Journal of Transportation Safety & Security

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utss20

Pattern recognition in speeding related motorcycle crashes

Subasish Das , Seyedeh Maryam Mousavi & Maryam Shirinzad

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2021.1877228

Pattern recognition in speeding related motorcycle crashes

Subasish Das[a], Seyedeh Maryam Mousavi[b], and Maryam Shirinzad[a]

[a]Texas A&M Transportation Institute, College Station, Texas, USA; [b]Texas A&M University, College Station, Texas, USA

ABSTRACT

Motorcycle crash is considered as one of the critical safety issues. This study targets to examine the role of geometric design and other factors that contribute to the decision-making of motorcyclists, such as speeding, which could potentially result in a crash. This study analyzed Louisiana crash data from 2010 to 2016 to determine the patterns associated with speeding-related motorcycle crashes. The collected data contained information regarding crash characteristics and circumstances, characteristics of vehicles and drivers, crash locations, roadway types, traffic volume, segment length, and other relevant geometric information. This study employed a relatively new categorical data analysis method, which combines cluster and correspondence analysis. This study identified several high-risk scenarios in which a speeding-related motorcycle crash is more likely to occur through interactions with other related factors. The recommendations on the countermeasures can be used as a resource for policymakers to reduce speeding associated motorcycle crashes in Louisiana.

KEYWORDS

cluster correspondence analysis; clusters; motorcycle crashes; speeding;

1. Introduction

Speeding is referred to as the driving behavior associated with exceeding the posted speed limit or driving too fast for the given conditions. Speeding is an influential factor in both fatal and non-fatal crashes. The Speed Management Strategic Initiative was developed by the US Department of Transportation (USDOT) in search of methods to manage the crash-related effects of speeding efficiently. In a 2007 National Highway Traffic Safety Administration (NHTSA) study (NHTSA, 2007), speeding is defined "as a factor associated with nearly seven times more for motorcycle operators than for the driver of the passenger vehicle in two-vehicle crashes (27 vs. 4%)".

In many countries worldwide, inappropriate, and excessive speed is the leading cause of road trauma. A vehicle traveling at higher speeds requires

CONTACT Subasish Das s-das@tti.tamu.edu Texas A&M University Transportation Institute, College Station, TX 77843-0001, USA

more time to stop. While majority of motorcycle crashes happen at lower speeds, serious and fatal injuries are mostly accompanied by higher speeds. Since motorcyclists are not offered the same level of crash protection as those in motor vehicles, those riding motorcycles are particularly vulnerable to serious or fatal injuries related to excessive speed. When riding a motorcycle, age has been proven to be a significant indicator of the intention to engage in risky speed-related behaviors. Research involving car drivers indicates that young riders were more likely to perform speeding. The study conducted by Huang and Preston (2004) supports this finding, which suggested that motorcyclists displayed aggressive or risky behavior by overtaking more frequently than other road users, choosing higher speeds, and pulling into smaller gaps in the traffic. In particular young male motorcycle riders were more likely to show these kinds of risky behavior. However, age was not the only indicator of risky behavior. Other factors played roles as well, such as the types of riding that a rider engages in, the experience, and the size of their engine. Some studies (Das, 2020; Das, Dutta et al., 2018; Pour-Rouholamin, Jalayer, & Zhou,2017) examined the influences of key geometric factors that are associated with motorcycle crashes. Prior studies show that both geometric factors and rider behavior are associated with motorcycle crashes (Das et al., 2018; Das, Dutta, & Tsapakis,2021; Haworth, Greig, & Nielson, 2009; Pour-Rouholamin et al., 2017; Shaheed & Gkritza, 2014). However, exploration of speeding-related motorcycle crashes is limited.

Motorcycle crashes in Louisiana reflect the national trend. From 2014 to 2017, fatal, injury, and property damage only (PDO) motorcycle crashes increased by 20% in Louisiana (Highway Safety Research Group, 2019). Traffic safety research encompasses a wide variety of research areas, and one of the most prominent focus areas is crash data analysis. The most common trend of analysis is to establish an association between crash count and the roadway characteristics, spatial and temporal conditions, and traffic operation characteristics. This study assumes that crashes are caused by dangerous decisions made by the motorcyclist in a location resulting from surrounding conditions and roadway geometry. This study used a wide range of variables to determine the key risk factors associated with motorcycle-related crashes.

2. Literature review

Speeding involvement is defined in the crash data if the driver is charged with a speeding-related offense or reported by a law enforcement officer for operating the motorcycle too fast, racing, or exceeding the designated posted speed limit on a particular roadway. In 2016, 33%of motorcycle fatal crashes in the US were associated with speeding issues, in comparison to "19 percent for passenger car drivers, 15 percent for light-truck drivers, and 7 percent for large-truck drivers" (NHTSA, 2020). Haworth et al. (2009) examined the role of risk-taking in moped crashes compared to motorcycle crashes. They presented the results of police-reported crashes analyses in Queensland, Australia. Excessive speed is not associated with majority of moped crashes, but this finding may reflect the limitations regarding vehicle performance as much as a decision not to speed. Motorcycles are different with regards to design and performance capability. Based on their driving preferences, motorcyclists can choose specific types of motorcycles. On the other hand, the capabilities of motorcycle performance are associated with the possibility of risky driving behaviors like speeding. When examining by motorcycle type, both mechanisms have the potential to influence the risk of fatal crashes. Although the impact of each mechanism cannot be estimated, fatal crash data was analyzed for evidence of motorcycle brand differences in the risk of driver death and risky driving behaviors and. Teoh and Campbell (2010) found robust impacts of motorcycle type on fatal motorcycle crashes and explored high risk operating behaviors such as alcohol impairment and speeding. The findings showed the significant impact of both motorcycle type and rider age on the likelihood of risky driving behaviors.

The principal contributing factors of crashes were identified as speeding, improper evasive action, and improper handling by Vachal, Malchose, and Benson(2013). For citation data, careless riding and alcohol involvement were found as the key factors. With crash data from 3644 single-vehicle motorcycle crashes in Iowa from 2001-2008, Shaheed and Gkritza (2014) applied a latent class multinomial logit model notable to determine the association between crash injury patterns and key contributing factors (such as run off-road, speeding), riding without a helmet, riding on rural roads, impaired riding (riders under the influence of alcohol, drug, or medication), and rider's age (older than 25 years old). The speed evaluation conducted by Manan, Ho, Arif, Ghani, and Várhelyi(2017) demonstrated the excessive speed of motorcyclists compared to other vehicles. Overall, 42.2% of the motorcycles showed speeding behavior, and 28.6% of them surpass the 85th percentile of traffic speed. To better comprehend the relation of motorcycle type and their motive to exceed the speed limit on a 55-mph roadway, Eyssartier, Meineri, & Gueguen (2017) developed a model of planned behavior. The predictive factors of surpassing the speed limit differ for every motorcyclist. Pour-Rouholamin et al. (2017) exhibited that alcohol-involved riding, older riders, crashes during weekends and summer, not wearing helmets, darkness, presence of roadside fixed objects, speeding, and reckless riding increased the severity of injuries. Das et al. (2018)

showed that curved roadways and weekends were significant contributors to motorcycle crashes. Ding, Rizzi, Strandroth, Sander, and Lubbe(2019) also found a significant and strong correlation between the injury severity and associated speed in motorcycle crashes.

The literature review reveals a necessity for an extensive study concentrating on speeding-related motorcycle crashes. This study aims to mitigate the current research gap by employing a clustering technique to determine the patterns of influencing factors associated with speed-related motorcycle crashes with statistical evidence. Moreover, this study identifies countermeasures for each cluster of crashes to help lower the number of crashes in each category.

3. Methodology

3.1. Data integration

This study collected traffic crash data from the Louisiana Department of Transportation and Development (LADOTD). The database contains three major datasets: (1) crash file, (2) vehicle file, and (3) roadway inventory file (known as DOTD file). In Louisiana, police complete a uniform motor vehicle traffic crash report for any traffic crash; the reports are later transcribed into Access- or Excel-based databases (State of Louisiana, 2020). Figure 1 shows the overall data assimilation framework. Seven years, from 2010 to 2016, of crash data were used in this study. In these years, a total of 14,084 motorcyclists were involved in 13,765 reported traffic crashes. It indicates that majority of the crashes are associated with a single motorcyclist. Speed related motorcycle crashes were identified by filtering the crash data that contains speeding as a violation. A total of 318 motorcyclists were involved in speeding-related traffic crashes. To obtain a complete picture of the scenarios associated with speeding-related motorcycle crashes, crash data and DOTD data were merged with motorcyclist level data.

3.2. Exploratory data analysis

The scenarios associated with motorcycle crashes have been recognized as multifaceted events involving interactions between many significant factors such as operating patterns of the motorcyclists, traffic characteristics, roadway properties, roadside objects/treatments, and surroundings. However, it is critical to evaluate the individual influence of a factor because these events occurred with the presence of many variables. The analysis mandates data spanning a considerable number of variables and a large sample size to provide precise risk estimates and to control for various potential confounders.

Exploratory data analysis was initially conducted to detect the crucial factors that may weigh in speeding related motorcycle crash events. Variables

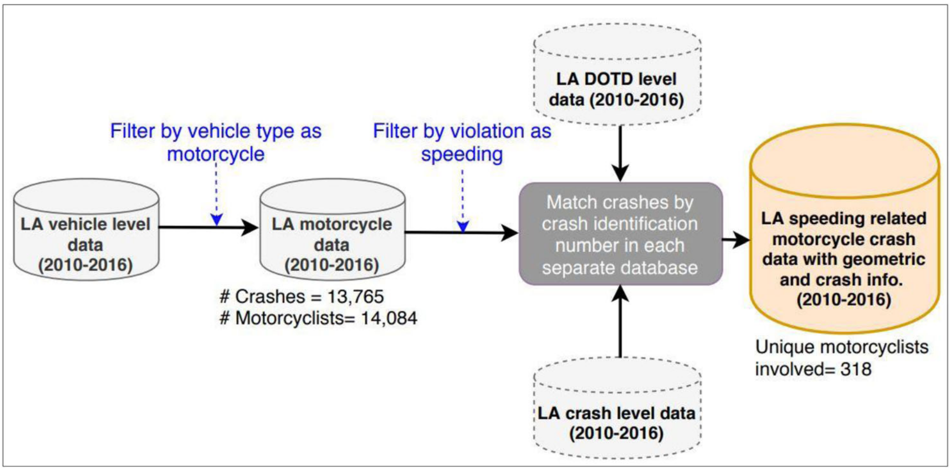

Figure 1. Flowchart of the data integration.

such as driver condition, vehicle make, crash severity, driver age, road type, the reason for the crash occurrence, and collision type are listed in the table. However, some important variables, such as cylinder capacity of the motorcycle is missing in the collected dataset. Additionally, some variables, such as gender, are not considered in the analysis due to the skewed nature of the variable. For example, male riders represent over 95% of the collected crash database and it is thus excluded in the final variable selection. Table 1 lists the frequency distribution of the selected variables for the analysis. Table 1 shows that crashes in residential locations experience the most speeding-related crashes, and the most frequent collision type involve single motorcycles. Motorcycle brands known to perform at higher speeds were found to have a higher crash frequency. Almost 93%of speeding-related motorcycle crashes occur on dry surface conditions, and a majority of crashes take place during the night. Note that motorcycle crash characteristics such as lane-splitting is an urban safety issue rather than a rural safety issue. However, the variable selection procedure does not find facility type (urban vs. rural) as a critical factor because another variable location type is more influential than facility type. The further exploration shows that business and industrial are mostly associated with urban facility types and open country is mostly associated with rural facility types.

3.3. Cluster correspondence analysis

Correspondence analysis (CA), a categorical data analysis technique, analyzes simple two-way and multi-way data tables containing a link between the columns and rows from an intricate dataset by conducting dimension reduction. Recently, a handful of transportation safety investigations have utilized different forms of CA techniques (for example, multiple correspondence

Table 1. Percentage distribution of the key attributes.

Category	Frequency	%	Category	Frequency	%	Category	Frequency	%
Crash severity			*Collision type*			Industrial	13	4.09%
Fatal (K)	64	20.13%	Single motorcycle	189	59.43%	Others	12	3.77%
Incapacitating injury (A)	50	15.72%	Right angle	37	11.64%	Residential	132	41.51%
Non incapacitating injury (B)	109	34.28%	Left turn	22	6.92%	*Roadway type*		
Minor injury (C)	55	17.30%	Rear end	24	7.55%	Two way with no separation	199	62.58%
No injury (O)	40	12.58%	Sideswipe	17	5.35%	Two way with separation	68	21.38%
Driver condition			Head on	8	2.51%	Two way with barrier	12	3.77%
Normal	103	32.39%	Others	21	6.60%	One way	34	10.69%
Inattentive	86	27.04%	*Vehicles make*			Others	5	1.57%
Distracted	8	2.52%	Harley Davidson	70	22.01%	*Surface condition*		
Impaired	15	4.72%	Honda	56	17.61%	Dry	295	92.77%
Others	106	33.33%	Kawasaki	53	16.67%	Wet	23	7.23%
Drive age			Suzuki	85	26.73%	*Occurrence at intersection*		
15–24	77	24.21%	Yamaha	34	10.69%	N	225	70.75%
25–34	113	35.53%	Others	20	6.29%	Y	93	29.25%
35–44	56	17.61%	*Access control type*			*Annual daily traffic*		
45–54	35	11.01%	No control	254	79.87%	0–1000	88	27.67%
55–64	19	5.97%	Full control	39	12.26%	1001–3000	93	29.25%
>65	18	5.66%	Partial control	25	7.86%	3001–8000	32	10.06%
Reason for the crash occurrence			*Road alignment*			8001–20,000	39	12.27%
Violation	164	51.57%	Straight level	171	53.77%	>20,000	66	20.75%
Normal	57	17.92%	Curve	105	33.02%	*Lighting condition*		
Avoid vehicle/animal	20	6.29%	On grade	26	8.18%	Daylight	125	39.31%
Loss of control	25	7.86%	Others	16	5.03%	Dark	193	60.69%
Driver condition	8	2.52%	*Locality type*			*Day of week*		
Others	44	13.84%	Business	105	33.02%	FSS	185	58.18%
			Open country	56	17.61%	MTWT	133	41.82%

Note: Seve = Crash Severity, Cond = Driver Condition, Age = Driver Age, Reason = Reason for the crash occurrence, Coll = Collision Type, Make = Vehicle Make, Access = Access Control Type, Align = Roadway Alignment, Loc = Locality Type, Road = Roadway Type, Surf = Surface Condition, Int = Occurrence at Intersection, adt = Annual Daily Traffic, Light = Lighting Condition, DOW = Day of Week.

analysis, taxicab correspondence analysis, and factor analysis) to identify the trends of the key influential factors (Das & Sun, 2015; 2016; Das et al., 2018; Das, Jha, Fitzpatrick, Brewer, & Shimu, 2019; Jalayer, Pour-Rouholamin, & Zhou, 2018;). Cluster correspondence analysis includes both cluster analysis dimension reduction for categorical data by concurrently allocating individuals to clusters and standard scaling measures to different variable categories. The concept of cluster CA is described below, which is mostly based on the study by Van de Velden, Enza, and Palumbo (2017). Interested readers can consult them for a comprehensive study on cluster correspondence analysis.

Consider a dataset that contains information of n individuals (for example, motorcyclists) by providing additional details of p categorical variables (for example, roadway type). The dataset can be expressed in a super indicator matrix Z with the dimensionality of $n \times Q$, where $Q = \sum_{j=1}^{p} q_j$. Cluster membership can be coded as an indicator matrix Z_K (where K is the number of dimensions or axes). In this way, the user produces a table to cross-tabulate cluster memberships. For example, $F = Z'_K Z$, where Z_K is the $n \times K$ indicator matrix signifying cluster membership. Applying CA framework to this matrix produces optimal scaling values for rows (as clusters) and columns (as categories). Based on variable distribution on the place, the clusters are optimally divided. Likewise, the categories with varying distributions over the clusters are optimally divided. The optimal cluster allocation Z_K can be expressed as (Van de Velden et al., 2017):

$$max\emptyset_{clusca}(Z_K, B^*) = \frac{1}{p} traceB^{*'} D_z^{-1/2} Z' M Z_K D_Z^{-1} Z'_K M Z D_z^{-1/2} B^* \quad (1)$$

For fixed B^*, the optimization issue can be considered as K-means clustering problem. The maximization of $\emptyset(Z_K, B^*)$ with regards to Z_K can resolve the optimization issue *:*

$$min\emptyset'_{clusca}(Z_K, G) = \left|\left|\sqrt{\frac{n}{p}} M Z D_z^{\frac{-1}{2}} B^* - Z_K G\right|\right|^2, \quad (2)$$

where:

G = matrix having cluster average measures

B = column coordinate matrix of rank k, where k is approximation dimensionality.

$M = I_n - 1_n 1'_n / n.$

D_z = diagonal matrix with the assumption $D_z 1_Q = Z' 1_n$

4. Results and discussions

Table 2 lists the location of cluster centroids with 'sum of squares' and 'size' information. Figure 2 portrays the biplots (a simplistic visualization in

Table 2. Location of the cluster centroids.

Cluster	Size (percentage)	Sum of squares	Dimension 1	Dimension 2
Cluster 1	81 (25.5%)	6.7138	−0.3454	−0.2482
Cluster 2	66 (20.8%)	7.1362	−0.4396	−0.809
Cluster 3	49 (15.4%)	7.3275	−0.0825	1.1636
Cluster 4	46 (14.5%)	4.2171	−0.5348	0.3479
Cluster 5	38 (11.9%)	7.3757	0.6352	0.4796
Cluster 6	38 (11.9%)	8.0998	1.6182	−0.4672

the form of a scatter plot with attribute labels) of the attributes. After performing several trials, a two-dimensional, six cluster solution was found to be appropriate for this evaluation. The solution illustrating the clusters and attributes is shown in Figure 3. Individual subject points can be projected into this biplot visualization. This study used valence ratio criterion as the cluster validity measure. This value is generated from k means clustering. These k means were operated multiple times to confirm the best possible number of clustering. The cluster validity measure for this study is 20.966. In the cluster CA framework, the origin represents the mean profile, and all other coordinates imply deviations from this profile. Figure 3 shows that two clusters (Clusters 1 and 4) located near the origin. The adjacent other two clusters are Clusters 3 and 5. The outstanding two clusters are far from the location of the origin.

Figures 3 and 4 show association patterns of the key variable categories or attributes (for example, "fatal crash" is an attribute of the "crash injury" variable). For insightful interpretation of the clusters, attributes with the most deviation from the independence condition can be examined. Six clusters and associated attribute contributions are shown in Figures 3 and 4. In each of these six cluster plots, the bars represent the highest standardized residuals (positive or negative). A positive residual means that the attribute has an above average frequency within the cluster. In a similar context, a negative value indicates the below average frequency with the cluster. This approach has an advantage in generating "in cluster" proportions of the attributes. Table 3 provides visual guidance to understand the "between cluster" proportion of the attributes. The darkest shading indicates the cluster with the most crashes for that attribute. Table 3, along with Figures 3–4, provides techniques for identifying the primary attributes within a cluster.

4.1. Cluster 1

This cluster represents seven attributes with positive residual means: residential, single motorcycle, two way undivided, low annual average daily traffic (AADT), segment (no intersection exists), no access control, and alignment others. It indicates the association between single vehicle

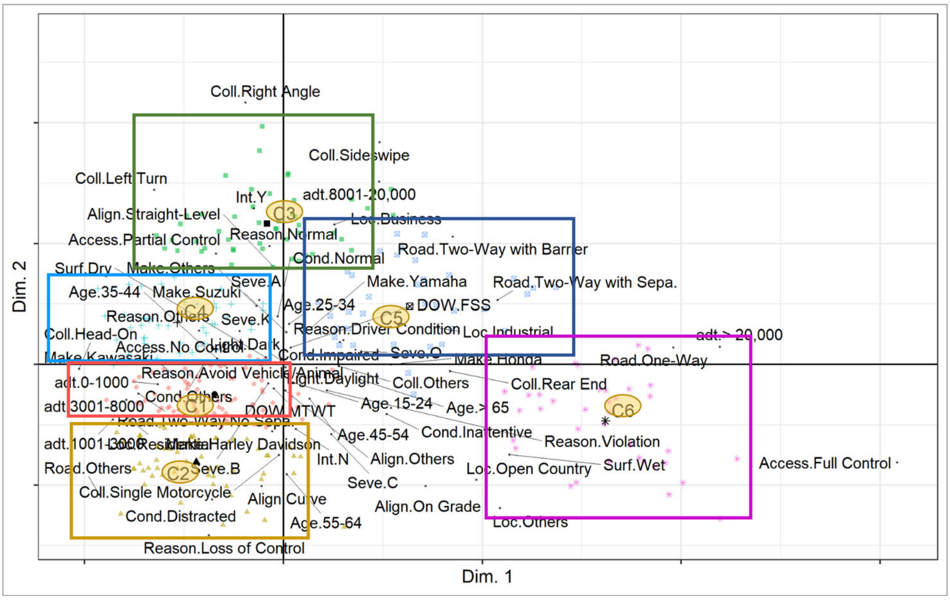

Figure 2. Biplot of the attributes.

segment related speeding related crashes on residential two-lane undivided roadways with low AADT. Improving curves, installing motorcyclist-friendly barriers on the curves, and improving two-lane undivided roadways through widening the roadways are the potential countermeasures to improve the safety of the motorcyclists. Moreover, since the residential areas are associated with a fewer number of access points and intersections, as indicated by Brimley, Mousavi, Carlson, and Dixon (2017), providing better roadway lighting, especially at intersections and access points, enhances the safety of the motorcyclists. The safety training of the motorcyclists was also found to be beneficial.

4.2. Cluster 2

This cluster has several attributes with positive residual means such as curve, loss of control, single motorcycle, two way undivided, residential, low AADT, 55 to 64 years old motorcyclist with Harley Davidson motorcycle, and segment. This cluster indicates that loss of control single motorcycle speeding involved crashes at curves on residential two-way undivided roadways with low AADT. One exogenous variable attribute is 55 to 64 years old motorcyclist with Harley Davidson motorcycles. Curve improvement, installation of motorcyclist-friendly barriers on the curves, better road space allocation, well-lit roadways, pavement marking retro-reflectivity improvement or using wider pavement markings especially aiming older riders, and improvement of two-lane undivided roadways are the

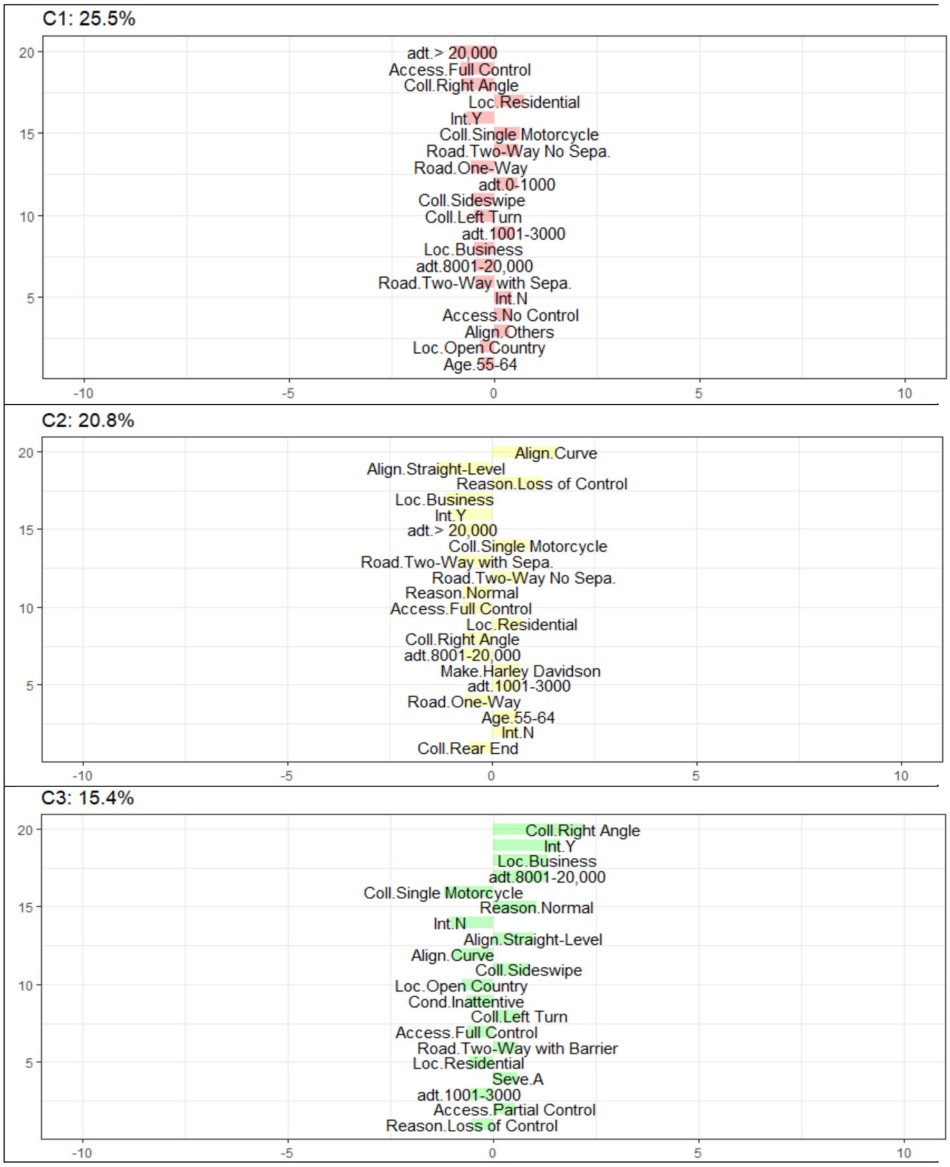

Figure 3. Residual representations in Clusters 1 to 3.

potential countermeasures. Older riders need to take additional precautions and safety gears, so the motorcyclists can be seen ahead by the drivers.

4.3. Cluster 3

The most frequent motorcycle crashes involve another vehicle violating the motorcycle's right-of-way at an intersection. cluster represents nine attributes with positive residual means: right angle, intersection, business area, moderate AADT, normal condition, incapacitating injury, two way with a barrier,

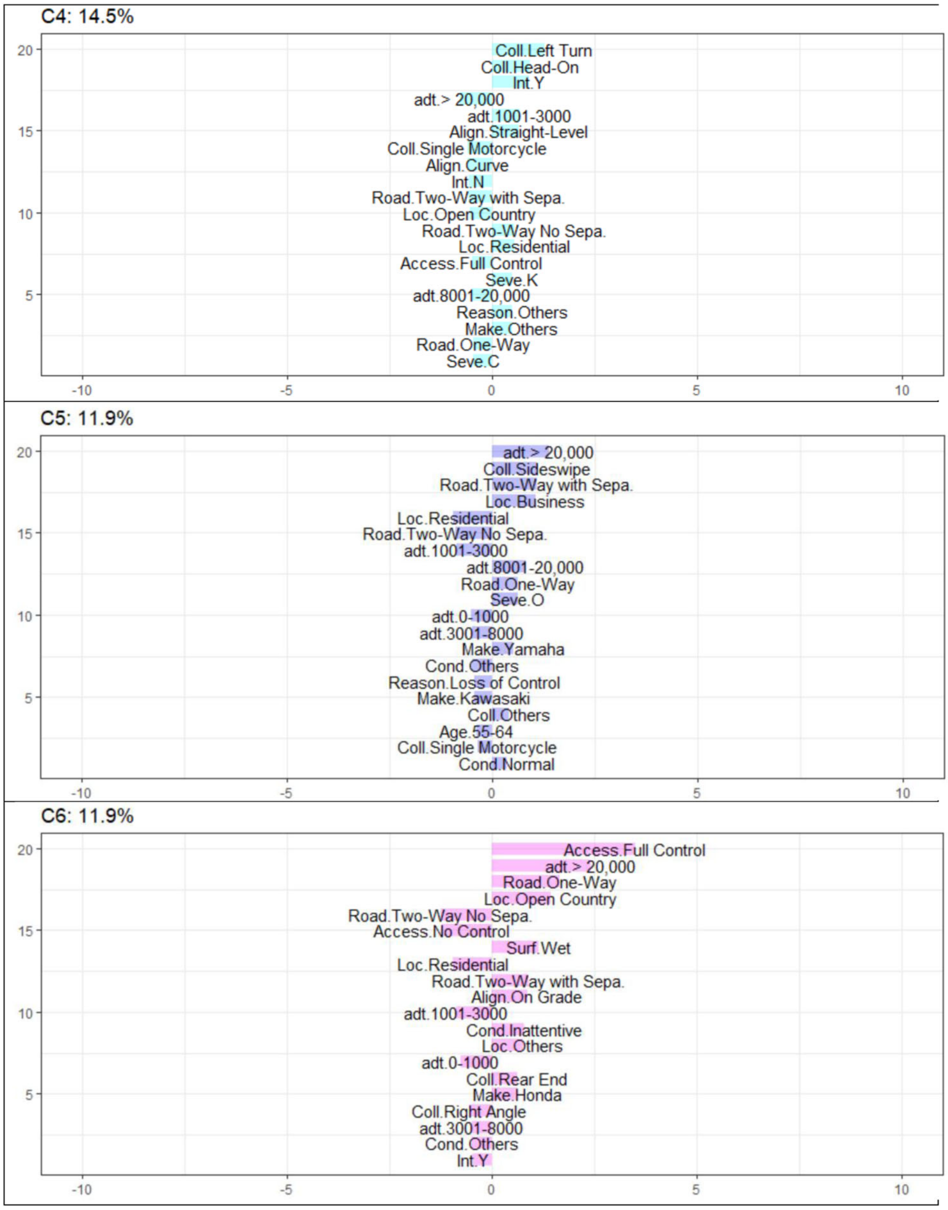

Figure 4. Residual representations in Clusters 4 to 6.

sideswipe and left turn, and partial access control. This cluster highlights intersection related right-angle crashes due to speeding motorcyclists. This cluster is associated with incapacitating injury of the motorcyclists. Other intersection crash types in this cluster are sideswipe and left turn. Countermeasures like signal ahead sign and supplemental signal face per approach for signalized intersections enhance the safety of the motorcyclists. In addition, adequate stopping sight distance, cleared vision triangle, vehicle to motorcycle

Table 3. Between clusters proportions of the attributes.

Variable	Attribute	Count	Cluster 1 (81)	Cluster 2 (66)	Cluster 3 (49)	Cluster 4 (46)	Cluster 5 (38)	Cluster 6 (38)
Crash severity	Fatal	64	20.31	18.75	21.88	**23.44**	6.25	9.38
	Incapacitating injury	50	22.00	16.00	**28.00**	14.00	14.00	6.00
	Non incapacitating injury	109	**28.44**	22.94	12.84	14.68	9.17	11.93
	Minor injury	55	**29.09**	**29.09**	5.45	5.45	12.73	18.18
	No injury	40	**25.00**	12.50	10.00	12.50	**25.00**	15.00
Driver condition	Normal	103	**27.18**	12.62	21.36	14.56	16.50	7.77
	Inattentive	86	**25.58**	20.93	4.65	9.30	16.28	23.26
	Distracted	8	25.00	**50.00**	0.00	12.50	0.00	12.50
	Impaired	15	**33.33**	6.67	20.00	13.33	6.67	20.00
	Others	106	22.64	**28.30**	18.87	18.87	5.66	5.66
Driver age	15–24	77	**31.17**	22.08	12.99	7.79	14.29	11.69
	25–34	113	**23.89**	12.39	19.47	19.47	13.27	11.50
	35–44	56	25.00	**26.79**	17.86	14.29	12.50	3.57
	45–54	35	**28.57**	22.86	11.43	11.43	8.57	17.14
	55–64	19	10.53	**47.37**	5.26	15.79	0.00	21.05
	>65	18	**22.22**	16.67	11.11	16.67	11.11	22.22
Reason for the crash occurrence	Violation	164	**24.39**	23.78	10.98	9.76	15.24	15.85
	Normal	57	21.05	3.51	**36.84**	21.05	8.77	8.77
	Loss of control	25	20.00	**64.00**	0.00	4.00	0.00	12.00
	Avoid vehicle/animal	20	**40.00**	10.00	5.00	25.00	15.00	5.00
	Driver condition	8	**25.00**	12.50	12.50	12.50	25.00	12.50
	Others	44	**31.82**	13.64	18.18	25.00	6.82	4.55
Collision type	Single motorcycle	189	**34.39**	32.28	2.12	7.94	8.47	14.81
	Right angle	37	0.00	0.00	**70.27**	18.92	10.81	0.00
	Rear end	24	**37.50**	0.00	8.33	8.33	16.67	29.17
	Left turn	22	4.55	0.00	36.36	**54.55**	4.55	0.00
	Sideswipe	17	0.00	0.00	**47.06**	5.88	**47.06**	0.00
	Head on	8	12.50	25.00	0.00	**62.50**	0.00	0.00
	Others	21	**23.81**	14.29	4.76	19.05	**23.81**	14.29
Lighting condition	Daylight	193	**25.39**	20.21	14.51	16.06	10.88	12.95
	Dark	125	**25.60**	21.60	16.80	12.00	13.60	10.40
Vehicles Make	Harley Davidson	70	22.86	**35.71**	8.57	11.43	7.14	14.29
	Honda	56	**25.00**	12.50	14.29	10.71	14.29	23.21
	Kawasaki	53	**33.96**	22.64	13.21	22.64	3.77	3.77
	Others	20	20.00	10.00	20.00	**30.00**	10.00	10.00
	Suzuki	85	21.18	20.00	**22.35**	11.76	15.29	9.41
	Yamaha	34	**32.35**	8.82	14.71	11.76	23.53	8.82
Access control type	Full control	39	0.00	0.00	0.00	2.56	10.26	**87.18**
	No control	254	**30.31**	24.80	16.14	15.35	11.81	1.57
	Partial control	25	16.00	12.00	**32.00**	24.00	16.00	0.00
	Curve	105	23.81	**47.62**	0.95	5.71	8.57	13.33
	On grade	26	15.38	**38.46**	3.85	3.85	3.85	34.62
	Others	16	**43.75**	12.50	6.25	12.50	12.50	12.50
	Straight level	171	26.32	2.34	**26.90**	21.64	15.20	7.60
Locality type	Business	105	16.19	0.95	**35.24**	14.29	25.71	7.62
	Industrial	13	**30.77**	0.00	15.38	7.69	23.08	23.08
	Open country	56	16.07	32.14	0.00	3.57	10.71	**37.50**
	Others	12	8.33	**41.67**	0.00	0.00	8.33	**41.67**
	Residential	132	**37.88**	31.82	7.58	21.21	0.76	0.76
Roadway type	Two way with no separation	199	**33.67**	30.65	12.06	20.10	3.52	0.00
	Two way with separation	68	14.71	2.94	22.06	4.41	**29.41**	26.47
	One way	34	5.88	0.00	14.71	2.94	26.47	**50.00**
	Two way with barrier	12	8.33	0.00	**41.67**	8.33	16.67	25.00

(*continued*)

Table 3. Continued.

Variable	Attribute	Count	Cluster 1 (81)	Cluster 2 (66)	Cluster 3 (49)	Cluster 4 (46)	Cluster 5 (38)	Cluster 6 (38)
	Others	5	20.00	**60.00**	0.00	20.00	0.00	0.00
Surface type	Dry	295	**25.42**	21.02	16.27	15.25	12.54	9.49
	Wet	23	26.09	17.39	4.35	4.35	4.35	**43.48**
Occurrence at intersection	N	225	**31.11**	28.44	4.89	8.89	12.00	14.67
	Y	93	11.83	2.15	**40.86**	27.96	11.83	5.38
Day of week	FSS	185	24.32	17.30	18.38	16.22	10.81	12.97
	MTWT	133	27.07	25.56	11.28	12.03	13.53	10.53
Annual average daily traffic	0–1000	88	**37.50**	26.14	13.64	17.05	4.55	1.14
	1001–3000	93	**35.48**	33.33	6.45	24.73	0.00	0.00
	3001–8000	32	28.13	**37.50**	15.63	18.75	0.00	0.00
	8001–20,000	37	10.81	0.00	**48.65**	2.70	29.73	8.11
	>20,000	68	2.94	0.00	11.76	1.47	33.82	**50.00**

The bold text indicates the highest percentage in each variable category.

communication, innovative motorcycle headlight design, and advanced detection control system.

4.4. Cluster 4

The attributes having positive residual means are left turn, head on, low AADT, intersection, straight, two way with no separation, fatal, and residential. This cluster shows association between fatal speeding related motorcyclists crashes with intersection head on or left turn crashes at residential two way roadways with no separation and low AADT. The findings showed that poor lighting at the roadway intersections is associated with high number of fatal crashes. Sufficient street lighting and necessary pavement marking retro-reflectivity in low AADT residential roadways can mitigate the severity of crashes. Additionally, sight distance can be improved by removing vegetation near and within intersections. The countermeasures discussed in Cluster 3 can also help in reducing crashes associated with this cluster. Additionally, usage of helmets can reduce the intensity of severity of the motorcyclists.

4.5. Cluster 5

Sideswipe motorcyclists crashes occur mainly for lane splitting (riding between cars that are in two lanes and driving next to each other). The seven attributes with positive residual means are: moderate to high AADT, sideswipe, two way with separation, one way, property damage only (PDO) crash, normal condition, and Yamaha brand. This cluster represents PDO crashes and their association with roadway features like moderate to high AADT, one way roadway, and a two way roadway with separation. Although lane splitting is not allowed in Louisiana, around 6% of the crashes are caused by sideswiping. Safety training and strict lane-splitting law enforcement can be helpful in reducing crashes associated with the attributes in this cluster. Displaying advisory variable message signs at high AADT roads is another countermeasure that brings the attention of motorcyclist to the risks associated with lane splitting.

4.6. Cluster 6

The attributes with positive residual means in this cluster are- full access control, high AADT, one way road, open country, wet surface, a two-way roadway with separation, alignment on grade, inattentive, Honda brand, and rear end. It indicates speeding-related motorcyclist crashes on open country wet surface roadways with high AADT at on grade locations. Visibility during adverse weather condition is a big problem for motorcyclists. Maintaining traction during wet weather condition is problematic. Safety training can help the motorcyclist by informing about the potential hazards and necessary precautions. The LADOTD can also consider applying high friction surface treatments (HFST) at locations with a high likelihood of occurrences of motorcycle crashes. Using permeable paving (porous concrete) is another method that can help keep high AADT open country roads dryer during precipitation and ultimately lead to decreasing motorcycle crashes in rural areas. Another countermeasure is to display advisory messages on variable message signs to inform motorcyclists of wet conditions and advise them to drive attentively. Other studies (Jones, Janssen, & Mannering1991; Theofilatos &Yannis, 2014) suggested additional motorcyclist education regarding safety maintenance in improper weather conditions.

One of the major advantages of this analysis is the ability to generate "in cluster" proportions of the attributes. The color heatmap format of Table 3 provides a quick glance of the "between cluster" proportion of the attributes.

Table 3 shows the "between cluster" proportions of the exogenous variables. This table lists the distribution of the variable attributes in each cluster. Instead of normalizing the attributes by variable group, this approach effectively represents which cluster is dominant in each attribute type. The highest number in each row is displayed in bold text with darker red call color. The key findings from this table are stated below:

- Fatal speeding related motorcycle crashes are higher in proportions in Clusters 3 and r 4.
- Around 50% of the distracted drivers in the speeding-related motorcycle crashes were present in Cluster 2, while Clusters 3 and 4 did not have any distracted drivers.
- Age group 55 to 64 is higher in proportion in Cluster 2, where distracted driving conditions were also more dominant and loss of control was the most significant reason for the crash. This means that there

may be an association between this age group, loss of control, and distracted driving.

- Cluster 3 is dominant in right-angle crashes. Over 2/3 of crashes were classified as right-angle. Similar to the right-angle crashes, other intersection-related crash types were also dominant in this cluster.
- Three clusters show zero (Cluster 1 to 3) percentage of crashes on the roadways with full access control. Cluster 6 is dominant in full access control relevant motorcycle crashes. Cluster 6 contributes to 43% of the crashes when the surface is wet. This cluster is also highly representative in roadways with high AADT values.
- For the locations with business as the prominent land-use, Cluster 3 has the highest percent of the speeding-related motorcycle crashes, and for industrial and residential locations, Cluster 1 represents the highest percentage of the crashes with approximately 31 and 38%, respectively.

5. Conclusion

To identify the key contributing lusters in speeding-related motorcycle crashes, the present study applied a comparatively new categorical data analysis method that incorporates correspondence analysis and cluster analysis. By variable association table and yields, this method operates as a correspondence analysis of a cluster. In conjunction with a low dimensional estimation depicting variables and clusters, this study explained the clusters in respect to individual crash attributes. This method exceeds the performance of CA and multiple correspondence analysis (MCA) in determining the nature of the underlying cluster structures. With seven years of Louisiana crash data, this study analytically determined the relative contribution of key factors for various cluster groups.

Typically, a crash is the interrelated and complex consequence of human, roadway, vehicle, and environmental factors. This study identifies high-risk scenarios where speeding related motorcycle crashes are more likely to occur through interactions with the presence of associated factors. This study identified six clusters with groups of variables that are associated with the high likelihood of speeding-related motorcycle crashes. Some of the key groups are single motorcycle crashes on low volume two lane roadways (Cluster 1), two way undivided curve related crashes with older motorcyclists as the rider (Cluster 2), intersection related right angle, sideswipe, and left turn crashes (Cluster 3), fatal crashes on left turn related intersection crashes (Cluster 4), lane splitting related sideswipe crashes on moderate to high traffic volume roadways (Cluster 5), and crashes on open country at-grade locations during wet pavement condition (Cluster 6). Naqvi and Tiwari (2018) showed that the likelihood of single-vehicle crash occurrences was three times greater than the likelihood of two or more vehicle crashes, which is in line with the findings of Cluster 1. The findings of Cluster 2 are in line with other studies (Das et al., 2018; Xin et al., 2019). Xin et al. (2019) showed that single-motorcycle crashes are overrepresented on rural two lane undivided highways in the vicinity of curves. Findings from Clusters 3 and 5 are in line with the findings of Vachal et al. (2013), Baldi, Baer, and Cook (2005), and Li, Doong, Huang, Lai, and Jeng (2009). The study also provided "between cluster" proportions by the attributes. It is expected that these results are anticipated to allow an insightful comprehension of speeding related motorcycle crashes.

This study has several limitations. The major limitation is that the current speeding related factors are identified from the police reported filtering option that defined a crash as speeding related based on the violation type. Police reported crash data was transcribed into access and excel based crash databases in Louisiana. The database has a variable known as "vehicle speed," which indicates the operating speed of the vehicle at the time of the crash. However, the measure of this variable is mostly missing in the crash data. The data from police reports provide information about whether the crash was related to speeding. This study used that variable to acquire speeding related motorcycle crashes. Louisiana contains police documented crash narratives in database format. The definition of "speeding-related crashes" can be validated by performing text mining of the crash narratives, which is not performed in this study. Future research can explore speeding related contexts and include a wide range of variables.

Acknowledgments

The authors like to thank three anonymous reviewers for their excellent feedback.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753
Seyedeh Maryam Mousavi http://orcid.org/0000-0001-8188-4466

References

Baldi, S., Baer, J. D., & Cook, A. L. (2005). Identifying best practices states in motorcycle rider education and licensing. *Journal of Safety Research*, *36*(1), 19–32. doi:10.1016/j.jsr.2004.11.001

Brimley, B. K., Mousavi, S. M., Carlson, P. J., & Dixon, K. K. (2017). *Safety effects of traffic sign upgrades in Albuquerque, New Mexico*. Transportation Research Board 96th Annual Meeting, Washington, DC, USA.

Das, S. (2020). Identifying key patterns in motorcycle crashes: Findings from taxicab correspondence analysis. *Transportmetrica A: Transport Science*.

Das, S., & Sun, X. (2015). Factor association with multiple correspondence analysis in vehicle-pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2519*(1), 95–103. doi:10.3141/2519-11

Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research*, *39*(2), 146–155. doi:10.1016/j.iatssr.2015.07.001

Das, S., Avelar, R., Dixon, K., & Sun, X. (2018). Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident; Analysis and Prevention*, *111*, 43–55. doi:10.1016/j.aap.2017.11.016

Das, S., Dutta, A., Dixon, K., Minjares-Kyle, L., & Gillette, G. (2018). Using deep learning in severity analysis of at-fault motorcycle rider crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(34), 122–134. doi:10.1177/0361198118797212

Das, S., Dutta, A., & Tsapakis, I. (2021). Motorcycle crash causation study: Exploratory topic models from crash narrative reports. *Transportation Research Record: Journal of Transportation Research Board*.

Das, S., Jha, K., Fitzpatrick, K., Brewer, M., & Shimu, T. H. (2019). Pattern identification from older bicyclist fatal crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(6), 638–649. doi:10.1177/0361198119841863

Ding, C., Rizzi, M., Strandroth, J., Sander, U., & Lubbe, N. (2019). Motorcyclist injury risk as a function of real-life crash speed and other contributing factors. *Accident; Analysis and Prevention*, *123*, 374–386. doi:10.1016/j.aap.2018.12.010

Eyssartier, C., Meineri, S., & Gueguen, N. (2017). Motorcyclists' intention to exceed the speed limit on a 90 km/h road: effect of the type of motorcycles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *45*, 183–193. doi:10.1016/j.trf.2016.12.009

Haworth, N., Greig, K., & Nielson, A. (2009). Comparison of risk taking in moped and motorcycle crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2140*(1), 182–187. doi:10.3141/2140-20

Highway Safety Research Group. (2020). Louisiana crash reports. Retrieved from http://datareports.lsu.edu/Reports.aspx?yr=2019&rpt=A4b&p=ci.

Huang, B., & Preston, J. (2004). A literature review on motorcycle collisions: Final report. Transport Studies Unit, University of Oxford, Oxford, UK.

Jalayer, M., Pour-Rouholamin, M., & Zhou, H. (2018). Wrong-way driving crashes: A multiple correspondence approach to identify contributing factors. *Traffic Injury Prevention*, *19*(1), 35–41. doi:10.1080/15389588.2017.1347260

Jones, B., Janssen, L., & Mannering, F. (1991). Analysis of the frequency and duration of freeway accidents in Seattle. *Accident; Analysis and Prevention*, *23*(4), 239–255. doi:10.1016/0001-4575(91)90003-N

Li, M., Doong, J., Huang, W., Lai, C., & Jeng, M. (2009). Survival hazards of road environment factors between motor-vehicles and motorcycles. *Accident; Analysis and Prevention*, *41*(5), 938–947. doi:10.1016/j.aap.2009.05.009

Manan, M. M. A., Ho, J. S., Arif, S. T. M. S. T., Ghani, M. R. A., & Várhelyi, A. (2017). Factors associated with motorcyclists' speed behaviour on Malaysian roads. *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 109–127. doi:10.1016/j.trf.2017.08.006

Naqvi, H., & Tiwari, G. (2018). Factors contributing to motorcycle fatal crashes on National Highways in India. *International Journal of Injury Control and Safety Promotion*, *25*(3), 319–328. doi:10.1080/17457300.2018.1431937

NHTSA (2007). Fatal two-vehicle motorcycle crashes (DOT HS 810 834). Washington, DC: NHTSA

NHTSA (2020). Speeding: traffic safety facts. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812932.

Pour-Rouholamin, M., Jalayer, M., & Zhou, H. (2017). Modelling single-vehicle, single-rider motorcycle crash injury severity: An ordinal logistic regression approach. *International Journal of Urban Sciences*, *21*(3), 344–363. doi:10.1080/12265934.2017.1311801

Shaheed, M. S., & Gkritza, K. (2014). A latent class analysis of single-vehicle motorcycle crash severity outcomes. *Analytic Methods in Accident Research*, *2*, 30–38. doi:10.1016/j.amar.2014.03.002

State of Louisiana. (2020). Uniform motor vehicle traffic crash report. Retrieved from https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/la_crash_report_guide01_2005.pdf.

Teoh, E. R., & Campbell, M. (2010). Role of motorcycle type in fatal motorcycle crashes. *Journal of Safety Research*, *41*(6), 507–512.. doi:10.1016/j.jsr.2010.10.005

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident; Analysis and Prevention*, *72*, 244–256. doi:10.1016/j.aap.2014.06.017

Vachal, K., Malchose, D., & Benson, L. (2013). Motorcycles: Crash trends, conspicuity, and intervention (DP-266). North Dakota State University, Fargo: Upper Great Plains Transportation Institute. Van de Velden, M., Enza, A. I. D., & Palumbo, F. (2017). Cluster correspondence analysis. Psychometrika, *82*(1), 158–185.

Xin, C., Wang, Z., Lee, C., Lin, P., Chen, T., Guo, R., & Lu, Q. (2019). Development of crash modification factors of horizontal curve design features for single-motorcycle crashes on rural two-lane highways: A matched case-control study. *Accident Analysis & Prevention*, *123*, 51–59. doi:10.1016/j.aap.2018.11.008