, including hierarchical and k-means, utilise a subjective distance measure to recognise clusters. Therefore, determining the appropriate number of clusters is not based on any prior assumptions (Sinha, Calfee, and Delucchi 2021).

In addition to the basic cluster-regression approach, an emerging area of interest is the desire to identify the latent patterns in the various homogeneous data groups identified. The data clustering approach in this desired method, which can be accomplished using the novel Cluster Correspondence Analysis (CCA) and the traditional Multiple Correspondence Analysis (MCA), also provides new insights into the associations between factors in the clusters that could be used to inform policy decisions. More recently, the MCA and the ordered logistic regression were used to explore factors affecting motorcycle crashes in Thailand (Champahom et al. 2022). However, as the analyst's judgment plays a key role in selecting clusters in the MCA approach (Das, Hossain, et al. 2023), it may be subject to some level of bias. The unsupervised machine learning, CCA, which employs both dimension reduction and clustering methods for automatically recognising patterns of influential factors with relative contribution measures, is gaining traction in the traffic safety field (Ashifur Rahman, Das, and Sun 2022; Das, Hossain, et al. 2023; Das, Tran, and Theel 2021).

Study objective and contribution

Although the previous studies shed light on the factors influencing motorcycle-barrier crashes, research regarding the discovery of motorcycle-barrier crash contributory factors associated with homogeneous clusters in a dataset and the variables affecting the severity of these crashes for each latent homogeneous cluster is non-existent. This study aims to fill the literature gap by conducting a comprehensive investigation that identifies latent patterns/associations in critical clusters in motorcycle-barrier crash data together with the individual effect of the variables influencing injury severity outcomes in the various clusters identified. Compared with the previous studies, this research contributes to the literature by applying an innovative cluster-regression approach to motorcycle-barrier crash data. The method involves combining the unsupervised machine learning algorithm, CCA, with the ordered probit regression model (OPRM) to automatically determines plausible clusters, the key associations between factors in these clusters, and the individual impact of factors on injury severity outcomes which will contribute to the literature by providing knowledge on the specific groups of motorcycle-barrier crashes. Consequently, this information could be used to inform policy decisions on improving motorcycle safety.

Data description

A five-year highway crash dataset (2016–2020) containing crashes in Massachusetts was utilised for this study. The data were retrieved from the Massachusetts Department of Transport database. In the event of a crash, the police are dispatched to the scene to record the incident. These data are aggregated, quality checked and stored for analysis. In the dataset, injury severity levels are defined into three categories for this study. These are fatal

(K), non-fatal (ABC), and no injury (O). Fatal crashes are those crashes that result in death at the scene or within 30 days of the crash. Non-fatal are those crashes that result in an injury (no death), whereas no-injury crashes are those that only cause property damage only (PDO). To achieve the aim of the study, motorcycle-barrier crashes were extracted for analysis. In total, as shown in Table 1, 201 crash observations were used for this study. The data contained variables classified into roadway, environmental, rider, temporal, crash location, and crash characteristics. Under the crash characteristics category, rider injury severity and the number of vehicles involved in the crash were presented. The distribution of crashes by injury severity outcome indicates that 90% of crashes resulted in an injury (fatal: 9.9%; non-fatal: 80.1%) and 10% resulted in no injury, which is likely as the traffic speeds on the highway is high – increasing the chances of injury when a motorcycle-barrier crash occurs. This finding highlights the need to identify factors influencing injury severity outcomes of motorcycle-barrier crashes. It is noteworthy that most of the crashes (92%) were single-vehicle crashes.

Notably, the low number of PDO crashes is likely due to underreporting of such crashes. Previous studies have established that less severe injury crashes are often underreported (Salifu and Ackaah 2012; Watson, Watson, and Vallmuur 2015). This problem leads to significant uncertainty regarding the extent and consequences of no-injury crashes. However, due to the difficulty in capturing all the variables associated with unreported crashes, researchers only rely on the available data in the police-reported crash database.

Regarding the crash location/roadway features, the variable pertaining to the location of motorcycle-barrier crash impact, which has not been used in previous studies, was considered. Table 1 shows that motorcycles mostly crash into barriers located near the roadway (44%) or the roadside (29%). Besides, the crashes frequently occurred on mainline segments (68%) and at segments with no traffic control (87%). The road surfaces were mostly dry (85%), and the crash locations mostly had level/flat pavements (84%). Collisions on no/partial access-controlled segments were most common (48%), followed by crashes on full access-controlled segments (45%). This shows that these locations are critical to motorcycle safety, warranting a deeper investigation into the factors influencing crashes at such locations. Essentially, an uncontrolled access road is also a highway that allows vehicles to enter and leave the highway at any point. This can create confusion, cause slow traffic flow, and can be dangerous. Full access-controlled segments are those highways designed for high-speed operation in a controlled manner such that traffic ingress/egress via ramps. Highway-crossing traffic is also directed through bridges or tunnels. These highways provide unhindered traffic flow without interruption from traffic signals (FHWA Safety 2000).

Concerning environmental factors, it was observed that 88% of the crashes occurred during clear weather, and 67% of the collisions took place during daylight conditions. Looking at the temporal factors revealed that motorcycle crashes are common on weekdays (65%) and during summer (47%). For the rider's features, only the variable for the rider's age was available in the dataset. Early mid-aged drivers (25–44 years) were found to have been involved in the majority of the crashes, followed by late mid-aged riders (45–64 years) and younger riders (16–24 years). The high involvement rate reflects that young and mid-aged riders in Massachusetts are common victims of motorcycle crashes.

Table 1. Summary of variables used in the study.

Category	Variable	N	%	Cum. %	Category	Variable	N	%	Cum. %
Crash location and roadway characteristics					***Crash characteristics***				
First harmful event (FHE) location	Median (1 if the motorcycle hit an object at the median; 0 otherwise)	26	12.9	12.9	Severity (sev)	Fatal (1 if the crash was fatal; 0 otherwise)	20	9.9	9.9
	Roadside (1 if the motorcycle hit an object at the roadside; 0 otherwise)	59	29.4	42.3		Non-fatal (1 if the crash was non-fatal; 0 otherwise)	161	80.1	90.1
	Roadway (1 if the motorcycle hit an object near the roadway; 0 otherwise)	88	43.8	86.1		PDO (1 if the crash resulted in no injuries; 0 otherwise)	20	10.0	100.0
	Shoulder – paved (1 if the motorcycle hit an object at a paved shoulder location; 0 otherwise)	16	8.0	94.0		*Total*	*201*	*100.0*	*NA*
	Shoulder – unpaved (1 if the motorcycle hit an object at an unpaved shoulder location; 0 otherwise)	11	5.5	99.5	Number of vehicles involved (numv)	Multiple (1 if the crash occurred involved multiple vehicles; 0 otherwise)	16	8.0	8.0
	Unknown (1 if the motorcycle hit an object at an unpaved shoulder; 0 otherwise)	1	0.5	100.0		Single (1 if the crash involved only the motorcycle; 0 otherwise)	185	92.0	100.0
	Total	*201*	*100.0*	*Na*		*Total*	*201*	*100.0*	*NA*
Segment (sgm)	Intersection/roundabout (1 if crash occurred at an intersection/roundabout; 0 otherwise)	31	15.4	15.4	***Environmental characteristics***				
	Not at a junction (1 if the crash occurred on the mainline segment; 0 otherwise)	136	67.7	83.1	Weather (wth)	Clear (1 if the crash occurred in clear weather; 0 otherwise)	177	88.1	88.1
	On/off-ramp (1 if the crash occurred at the entrance or exit or a ramp; 0 otherwise)	32	15.9	99.0		Inclement (1 if the crash occurred in inclement weather; 0 otherwise)	21	10.4	98.5
	Unknown (1 if the crash location is unknown; 0 otherwise)	2	1.0	100.0		Unknown (1 if the weather condition is unknown; 0 otherwise)	3	1.5	100.0
	Total	*201*	*100.0*	*NA*		*Total*	*201*	*100.0*	*NA*

(*continued*).

Table 1. Continued.

Category	Variable	N	%	Cum. %	Category	Variable	N	%	Cum. %
Traffic control device (tcd)	Present (1 if the traffic control device is present; 0 otherwise)	26	12.9	12.9	Lighting condition (lgh)	Dark-lighted (1 if the crash occurred in dark-lighted condition; 0 otherwise)	38	18.9	18.9
	Absent (1 if the traffic control device is absent; 0 otherwise)	174	86.6	99.5		Dark – not lighted (1 if the crash occurred in dark-unlighted condition; 0 otherwise)	17	8.5	27.4
	Unknown (1 if the traffic control device is unknown; 0 otherwise)	1	0.5	100.0		Dark – unknown lighting (1 if the crash occurred in dark-unknown condition; 0 otherwise)	2	1.0	28.4
	Total	*201*	*100.0*	*NA*		Dawn/dusk (1 if the crash occurred in dawn or dusk condition; 0 otherwise)	9	4.5	32.9
Road surface condition (srf)	Dry (1 if the pavement is dry; 0 otherwise)	171	85.1	85.1		Daylight (1 if the crash occurred in daylight condition; 0 otherwise)	135	67.2	100.0
	Snow/mud/dirt/ice (1 if pavement is covered in snow/mud/dirt/ice; 0 otherwise)	10	5.0	90.1	*Total*		*201*	*100.0*	*NA*
	Unknown (1 if pavement condition is unknown; 0 otherwise)	10	5.0	95.0	***Riders' characteristics***				
	Wet (1 if pavement condition is wet; 0 otherwise)	10	5.0	100.0	Age	16–24 years (1 if motorcycle rider was aged between 16–26; 0 otherwise)	37	18.4	18.4
	Total	*201*	*100.0*	*NA*		25–44 years (1 if motorcycle rider was aged between 25–44; 0 otherwise)	96	47.8	66.2
Terrain (trr)	Level (1 if the terrain is 'level'; 0 otherwise)	168	83.6	83.6		45–64 years (1 if motorcycle rider was aged between 45–64; 0 otherwise)	55	27.4	93.5
	Mountainous (1 if the terrain is 'mountainous'; 0 otherwise)	2	1.0	84.6		65–84 years (1 if motorcycle rider was aged between 65–84; 0 otherwise)	13	6.5	100.0
	Rolling (1 if the terrain is 'rolling'; 0 otherwise)	20	10.0	94.6	*Total*		*201*	*100.0*	*NA*
	Unknown (1 if terrain type is unknown; 0 otherwise)	11	5.5	100.0	***Temporal characteristics***				
	Total	*201*	*100.0*	*NA*	Day of week (dow)	Weekday (1 if the crash occurred on a weekday; 0 otherwise)	130	64.7	64.7

Functional classification (fnc)	Interstate (1 if the crash occurred on an interstate road; 0 otherwise)	40	19.9	19.9		Weekend (1 if the crash occurred on a weekend; 0 otherwise)	71	35.3	100.0
	Local street/road (1 if the crash occurred on a local street/road; 0 otherwise)	16	8.0	27.9		*Total*	*201*	*100.0*	*NA*
	Major collector (1 if the crash occurred on a major collector road; 0 otherwise)	24	11.9	39.8	Season (ssn)	Autumn (1 if the crash occurred in autumn; 0 otherwise)	53	26.4	26.4
	Minor arterial (1 if the crash occurred on a minor arterial road; 0 otherwise)	41	20.4	60.2		Spring (1 if the crash occurred in spring; 0 otherwise)	47	23.4	49.8
	Minor collector (1 if the crash occurred on a minor arterial road; 0 otherwise)	1	0.5	60.7		Summer (1 if the crash occurred in summer; 0 otherwise)	94	46.8	96.5
	Principal arterial (1 if the crash occurred on a principal arterial road such as freeways/expressways; 0 otherwise)	69	34.3	95.0		Winter (1 if the crash occurred in winter; 0 otherwise)	7	3.5	100.0
	Unknown (1 if the functional class of road is unknown; 0 otherwise)	10	5.0	100.0		*Total*	*201*	*100.0*	*NA*
	Total	*201*	*100.0*	*NA*	Year	2016 (1 if the crash occurred in 2016; 0 otherwise)	46	22.9	22.9
Access control (acc)	Full (1 if the road is full access-controlled; 0 otherwise)	90	44.8	44.8		2017 (1 if the crash occurred in 2017; 0 otherwise)	36	17.9	40.8
	None (1 if the road is not full access-controlled; 0 otherwise)	97	48.2	93.0		2018 (1 if the crash occurred in 2018; 0 otherwise)	27	13.4	54.2
	Partial (1 if the road has partial access-control; 0 otherwise)	4	2.0	95.0		2019 (1 if the crash occurred in 2019; 0 otherwise)	39	19.4	73.6
	Unknown (1 if road control type is unknown; 0 otherwise)	10	5.0	100.0		2020 (1 if the crash occurred in 2020; 0 otherwise)	53	26.4	100.0
	Total	*201*	*100.0*	*NA*		*Total*	*201*	*100.0*	*NA*

Note: n = number of crash observations; % = percentage; cum. % = cumulative percentage.

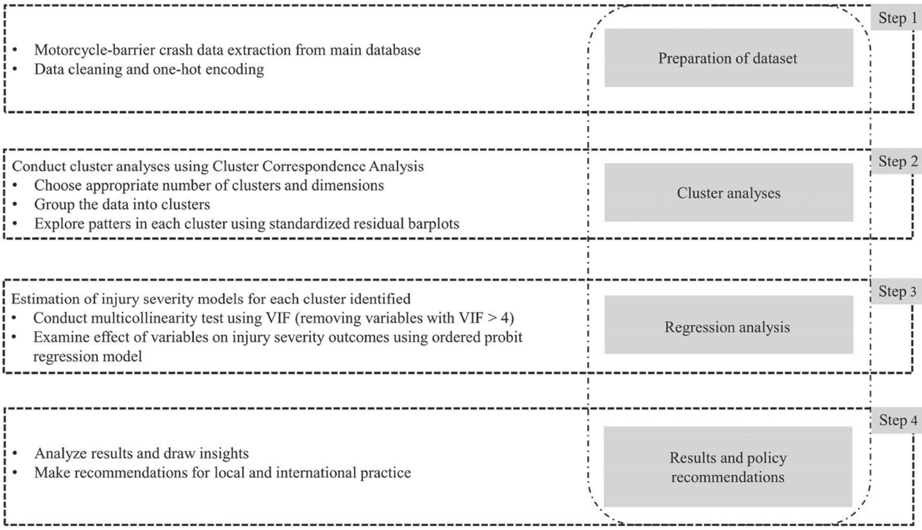

Figure 1. Proposed study procedure.

Methodology

The procedure outlined in Figure 1 was employed to achieve this study's goal. First, the dataset was prepared by removing observations with missing variable information. Afterward, CCA was used to automatically classify the observations in the crash dataset into meaningful groups and to identify patterns and associations between the factors in each group. Next, an OPRM was applied to explore the effect of risk factors on crash severity outcomes. Further, the results were explained, and policy recommendations were provided based on the findings. Note that the approach applied does not seek to improve the performance of the regression model but to provide additional insights into factors influencing the severity of motorcycle-barrier crashes for unique homogeneous clusters identified in the dataset. The methods employed for the study are presented in the sub-sections below.

Cluster correspondence analysis

CCA is a machine learning method that combines K-means cluster analysis and simple correspondence analysis to partition individual attributes into similar groups. Cluster analysis is an unsupervised machine learning technique researchers employ to allocate observations into homogeneous groups with high within-cluster similarity and low intra-cluster similarity (Das, Sun, et al. 2022b). When the number of observations is significantly large, the computation of cluster dissimilarities becomes cumbersome. CCA addresses this setback by simultaneously combining dimension reduction, cluster analysis for categorical data explorations, and correspondence analysis, with the primary goal of identifying a reasonable allocation of crash observations into different similar groups revealing patterns of critical factor associations in each group. It does this by simultaneously placing individuals into clusters and then optimally scaling values to categories such that the objective of maximising the single between variance is obtained (van de Velden, D'Enza, and Palumbo

2017). The CCA is more advantageous than the popular K-means algorithm, which only partitions data into k clusters with similar features. The K-means algorithm does not consider dimension reduction. Besides, the researcher cannot obtain patterns of influencing factors associated with each cluster by running the K-means algorithm. Research has identified that the CCA approach outperformed other joint dimension reduction methods and cluster analysis methods while revealing patterns of contributory factors associated with each cluster (Kong et al. 2022; Markos, D'Enza, and van de Velden 2019). Due to this, we decided to use the CCA approach for this study. The technique is described below.

Suppose that the motorcycle-barrier crash dataset contains n crash observations that are distributed on p categorical variables and gathered in a super indicator matrix $\boldsymbol{Z}$with dimensionality $n \times Q$ such that $Q = \sum_{j=1}^{p} q_j$ and $\boldsymbol{Z}_j 1_{q_j} = 1_n$. One way to conceptualise cluster membership is as a categorical variable, which can be represented by utilising the indicator matrix $\boldsymbol{Z}_K$ which is coded as an $n \times K$ indicator matrix that represents cluster membership. In order to examine the relationship between the clusters and the categorical variables, we create a table $\boldsymbol{F}$ that displays the distribution of cluster memberships across the different categories of the variables. The table $\boldsymbol{F}$ that shows the cross-tabulation cluster memberships with the categorical variables can be constructed as $\boldsymbol{F} = \boldsymbol{Z}'_K\boldsymbol{Z}$. By utilising correspondence analysis on this matrix, the optimal scaling values for rows (represented as clusters) and columns (represented as categories) that ensure maximum between-cluster variance is achieved. In a two-dimensional space, clusters are split optimally based on distributions over the categorical variables. The objective function for optimal cluster allocation can be expressed as:

$$max\emptyset_{clusca}(\boldsymbol{Z}_K, \boldsymbol{B}^*) = \frac{1}{p}\text{trace}\boldsymbol{B}^{*'}\boldsymbol{D}_z^{-\frac{1}{2}}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}_K\boldsymbol{D}_K^{-1}\boldsymbol{Z}'_K\boldsymbol{M}\boldsymbol{Z}\boldsymbol{D}_z^{-\frac{1}{2}}\boldsymbol{B}^* \tag{1}$$

subject to $\boldsymbol{B}^{*'}\boldsymbol{B}^* = \boldsymbol{I}_k$, where $\boldsymbol{B}$ represents the column coordinate matrix, $\boldsymbol{B}^* = \frac{1}{\sqrt{np}}\boldsymbol{D}_z^{\frac{1}{2}}\boldsymbol{B}$, $\boldsymbol{D}_K$ represents the diagonal matrix of $\boldsymbol{Z}'_K\boldsymbol{Z}_K$, $\boldsymbol{M} = \boldsymbol{I}_n - 1_n 1'_n/n$, and $\boldsymbol{D}_z$ connotes a diagonal matrix such that $\boldsymbol{D}_z 1_Q = \boldsymbol{Z}' 1_n$.

Optimal cluster allocation, $\boldsymbol{Z}_K$, is determined in such a way that Equation 1 is a maximum. This optimisation problem involves maximising $\emptyset(\boldsymbol{Z}_K, \boldsymbol{B}^*)$ with respect to $\boldsymbol{Z}_K$ can be transformed into a K-means clustering problem for a fixed $\boldsymbol{B}^*$. Where $\boldsymbol{G}$ is a matrix with cluster means, the resulting K-means objective needed to be solved is presented as follows (van de Velden, D'Enza, and Palumbo 2017):

$$min\emptyset_{clusca}(\boldsymbol{Z}_K, \boldsymbol{G}) = \boldsymbol{Y} - \boldsymbol{Z}_K\boldsymbol{G}^2 \tag{2}$$

where $\boldsymbol{Y}$, which contains subject coordinates obtained after optimal scaling, is computed as $\sqrt{\frac{n}{p}}MZ\boldsymbol{D}_z^{-\frac{1}{2}}\boldsymbol{B}^*$, and solving the K-means problem yields $\boldsymbol{G} = (\boldsymbol{Z}'_K\boldsymbol{Z}_K)^{-1}\boldsymbol{Z}'_K\boldsymbol{Y}$. Minimising the resulting K-means objective function with respect to both $\boldsymbol{G}$ and $\boldsymbol{Z}_K$ then amounts to maximising n trace $\frac{1}{p}\boldsymbol{B}^{*'}\boldsymbol{D}_z^{-\frac{1}{2}}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}_K\boldsymbol{D}_K^{-1}\boldsymbol{Z}'_K\boldsymbol{M}\boldsymbol{Z}\boldsymbol{D}_z^{-\frac{1}{2}}\boldsymbol{B}^*$ with respect to only $\boldsymbol{Z}_K$. Thus, for a fixed $\boldsymbol{B}^*$, a cluster allocation is obtainable by applying the K-means algorithm to $\boldsymbol{Y}$.

The iterative procedure for CCA is outlined as follows (van de Velden, D'Enza, and Palumbo 2017):

(1) First, an initial cluster allocation $\boldsymbol{Z}_K$ is generated through random assignments subject to clusters.
(2) Quantifications for $\boldsymbol{B}^*$ are determined through eigenvalue decomposition of $\frac{1}{p}\boldsymbol{D}_z^{-\frac{1}{2}}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}_K\boldsymbol{D}_K^{-1}\boldsymbol{Z}_K'\boldsymbol{M}\boldsymbol{Z}\boldsymbol{D}_z^{-\frac{1}{2}} = \boldsymbol{B}^*\boldsymbol{\Lambda}^2\boldsymbol{B}^{*'}$.
(3) Construction of an initial configuration for $\boldsymbol{Y}$.
(4) Finding an initial matrix of cluster updates and applying K-means clustering algorithm to $\boldsymbol{Y}$ to determine updates for $\boldsymbol{Z}_K$.
(5) Return to step 2 and repeat the procedure by using $\boldsymbol{Z}_K$ as the cluster allocation matrix until the algorithm converges.

Interested readers may find a detailed explanation of the CCA technique in the paper by van de Velden, D'Enza, and Palumbo (2017).

Ordered probit regression model

The OPRM is selected for modelling the impact of risk factors on injury severity outcomes in each cluster identified. Since the dataset has an ordered dependent variable (fatal, non-fatal injury, and PDO/no injury), the OPRM is appropriate for use. If S_i represents a motorcycle-barrier severity index expressed as a linear function of K factors such that $k = 1, 2, \ldots, K$, whose values for each crash observation i $(i = 1, 2, \ldots, N)$ are X_{ik}, then the structure of the model can be written as (Borooah 2002; Mohamed et al. 2013b):

$$S_i = \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i = Z_i + \varepsilon_i \tag{3}$$

where x_{ik} represents the row vector of explanatory variables, β_k denotes a vector of estimable parameters, and ε_i is the normally distributed error component. Let Y_i be a variable that can be associated with the severity levels of a crash. Then, three categories in the dependent variable in this study are given as (Mohamed et al. 2013b):

$$Y_i = \begin{cases} 1 & if\ S_i \leq \varphi_1 \\ 2 & if\ \varphi_1 \leq S_i \leq \varphi_2 \\ 3 & if\ S_i \geq \varphi_2 \end{cases} \tag{4}$$

where φ_1 and φ_2 represent the cut-off/threshold points of the severity categories computed from the dataset. Consequently, the probability of crash observation Y_i sustaining an injury of severity 1 (PDO), 2 (non-fatal), or 3 (fatal) can be expressed as (Borooah 2002):

$$\begin{aligned} \Pr(Y_i = 1) &= \Pr(\varepsilon_i \leq \varphi_1 - Z_i) \\ \Pr(Y_i = 2) &= \Pr(\varphi_1 - Z_i < \varepsilon_i \leq \varphi_2 - Z_i) \\ \Pr(Y_i = 3) &= \Pr(\varepsilon_i \geq \varphi_2 - Z_i) \end{aligned} \tag{5}$$

By treating each motorcycle-barrier crash observation as a single draw sampled from a multinomial distribution with three possible outcomes (fatal, non-fatal injury, and PDO/no injury), we can have that, for each of the N crash observations indexed as $i = 1, 2, \ldots, N$, N_1

were PDO, N_2 were non-fatal, and N_3 were fatal. Based on this, the likelihood function for observing the sample can be formulated as follows (Borooah 2002):

$$L = [\Pr(y_i = 1)]^{N_1}[\Pr(y_i = 2)]^{N_2}[\Pr(y_i = 3)]^{N_3}$$
$$L = [F(\varphi_1 - Z_i)]^{N_1}[F(\varphi_2 - Z_i) - F(\varphi_1 - Z_i)]^{N_2}[1 - F(\varphi_1 - Z_i)]^{N_3} \quad (6)$$

Where the cumulative probability distribution of the error terms is represented by $F(x) = \Pr(\varepsilon_i < x)$. To understand the impact of each independent variable on the dependent variable, marginal effect (ME) estimates are computed. Before running the models, Variance Inflation Factor (VIF) tests were computed to identify possible multicollinearity issues between the independent variables. The variables selected for the final models had VIFs less than 4, as suggested in the literature (O'brien 2007). The computations were all carried out using STATA 17.

Results and discussions

Discussion of results from the cluster correspondence analysis

The CCA technique was employed to explore the motorcycle-barrier crash data. All the variables in the data were input into the CCA model. Determining an optimal number of clusters and dimensions requires careful assessment to guarantee the results' accuracy. Functions in the R software package 'clustrd' were employed to facilitate the quantitative appraisal of solutions obtained from different parameter settings (Markos, D'Enza, and van de Velden 2019). This study applied an MCA K-means approach and the average silhouette width (ASW) criterion, a distance-based statistic ranging from −1 to 1 used to assess the quality of clustering solutions, segregate the data into clusters, and determine the optimal number of clusters. A possible range of clusters between 2 and 10 and dimensions ranging from 1 to 9 with 100 random starts (to prevent local minima issues) was specified. After several runs, it was identified that 3 clusters and 2 dimensions produced the best solution with the highest ASW value (0.178). The cluster-specific ASW values corresponding to the selected solution are 0.17, 0.17, and 0.37. Note that the higher the ASW value, the higher the compactness of the cluster depicting a better cluster separation (Markos, D'Enza, and van de Velden 2019). The objective criterion value of the analysis was found to be 5.04.

Figure 2 illustrates the biplots for the first 2 dimensions (the abbreviations in the figure are described in Table 1). As presented on the biplots, the algorithm segregated the data into three clusters. The first two clusters represent 95% of the data. As shown in Table 2, the centroid of Cluster 1 (C1) is found in the third quadrant. Essentially, the plot points represent the attributes' location in two-dimensional space. From this, one can observe the distances from the points to the origin of the biplot. Notably, the origin of the biplot represents the mean profile, and the points reflect variations from the average profile.

The centroids of Clusters 2 and 3 (C2 and C3) are located in quadrants 2 and 1, respectively. Since the CCA algorithm minimises the distances within each cluster and maximises the distances between clusters, the features in each of the three clusters are strongly associated with each other. From Figure 2, the centroids of C1 and C2 are identified to be closer to the plot's centre than C3, signifying that the features in C1 and C2 show characteristics that align closely with most of the dataset. As shown in Table 2, these clusters explain the majority of the data compared to C3.

Figure 2. Biplot showing the clusters identified in the motorcycle-barrier crash data.

Table 2. Measures associated with the clusters identified.

	Cluster centroid locations				
Cluster	Dimension 1	Dimension 2	Within cluster sum of Squares	Size	Coverage (%)
C1	−0.0078	−0.0624	0.0020	100	50
C2	−0.0222	0.0681	0.0015	91	45
C3	0.2777	0.0307	0.0004	10	5
Sum				*201*	*100*

To understand the associations between the features in the dataset, bar plots for each cluster containing the top 30 variables with the highest standardised residuals were plotted and presented in Figures 2 through 4. For each bar plot, the bars' length shows the variable's dominance in the cluster. Again, for each bar plot, those variables with positive (negative) residuals have above (below) average frequency within the cluster being considered. According to the bar plots, it appears that the motorcycle-barrier crash data was clustered based on the access control type of the road since access control-related variables happened to be the top standardised variables in each plot (C1 = no/partial access control; C2 = full access-controlled; C3 = access-control unknown). Following previous studies,

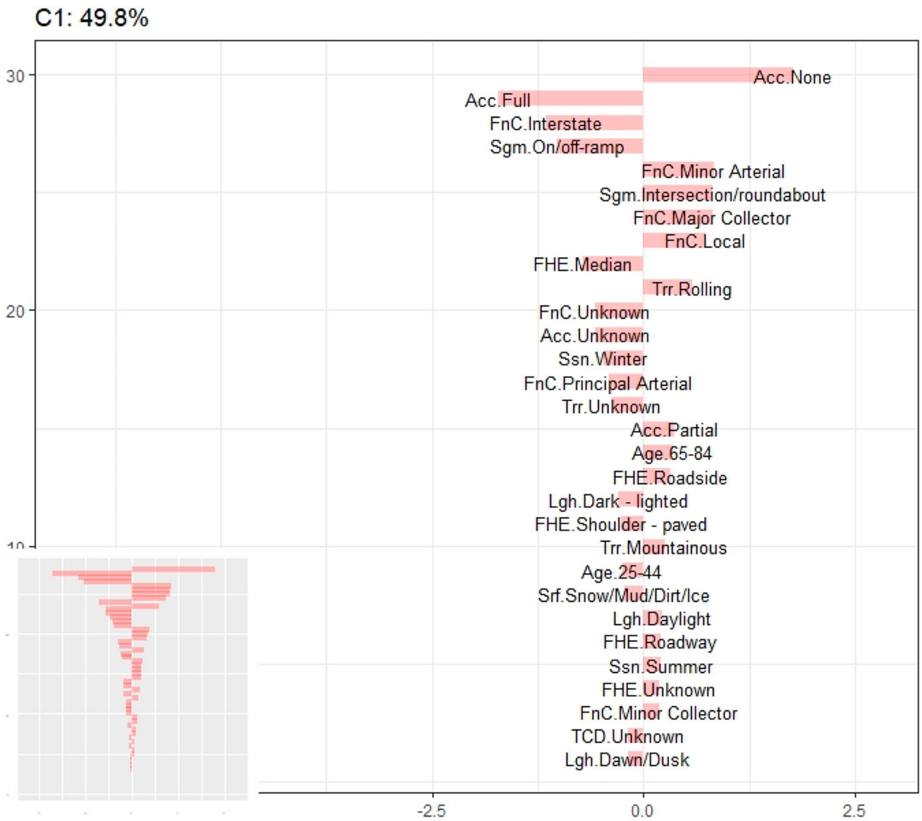

Figure 3. Top 30 largest standardised residuals in Cluster 1.

the cluster factor associations identified are explained using the bar plots with positive residuals (Das, Sun, et al. 2022b; Ashifur Rahman, Das, and Sun 2022).

Cluster 1 (no/partial access-controlled segment)

From the bar plot for C1, it can be observed that the top six (6) positive residual mean variables are no/partial access control, minor arterial, intersection/roundabout, major collectors, local street/road, and rolling terrain (see Figure 3; the abbreviations in the figure are provided in Table 1). The cluster demonstrates strong associations between intersection/roundabout or collector/local/minor arterial road-related motorcycle crashes with rolling terrains. From the remaining variables, it can be seen that these crashes are likely to occur by aged riders (65–84 years) hitting barriers/guardrails located around the roadside during the daylight conditions in summer.

Intersections/roundabouts are complex in nature. When older riders face this complexity, coupled with slow reaction times and the short-sight distance problem associated with rolling terrains, they are likely to collide with barriers while trying to prevent rear-end crashes. Consistent with the literature, rolling terrain has been found to increase the likelihood of crash occurrences (Agbelie 2016; Ye et al. 2009).

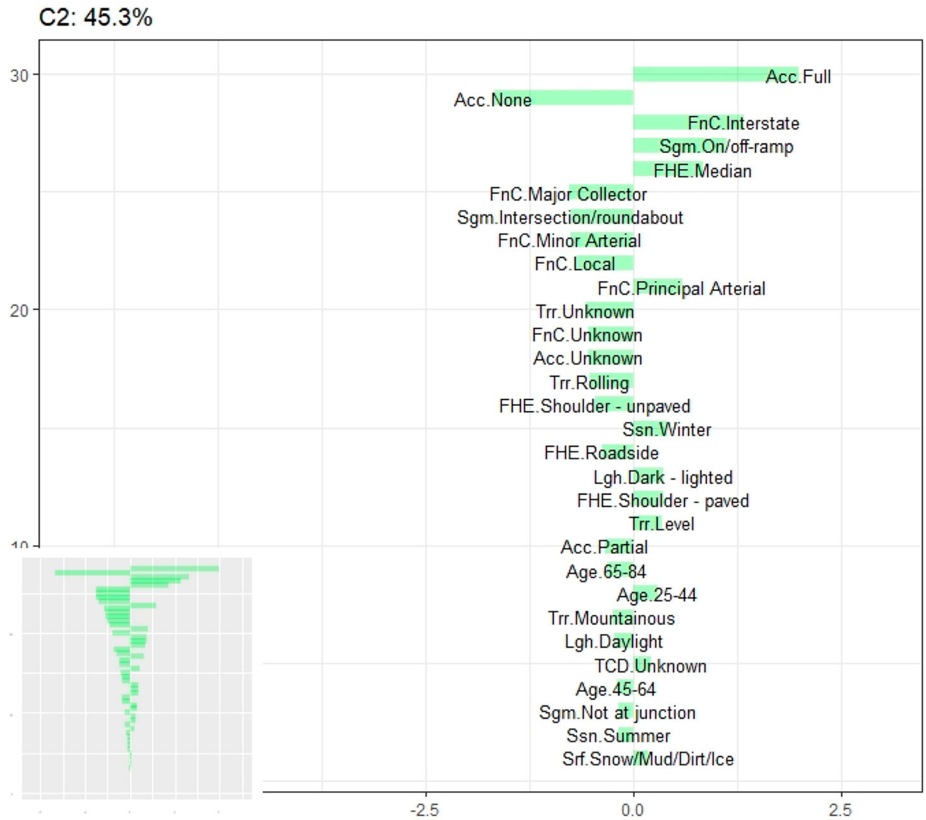

Figure 4. Top 30 largest standardised residuals in Cluster 2.

Cluster 2 (full access-controlled segment)

From C2, the top six (6) variables with the highest positive residuals are full access control, interstate roads, on/off ramps, medians, principal arteries, and winter (see Figure 4; Table 1 provides a description of the abbreviations illustrated in the figure). Essentially, motorcycle-barrier crashes occurring at full access control segments are likely to occur around the medians on interstates or principal arterial roads during winter. Considering the remaining variables, these crashes are likely to be associated with early mid-aged riders (25–44 years) riding on level terrain roads and hitting barriers/guardrails around the median separation area during the night on dark-but-lighted roads. This finding is plausible as riders will likely be less careful when riding on level terrain roads and dark-but-lighted segments due to their increased safety perception towards these roads. Besides, riders are likely to have consumed alcohol at night, which has been noted to lead to less compliance with road safety regulations (Das, Hossain, et al. 2023; Kasantikul et al. 2005). The careless behaviour of some riders on good/level roads coupled with high speeds and slippery pavements on interstate/main arterial roads during the winter are likely to cause these crashes (Hong, Tamakloe, and Park 2020).

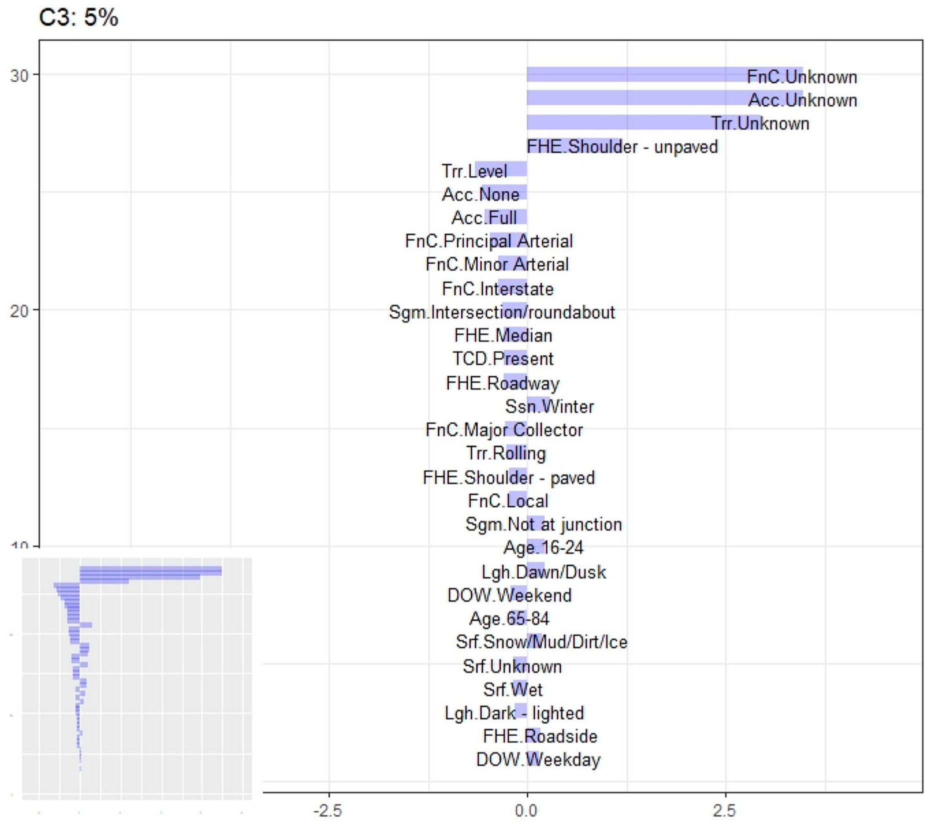

Figure 5. Top 30 largest standardised residuals in Cluster 3.

Cluster 3 (unknown access control status)

Figure 5 shows that C3 contains motorcycle-barrier crashes occurring on roads with unknown functional classes, access controls, and terrain types (Table 1 provides a description of the abbreviations in the figure). From the results, much younger riders (16–24 years) are likely to be involved in crashes that occur during the weekdays on mainline segments during the winter season at dawn/dusk. This cluster represents only 5% of the total data. Notably, these crashes are likely to occur because of hitting barriers/guardrails near unpaved shoulders or around the roadside. Young riders are known to engage in risky behaviours such as speeding and disregarding traffic regulations, likely to cause barrier-related crashes on snowy mainline segments (Tamakloe, Das, et al. 2022).

Discussion of results from the ordered probit regression model

Ordered probit models were estimated to observe how risk factors influence motorcycle-barrier crash injury severity in the clusters identified. Notably, only the first two clusters were considered for this analysis, as the sample size for the third cluster was insufficient (observations = 14). Based on the rules of thumb, the data size for the other clusters is sufficient for analysis (Wilson Van Voorhis and Morgan 2007). The estimation results and the ME estimates are summarised in Table 3. Only variables found to be statistically significant at a 90% confidence level were maintained in the model. From the table, most variables that were found to be significant in the 'full access control' cluster models were not significant in the 'no/partial access control' cluster model (and vice-versa).

Crash location and roadway characteristics

The variable connoting motorcycle crashes into barriers near the roadway was found to be significant only in the no/partial access control cluster data. The results presented in Table 3 showed that motorcycles crashing into barriers/guardrails near the roadway area on no/partial access-controlled segments are likely to have low severity levels. The ME estimates show that when a crash occurs on such roads, there is about a 10.6% chance of it being PDO. The propensity for observing non-fatal and fatal injuries decreased on road segments with no/partial access control by 1.9% and 8.6%, respectively.

Regarding the variable for the functional class, the variable for motorcycle-barrier crashes occurring on major collector and principal arterial road segments with no/partial access controls was significant, and motorcyclists in these crashes were observed to have a high chance of sustaining higher injury severity outcomes. The variable increases the chance of fatalities by 11% and 8.6%, respectively, and reduces the propensity of non-fatal or PDO. These crashes are likely to be associated with higher speeds. A previous study also showed a high likelihood of a fatal crash on non-interstate road segments with no/partial access controls (Penmetsa and Pulugurtha 2019).

Concerning the pavement condition, motorcycle-barrier crashes occurring on dry paved roads were found to be more likely to result in injuries, irrespective of the type of access control present. Based on the ME estimates, it was observed that the likelihood of fatal and non-fatal injuries increases when a crash occurs on a dry pavement road (see Table 3). However, the probability of no injury reduces when a motorcycle-barrier crash occurs on a road with dry pavement. Generally, dry pavements have been identified as associated with higher motorcycle crash severity outcomes (Shaheed and Gkritza 2014). Motorcyclists are less likely to ride carefully when they perceive the road conditions as good/perfect. Since dry roads provide ideal conditions for riders, they are likely to overspeed and be less attentive (Tamakloe, Hong, et al. 2022). A motorcycle-barrier crash caused by these conditions is likely to result in minor or fatal injuries.

The influence of variables pertaining to the traffic control device on motorcycle-barrier crash injury severity was explored. The findings shown in Table 3 depict that injuries are expected when the collision occurs in a segment with no traffic control device. Essentially, the probability of PDO decreases by 13.4% while the propensity for non-fatal and fatal injuries increases by 7.8% and 5.6%, respectively. This result was in line with other studies in the literature (Aidoo and Amoh-Gyimah 2020) and highlighted the need to consider providing traffic control devices where necessary to reduce the severity of crashes.

Motorcycle-barrier crashes occurring on full access-controlled segments with level/flat terrains were found to have a high chance of being associated with injury. The marginal effect estimates show that the probability of observing non-fatal injuries increases highest (56.5%) followed by fatalities (6.0%). This result is consistent with the findings in the literature on motorcycle crashes that showed that non-fatal injuries are more likely when a crash

Table 3. Ordered probit model estimation results for motorcycle-barrier crashes no access and full access-controlled clusters.

	No/partial access-controlled segment cluster					Full access-controlled segment cluster				
			Marginal effect estimates					Marginal effect estimates		
Variable	Coefficient	z stats	PDO	Non-fatal injury	Fatal injury	Coefficient	z stats	PDO	Non-fatal injury	Fatal injury
Crash location and roadway characteristics										
First harmful event location										
Roadway	−0.721	−2.42	0.1061	−0.0200	−0.0862	–	–	–	–	–
Functional classification										
Major collector	0.700	1.96	−0.0773	−0.0328	0.1101	–	–	–	–	–
Principal arterial	0.574	1.65	−0.0660	−0.0197	0.0857	–	–	–	–	–
Road surface condition										
Dry	1.144	2.52	−0.2742	0.2016	0.0726	1.455	3.07	−0.2292	0.1442	0.0850
Traffic control device										
Absent	0.681	1.65	−0.1340	0.0782	0.0557	–	–	–	–	–
Terrain										
Level	–	–	–	–	–	2.336	2.72	−0.6254	0.5653	0.0601
Environmental characteristics										
Lighting condition										
Dark – lighted	−1.074	−2.66	0.2436	−0.1689	−0.0747	–	–	–	–	–
Season										
Summer	–	–	–	–	–	0.7620	1.90	−0.0503	−0.0410	0.0914
Spring	–	–	–	–	–	1.3820	2.92	−0.0591	−0.1992	0.2583
Thresholds										
/Cut1	−0.067	−0.11	–	–	–	2.213	2.20	–	–	–
/Cut2	2.933	4.31	–	–	–	5.715	4.49	–	–	–
Model statistics										
Number of observations			97					90		
Number of parameters			6					4		
Pseudo r-squared			0.190					0.209		
Log likelihood at conv.			−54.379					−42.01		

occurs on the road with a levelled surface compared to inclines (Agyemang, Adanu, and Jones 2021; Aidoo and Amoh-Gyimah 2020).

Environmental characteristics

Variables for lighting conditions were significant in models for crashes on no/partial access control road segments only. For example, crashes on no/partial access control road segments were more likely to have low severity outcomes, which is reasonable as the riders have improved visibility compared to cases with no lighting present (Agyemang, Adanu, and Jones 2021). The ME estimates show that PDO propensity increases by 24.4% when a motorcycle-barrier crash occurs on a full access control road.

The estimated models showed that season-related variables, namely summer and spring, significantly impact full access-controlled segment motorcycle-barrier crash injury severity outcomes. Relative to winter, the results showed that it would likely have higher injury severity outcomes when a crash occurs in summer or spring. From the ME estimates, it was observed that the likelihood of fatal injuries increases by 9.1% and 25.8% when the crash occurs in summer and spring, respectively, while that of non-fatal and PDO reduces by 4.1% and 5.0%, and 19.9% and 5.9%, respectively. Unlike in winter, motorcyclists are likely to overspeed in the summer and spring seasons on full access-controlled segments. This finding was also identified in previous motorcycle safety research (Pai and Saleh 2008).

Conclusion

Motorcycles have become an important means of commute. The high fatalities associated with crashes involving motorcycles, particularly motorcycle-barrier collisions, have attracted the attention of many researchers; nevertheless, the existing studies failed to explore key factor associations and the impact of explanatory variables on injury severity outcomes for critical groups of motorcycle-barrier crashes. This study contributes to the motorcycle-barrier safety literature by employing a two-step approach combining the robust CCA and an OPRM to homogeneously cluster motorcycle-barrier crash data, reveal associations between risk factors in each identified cluster, and examine the effect of risk factors on crash injury severity outcomes to help identify useful insights for improving motorcycle safety.

The CCA technique clustered the motorcycle data into homogeneous groups based on the type of access control present – none/partial, full, and unknown. Some critical observations were made by assessing the patterns and associations among variables in each cluster and the factors affecting injury severity outcomes. First, it was deduced that the groups of factors likely to cause motorcycle-barrier crashes differed based on the access-control type of the road on which the crash occurred. Essentially, while the first cluster revealed that motorcycle-barrier crashes at no/partial access-controlled segments are strongly associated with older riders crashing into barriers/guardrails on low-speed road segments with rolling terrains during daylight conditions in summer, the second cluster showed that motorcycle-barrier crashes at full access-controlled segments are strongly associated with mid-aged riders colliding with barriers/guardrails on dark-but-lighted median-separated high-speed roads with levelled and slippery pavements. The third cluster showed associations of factors influencing crashes on roads with unknown access control status.

The severity analysis using the OPRM showed that different factors influence motorcycle-barrier crash injury severity outcomes differently depending on the type of access control present. As observed, dry pavements increased the risk of non-fatal injury on both full and no/partial access-control highways. However, traffic control absence and level pavements increased the non-fatality risk for crashes on no/partial and full access-controlled segments, respectively. Spring and summer variables increased the chance of fatalities on the high-speed full access-controlled segments. Besides, the variables for major collector/principal arterial roads increased the fatality risk of no access control roads.

Overall, the key insight drawn from this study is that the factors correlated with motorcycle-barrier crashes on highways are differentiated based on the access control type present at the crash segment. Specifically, different factors are strongly associated with crashes occurring at each segment type. Besides, the determinants of injury severity outcomes for each cluster varies. These interesting findings would not have been identified without the application of the cluster-regression technique employed in this paper. The information regarding the patterns of groups of factors identified and the effect of factors influencing injury severity of motorcycle-barrier crashes in each cluster could be used to inform engineering, education, and enforcement policy decisions to improve motorcycle safety. For instance, educating older on the need to ride carefully at low-speed sections on no/partial access-controlled road segments and installing clearly visible road signs to warn riders of dangers on rolling terrains could reduce their involvement in crashes. Enforcement could be heightened to check overspeeding on levelled roads and ramps located on interstates/arterials to minimise the crashes involving 25–44-year-old riders. It is also crucial to further increase visibility on dark-lighted roads and institute measures to discourage drunk riding, which is common at night. Finally, since motorcycle-barrier crashes on both full access and no/partial access-controlled segments mostly involve crashes into median/shoulder barriers and roadside barriers, respectively, considering the redesigning of these segments by placing median/shoulder/roadside barriers away from the through travel lanes could provide some recovery area for motorcyclists. Regarding injury severity, educating and encouraging riders to ride carefully in good conditions during summer and spring on level roads would be worthwhile to reduce fatalities on full access-controlled segments. Installing road signs and traffic control devices where necessary would effectively reduce fatalities among 65-84-year-old riders at no/partial access control segments on principal arterials and major collector highways. To reduce the severity of motorcycle-barrier crashes, it is recommended to educate riders on the need to ride carefully on dry pavement roads, wear protective gear as they ride, and consider measures such as installing paddings on barriers, especially wire rope safety barriers.

The current study is not without limitations. First, a major limitation is that the dataset did not contain other key variables related to the rider, the roadway geometry, and the weather. Besides, the variables, such as the violations leading to the crash, were missing from the dataset. Including these variables in future studies could provide further information leading to a better understanding of motorcycle-barrier crash mechanisms. Further, the crash data employed for the analysis included crashes occurring during the COVID-19 pandemic. The lower traffic volumes due to weeks of the lockdown will likely result in lower severity outcomes. In future studies, it would be worthwhile to consider performing analysis for the individual years and comparing the stability of factors across them.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Reuben Tamakloe http://orcid.org/0000-0002-9605-8439
Subasish Das http://orcid.org/0000-0002-1671-2753
Emmanuel Kofi Adanu http://orcid.org/0000-0002-1950-022X
Dongjoo Park http://orcid.org/0000-0002-6330-4437

References

Abdul Manan, Muhammad Marizwan, Thomas Jonsson, and András Várhelyi. 2013. “Development of a Safety Performance Function for Motorcycle Accident Fatalities on Malaysian Primary Roads.” *Safety Science* 60 (December): 13–20. https://doi.org/10.1016/j.ssci.2013.06.005.

Agbelie, Bismark R.D.K. 2016. “Random-Parameters Analysis of Highway Characteristics on Crash Frequency and Injury Severity.” *Journal of Traffic and Transportation Engineering (English Edition)* 3 (3): 236–242. https://doi.org/10.1016/j.jtte.2015.09.006.

Agyemang, William, Emmanuel Kofi Adanu, and Steven Jones. 2021. “Understanding the Factors That Are Associated with Motorcycle Crash Severity in Rural and Urban Areas of Ghana.” *Journal of Advanced Transportation* 2021 (December): 1–11. https://doi.org/10.1155/2021/6336517.

Aidoo, Eric Nimako, and Richard Amoh-Gyimah. 2020. “Modelling the Risk Factors for Injury Severity in Motorcycle Users in Ghana.” *Journal of Public Health* 28 (2): 199–209. https://doi.org/10.1007/s10389-019-01047-7.

Ashifur Rahman, M., Subasish Das, and Xiaoduan Sun. 2022. “Using Cluster Correspondence Analysis to Explore Rainy Weather Crashes in Louisiana.” *Transportation Research Record: Journal of the Transportation Research Board* March: 036119812210825. https://doi.org/10.1177/03611981221082582.

Bambach, M. R., R. H. Grzebieta, and A. S. McIntosh. 2012. “Injury Typology of Fatal Motorcycle Collisions with Roadside Barriers in Australia and New Zealand.” *Accident Analysis & Prevention* 49 (November): 253–260. https://doi.org/10.1016/j.aap.2011.06.016.

Bambach, M. R., R. H. Grzebieta, and A. S. McIntosh. 2013. “The Crash Mechanics of Fatal Motorcycle–Barrier Collisions in Australasia.” *Journal of Transportation Safety & Security* 5 (1): 66–77. https://doi.org/10.1080/19439962.2012.700389.

Borooah, V. K. 2002. *Logit and Probit: Ordered and Multinomial Models*. California: Sage Publications. https://books.google.co.kr/books?hl = en&lr = &id = H2mxYlRrcCoC&oi = fnd&pg = PP6&ots = xESuY6g07M&sig = dsWswcf34uZP5XLE-ykDw0xGPis&redir_esc = y#v = onepage&q&f = false.

Champahom, Thanapong, Panuwat Wisutwattanasak, Kattreeya Chanpariyavatevong, Napat Laddawan, Sajjakaj Jomnonkwao, and Vatanavongs Ratanavaraha. 2022. “Factors Affecting Severity of Motorcycle Accidents on Thailand’s Arterial Roads: Multiple Correspondence Analysis and Ordered Logistics Regression Approaches.” *IATSS Research* 46 (1): 101–111. https://doi.org/10.1016/j.iatssr.2021.10.006.

Chen, Huey-Kuo, Huey-Wen Chou, and Sam-Chu Hung. 2019. “Interrelationships between Behaviour Intention and Its Influential Factors for Consumers of Motorcycle Express Cargo Delivery Service.” *Transportmetrica A: Transport Science* 15 (2): 526–555. https://doi.org/10.1080/23249935.2018.1509401.

Daniello, Allison, Danielle Cristino, and Hampton C. Gabler. 2013. “Relationship between Rider Trajectory and Injury Outcome in Motorcycle-to-Barrier Crashes.” *Transportation Research Record: Journal of the Transportation Research Board* 2388 (1): 47–53. https://doi.org/10.3141/2388-07.

Daniello, Allison, and Hampton C. Gabler. 2011a. "Effect of Barrier Type on Injury Severity in Motorcycle-to-Barrier Collisions in North Carolina, Texas, and New Jersey." *Transportation Research Record: Journal of the Transportation Research Board* 2262 (1): 144–151. https://doi.org/10.3141/2262-14.

Daniello, Allison, and Hampton C. Gabler. 2011b. "Fatality Risk in Motorcycle Collisions with Roadside Objects in the United States." *Accident Analysis & Prevention* 43 (3): 1167–1170. https://doi.org/10.1016/j.aap.2010.12.027.

Das, Subasish. 2021. "Identifying Key Patterns in Motorcycle Crashes: Findings from Taxicab Correspondence Analysis." *Transportmetrica A: Transport Science* 17 (4): 593–614. https://doi.org/10.1080/23249935.2020.1802362.

Das, Subasish, Md Mahmud Hossain, M. Ashifur Rahman, Xiaoqiang Kong, Xiaoduan Sun, and G. M. Al Mamun. 2023. "Understanding Patterns of Moped and Seated Motor Scooter (50 Cc or Less) Involved Fatal Crashes Using Cluster Correspondence Analysis." *Transportmetrica A: Transport Science* 19 (2): 2029613. https://doi.org/10.1080/23249935.2022.2029613.

Das, Subasish, Xiaoduan Sun, Bahar Dadashova, M. Ashifur Rahman, and Ming Sun. 2022b. "Identifying Patterns of Key Factors in Sun Glare-Related Traffic Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2676 (2): 165–175. https://doi.org/10.1177/03611981211037891.

Das, Subasish, Ly-Na Tran, and Magdalena Theel. 2021. "Understanding Patterns in Marijuana Impaired Traffic Crashes." *Journal of Substance Use* 26 (1): 21–29. https://doi.org/10.1080/14659891.2020.1760381.

Farid, Ahmed, and Khaled Ksaibati. 2021. "Modeling Severities of Motorcycle Crashes Using Random Parameters." *Journal of Traffic and Transportation Engineering (English Edition)* 8 (2): 225–236. https://doi.org/10.1016/j.jtte.2020.01.001.

FHWA Safety. 2000. *Road Function Classifications*. https://safety.fhwa.dot.gov/speedmgt/data_facts/docs/rd_func_class_1_42.pdf.

Fountas, Grigorios, Achille Fonzone, Niaz Gharavi, and Tom Rye. 2020. "The Joint Effect of Weather and Lighting Conditions on Injury Severities of Single-Vehicle Accidents." *Analytic Methods in Accident Research* 27 (September): 100124. https://doi.org/10.1016/j.amar.2020.100124.

Gabauer, Douglas J. 2016. "Characterization of Roadway Geometry Associated with Motorcycle Crashes Into Longitudinal Barriers." *Journal of Transportation Safety & Security* 8 (1): 75–96. https://doi.org/10.1080/19439962.2014.984886.

Gabauer, Douglas J., and Xiaolong Li. 2015. "Influence of Horizontally Curved Roadway Section Characteristics on Motorcycle-to-Barrier Crash Frequency." *Accident Analysis & Prevention* 77 (April): 105–112. https://doi.org/10.1016/j.aap.2015.02.006.

Grzebieta, Raphael, Mike Bambach, and Andrew McIntosh. 2013. "Motorcyclist Impacts into Roadside Barriers." *Transportation Research Record: Journal of the Transportation Research Board* 2377 (1): 84–91. https://doi.org/10.3141/2377-09.

Hong, Jungyeol, Reuben Tamakloe, and Dongjoo Park. 2020. "Application of Association Rules Mining Algorithm for Hazardous Materials Transportation Crashes on Expressway." *Accident Analysis & Prevention* 142 (July): 105497. https://doi.org/10.1016/j.aap.2020.105497.

Jama, Hussein H., Raphael H. Grzebieta, Rena Friswell, and Andrew S. McIntosh. 2011. "Characteristics of Fatal Motorcycle Crashes into Roadside Safety Barriers in Australia and New Zealand." *Accident Analysis & Prevention* 43 (3): 652–660. https://doi.org/10.1016/j.aap.2010.10.008.

Kasantikul, Vira, James V. Ouellet, Terry Smith, Jetn Sirathranont, and Viratt Panichabhongse. 2005. "The Role of Alcohol in Thailand Motorcycle Crashes." *Accident Analysis & Prevention* 37 (2): 357–366. https://doi.org/10.1016/j.aap.2004.07.006.

Kong, Xiaoqiang, Subasish Das, Yunlong Zhang, Lingtao Wu, and Jason Wallis. 2022. "In-Depth Understanding of Near-Crash Events Through Pattern Recognition." *Transportation Research Record: Journal of the Transportation Research Board* June: 036119812210973. https://doi.org/10.1177/03611981221097395.

Li, Zhenning, Cong Chen, Yusheng Ci, Guohui Zhang, Qiong Wu, Cathy Liu, and Zhen (Sean) Qian. 2018b. "Examining Driver Injury Severity in Intersection-Related Crashes Using Cluster Analysis and Hierarchical Bayesian Models." *Accident Analysis & Prevention* 120 (November): 139–151. https://doi.org/10.1016/j.aap.2018.08.009.

Li, Ning, Byungkyu Brian Park, and James H. Lambert. 2018. "Effect of Guardrail on Reducing Fatal and Severe Injuries on Freeways: Real-World Crash Data Analysis and Performance Assessment." *Journal of Transportation Safety & Security* 10 (5): 455–470. https://doi.org/10.1080/19439962.2017.1297970.

Liu, Shaojie, Zijing Lin, and Wei (David) Fan. 2021. "Investigating Contributing Factors to Injury Severity Levels in Crashes Involving Pedestrians and Cyclists Using Latent Class Clustering Analysis and Mixed Logit Models." *Journal of Transportation Safety & Security* July: 1–28. https://doi.org/10.1080/19439962.2021.1958037.

Mannering, Fred. 2018. "Temporal Instability and the Analysis of Highway Accident Data." *Analytic Methods in Accident Research* 17 (March): 1–13. https://doi.org/10.1016/j.amar.2017.10.002.

Markos, Angelos, Alfonso Iodice D'Enza, and Michel van de Velden. 2019. "Beyond Tandem Analysis: Joint Dimension Reduction and Clustering in R." *Journal of Statistical Software* 91 (10), https://doi.org/10.18637/jss.v091.i10.

Mehrara Molan, Amirarsalan, Mahdi Rezapour, and Khaled Ksaibati. 2020. "Modeling the Impact of Various Variables on Severity of Crashes Involving Traffic Barriers." *Journal of Transportation Safety & Security* 12 (6): 800–817. https://doi.org/10.1080/19439962.2018.1547995.

Mohamed, Mohamed Gomaa, Nicolas Saunier, Luis F. Miranda-Moreno, and Satish V. Ukkusuri. 2013a. "A Clustering Regression Approach: A Comprehensive Injury Severity Analysis of Pedestrian–Vehicle Crashes in New York, US and Montreal, Canada." *Safety Science* 54 (April): 27–37. https://doi.org/10.1016/j.ssci.2012.11.001.

Mohamed, Mohamed Gomaa, Nicolas Saunier, Luis F. Miranda-Moreno, and Satish V. Ukkusuri. 2013b. "A Clustering Regression Approach: A Comprehensive Injury Severity Analysis of Pedestrian–Vehicle Crashes in New York, US and Montreal, Canada." *Safety Science* 54 (April): 27–37. https://doi.org/10.1016/j.ssci.2012.11.001.

Obaid, Ihsan, Ali Alnedawi, Ghufraan Mohammed Aboud, Reuben Tamakloe, Hamsa Zuabidi, and Subasish Das. 2022. "Factors Associated with Driver Injury Severity of Motor Vehicle Crashes on Sealed and Unsealed Pavements: Random Parameter Model with Heterogeneity in Means and Variances." *International Journal of Transportation Science and Technology* April. https://doi.org/10.1016/j.ijtst.2022.04.002.

O'brien, Robert M. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality & Quantity* 41 (5): 673–690. https://doi.org/10.1007/s11135-006-9018-6.

Pai, Chih-Wei, and Wafaa Saleh. 2008. "Modelling Motorcyclist Injury Severity by Various Crash Types at T-Junctions in the UK." *Safety Science* 46 (8): 1234–1247. https://doi.org/10.1016/j.ssci.2007.07.005.

Penmetsa, Praveena, and Srinivas S. Pulugurtha. 2019. "Risk Factors Attributed to Fatal Fixed-Object Crashes on Noninterstate Roads." *Journal of Transportation Safety & Security* 11 (1): 102–116. https://doi.org/10.1080/19439962.2017.1356414.

Salifu, Mohammed, and Williams Ackaah. 2012. "Under-Reporting of Road Traffic Crash Data in Ghana." *International Journal of Injury Control and Safety Promotion* 19 (4): 331–339. https://doi.org/10.1080/17457300.2011.628752.

Schneider, William H., Peter T. Savolainen, and Darren N. Moore. 2010. "Effects of Horizontal Curvature on Single-Vehicle Motorcycle Crashes along Rural Two-Lane Highways." *Transportation Research Record: Journal of the Transportation Research Board* 2194 (1): 91–98. https://doi.org/10.3141/2194-11.

Shaheed, Mohammad Saad, and Konstantina Gkritza. 2014. "A Latent Class Analysis of Single-Vehicle Motorcycle Crash Severity Outcomes." *Analytic Methods in Accident Research* 2 (April): 30–38. https://doi.org/10.1016/j.amar.2014.03.002.

Sinha, Pratik, Carolyn S. Calfee, and Kevin L. Delucchi. 2021. "Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls." *Critical Care Medicine* 49 (1): e63–e79. https://doi.org/10.1097/CCM.0000000000004710.

Sivasankaran, Sathish Kumar, and Venkatesh Balasubramanian. 2020. “Exploring the Severity of Bicycle–Vehicle Crashes Using Latent Class Clustering Approach in India.” *Journal of Safety Research* 72 (February): 127–138. https://doi.org/10.1016/j.jsr.2019.12.012.

Tamakloe, Reuben, Subasish Das, Eric Nimako Aidoo, and Dongjoo Park. 2022. “Factors Affecting Motorcycle Crash Casualty Severity at Signalized and Non-Signalized Intersections in Ghana: Insights from a Data Mining and Binary Logit Regression Approach.” *Accident Analysis & Prevention* 165 (February): 106517. https://doi.org/10.1016/j.aap.2021.106517.

Tamakloe, Reuben, J. Hong, J. Kim, and D. Park. 2022. “Factors Affecting Fatal PTW At-Fault Crash Outcome Metrics at Intersections and Non-Intersections in South Korea.” *Journal of Transportation Safety & Security* September: 1–36. https://doi.org/10.1080/19439962.2022.2123582.

Tamakloe, Reuben, Sungho Lim, Enoch F. Sam, Shin Hyoung Park, and Dongjoo Park. 2021. “Investigating Factors Affecting Bus/Minibus Accident Severity in a Developing Country for Different Subgroup Datasets Characterised by Time, Pavement, and Light Conditions.” *Accident Analysis & Prevention* 159 (September): 106268. https://doi.org/10.1016/j.aap.2021.106268.

Trinh, Linh Thanh, Kazushi Sano, and Kiichiro Hatoyama. 2021. “Modelling and Simulating Head-on Conflict-Solving Behaviour of Motorcycles under Heterogeneous Traffic Condition in Developing Countries.” *Transportmetrica A: Transport Science* 17 (4): 921–945. https://doi.org/10.1080/23249935.2020.1819911.

van de Velden, M., A. Iodice D’Enza, and F. Palumbo. 2017. “Cluster Correspondence Analysis.” *Psychometrika* 82 (1): 158–185. https://doi.org/10.1007/s11336-016-9514-0.

Venkatraman, V., C. M. Richard, K. Magee, and K. Johnson. 2021. *Countermeasures That Work: A Highway Safety Countermeasures Guide for State Highway Safety Offices*. Washington, DC: National Highway Traffic Safety Administration.

Vlahogianni, Eleni I., George Yannis, and John C. Golias. 2012. “Overview of Critical Risk Factors in Power-Two-Wheeler Safety.” *Accident Analysis & Prevention* 49 (November): 12–22. https://doi.org/10.1016/j.aap.2012.04.009.

Waseem, Muhammad, Anwaar Ahmed, and Tariq Usman Saeed. 2019. “Factors Affecting Motorcyclists’ Injury Severities: An Empirical Assessment Using Random Parameters Logit Model with Heterogeneity in Means and Variances.” *Accident Analysis & Prevention* 123 (February): 12–19. https://doi.org/10.1016/j.aap.2018.10.022.

Watson, Angela, Barry Watson, and Kirsten Vallmuur. 2015. “Estimating Under-Reporting of Road Crash Injuries to Police Using Multiple Linked Data Collections.” *Accident Analysis & Prevention* 83 (October): 18–25. https://doi.org/10.1016/j.aap.2015.06.011.

Wilson Van Voorhis, Carmen R., and Betsy L. Morgan. 2007. “Understanding Power and Rules of Thumb for Determining Sample Sizes.” *Tutorials in Quantitative Methods for Psychology* 3 (2): 43–50. https://doi.org/10.20982/tqmp.03.2.p043.

Xin, Chunfu, Zhenyu Wang, Pei-Sung Lin, Chanyoung Lee, and Rui Guo. 2017. “Safety Effects of Horizontal Curve Design on Motorcycle Crash Frequency on Rural, Two-Lane, Undivided Highways in Florida.” *Transportation Research Record: Journal of the Transportation Research Board* 2637 (1): 1–8. https://doi.org/10.3141/2637-01.

Ye, Xin, Ram M. Pendyala, Simon P. Washington, Karthik Konduri, and Jutaek Oh. 2009. “A Simultaneous Equations Model of Crash Frequency by Collision Type for Rural Intersections.” *Safety Science* 47 (3): 443–452. https://doi.org/10.1016/j.ssci.2008.06.007.