JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2016; **50**:2089–2099
Published online 8 February 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.1447

A multiple correspondence analysis of at-fault motorcycle-involved crashes in Alabama

Mohammad Jalayer[1]* and Huaguo Zhou[2]

[1]*Center for Advanced Infrastructure and Transportation (CAIT), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, U.S.A.*
[2]*Department of Civil Engineering, Auburn University, Auburn, AL 36849, U.S.A.*

SUMMARY

According to the U.S. National Highway Traffic Safety Administration, in 2012, more than 4950 motorcyclists were killed in traffic accidents. Compared to passenger car occupants, mile for mile, motorcyclists are more than 26 times more at risk to dying in crashes. Due to the high fatality rate associated with motorcycle crashes, factors contributing to this type of crash must be identified in order to implement effective safety countermeasures. Given that the available datasets are large and complex, identifying the key factors contributing to crashes is a challenging task. Using multiple correspondence analysis, as an exploratory data analysis technique to determine the dataset structure, we identified the roadway/environmental, motorcycle, and motorcyclist-related variables influencing at-fault motorcycle-involved crashes. This study used the latest available dataset (2009 to 2013) from the Critical Analysis Reporting Environment database to study motorcycle crashes in the state of Alabama. The most significant contributors to the frequency and severity of at-fault motorcycle-involved crashes were found to be light conditions, time of day, driver condition, and weather conditions. Copyright © 2017 John Wiley & Sons, Ltd.

KEY WORDS: motorcycle crash; CARE database; multiple correspondence analysis; exploratory data analysis

1. INTRODUCTION

Motorcycles represent an increasing proportion of traffic casualties in the United States. In 2012, according to the National Highway Traffic Safety Administration (NHTSA), per registered vehicle, the fatality rate associated with motorcycles was 6 times that of passenger cars [1]. An analysis of motorcycle-crash statistics from the Fatality Analysis Reporting System database for the period from 2003 to 2012 reveals that an average of 4620 motorcyclist fatalities occurred nationally, accounting for more than 12% of the total traffic fatalities for that time period [2]. In order to get a better understanding of the most significant contributing factors, and then develop more effective safety countermeasures, these numbers require further analysis. In the motorcycle crash study domain, separating at-fault and not-at-fault motorcycle crashes is advisable due to different crash causes [3, 4]. Additionally, Elliott et al. [5, 6] demonstrated that riding errors were the main contributor of at-fault crash involvement, which was the crash type examined in this paper.

This study intends to utilize multiple correspondence analysis (MCA), as an exploratory data analysis (EDA) technique that identifies patterns in large and complex datasets, to identify key factors contributing to at-fault motorcycle-involved crashes. EDA can extract specific information from datasets and transform it into an understandable structure. To employ this method, we obtained datasets from 2009 to 2013 for the state of Alabama from the Critical Analysis Reporting Environment (CARE) database. The results of this study can help policymakers to gain a better understanding of the

*Correspondence to: Dr. Mohammad Jalayer, Center for Advanced Infrastructure and Transportation (CAIT), Rutgers, The State University of New Jersey, Piscataway, NJ 08854-8058, U.S.A. E-mail: mohammad.jalayer@rutgers.edu

major contributing factors to at-fault motorcycle-involved crashes and to then develop effective safety countermeasures accordingly.

2. LITERATURE REVIEW

2.1. *Contributing factors to motorcycle-involved crashes*

Several previous studies identified various contributing factors to motorcycle crashes through different tools. To examine the prevalence of alcohol and drugs among motorcycle riders killed in road crashes in Norway, Christophersen and Gjerde [7] gathered crash data from 2001 to 2010, along with information on gender, age, and day and time of crash. Using the Person chi-squared test, the authors compared the levels of alcohol, illicit drugs, and medications in blood samples of motorcyclists and car drivers killed during the same time period. The results show that the presence of drugs or alcohol in the blood was mostly found among riders between 25 and 34 years of age and in crashes that occurred during weekends and at night. Compared with blood samples from car drivers killed, those from killed motorcyclists showed lower percentages of alcohol and drugs. Shaheed et al. [8] investigated the factors contributing to crash severity in the state of Iowa between two vehicles, one of which was a motorcycle. These factors include roadway and environmental conditions, driver and vehicle conditions, location, and time. This study found that roadway-surface conditions, light conditions, speed limit, and the use of a helmet significantly influenced crash severity outcomes.

In another study conducted by Rome and Senserrick [9], roadway-surface conditions, road curve, driver license type, and driver age were identified as the most significant factors contributing to motorcycle crashes. Geedipally et al. [10] found that alcohol, gender, light conditions, and the presence of horizontal and vertical curves significantly contributed in motorcycle crash severity. Using Bayesian hierarchical models, Haque et al. [11] determined that four-way and T-type signalized intersections were significantly correlated with motorcycle crashes. The results demonstrated that the greater the number of lanes at four-way intersections, the higher the number of motorcycle crashes. Additionally, the presence of red light cameras mitigated the likelihood of motorcycle crashes at both intersection types. Kashani et al. [12] used 4-year (2009–2012) crash data to investigate the contributing factors to motorcycle pillion passengers' crash severity in Iran. The authors employed classification and regression trees as a data mining technique to identify the most influential factors in fatal motorcycle crashes. Their results indicated that land use, area type, and the part of the body affected were the most significant factors. In another study, Haque et al. [13] used log-linear models to investigate multi-vehicle motorcycle crashes in Singapore, and found that nighttime, wet road surface conditions, and intersections with many hazardous interactions and access point increased the probability of motorcycle-involved crashes. Taken together, the results of aforementioned studies demonstrate that roadway surface condition, light condition, motorcyclist condition, and crash time contributed significantly to motorcycle-involved crashes.

2.2. *Multiple correspondence analysis*

Regarding the methodology utilized in this study, while there is a substantial body of literature on the application of statistical methods in transportation research [14–19], a few of previous studies focused on application of MCA in transportation research. Mitchell et al. [20] compared novice and full-license driver common crash types in Australia using corresponding analysis. Factors such as vehicle speed, fatigue, and impaired driving were identified as the risk factors associated with novice driver crashes. Using MCA, Chauvin et al. [21] identified the human and organizational factors in maritime accidents. The results of the analysis revealed that the source of most accidents was a decision error. In another study, Mabunda et al. [22] utilized the MCA method to explore associations between age, sex, blood alcohol concentration, and time and day of death in pedestrian fatalities in South Africa. Nallet et al. [23] applied MCA to analyze the results of a questionnaire-based survey of a sample of drivers who took courses with the aim of awareness-raising about the causes and consequences of traffic accidents. In a recently published paper, Das and Sun [24] analyzed 8 years (2004–2011) of vehicle-pedestrian crash data in Louisiana using MCA. According to the obtained results, factors such as roadway alignment, lighting and weather conditions, and gender increase the likelihood of vehicle-pedestrian crashes. In another study, Das and Sun [25] used MCA to explore the contributing factors regarding

DOI: 10.1002/atr

fatal run-off-road crashes in Louisiana. Kim and Yamashita [26] also used the MCA method to investigate the characteristics of pedestrian-involved collisions in Hawaii. The targeted groups for training programs were found to be men, children, and young adults.

We note that although there are a considerable number of studies of the factors contributing to motorcycle crashes [7–13, 27–33], very few have used graphical EDA techniques for crash analysis. To our knowledge, no previous analysis of the CARE database has investigated contributing factors in at-fault motorcycle-involved crash frequency and severity, which we address in this paper.

3. METHOD AND DATA

3.1. Multiple correspondence analysis

MCA is a powerful technique for analysis and graphical presentation of categorical data in large and complex datasets [24, 25, 34–36]. MCA graphical overviews, which are more conventional rather than log-liner models, simplify the expression of the relationships between variables without the necessity of any preconditions, thereby making interpretation easier [25]. Additionally, very small and very large sample sizes significantly influence the performance of both count data and crash severity models [37]. MCA also has the capability to look at multiple types of data and dimensions simultaneously, which is in contrast to running countless bivariate analysis [26]. Detailed descriptions of this method and its development history can be found in the Das and Sun [24], Greenacre and Blasius [36], and LeRoux and Rouanet [38].

MCA is performed on an $I \times J$ indicator matrix in which I is the set of i individual records, motorcycle crashes, and J is the set of categories of all variables, crash contributing factors. Given this, the component in the cell (i, j) consists of the individual record i and category j [38]. For instance, "*Roadway Surface Condition*" is a nominal variable with two categories, dry vs. wet, with "0 1" for the dry and "1 0" for the wet. Associated categories in MCA are placed close together in a Euclidean space, leading clouds, or a combinations of points that have similar distributions [24, 25]. Notably, MCA produces two point clouds (i.e., individuals and categories), which are usually defined by two-dimensional graphs [24]. The cloud of individuals is based on the set of all distances between individual records for a variable, for which different categories of variable have been chosen [25]. For each variable, the squared distance between individuals associated with each category is calculated, based on Equation (1) shown in Table I [24, 34, 38]. The relative frequency

Table I. Equations of clouds of individuals and categories [24, 25].

Cloud	Equation
Individuals	1) $d_m^2(i, i') = \frac{1}{f_j} + \frac{1}{f_{j'}}$
	2) $D^2(i, i') = \frac{1}{M} \sum_{m \in M} d_m^2(i, i')$
Categories	3) $\left(N^j N^{j'}\right)^2 = \frac{n_j + n_{j'} - 2n_{jj'}}{n_j n_{j'}/n}$
Description of parameters	
$d_m^2(i, i')$	Squared distance between individuals i and i' for variable m
$D^2(i, i')$	Overall squared distance between individuals i and i'
$\left(N^j N^{j'}\right)^2$	Squared distance between categories j and j'
f_j	Relative frequency of individual records that selected category j
$f_{j'}$	Relative frequency of individual records that selected category j'
n_j	Number of individuals that selected category j
$n_{j'}$	Number of individuals that selected category j'
$n_{jj'}$	Number of individuals that selected both categories j and j'
n	Total number of individual records in database
M	Set of all variables

DOI: 10.1002/atr

of each category is determined as the total number of individual records that chose that particular category divided by the total number of individual records in the database. In order to determine the overall squared distance between two individual records, all individual squared distances must be added together, as shown in Equation (2) in Table I [24, 34, 38]. Note that the cloud of categories and the cloud of individuals have the same dimension, and each category in this cloud is defined by a point and a weight [25]. Equation (3) in Table I shows the squared distance between categories j and j'. We note that $n_{jj'}$ will be zero when categories j and j' are two categories of the same variable [24, 34, 38].

3.2. Data

To employ the MCA method, we compiled historical motorcycle crash data from the CARE database, an Internet-based online data analysis source, for a 5-year time period from 2009 through 2013. This database is an efficient tool that can analyze and categorize crash data from the standpoints of transportation safety engineers and policymakers. The CARE database was designed at the University of Alabama, in coordination with the State of Alabama and the NHTSA, and contains data pertaining to vehicles, roadways, and drivers—all factors that potentially contribute to crashes [39]. Specifically, this data includes crash location, crash type, crash severity, weather conditions, time, vehicle type, and driver age [39, 40]. Each state contributing to the CARE database gathers data from various sources (e.g., police crash reports, and emergency medical service reports) and converts them into a common format for transmission to the CARE database [40]. CARE software performs several major functions such as statistical generation, narrative data searching, hotspot determination, report generation, GIS integration, and collision diagram generation that can be used in crash pattern analysis [39]. More detailed descriptions of this database and its development history can be found in CARE [39] and Parish et al. [41]. Twelve state departments of transportation (i.e., Alabama, Delaware, Florida, Georgia, Iowa, Maryland, Michigan, Nevada, North Carolina, Rhode Island, Tennessee, and Wyoming) currently contribute to and use the CARE database for their jurisdictions [39], and, as such, the method set forth in this paper will also provide helpful guidance for states outside Alabama. For the purposes of this study, based on engineering study results gleaned from a comprehensive literature review, we nominated a set of key variables for investigation from the parameters included in the crash database, as shown in Table II. Since Alabama has established motorcycle helmet law, in this study, the effect of helmet use on injury severity was not considered due to the small sample size of motorcycle crashes without helmet, about 1% of crashes. We note that from the initial total sample of 5969 crash records within a 5-year time period in Alabama, we excluded 576 crashes due to missing values in one or more of the class variables. Table I lists all the contributing variables and categories, along with their frequencies and percentages. When looking at this table, a few points are worth mentioning. For some of these variables, the majority of crashes fall into one or two categories. For instance, 70% of crashes occurred during daylight hours, and 86% of the motorcyclists tested as being in normal physiological condition. Moreover, the majority of crashes happened on the weekends.

4. RESULTS AND DISCUSSIONS

As mentioned in the previous section, a set of selected categorical variables relating to motorcycle crashes were established to conduct the MCA analysis. In order to identify the key contributing factors, we used *R Version 3.02* statistical software and the *FactoMineR* package to analyze the dataset and plot the two-dimensional graphs. The MCA graphical representations help simplify the process of interpreting the relationships among variables [24]. In a two-dimensional graphical display of the data, categories sharing similar characteristics are located close together, forming point clouds [24, 34, 38]. The magnitude of information associated with each dimension is called eigenvalue [24, 25]. The eigenvalue of each dimension, which is a value between 0 and 1, indicates the total variance between variables. We note that the first and second dimensions had higher eigenvalues compared to other dimensions, so a two-dimensional graph includes most of the information, as shown in Figure 1. Every point on each plot is uniquely coordinated for all dimensions, and, obviously, the scale of the plot

DOI: 10.1002/atr

Table II. At-fault motorcycle-involved crash distributions based on study category.

Variable	Category	Frequency	Percentage (%)
Month	January	215	4.0
	February	275	5.1
	March	482	8.9
	April	633	11.7
	May	574	10.6
	June	561	10.4
	July	553	10.3
	August	530	9.8
	September	572	10.6
	October	507	9.4
	November	320	5.9
	December	171	3.2
Day of the week	Monday	557	10.3
	Tuesday	595	11.0
	Wednesday	628	11.6
	Thursday	648	12.0
	Friday	839	15.6
	Saturday	1172	21.7
	Sunday	954	17.7
Time of the day	0:00 to 1:59	181	3.4
	2:00 to 3:59	115	2.1
	4:00 to 5:59	95	1.8
	6:00 to 7:59	243	4.5
	8:00 to 9:59	265	4.9
	10:00 to 11:59	519	9.6
	12:00 to 13:59	724	13.4
	14:00 to 15:59	880	16.3
	16:00 to 17:59	971	18.0
	18:00 to 19:59	661	12.3
	20:00 to 21:59	464	8.6
	22:00 to 23:59	275	5.1
Area type	Rural	2528	46.9
	Urban	2865	53.1
Light condition	Dark	1338	24.8
	Dawn	45	0.8
	Dusk	184	3.4
	Daylight	3826	70.9
Weather condition	Clear	4306	79.8
	Cloudy	861	16.0
	Fog	23	0.4
	Mist	27	0.5
	Rain	167	3.1
	Severe winds	9	0.2
Roadway surface condition	Dry	5047	93.6
	Wet	346	6.4
Roadway curvature and grade	Curve and level	841	15.6
	Curve at hillcrest	36	0.7
	Curve with downgrade	596	11.0
	Curve with upgrade	403	7.5
	Straight and level	2565	47.6
	Straight at hillcrest	65	1.2
	Straight with downgrade	520	9.6
	Straight with upgrade	367	6.8
One-way street	Yes	129	2.4
	No	5264	97.6
Motorcyclist residence distance (mile)	25 and less	4152	77.0
	Greater than 25	1241	23.0
Intersection-related crash	Yes	666	12.3
	No	4727	87.7

(*Continues*)

DOI: 10.1002/atr

Table II. (Continued)

Variable	Category	Frequency	Percentage (%)
Work zone-related crash	Yes	5291	98.1
	No	102	1.9
Traffic lanes	One	114	2.1
	Two	3561	66.0
	Three	143	2.7
	Four	1206	22.4
	Five	101	1.9
	Six and more	268	5.0
Speed limit (mph)	1 to 19	32	0.6
	20 to 29	485	9.0
	30 to 39	1141	21.2
	40 to 49	2044	37.9
	50 to 59	1240	23.0
	60 to 69	260	4.8
	70 and greater	191	3.5
Motorcyclist gender	Male	5064	93.9
	Female	359	6.7
Motorcyclist age	Less than 20	347	6.4
	20 to 29	1352	25.1
	30 to 39	1072	19.9
	40 to 49	1135	21.0
	50 to 59	963	17.9
	60 to 99	521	9.7
	100 and greater	3	0.1
Motorcyclist condition	Alcohol/drug/medicine	313	5.8
	Apparently normal	4686	86.9
	Asleep/fainted/fatigued	22	0.4
	Physical impairment	8	0.1
	Unknown	364	6.7
Motorcycle model	1964 to 1973	17	0.3
	1974 to 1983	119	2.2
	1984 to 1993	268	5.0
	1994 to 2003	1618	30.0
	2004 to 2013	3371	62.5
Primary contributing factors	Aggressive operation	493	9.2
	Defective equipment	142	2.6
	Distraction	18	0.3
	Driving under the influence (DUI)	180	3.3
	Failed to yield right-of-way	145	2.7
	Fatigued/asleep	15	0.3
	Improper action	477	8.8
	Over speed limit	687	12.7
	Swerved	802	14.9
	Traveling wrong way/wrong side	30	0.6
	Unseen object/person/vehicle	289	5.4
	Other	684	12.7
	Unknown	1431	26.5
Crash severity	Fatal	265	4.9
	Incapacitating injury	1782	33.0
	Non-incapacitating injury	1716	31.8
	Possible injury	368	6.8
	Property damage only	1262	23.4

depends heavily on the total amount of contributions by each dimension. Figure 2 maps all the study variables and their relative proximities.

To interpret an MCA plot, we compared individual records, variables, and categories within a variable by gauging the distances between the points on the map [42]. Figure 2 shows that many variables are placed near each other, thus making roughly the same contribution to all the

DOI: 10.1002/atr

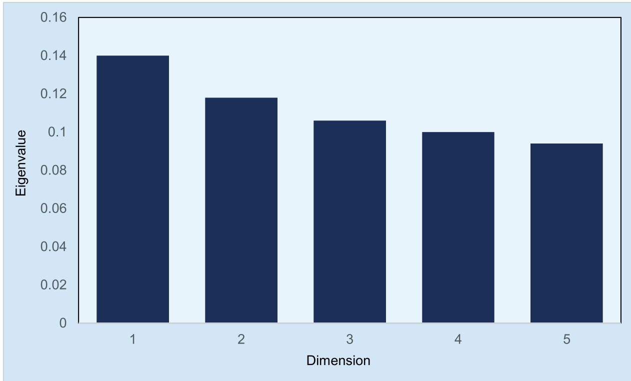


Figure 1. Eigenvalues of the top five dimensions.

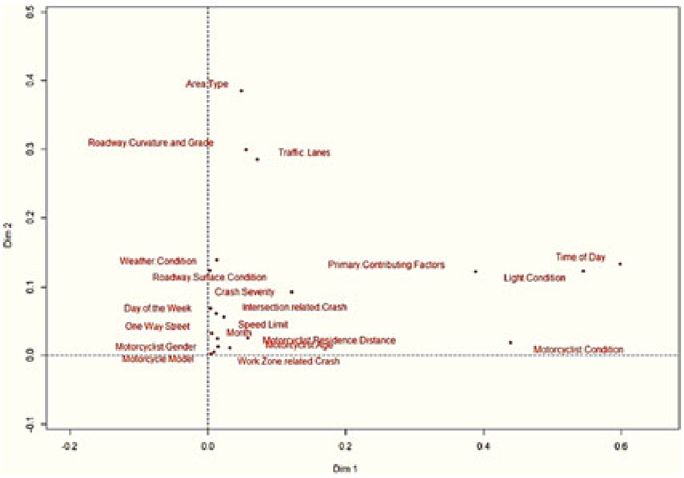


Figure 2. MCA plot of all study variables.

variances. Additionally, the points close to the centroid of the map, for one dimension, contribute less to the eigenvalue of that particular dimension [24, 34, 38]. Therefore, for dimension one factors such as time of day, light conditions, and motorcyclist condition and for dimension two, area type, roadway alignment, and number of traffic lanes contributed the most in motorcycle crash frequency and severity, which is consistent with the findings of a majority of the existing literature [27–32]. Considering the coefficient of determination (R^2) and the p-value of the F-test, all the at-fault motorcycle-involved contributing factors in this study were identified in descending order of significance, as shown in Table III. The range of the coefficient of determination is 0 to 1. A value of zero indicates no correlation between the qualitative variable and the MCA dimension and an absolute value of 1.0 indicates a perfect correlation [24, 25]. As we can see, compared to the time of day and motorcyclist condition, the risk of motorcycle crashes is not strongly associated with motorcyclist gender, work zone, or intersection. Four other similar studies by Haque et al. [3], Patil et al. [43], Schneider IV and Savolainen [44], and Indupuru [45] also show the same trends.

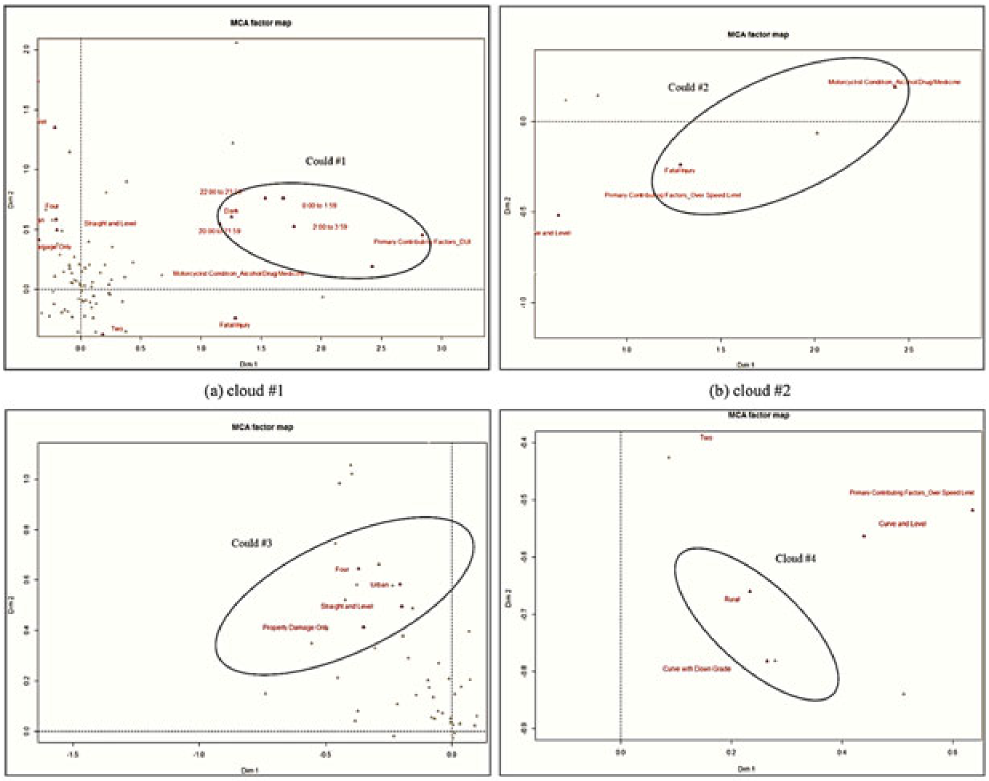
Figure 3 illustrates a two-dimensional plot of the top 20 categories that contributed most to at-fault motorcycle-involved crashes. According to this figure, several point clouds can be created based on the relative proximity of point combination. For instance, Cloud 1 associates driving under the influence of drugs or alcohol as the primary contributing factor with the time of day between

DOI: 10.1002/atr

Table III. Significance of test results for key at-fault motorcycle-involved contributing factors in top two dimensions.

	Variable	R^2	p-Value
Dimension 1	Time of day	0.598	<0.001
	Light condition	0.545	<0.001
	Motorcyclist condition	0.439	<0.001
	Primary contributing factors	0.389	<0.001
	Crash severity	0.121	<0.001
	Traffic lanes	0.072	<0.001
	Motorcyclist age	0.058	<0.001
	Roadway curvature and grade	0.056	<0.001
	Area type	0.048	<0.001
	Motorcyclist residence distance	0.032	<0.001
	Speed limit	0.023	<0.001
	Motorcyclist gender	0.015	<0.001
	Intersection-related crash	0.012	<0.001
	Weather condition	0.013	<0.001
	Month	0.014	<0.001
	Motorcycle model	0.009	<0.001
	Work zone-related crash	0.004	<0.001
	Roadway surface condition	0.003	<0.001
	Day of the week	0.004	0.01
Dimension 2	Area type	0.385	<0.001
	Roadway curvature and grade	0.299	<0.001
	Traffic lanes	0.285	<0.001
	Weather condition	0.139	<0.001
	Time of day	0.134	<0.001
	Roadway surface condition	0.125	<0.001
	Light condition	0.124	<0.001
	Primary contributing factors	0.122	<0.001
	Crash severity	0.093	<0.001
	Day of the week	0.070	<0.001
	Intersection-related crash	0.062	<0.001
	Speed limit	0.056	<0.001
	One-way street	0.032	<0.001
	Motorcyclist age	0.026	<0.001
	Month	0.024	<0.001
	Motorcyclist condition	0.019	<0.001
	Motorcyclist gender	0.013	<0.001
	Motorcyclist residence distance	0.012	<0.001
	Motorcycle model	0.006	<0.001
	Work zone-related crash	0.003	<0.001

22:00 pm and 4:00 am. This result is consistent with the NHTSA report that shows motorcyclists who were killed in nighttime crashes were more than 3 times more likely to be drunk, with blood alcohol concentrations of 0.08 g/dL or more, compared to those who were killed during the day [1]. Cloud 2 combines factors such as fatal crashes, speeding, and impaired conditions. This means that impaired motorcyclist condition and driving over the speed limit significantly increase the severity of at-fault motorcycle-involved crashes. These results are in good agreement with the findings of Barrette et al. [46], demonstrating the stronger effect of an impaired motorcyclist and speeding on the severity of crashes compared to not wearing a helmet. In another combination (Cloud 3), crashes with lesser outcomes, known as property damage only (PDO) crashes, occurred more often in four-lane highways in urban areas. This result appears to be in line with the findings of another study on motorcycle crash prediction model [33]. According to Cloud 4, the possibility of a motorcyclist being involved in a fatal crash increases if the crash occurs on a curve with down-grade superelevation in a rural area which is consistent with previous studies [44, 47]. Adinegoro et al. [48] also demonstrated that motorcycle crashes in rural areas were more likely to result in fatal/severe injury outcomes.



(a) cloud #1 (b) cloud #2

(c) cloud #3 (d) cloud #4

Figure 3. MCA plot of top 20 key categories: (a) cloud #1; (b) cloud #2; (c) cloud #3; (d) cloud #4.

5. CONCLUSIONS AND RECOMMENDATIONS

In this paper, the MCA method is used to evaluate the roadway/environmental, motorcycle, and motorcyclist-related variables that affect the severity and frequency of at-fault motorcycle-involved crashes. We gathered 5 years of at-fault motorcycle-involved crash data from the CARE database for the years 2009 to 2013 for the state of Alabama. According to the obtained results, the main contributing factors to at-fault motorcycle-involved crashes are light conditions, time of day, driver condition, roadway curvature and grade, and weather conditions. During the nighttime, i.e., from 8 pm to 6 am, there is usually no light in the sky, which results in an interrelation of factors contributing to crashes. Motorcyclists with higher blood-alcohol levels are involved in crashes with more severe outcomes. The likelihood of a collision on a curve with downgrade superelevation in a rural area is associated with increased severity of motorcycle-involved crashes. Additionally, the likelihood of motorcycle-involved crashes is increased in rainy and wet roadway-surface conditions, when friction between the wheels and the road surface is reduced. The provision of more travel lanes in urban areas, with a minimum of four lanes, is associated with reduced severity of motorcycle-involved crashes. We note that the risk of motorcycle-involved crashes is not strongly associated with motorcyclist demographics (i.e., gender, age) or motorcycle model.

With respect to the total explained variances in the study variables, eigenvalue correction by analyzing the MCA on the Burt matrix can be conducted to increase the variances [24, 25, 35]. The ability to provide the visualization of variable clusters makes MCA a strong candidate, compared to conventional logit models, for analyzing patterns among categories of qualitative variables without making any prior assumptions [24, 49, 50]. Although the approach employed in this study does not compute marginal effects of variables, this method is exceptional in its ability to identify the most statistically significant combinations of factors. Given this fact, MCA has clear potential to help state departments of transports prioritize crash mitigation strategies with multiple benefits based on their large crash databases.

DOI: 10.1002/atr

6. LIST OF ABBREVIATIONS

CARE	Critical Analysis Reporting Environment
EDA	Exploratory Data Analysis
MCA	Multiple Correspondence Analysis
NHTSA	National Highway Traffic Safety Administration
PDO	Property Damage Only

REFERENCES

1. National Highway Traffic Safety Administration (NHTSA). Traffic safety facts 2012 data. U.S. Department of Transportation. 2014. Washington D.C.
2. National Highway Traffic Safety Administration (NHTSA). Fatality Analysis Reporting System (FARS). http://nhtsa.gov/FARS [1 October 2015].
3. Haque MM, Chin HC, Huang H. Modeling fault among motorcyclists involved in crashes. *Journal of Accident Analysis and Prevention* 2009; **41**(2): 327–335.
4. Savolainen P, Mannering F. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. *Journal of Accident Analysis and Prevention* 2007; **39**: 955–963.
5. Elliott M, Baughan C, Broughton J, Chinn B, Grayson G, Knowles J, Smith L, Simpson H. Motorcycle safety: a scoping study. TRL Report 581. Crowthorne: Transport Research Laboratory.
6. Elliott M, Baughan CJ, Sexton BF. Errors and violations in re relation to motorcyclists' crash risk. *Journal of Accident Analysis and Prevention* 2007; **39**(3): 491–499.
7. Christophersen A, Gjerde H. Prevalence of alcohol and drugs among motorcycle riders killed in road crashes in Norway during 2001–2010. *Journal of Accident Analysis and Prevention* 2015; **80**: 236–242.
8. Shaheed MSB, Gkritza K, Zhang W, Hans Z. A mixed logit analysis of two-vehicle crash severities involving motorcycle. *Journal of Accident Analysis and Prevention* 2013; **61**: 119–128.
9. Rome LD, Senserrick T. Factors associated with motorcycle crashes in New South Wales, Australia, 2004 to 2008. *Transportation Research Record: Journal of Transportation Research Board* 2011; **2265**: 54–61.
10. Geedipally SR, Turner PA, Patil S. Analysis of motorcycle crashes in Texas with multinomial logit model. *Transportation Research Record: Journal of Transportation Research Board* 2011; **2265**: 62–69.
11. Haque MM, Chin HC, Huang H. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Journal of Safety Science* 2010; **42**: 203–212.
12. Kashani AT, Rabieyan R, Besharati MM. A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers. *Journal of Safety Research* 2014; **51**: 93–98.
13. Haque MM, Chin HC, Debnath AK. An investigation on multi-vehicle motorcycle crashes using log-linear models. *Journal of Safety Science* 2012; **50**: 352–362.
14. Khalilikhah M, Heaslip K. Analysis of factors temporarily impacting traffic sign readability. *International Journal of Transportation Science and Technology* 2016; **5**: 60–67.
15. Sharifi MS, Shabaniverki H. Modeling crash delays in a route choice behavior model for two way road network. *Journal of Geotechnical and Transportation Engineering* 2016; **2**(1): 19–25.
16. Jalayer M, Zhou H, Williamson M, LaMondia J. Developing calibration factors for crash prediction models with consideration of crash recording threshold change. *Transportation Research Record: Journal of Transportation Research Board* 2015; **2515**: 57–62.
17. Soltani-Sobh A, Heaslip K, Bosworth R, Barnes R. Effect of improving vehicle fuel tax revenue and greenhouse gas emissions. *Transportation Research Record: Journal of Transportation Research Board* 2015; **2502**: 71–79.
18. Khalilikhah M, Heaslip K. Important environmental factors contributing to the temporary obstructing of the sign message. In *Transportation Research Board 95th Annual Meeting.* 2016.
19. Jalayer M, Zhou H. Evaluating the safety risk of roadside features for rural two-lane roads using reliability analysis. *Journal of Accident Analysis and Prevention* 2016; **93**: 101–112.
20. Mitchell RJ, Senserrick T, Bambach MR, Mattos G. Comparison of novice and full-licenced driver common crash types in New South Wales, Australia, 2001–2011. *Journal of Accident Analysis and Prevention* 2015; **81**: 204–210.
21. Chauvin C, Lardjane S, Morel G, Clostermann J, Langrad B. Human and organisational factors in maritime accidents: analysis of collisions at sea using the HFACS. *Journal of Accident Analysis and Prevention* 2013; **59**: 26–37.
22. Mabunda MM, Swart L, Seedat M. Magnitude and categories of pedestrian fatalities in South Africa. *Journal of Accident Analysis and Prevention* 2008; **40**: 586–593.
23. Nallet N, Bernard M, Chiron M. Individuals taking a French driving licence points recovery course: their attitudes towards violations. *Journal of Accident Analysis and Prevention* 2008; **40**: 1836–1843.
24. Das S, Sun X. Factor association using multiple correspondence analysis in vehicle-pedestrian crashes. *Transportation Research Record: Journal of Transportation Research Board* 2015; **2519**: 95–103.
25. Das S, Sun X. Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research* 2016; **39**(2): 146–155.

DOI: 10.1002/atr

26. Kim K, Yamashita E. Corresponding characteristics and circumstances of collision-involved pedestrian in Hawaii. *Transportation Research Record: Journal of Transportation Research Board* 2008; **2424**: 18–24.
27. Fagnant DJ, Kockelman KM. Motorcycle use in the United States: crash experiences, safety perspectives, and countermeasures. *Journal of Transportation Safety and Security* 2015; **7**(1): 20–39.
28. Simpson JC, Wilson S, Currey N. Motorcyclists' perceptions and experiences of riding and risk and their advice for safety. *Journal of Traffic Injury Prevention* 2015; **16**(2): 159–167.
29. Craen SD, Doumen MJ, Norden YV. A different perspective on conspicuity related motorcycle crashes. *Journal of Accident Analysis and Prevention* 2014; **63**: 133–137.
30. Flask T, Schneider WH, Lord D. A segment level analysis of multi-vehicle motorcycle crashes in Ohio using Bayesian multi-level mixed effects models. *Journal of Safety Science* 2014; **66**: 47–53.
31. Abbas AK, Hefny AF, Abu-Zidan FM. Does wearing helmets reduce motorcycle-related death? A global evaluation. *Journal of Accident Analysis and Prevention* 2012; **49**: 249–252.
32. Jou RC, Yeh TH, Chen RS. Risk factors in motorcyclist fatalities in Taiwan. *Journal of Traffic Injury Prevention* 2012; **13**(2): 155–162.
33. Harnen S, Wong SV, Umar RSR, Hashim WIW. Motorcycle crash prediction model for non-signalized intersections. *Journal of IATSS Research* 2003; **27**(2): 59–65.
34. Das S, Sun X. Exploring clusters of contributing factors for single-vehicle fatal crashes through multiple correspondence analysis. In *Transportation Research Board 93rd Annual Meeting* Washington D.C. 2014.
35. Abdi H, Valentin D. Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics* 2007; 651–657.
36. Greenacre M, Blasius J. *Multiple Correspondence Analysis and Related Methods* (1st edn), Chapman and Hall/CRC: United Kingdom, 2006.
37. Ye F, Lord D. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. *Journal of Analytic Methods in Accident Research* 2014; **1**: 72–85.
38. LeRoux B, Rouanetm H. *Multiple Correspondence Analysis*, SAGE Publications, Inc.: United State, 2010.
39. Critical Analysis Reporting Environment (CARE). http://caps.ua.edu/care.aspx [1 October, 2015].
40. Jalayer M, Baratian-Ghorghi F, Zhou H. Identifying and characterizing secondary crashes on the Alabama state highway systems. *Journal of Advances in Transportation Studies* 2015; **37**: 129–140.
41. Parrish A, Brown D, Stricklin R, Turner D. Critical analysis reporting environment (CARE). *WIT Transactions on Information and Communication Technologies* 2003; **29**: 119–128.
42. Correspondence Analysis. http:// www.wiki.q-researchsoftware.com/wiki/Correspondence_Analysis [1 October, 2015].
43. Patil S, Geedipally SR, Lord D. Analysis of crash severities using nested logit model—accounting for the underreporting of crashes. *Journal of Accident Analysis and Prevention* 2012; **45**: 646–653.
44. Schneider WH, Savolainen PT. Comparison of severity of motorcyclist injury by crash types. *Transportation Research Record: Journal of Transportation Research Board* 2011; **2265**: 70–80.
45. Indupuru VK. Identification of factors related to motorcycle fatal injuries in Ohio. Doctoral dissertation, *University of Dayton*. 2010.
46. Barrette T, Kirsch T, Savolainen P, Russo B, Gates T. Disaggregate-level assessment of changes to Michigan's motorcycle helmet use law: effects on motorcyclist injury outcomes. *Transportation Research Record: Journal of the Transportation Research Board* 2014; **2468**: 131–137.
47. Abegaz T, Berhane Y, Worku A, Assrat A, Assefa A. Effects of excessive speeding and falling asleep while driving on crash injury severity in Ethiopia: a generalized ordered logit model analysis. *Journal of Accident Analysis and Prevention* 2014; **71**: 15–21.
48. Adinegoro Y, Haworth N, Debnath AK. Characteristics of road factors in multi and single vehicle motorcycle crashes in Queensland. In *2015 Australasian Road Safety Conference.* Queensland, Australia 2015.
49. Chatfield C. *Problem Solving: A Statistician's Guide* (2nd edn), Chapman and Hall/CRC: United Kingdom, 1995.
50. Cook D, Swayne DF. *Interactive and Dynamic Graphics for Data Analysis*, Springer-Verlag: New York, 2007.

DOI: 10.1002/atr