RAIRO-Oper. Res. 55 (2021) 1257–1278
https://doi.org/10.1051/ro/2021055

RAIRO Operations Research
www.rairo-ro.org

PREDICTION OF MOTORCYCLIST TRAFFIC CRASHES IN CARTAGENA (COLOMBIA): DEVELOPMENT OF A SAFETY PERFORMANCE FUNCTION

Holman Ospina-Mateus[1,2], Leonardo Augusto Quintana Jiménez[2], Francisco J. Lopez-Valdes[3] and Shib Sankar Sana[4,*]

Abstract. Motorcyclists account for more than 380 000 deaths annually worldwide from road traffic accidents. Motorcyclists are the most vulnerable road users worldwide to road safety (28% of global fatalities), together with cyclists and pedestrians. Approximately 80% of deaths are from low- or middle-income countries. Colombia has a rate of 9.7 deaths per 100 000 inhabitants, which places it 10th in the world. Motorcycles in Colombia correspond to 57% of the fleet and generate an average of 51% of fatalities per year. This study aims to identify significant factors of the environment, traffic volume, and infrastructure to predict the number of accidents per year focused only on motorcyclists. The prediction model used a negative binomial regression for the definition of a Safety Performance Function (SPF) for motorcyclists. In the second stage, Bayes' empirical approach is implemented to identify motorcycle crash-prone road sections. The study is applied in Cartagena, one of the capital cities with more traffic crashes and motorcyclists dedicated to informal transportation (motorcycle taxi riders) in Colombia. The data of 2884 motorcycle crashes between 2016 and 2017 are analyzed. The proposed model identifies that crashes of motorcyclists per kilometer have significant factors such as the average volume of daily motorcyclist traffic, the number of accesses (intersections) per kilometer, commercial areas, and the type of road and it identifies 55 critical accident-prone sections. The research evidences coherent and consistent results with previous studies and requires effective countermeasures for the benefit of road safety for motorcyclists.

Mathematics Subject Classification. 90B06, 62J05, 62M10, 62C12.

Received November 20, 2020. Accepted April 5, 2021.

Keywords. Motorcycle, crashes, prone-section, safety performance function, negative binomial regression, empirical Bayesian approach.

[1] Universidad Tecnológica de Bolívar, Department of Industrial Engineering, Cartagena, Colombia.

[2] Pontificia Universidad Javeriana, Department of Industrial Engineering, Carrera 7 # 40-62, Bogotá, Colombia.

[3] Universidad Pontificia Comillas, Instituto de Investigacion Tecnológica (IIT), ICAI Engineering School, c/Alberto Aguilera 25, 28250 Madrid, Spain.

[4] Department of Mathematics, Kishore Bharati Bhagini Nivedita College, 148, Ramkrishna Sarani, Behala, Kolkata 700060, India.

*Corresponding author: shib_sankar@yahoo.com

Article published by EDP Sciences

1. Introduction

The World Health Organization (WHO) estimates 1.35 million fatalities in road traffic crashes [63]. Motorcyclists account for more than 380 000 annual deaths worldwide (28% of global fatalities). The 80% of deaths are from low- or middle-income countries. The most vulnerable victims in these countries are in the age range of 15–35 years. Motorcycle users are at greater risk due to their level of exposure and lack of an efficient protection system [35, 45]. The growth of the vehicle fleet, the deficit of roads, and the increase in journeys are important aspects of the accident of motorcyclists [62]. In South America, the Dominican Republic, Paraguay, and Colombia have the highest proportions of fatalities in motorcyclists with more than 50% of road users (see Tab. 1). Colombia faces a rate of 9.7 fatal motorcyclists per 100 000 inhabitants, ranking tenth worldwide, third in the region, and second in South America [46, 48].

In 2019, motorcycles in Colombia correspond to 57% of the vehicle fleet (8.6 million) [43] and motorcycles make approximately 50 million daily trips [12]. Between 2012 and 2018, an average of 3100 motorcyclist fatalities have occurred, corresponding to 51% of road fatalities [44]. In Colombia, a motorcyclist dies every 2.5 h, six motorcyclists injured in an hour [48]. Global road traffic crashes are critical, and Colombia is no exception when analyzing vulnerable users such as motorcyclists. The present study seeks to identify environmental factors and prioritize sections prone to crashes in motorcyclists. In a practical context, collisions can be reduced by implementing specific actions in strategic places with a high potential for road crashes.

The objective of this study is to identify significant factors of the environment, infrastructure, road flow, and road conditions, as well as critical sections prone to motorcyclist accidents in Cartagena. The prediction model used a negative binomial regression for the definition of a Safety Performance Function (SPF) of motorcyclists. In the second stage, Bayes' empirical approach is implemented to identify critical sections of motorcycle crashes. This study includes 7 sections. Background and context relating to the development of safety performance functions are described in Section 2. Section 3 describes the theoretical framework and settings. Section 4 describes the methodology used, the data, and the variables. Section 5 contains the results of the predictive models developed. Finally, the discussion and conclusions are presented in Sections 6 and 7, respectively.

2. Background and literature review

Road safety as a public health problem requires techniques to identify multidimensional aspects of the environment, individuals, and vehicles [42]. The identification of critical sections is one of the main strategies and analyzes carried out for the benefit of road safety programs [54]. From an environmental perspective, the development of prediction models for crash frequency analysis has made substantial progress in recent years, these are known as Safety Performance Functions (SPF) [53]. Prediction models based on SPFs are statistical analysis tools to identify the association between accident risks and accident conditions [25]. The SPFs relate the frequency of crashes with characteristics of the road (quality, size, lanes, separators, accesses, curves, among others), traffic flow (volume), speed, as well as environmental conditions (signaling, traffic lights, type of area) [1]. These models have been estimated with Poisson regression techniques, negative binomial, generalized linear models, and the Bayesian approach [34]. Currently, predictive models based on the development of an SPF and the empirical Bayesian approach are widely recommended [53]. These analyzes allow detailed interaction of the environment variables and identifying critical road sections with a high probability of crashes [1].

Road crashes are random, non-negative, and discrete events, these can be represented using the Poisson probability distribution or the non-negative binomial distribution, if there is "over dispersion" [34]. Studies by Hauer *et al.* [26], Cheng and Washington [14], and Montella [41] have showed that the number of traffic crashes follows a non-negative binomial distribution, based on excessive dispersion compared to a Poisson model. Several studies have used and recommended non-negative binomial regressions as best-fit models to analyze the association between traffic crashes and environmental, infrastructure, and operational conditions [2, 4, 37, 56, 57]. Furthermore, the Bayesian empirical approach has been widely used in road safety analysis to identify prone sections and black spots [50]. Elvik [18] has described a black spot as "a place that has a higher

TABLE 1. Ranking of countries with the highest rate of motorcycle accidents per 100 000 inhabitants in 2016 [63].

Country	Income level	Population numbers (million)	Total registered vehicles (million)	Registered motorcycle		Fatalities			Fatalities per 100 000 population	Fatalities per 100 000 population
				Total (Millions)	(%)	Total	Motorcycle	(%)		
Thailand	Middle	68.9	37.3	20.5	55%	21 745	16 178	74%	32.7	24.3
Dominican Republic	Middle	10.6	3.9	2.1	54%	3118	2089	67%	34.6	23.2
Togo	Low	7.6	0.1	0.0	71%	514	368	72%	29.2	20.9
Benin	Low	10.9	0.5	0.2	42%	637	360	57%	27.5	15.5
Cook Islands	High	0.02	0.01	0.01	55%	5	4	80%	17.3	13.8
Cambodia	Middle	15.8	3.8	2.7	72%	1852	1361	74%	17.8	13.1
Myanmar	Middle	52.9	6.4	5.4	84%	4887	3167	65%	19.9	12.9
Paraguay	Middle	6.7	1.9	0.6	33%	1202	627	52%	22.7	11.8
Mali	Low	18.0	0.3	0.1	16%	541	229	42%	23.1	9.8
Colombia	Middle	48.7	13.5	7.5	56%	7158	3758	53%	18.5	9.7

Notes. MC: motorcyclists.

expected number of accidents than other similar places as a result of a local risk factor". Several studies have implemented this approach to prioritize black spots, critical sections, accident-prone segments [5,8,20,58,65].

Road safety investigations are necessary to identify factors associated with the accident of a highly vulnerable user, such as motorcyclists in a specific context [47]. In the global context, motorcycle accident prediction models are limited. Analyses based solely on accident statistics are insufficient to evaluate the road safety performance of the environment. Designing road safety models from a predictive approach contributes to the definition of efficient policies. In the global context, there have been successful studies in the application of safety performance Functions such as those developed by Lord and Persaud [36], El-Basyouny and Sayed [17], Vogt and Bared [60], Tegge *et al.* [59], and Bauer and Harwood [7], among others.

When the SPF models for motorized vehicles are established within the 2010 Highway Safety Manual (HSM), it aims to promote this type of model for vulnerable actors such as motorcyclists and cyclists [13]. Those studies with motorcyclists are still ongoing. Among the studies are those developed by Abdul Manan *et al.* [3], de Lapparent [16], Xin *et al.* [64], Radin Umar *et al.* [55], Harnen *et al.* [30] and Lyon *et al.* [38].

Abdul Manan *et al.* [3] developed a negative binomial regression to predict fatal crashes in motorcyclists on Malaysian primary roads. The findings show that motorcyclist deaths per kilometer on major roads are statistically significantly influenced by the average daily number of motorcycles and the number of access points per kilometer. de Lapparent [16] has developed a model based on the empirical approach of Bayes to identify the severity of accidents involving motorcyclists in urban areas in France. Xin *et al.* [64] have quantified the effects of horizontal curve parameters and contributing factors on the occurrence of motorcycle crashes in specific sections of the road with a random parameter negative binomial regression model. Radin Umar *et al.* [55] have developed a multivariate analysis of motorcycle accidents and the effects of motorcycle exclusive lanes in Malaysia. Haque and Chin [23] have endeavored to identify the factors that affected motorcycle accidents at reported three- or four-legged intersections by developing Bayesian accident prediction models in Singapore. Finally, Lyon *et al.* [38] have developed an analysis of the total annual average daily traffic as a surrogate for motorcycle volume in estimating safety performance functions for motorcycle accidents in the USA.

In Colombia, studies on road safety for motorcyclists are limited; therefore, detailed analyzes are required on factors associated with motorcycle accidents [32,48,49]. Many studies that evaluate the performance of road safety at a global level are applied with a generalized focus on all road actors [3,16]. Due to the limited studies focused on motorcycling, the need for this type of research is evident, and more so in a developing country (LMIC) such as Colombia.

This research is a pioneer in the development of a predictive model in the capital city with high accident rates for motorcyclists and where informal transport on motorcycles is most practiced in Colombia (Motorcycle taxi riders). Developing a reliable accident prediction model is a demanding task. Therefore, it is an opportunity to expand the line of knowledge to vulnerable road users such as motorcyclists. In addition to being a new model in road safety for motorcycles in Colombia, the model considers aspects of infrastructure, environmental conditions, and the road. Likewise, the proposed model assesses the impact of road accidents on the total traffic volume and the volume exclusively of motorcyclists.

3. Theoretical framework and setting

3.1. Negative Binomial Regression

The Negative Binomial Regression is an adjusted model in which the dependent variable Y consists of counts or frequencies. The model relates Y to one or more predictor variables X, which can be quantitative or categorical. This regression is like the Poisson regression process; however, the conditional variance of Y is greater than the mean. Thus, the model is relevant in terms of "over-dispersion" compared to a Poisson process. The statistical model is expressed as follows:

$$p\left(Y\right)=\frac{\Gamma\left(Y+\alpha^{-1}\right)}{\Gamma\left(Y+1\right)\Gamma(\alpha^{-1})}\left[\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right]^{\alpha^{-1}}\left[\frac{\mu}{\alpha^{-1}+\mu}\right]^{Y},\quad \mu>0,\ \alpha\geq 0$$

where the mean μ is the product of λ, the rate at which accidents occur, and the observation period t, thus:

$$E\left(Y\right)=\mu=\lambda t.$$

The variance of Y is given with the parameter of over-dispersion α, as follows:

$$\operatorname{Var}\left(Y\right)=\mu+\alpha\mu^{2}.$$

To relate to the frequency of accidents, there is a need to express it in terms of an exponential function representing the expected number of accidents with a positive value. The accident frequency can be estimated with predictive variables on one log-linear scale as follows:

$$E\left(Y\right)=\operatorname{Exp}\left[\sum_{i=0}^{k}\beta_{i}\operatorname{Ln}\left(X_{i}\right)\right]$$
$$\operatorname{Log}\left(Y\right)=\beta_{0}+\beta_{1}\operatorname{Ln}\left(X_{1}\right)+\beta_{2}\operatorname{Ln}\left(X_{2}\right)\ldots\beta_{k}\operatorname{Ln}\left(X_{k}\right)$$

where β is the estimation coefficient, and X_i are the independent variables.

3.2. Application of the Bayesian empirical approach

The application of the method is also known as "regression to the mean" or in road safety studies "before and after" [33]. The approach eliminates bias in the observed number of accidents due to random fluctuations. The method calculates the expected number of accidents with the observed number and the estimated number of accidents in a road section [27]. In the model, the number of accidents is normalized by sections and these are expressed as accidents per unit of length of the road. The result of predicting road accidents by sections is a linear combination of two numbers, like this:

$$E\left(A\right)=wY_{\text{pre}}+\left(1-w\right)C_{\text{obs}}$$

where $E(A)$ is the estimated number of expected accidents in a unit of time. Y_{pre} is the predicted number, and C_{obs} is the number of observed accidents in a unit of time. Finally, w is a statistical value-weighted and is calculated as follows:

$$w = \frac{1}{1 + N\alpha Y_{\text{pre}}}$$

where N is the number of observation periods. α is the over-dispersion parameter associated with the accident prediction model. The w value varies between 0 and 1. This value controls the relevance between the predictions of the model and the number of accidents. If the data used in the accident prediction model show little dispersion, w will be higher, since in this situation the proposed model will be more reliable. The final step in the definition of critical sections consists of calculating the excess in the frequency of expected accidents, which corresponds to the difference between the estimates predicted by the model and the estimation with the empirical Bayes adjustment and it is formulated as follows:

$$\Delta = E\left(A\right) - Y_{\text{pre}}.$$

The value (Δ) identifies the sections which have the highest frequency of expected accidents in contrast to the frequency of accidents predicted by the model. These sections have priority because they are considered to respond better to the proposed mitigation measures, as they have a significant excess that can be reduced.

In the methodological context, the application of the Bayes method considers three important aspects that make it more precise than others: the availability of data, the regression bias to the mean, and the performance threshold [27,33]. Bayes' approach allows the accident-prone sections that produce the lowest proportion of false negatives and false positives to be identified. These considerations emphasize that it is a widely recommended method for estimating traffic accidents [52].

3.3. Setting

The research is applied in the city of Cartagena, Colombia. The city has more than 1 million inhabitants and more than 130 000 vehicles in circulation. Cartagena has an extension of 650 km^2 and has a road network of more than 730 km. The city is in the north of Colombia in the Caribbean region. The city has a tourist, industrial, and port vocation, but with high informal underemployment. In the last 8 years, Cartagena has been considered among the most dangerous cities in road safety for motorcyclists, after Medellín, Cali, Bogotá, and Barranquilla [44]. In the country, due to the increase in commuting, informal transportation, and mobility difficulties, the use of motorcycles has increased, as well as the phenomenon of motorcycle taxi drivers as an informal activity [39]. Cartagena and the cities of Barranquilla, Monteria, and Sincelejo are the capital cities where this informal transport activity is most practiced in Colombia [31, 39]. Mototaxism is a typical and uncontrolled activity in Cartagena, which turns out to be an illegal practice and without minimum safety standards [40].

In Cartagena, in the last 4 years (2016–2019), 174 fatalities have occurred, and more than 1750 crashes in motorcyclists with serious injuries per year. Cartagena presents an official registry of 68.000 motorcycles until March 2019 [43]. It is estimated that more than 75.000 motorcycles circulate, coming from nearby cities and municipalities. According to official statistics, 60%–75% of motorcycles are engaged in informal passenger transportation. In 2018 in Cartagena, motorcycles represents the second most used means of transportation [11]. The high level of exposure and accidents that exist in the city is a favorable and representative setting to analyze the crashes of motorcycles on the roads.

4. Method

The research methodology is based on 4 stages as recommended by Elvik [18] and Polders and Brijs [53]. First, identify the road sections that will be included in the analysis. Second, obtain historical information on motorcyclist traffic crashes. Third, predict the number of crashes on the road sections with the Safety Performance Function. The predictive model will be estimated with the negative binomial regression. Fourth,

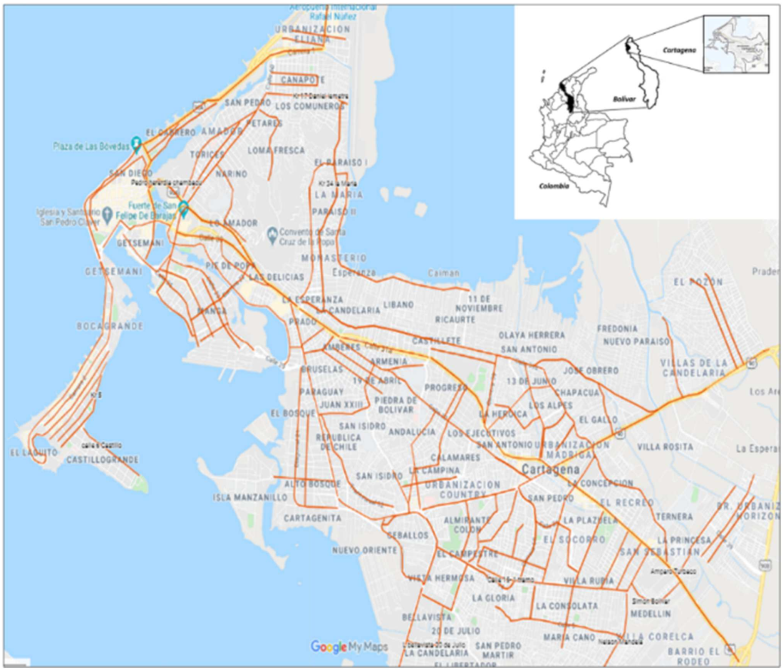


FIGURE 1. Location of Cartagena (Bolívar, Colombia), and the road sections with motorcycle crashes (Google maps).

estimate the expected number of crashes in road sections using Bayes' empirical approach, applying the number of observed and predicted accidents.

4.1. Collection and processing of information

The first two methodological stages that describe the environment information and the road accident data from Cartagena (Colombia) are presented below. The data of 2884 motorcycle crashes between 2016 and 2017 are analyzed. The database is provided by the Department of Traffic and Transportation (DATT) of Cartagena, which manages mobility and road safety in the city. The crash dataset includes information on the timing, type of collision, location, road users, and severity. The road crashes were manually geo-located to determine the analysis sections within the study. The records have identified the accident data in all localities. The traffic accident dataset ranges from property damage to minor injuries, serious injuries, and fatalities. Table A.1 provides statistical information from the dataset.

In total, 121 road sections are identified within the dataset. Figure 1 shows the location of the city and the road sections with motorcycle crashes. The sample size corresponds to 242 road sections for two years (2016–2017).

4.2. Definition of variables

In the development of the prediction model (SPF), a set of infrastructures, operational, environmental, and traffic volume variables (12 quantitative variables and 5 categorical variables) are considered. The number of motorcycle crashes per kilometer be the dependent variable within the model which represent the frequency of collisions on the road sections in equivalent units. The quantitative variables are the number of curves, number of accesses (intersections), the number of traffic lights, annual average daily traffic (AADT) by road users (cars, buses, heavy vehicles, and motorcycles) and the type of area (commercial, residential, industrial) in percentage coverage.

The categorical variables are sense of the road, type of road (arterial, collector, local, and rural), lane configuration, use of separators (median), and pavement condition (quality) of the road. The data related to the length, geometry, and infrastructure of the road sections are collected directly on the road and validated with satellite geo-referencing with Google Maps. Traffic volume is provided by studies of the consortium mobility and traffic of the city between 2016 and 2017.

The type of road refers to the categorization of roads in Colombia. Arterial roads refer to the main roads with greater interconnection, such as avenues. Collector roads are sections of roads that connect with main roads. Local roads refer to inter-neighborhood sections. Finally, rural sections are defined as interurban or perimeter connection routes between the city and the townships. The definition of pavement conditions (quality) is assessed considering a functional valuation of serviceability, as defined by Fuentes *et al.* [21]. These categories are good, regular (poor) and bad (very poor).

Table 2 shows the analysis variables of the road sections. The road sections are diverse in road characteristics and not homogeneous. Road sections range from 1 to 65 motorcycle crashes. The length of the sections ranges from 0.4 to 17.2 km. The total length of the sections corresponds to 220.1 km. The average volume of daily motorcycle traffic is 47%. Automobiles correspond to 46% of the total traffic. The residential coverage area has the highest proportion with 57%, followed by commercial areas (35%). Most of the sections are characterized by collector type (40%), double lanes (68%), and double direction (sense) (70%).

4.3. Development of the prediction model

In the third stage of the study, the dataset be processed and analyzed with the statistical software SSPS ver. 25. The prediction model has used a negative binomial regression to define a safety performance function (SPF) for motorcyclist accidents. The SPSS software is configured for negative binomial with log link analysis. The estimation of the parameters is considered with the "hybrid" method and the scale estimation (Pearson Chi-square). The effects of the model are estimated with the "type I and III" analysis, while chi-square statistics are set as "likelihood ratio". The model is considered with the intercept or constant because it allows a better fit [3,9]. The significance of the variable is to be examined with the chi-square test with a 95% confidence interval.

Before estimating the model, a Pearson correlation analysis is proposed on the identified continuous variables. This analysis avoids errors in the estimation of the parameters when the independent variables have a strong correlation. Additionally, this strategy allows reducing the number of variables for an adequate estimation of the final parameters of the model [2]. The variables considered in the model are those with correlations (Pearson's value) with values less than 0.5 [3]. Table A.2 shows the correlation matrix of the qualitative variables.

The continuous variables to consider in the model are the number of accesses (intersections), number of curves, type of areas commercial and industrial areas, average annual daily traffic (AADT), and average annual daily traffic of motorcyclists (AADT-MC). The ADDT and ADDT-MC variables are considered in alternative models because including them simultaneously would affect the significance and quality of the predictions [61]. Table 3 summarizes the total set of variables that are considered in the models.

The dependent variable Y is declared as the number of motorcycle crashes, which is accompanied by an offset variable "length of the road section". This fitting in the software allows the dependent variable to be interpreted as the number of motorcycle crashes per kilometer. The application of the "offset" serves so that the parameters of the linear component can be interpreted in terms of expected rates and not of expected counts. In general terms, the model can be expressed as follows:

$$Y = \exp\left(\beta_0 \cdot X_1^{\beta_1} \cdot X_3^{\beta_3} \cdot X_4^{\beta_4} \cdot X_5^{\beta_5} \cdot X_6^{\beta_6} \cdot X_7^{\beta_7} \cdot X_8^{\beta_8} \cdot X_9^{\beta_9} \cdot X_{10}^{\beta_{10}} \cdot X_{11}^{\beta_{11}}\right).$$

With the transformation of some quantitative variables (variables from X_1 to X_6) in logarithmic scale (Ln), the model can be expressed like as follows:

$$Y = \mathrm{Exp}\left(\beta_0\right) \cdot X_1^{\beta_1} \cdot X_2^{\beta_2} \cdot X_3^{\beta_3} \cdot X_4^{\beta_4} \cdot X_5^{\beta_5} \cdot X_6^{\beta_6} \cdot \mathrm{Exp}\left(X_7\beta_7 \cdot X_8\beta_8 \cdot X_9\beta_9 \cdot X_{10}\beta_{10} \cdot X_{11}\beta_{11}\right).$$

TABLE 2. Statistical summary of quantitative and categorical variables.

Quantitative variables							
Variables	Notation	N	Minimum	Maximum	Mean	Standard deviation	Units
Motorcycle crashes	Cra	242	1	65	12	11.7	number
Length (kilometers)	Le	242	0.4	17.2	1.8	1.9	kilometers
Motorcycle crashes per kilometers	Cra/Le	242	0.2	41.5	8.1	7.2	number/km
Number of accesses (intersections)	NA	242	4	57	18.2	10.5	number
Number de curves	Cu	242	0	12	2	2	number
Number of traffic lights	Li	242	0	7	1.1	1.3	Number
AADT	AADT	242	1181	80416	17185	13697	vehicles/day
AADT of cars	ADDT-C	242	18	92	46	18	% vehicles/day
AADT of buses	ADDT-B	242	18	15	3	3	% vehicles/day
AADT of heavy vehicles	ADDT-HV	242	0	10	2	2	% vehicles/day
AADT of motorcycles	ADDT-MC	242	0	76	49	19	% vehicles/day
Type of area – Commercial	% Com	242	6	80	35.2	20.1	%
Type of area – Residential	% Res	242	1	100	57.1	24.6	%
Type of area – Industrial	% Ind	242	1	80	6	15	%
Type of area – Rural	% Ru	242	1	100	3.5	13.1	%

Categorical variables				
Variables	Notation	Description	Total	%
Type of road	TR	Arterial	56	23%
		Collector	96	40%
		Local	80	33%
		Rural	10	4%
		Total	242	100%
Lane configuration	LC	Four lanes	22	9%
		Two Lanes	164	68%
		Two-four lanes	32	13%
		Three-six lanes	24	10%
		Total	242	100%
Direction of the road	DR	Double	170	70%
		Mixed	20	8%
		Simple	52	21%
		Total	242	100%
Median	Me	Without median	188	78%
		With Median	54	22%
		Total	242	100%
Pavement conditions	PC	Good	76	31%
		Bad	82	34%
		Regular-medium	84	35%
		Total	242	100%

TABLE 3. Summary of variables considered in the predictive model.

Type of variable	Variable	Notation		Coding
Dependent	Motorcycle crashes	Cra		Y
Off-set	Length (Kilometers)	–		Ln (Le)
Independent-quantitative	Number of accesses per kilometers	Nit/Le		X_1: Ln (NA/Le)
	Number of curves per kilometers	Cu/Le		X_2: Ln (Cu/Le)
	Type of area – Commercial	Com		X_3: Ln (Com)
	Type of area – Industrial	Ind		X_4: Ln (Ind)
	AADT	AADT		X_5: Ln(AADT)
	AADT-MC	AADT-MC		$X_{6:}$: Ln(AADT-MC)
Independent-categorical	Type of road – Arterial	TR	1	X_7
	Type of road – Collector		2	
	Type of road – Local		3	
	Type of road – Rural		4	
	Lane configuration – Four	LC	1	X_8
	Lane configuration – Two		2	
	Lane configuration – Two-four		3	
	Lane configuration – Three-six		4	
	Direction of the road – Double	DR	1	X_9
	Direction of the road – Mixed		2	
	Direction of the road – Simple		3	
	Median-without separator	Me	1	X_{10}
	Median-with separator		2	
	Pavement conditions – Good	PC	1	X_{11}
	Pavement conditions – Bad		2	
	Pavement conditions – Regular		3	

The quantitative variables are included within the model as covariates and categorical variables as factors. These factors are raised by levels and are created as a "dummy". Their participation in the model is binary, and the number of variables per factor is the number of levels minus one. In this study, it is important to highlight that the proposed dependent variable (motorcycle accidents per kilometer) is put back in motorcycle accidents to be executed within the SPSS Software. However, the introduction of an offset variable (Length, km) in the adjustment process in the software allows the final model to be interpreted as the number of motorcycle accidents per kilometer. This method follows the models of Harnen *et al.* [30] and Abdul Manan *et al.* [3].

5. Results

In this section, the results of the models proposed with the software to predict the number of motorcycle crashes per kilometer in the road sections of Cartagena are presented. Table 4 shows the results of the proposed models. In the validation of the models, the goodness-of-fit measure is considered as Deviance, Pearson Chi-Square, Akaike's Information Criterion (AIC), and Bayesian Information Criteria (BIC). Four models are proposed here. Models 1 and 2 contain the variable ADDT, while models 3 and 4 have the variable ADDT-MC. The four models are statistically significant within the omnibus test (P-value < 0.05). Models 1 and 3 are not representative. Some variables are not statistically significant with P-value > 0.05. Models 2 and 4 are found by reducing the variables to obtain statistical significance (P-value < 0.05) of all estimated parameters. The goodness-of-fit statistics of these models have very good results with the data. Pearson Chi-square values divided by the degrees of freedom are within the permissible and acceptable range (around 1.1) for a negative binomial distribution.

Table 4. Estimated results of the models.

Parameters	Model 1			Model 2			Model 3			Model 4		
	Beta	Std Error	P-value	Beta	Std Error	P-value	Beta	Std Error	P-value	Beta	Std Error	P-value
Constant	−2.95	0.96	0.00	−3.45	0.65	0.00	−2.11	0.67	0.00	−2.26	0.43	0.00
X_1 = Ln(Nit/Le)	0.49	0.14	0.00	0.36	0.12	0.00	0.42	0.14	0.00	0.30	0.13	0.02
X_2 = Ln(Cu/Le)	−0.09	0.09	0.28				−0.07	0.09	0.39			
X_3 = Ln(Com)	0.20	0.08	0.01	0.18	0.08	0.02	0.20	0.08	0.01	0.21	0.08	0.00
X_4 = Ln(Ind)	0.05	0.05	0.32				0.04	0.05	0.37			
X_5 = Ln(AADT)	0.25	0.10	0.01	0.31	0.08	0.00						
X_6 = Ln(AADT-MC)	1.25	0.32	0.00				0.21	0.07	0.00	0.20	0.05	0.00
X_7: TR-1	1.24	0.33	0.00	1.31	0.29	0.00	1.26	0.32	0.00	1.43	0.30	0.00
X_7: TR-2	1.08	0.34	0.00	1.04	0.29	0.00	1.26	0.33	0.00	1.02	0.29	0.00
X_7: TR-3	0^a			0.93	0.30	0.00	1.10	0.34	0.00	0.86	0.30	0.00
X_7: TR-4	−0.10	0.22	0.66	0^a			0^a			0^a		
X_8: LC-1	−0.38	0.24	0.11				−0.10	0.21	0.64			
X_8: LC-2	−0.27	0.27	0.30				−0.36	0.24	0.13			
X_8: LC-3	0^a						−0.33	0.26	0.20			
X_8: LC-4	−0.14	0.14	0.31				0^a					
X_9: DR-1	0^a						−0.23	0.15	0.11			
X_9: DR-2	−0.06	0.19	0.77				0^a					
X_{10}: Me-1	0^a						−0.09	0.19	0.62			
X_{10}: Me-2	−0.05	0.16	0.74				0^a					
X_{11}: PC-1	−0.15	0.12	0.20				−0.02	0.16	0.92			
X_{11}: PC-2	0^a						−0.18	0.11	0.12			
X_{11}: PC-3	0^a						0^a					
(Binomial negative)	0.31	0.04		0.33	0.04		0.31	0.04		0.33	0.04	
Omnibus test	df = 15 Sig = 0.000			df = 6 Sig = 0.000			df = 15 Sig = 0.000			df = 6 Sig = 0.000		
Likelihood ratio Chi-square	134.481			132.316			136.844			130.788		
Deviance	259.146, df = 225, value/df = 1.152			255.815, df = 234, value/df = 1.093			258.669, df = 225, value/df = 1.150			257.145, df = 234, value/df = 1.099		
Pearson Chi-square	250.073, df = 225, value/df = 1.111			247.385, df = 234, value/df = 1.057			249.495, df = 225, value/df = 1.109			246.699, df = 234, value/df = 1.054		
AIC	1562.967			1554.549			1570.569			1556.548		
BIC	1622.279			1582.461			1629.881			1584.459		

Notes. [(a)]Set to zero because this parameter is redundant.

Models 2 and 4 have the representation of three continuous variables and one categorical variable. Models 2 and 4 can be presented as follows:

Model 2:

$$Y = \exp(-3.45) \cdot \frac{\text{Nit}}{\text{Le}}^{0.36} \cdot \text{COM}^{0.18} \cdot \text{AADT}^{0.31} \times \exp(\text{TR1}(1.31) \cdot \text{TR2}(1.04) \cdot \text{TR3}(0.93)).$$

Model 4:

$$Y = \exp(-2.26) \cdot \frac{\text{Nit}}{\text{Le}}^{0.30} \cdot \text{COM}^{0.21} \cdot \text{AADT}^{0.20} \times \exp(\text{TR1}(1.43) \cdot \text{TR2}(1.02) \cdot \text{TR3}(0.86)).$$

Models 2 and 4 have in common the variable number of accesses (intersections) per kilometer, commercial area, and type of road. These two models include average daily traffic (AADT) and average daily traffic of motorcycles (AADT-MC) respectively. According to the models, the number of motorcyclist traffic accidents increases as the volume of traffic increases due the number of accesses per kilometer, commercial coverage, and changes in the type of road.

These two models represent the performance function in road safety to estimate annually the number of motorcycle crashes per kilometer, with the exploratory variables. The AADT and AADT-MC variables represent the measure of exposure, while the rest of the variables represent the risk factors in each of the models. For the deviation criteria, Pearson's Chi-square, Likelihood ratio Chi-square, Akaike's Information Criterion (AIC), and Bayesian Information Criterion (BIC), models 2 and 4 show a better statistical adjustment. In terms of goodness-of-fit, model 2 is statistically better than the model 4 in the Akaike's Information Criterion (AIC) and Bayesian Information Criteria (BIC). From this perspective, considering the two models to predict the frequency of accidents per kilometer can be adequately accepted.

Theoretically, the AIC and BIC values are affected by the number of explanatory variables in the model, so these metrics tend to be lower. However, the values between models 2 and 4 are very close. The AIC metric is typically used to compare various models, without necessarily being a formal inference [28]. In this condition, we have carried out an analysis of Cumulative Residuals (CURE) where the models share some continuous variables such as the number of accesses. A CURE plot that fluctuates closer to zero shows a model with a better fit between the ranges of the variables [19]. Figure 2 shows the CURE plot, which compares models 2 and 4. The graph shows that model 4 is closer to zero. At 65% of the points evaluated, model 4 oscillates closer to zero. Between 4 and 14 accesses per kilometer, model 4 shows the best fit in this range of the analyzed variables.

In the fourth stage of the study, the sections prone to traffic accidents in motorcyclists are identified. The fourth model is used as a reference because it has less residual variability in prediction with the number of accesses. The over-dispersion (α) of the model is 0.33. The weight (w_i) and the analysis between the predicted and observed values for the road sections are evidenced in Table A.3 for two years. Finally, 45 critical sections are obtained, corresponding to 45% of the sections (Tab. A.4). Figure 3 shows the results of predicting the annual frequency of traffic accidents per kilometer (density), and the sections prone to traffic accidents in motorcyclists are analyzed in the city with Google Maps.

6. Discussion

In the present study, a safety performance function of motorcyclists is developed with a negative binomial regression. The disaggregated analysis of exploratory variables of the environment, infrastructure, and operations has allowed identifying factors associated with the number of accidents per kilometer of motorcyclists in Cartagena. Risk factors include the number of accesses, traffic volume, land use, type of road, and pavement conditions. This study coincides with the findings of previous research in cities in emerging countries (LCIMs), where mobility solutions are limited, and informal transport practices are common [3, 28, 30, 55]. Also, this research has allowed the definition of road sections prone to accidents. These sections can be prioritized for the development of effective counter measures to the benefit of road safety of motorcyclists.

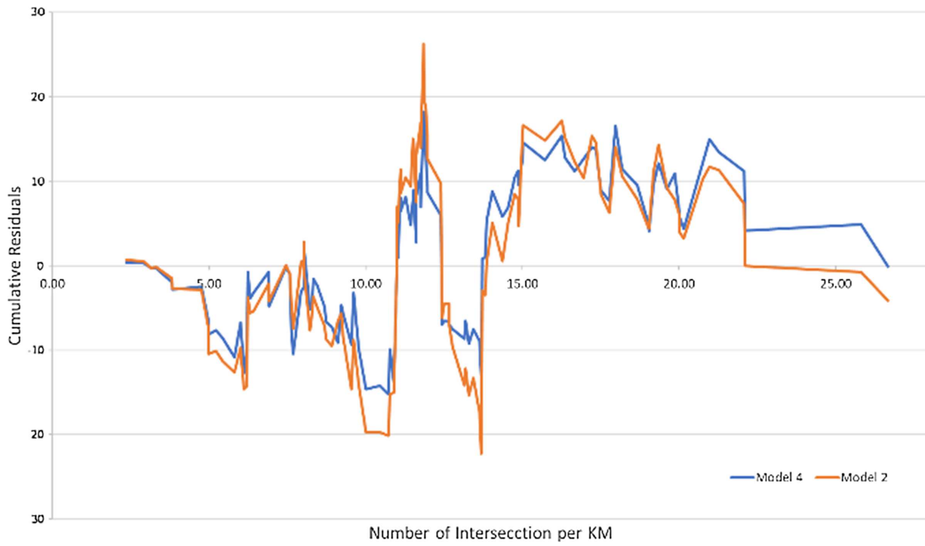
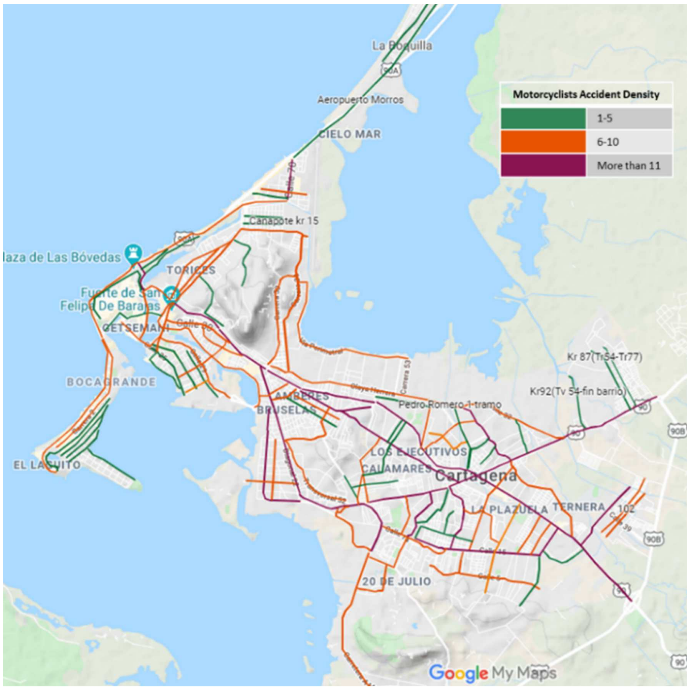


FIGURE 2. CURE plot for models 2 and 4.

In models 2 and 4, with better statistical significance, it is possible to identify that vehicle volume (AADT, ADDT-MC) is a key factor in the number of road accidents, as indicated by Elvik [18]. These factors are related to the level of exposure to motorcyclists. Considering the volume of motorcycles in traffic has a better fit in the prediction. This improves sensitivity and precision due to the increase or adjustment in the composition of the vehicle flow. These findings are also emphasized by Abdul Manan *et al.* [3], Lyon *et al.* [38] and Elvik *et al.* [19]. In this way, focusing the analysis on model 4 could be the most logical, as evidenced by the analysis of variance and error. Furthermore, the accident rate for motorcyclists is found to be slightly elevated when interacting with cars (58%). This condition is ratified by the most representative of daily traffic volume which are motorcyclists (47%) and automobiles (46%).

The number of (accesses) intersections is another of the important quantitative variables. The preliminary descriptive analysis has identified that the areas with the highest accidents correspond to intersections (77%), where the not signalized intersections correspond to 80%. Cartagena has territorial planning problems, which have not enabled an adequate demographic distribution. The city has most streets in a transversal and diagonal direction, increasing the number of accesses by road sections. The Literature has shown that intersections are more likely to cause serious accidents for motorcyclists [15, 16, 24]. Intersections are important due to being a high interaction that connects main roads with traffic at different speeds [23]. Among the most common accidents for motorcyclists is a violation of the right of way at intersections [16, 51]. Pai [51] in a literature review of the conditions in the crashes has evidenced problems in the visibility, and the crossing judgment (gap/time, distance/speed) of the motorcyclist. Speeding at urban intersections can influence error behaviors, such as failure to observe [15].

This study has identified that the commercial intensity of the road sections is a risk factor. The increase in the commercial intensity of a sector influences the density of vehicular traffic, as well as the presence of pedestrians. Within the identified critical sections, the commercial intensity of the sections is greater than 35%. It is common to see in the city environment, the deficiency of side-walks for pedestrian circulation. Pedestrian side-walks in commercial areas, and on arterial and collector roads very often become parking areas. This condition generates moving pedestrians on the road. Commercial areas on local roads are obstructed by the parking of vehicles on the road. This condition encourages motorcyclists to ride on the side-walks, or in the opposite direction. The



(a)

(b)

FIGURE 3. (a) Density and (b) sections prone to a road crash on motorcyclists in Cartagena.

indicated scenarios generate a high risk for the safety of all road users. These findings are consistent with the evidence by Harnen *et al.* [29] in Malaysia.

The type of road as a significant factor has a high relationship with the concentration of traffic. For example, arterial and collector roads are the largest connectivity and congestion. These allow greater displacement in population settings. Furthermore, local roads have showed high participation (55%) in crash-prone sections (see Tab. A.4). This condition is related to the lack of signaling and lighting, as well as the speeding of motorcyclists on inter-neighborhood roads. For example, motorcycle taxi drivers in the city take alternative roads to speed up their travel, evade authority, or mobilize areas with little connectivity. Additionally, most of the sections identified relate to areas with high socio-economic vulnerability. These areas are the most difficult in mobility due to the lack of efficient transport systems. These towns are the most prone to motorcycle taxi drivers. These socio-economic findings are consistent with the results found by Gutierrez and Mohan [22] on informal transportation by motorcyclists.

The estimated models for the prediction of accidents in motorcyclists correspond to the first analyzes in Colombia. Studies on the road safety of motorcyclists cannot be generalized, because these depend on the context and local conditions [6]. Under these aspects, this study is novel in the Latin American context. These types of models can be replicated in cities with similar conditions, developing adjustments or calibrations for future studies as recommended by Polders and Brijs [53] and Aashto [1]. As a recommendation and future work, it is planned to integrate GIS tools for spatial analysis. This study required manual and time-consuming work on the geo-location of crashes to increase the reliability of the analyzes.

Among the limitations, the road accident information recorded by the control entities may likely suffer from under-reporting. This phenomenon has been reported in other studies in motorcycling in Colombia [32, 48]. The results of this study allow prioritizing the accident-prone sections for the development of new studies. For example, identifying by direct observation, the behavioral risk factors are associated with motorcycle accidents.

6.1. Managerial insights

Road safety problems require a detailed analysis of time and causality. The development of the safety performance function is a demanding activity that seeks to locate all accidents to identify characteristics of the infrastructure, mobility, and interactions of road actors. Based on the results of this study, some measures focused on improving road safety for motorcyclists and road actors involved in the urban environment are recommended as follows:

- Prioritize solutions in local road sections, by improving signaling (stops and speed signs) and lighting at intersections related to arterial and collector roads.
- Improve pedestrian zone signage at all critical intersections.
- Implement traffic light studies on critical sections with high traffic density and accident frequency.
- Regulate and decrease speed on critical local roads with reducers and traffic agent control.
- Intensify traffic and mobility control in commercial environments.
- Develop road education campaigns focused on vulnerable age groups (*i.e.*, youth between 18 and 20 corresponds to 62% of victims).
- Carry out road tolerance campaigns between motorcyclists and automobiles. These road users have the highest relation in accidents (58%).
- In large road segments (arterial or collector) allow the possibility of exclusive lanes for motorcyclists.

The results of the proposed model help to improve specifically the urban planning, resource projection for road signs, traffic volume management, and operational support for mobility. All these benefits indicated in the literature are aspects in which it is expected to contribute to the research in the city of Cartagena, due to its complex problems with mobility, transport, and infrastructure. These findings are consistent with the results of the factors associated with road accidents in the city of Cartagena developed by Cantillo *et al.* [10].

To reproduce this model, adequate accident information, with all environmental characteristics, must be collected. Additionally, updated information is required for the vehicle capacity records on the different roads of the city. Nowadays, intelligent counting technology is recommended in traffic lights and specialist cameras. Otherwise, investments are needed to monitor vehicle capacity in significant time slots.

7. Conclusion

The methodological combination of the performance function and the empirical approach of Bayes has identified environmental risk factors associated with the accident of motorcyclists in Cartagena. The proposed model show that motorcycle accidents per kilometer have significant factors such as the daily volume of motorcyclists, the number of accesses per kilometer, commercial area, and the type of road. Bayes' analysis is the statistical identification of 55 high-risk sections, where local authorities can focus and prioritize solutions. The applied methodology shows the importance of analyzing road accidents to consider risk factors associated with the environment. The research evidenced coherent and consistent results with previous studies and demand effective countermeasures for the benefit of road safety for motorcyclists.

This study is the first model for the prediction of accidents in motorcyclists in Colombia with the declaration of a performance function for road safety. Additionally, this research is applied in Cartagena, a city with a high level of exposure, due to the number of motorcyclists who circulate, where the majority is dedicated to motorcycle taxis. The results of the proposed model give significance to the influence of the commercial environment of the road as well as the type of road. Reference models only highlight the importance of the volume of traffic and the number of intersections of the road section. Likewise, the results rule out that the pavement conditions have a significant influence on the accident rate. It is because of these findings that the results are found to contribute to urban planning as well as the definition of effective policies for mobility and transportation. The Cities with characteristics like Cartagena in infrastructure and social economy, such as Monteria, Sincelejo, Barranquilla, can be favorable scenarios to replicate and compare the results of this study.

Appendix A.

Table A.1. Descriptive analysis of motorcyclist accidents between 2016 and 2017 in Cartagena.

Trimester	2016	2017	Total	%
1	400	317	717	25%
2	521	279	800	28%
3	498	390	888	31%
4	250	229	479	17%
Day				
Monday	262	157	419	15%
Tuesday	244	195	439	15%
Wednesday	258	179	437	15%
Thursday	232	187	419	15%
Friday	176	129	305	11%
Saturday	275	198	473	16%
Sunday	222	170	392	14%
Motorcyclist Age				
18–20	1062	731	1793	62%
21–40	509	332	841	29%
50>	75	44	119	4%
No Register	23	108	131	5%

TABLE A.1. continued.

Trimester	2016	2017	Total	%
Non-motorcyclist age				
18–20	435	342	777	38%
21–40	516	341	857	42%
50>	160	114	274	13%
No Register	55	95	150	7%
Type of day				
Normal	1582	87	1669	58%
Festive	1169	46	1215	42%
Type of location				
Intersection	1276	939	2215	77%
Roundabout	41	24	65	2%
Continuous section	352	252	604	21%
Type of area				
Commercial	752	539	1291	45%
Rural	18	26	44	2%
Industrial	202	154	356	12%
Residential	439	350	789	27%
Commercial-residential	258	146	404	14%
Severity				
Minor injuries	348	242	590	20%
Severe injuries	1276	935	2211	77%
Fatalities	45	38	83	3%
Class of collision				
Rolled	106	71	177	6%
Vehicles fall	10	12	22	1%
Crash-shock	1547	1130	2677	93%
Dump	6	2	8	0%
Collision interaction				
Solo motorcycle	293	193	486	17%
Between-motorcycles	210	130	340	12%
Motorcycle-vehicle	979	708	1687	58%
Motorcycle-heavy vehicle	60	61	121	4%
Motorcycle-bus	111	101	212	7%
Motorcycle (others; IE Bikes)	16	22	38	1%
Type of road				
Arterial	763	505	1268	44%
Collector	603	463	1066	37%
Local	261	204	465	16%
Rural	42	43	85	3%

TABLE A.2. Correlation analysis between independent variables.

	Cu	Nit	% Res	% Com	% Ind	% Ru	AADT	AADT-A	AADT-MC	AADT- B	AADT-HV
Cu	1.00										
Nit	0.50	1.00									
% Res	−0.12	−0.03	1.00								
% Com	−0.12	−0.11	−0.61	1.00							
% Ind	0.26	0.09	−0.49	−0.16	1.00						
% Ru	0.11	0.13	−0.37	−0.20	0.00	1.00					
AADT	0.01	0.13	−0.46	0.46	0.19	−0.07	1.00				
AADT-C	0.05	0.04	−0.50	0.43	0.28	−0.05	0.94	1.00			
AADT-MC	−0.05	0.19	−0.30	0.43	−0.01	−0.10	0.91	0.73	1.00		
AADT-B	0.04	0.19	−0.46	0.44	0.21	−0.05	0.84	0.80	0.72	1.00	
AADT-HV	0.16	0.11	−0.52	0.05	0.77	0.01	0.51	0.56	0.30	0.43	1.00

TABLE A.3. Results of the proposed model 4 in Predicted vs. Observed values.

Road section	Condition	Observed (cobs)	Predicted (Y)	W	$E(A)$	Δ
1	No Critical	5	7	0.2	5.4	−1.6
2	No Critical	8	9	0.1	8.5	−0.5
3	No Critical	0	1	0.6	0.9	−0.1
4	No Critical	3	6	0.2	3.7	−2.3
5	No Critical	12	16	0.1	12.7	−3.3
6	Critical	22	19	0.1	21.3	2.3
7	No Critical	16	17	0.1	15.8	−1.2
8	Critical	27	19	0.1	25.9	6.9
9	No Critical	12	16	0.1	12.6	−3.4
10	No Critical	6	8	0.2	6.5	−1.5
11	No Critical	5	7	0.2	5.8	−1.2
12	No Critical	5	11	0.1	5.3	−5.7
13	No Critical	5	6	0.2	4.8	−1.2
14	Critical	9	8	0.2	8.9	0.9
15	No Critical	10	11	0.1	9.8	−1.2
16	Critical	5	5	0.2	5.3	0.3
17	No Critical	2	3	0.3	2.2	−0.8
18	No Critical	3	6	0.2	3.6	−2.4
19	No Critical	2	5	0.2	2.6	−2.4
20	No Critical	4	5	0.2	4.2	−0.8
21	Critical	7	6	0.2	6.9	0.9
22	Critical	5	3	0.3	4.5	1.5
23	Critical	10	6	0.2	9.4	3.4
24	No Critical	5	5	0.2	4.7	−0.3
25	No Critical	5	5	0.2	5.0	0.0
26	No Critical	9	11	0.1	9.4	−1.6
27	No Critical	8	10	0.1	8.1	−1.9
28	Critical	11	6	0.2	9.6	3.6
29	No Critical	9	9	0.1	9.0	0.0

TABLE A.3. continued.

Road section	Condition	Observed (cobs)	Predicted (Y)	W	$E(A)$	Δ
30	No Critical	7	11	0.1	7.1	−3.9
31	No Critical	4	7	0.2	4.2	−2.8
32	No Critical	1	3	0.3	1.3	−1.7
33	No Critical	1	5	0.2	1.8	−3.2
34	Critical	15	11	0.1	14.1	3.1
35	No Critical	3	7	0.2	4.1	−2.9
36	Critical	6	6	0.2	6.1	0.1
37	Critical	10	9	0.1	9.6	0.6
38	No Critical	9	10	0.1	8.8	−1.2
39	No Critical	2	5	0.2	2.7	−2.3
40	Critical	20	14	0.1	19.8	5.8
41	Critical	23	15	0.1	21.9	6.9
42	No Critical	2	5	0.2	3.1	−1.9
43	No Critical	6	7	0.2	6.0	−1.0
44	No Critical	5	8	0.2	5.4	−2.6
45	Critical	11	9	0.1	10.6	1.6
46	Critical	17	8	0.2	15.4	7.4
47	Critical	8	6	0.2	8.0	2.0
48	Critical	28	14	0.1	26.1	12.1
49	Critical	6	6	0.2	6.4	0.4
50	No Critical	7	7	0.2	6.8	−0.2
51	No Critical	2	2	0.4	1.7	−0.3
52	No Critical	4	7	0.2	4.5	−2.5
53	No Critical	2	8	0.2	3.3	−4.7
54	No Critical	6	8	0.2	6.2	−1.8
55	No Critical	5	6	0.2	5.5	−0.5
56	No Critical	5	6	0.2	5.3	−0.7
57	No Critical	5	7	0.2	5.5	−1.5
58	No Critical	4	5	0.2	4.6	−0.4
59	No Critical	3	6	0.2	3.3	−2.7
60	Critical	13	9	0.1	12.5	3.5
61	Critical	12	9	0.1	11.8	2.8
62	Critical	6	3	0.3	5.2	2.2
63	Critical	3	3	0.3	3.4	0.4
64	No Critical	3	8	0.2	3.5	−4.5
65	No Critical	1	6	0.2	2.2	−3.8
66	Critical	6	4	0.3	5.4	1.4
67	Critical	7	3	0.3	6.0	3.0
68	Critical	9	8	0.2	8.7	0.7
69	No Critical	1	7	0.2	2.1	−4.9
70	No Critical	2	3	0.3	2.4	−0.6
71	Critical	3	2	0.4	2.4	0.4
72	Critical	4	4	0.3	4.3	0.3
73	No Critical	2	6	0.2	2.9	−3.1
74	No Critical	2	4	0.3	2.8	−1.2
75	Critical	18	7	0.2	15.7	8.7
76	Critical	18	8	0.2	16.1	8.1

Table A.3. continued.

Road section	Condition	Observed (cobs)	Predicted (Y)	W	$E(A)$	Δ
77	No Critical	1	5	0.2	2.0	−3.0
78	Critical	12	8	0.2	11.6	3.6
79	No Critical	5	7	0.2	5.8	−1.2
80	Critical	5	5	0.2	5.3	0.3
81	No Critical	3	9	0.1	4.2	−4.8
82	No Critical	3	4	0.3	3.7	−0.3
83	Critical	8	5	0.2	6.9	1.9
84	No Critical	2	8	0.2	3.0	−5.0
85	Critical	10	2	0.4	7.0	5.0
86	Critical	7	6	0.2	6.7	0.7
87	Critical	7	5	0.2	6.1	1.1
88	Critical	6	6	0.2	6.2	0.2
89	Critical	7	6	0.2	7.0	1.0
90	Critical	21	8	0.2	18.6	10.6
91	Critical	6	3	0.3	5.1	2.1
92	Critical	14	8	0.2	13.1	5.1
93	Critical	13	11	0.1	12.6	1.6
94	No Critical	2	5	0.2	2.3	−2.7
95	No Critical	10	10	0.1	9.9	−0.1
96	No Critical	2	5	0.2	2.8	−2.2
97	Critical	14	9	0.1	13.5	4.5
98	Critical	17	10	0.1	16.1	6.1
99	Critical	15	10	0.1	14.8	4.8
100	No Critical	4	6	0.2	4.2	−1.8
101	No Critical	1	1	0.6	0.8	−0.2
102	No Critical	5	17	0.1	5.8	−11.2
103	Critical	36	14	0.1	33.5	19.5
104	Critical	30	19	0.1	29.4	10.4
105	Critical	17	17	0.1	17.4	0.4
106	No Critical	7	7	0.2	6.8	−0.2
107	No Critical	13	14	0.1	12.8	−1.2
108	Critical	16	10	0.1	14.9	4.9
109	Critical	11	9	0.1	11.0	2.0
110	Critical	27	16	0.1	26.0	10.0
111	No Critical	1	4	0.3	1.8	−2.2
112	No Critical	5	8	0.2	5.8	−2.2
113	Critical	11	8	0.2	10.5	2.5
114	No Critical	3	8	0.2	3.8	−4.2
115	Critical	6	4	0.3	5.6	1.6
116	No Critical	2	6	0.2	2.6	−3.4
117	Critical	5	4	0.3	4.9	0.9
118	Critical	7	4	0.3	6.5	2.5
119	No Critical	9	9	0.1	8.8	−0.2
120	Critical	3	1	0.6	2.1	1.1
121	No Critical	0	1	0.6	0.8	−0.2

TABLE A.4. Road section prone to accidents by type of road.

Type of road	Total	Critical road	%
Arterial	28	12	43%
Collector	48	20	42%
Local	40	22	55%
Rural	5	1	20%
Total	121	55	45%

Acknowledgements. The authors would like to express their thanks to the editors and the reviewers for their insightful comments for enhance the clarity of the earlier version of the present article. The first author is grateful to the Fundación Centro de Estudios Interdisciplinarios Básicos y Aplicados (CEIBA)–Gobernacin de Bolívar (Colombia) for providing funding to carry out this research works.

References

[1] AASHTO, The Highway Safety Manual. American Association of State Highway Transportation Professionals, Washington, DC, USA **529** (2010).

[2] M.A. Abdel-Aty and A.E. Radwan, Modeling traffic accident occurrence and involvement. *Accident Anal. Prev.* **32** (2000) 633–642.

[3] M.M. Abdul Manan, T. Jonsson and A. Várhelyi, Development of a safety performance function for motorcycle accident fatalities on Malaysian primary roads. *Safety Sci.* **60** (2013) 13–20.

[4] W. Ackaah and M. Salifu, Crash prediction model for two-lane rural highways in the Ashanti region of Ghana. *IATSS Res.* **35** (2011) 34–40.

[5] A.P. Afghari, M.M. Haque, S. Washington and T. Smyth, Effects of globally obtained informative priors on bayesian safety performance functions developed for Australian crash data. *Accident Anal. Prev.* **129** (2019) 55–65.

[6] M. Araujo, E. Illanes, E. Chapman and E. Rodrigues, Effectiveness of interventions to prevent motorcycle injuries: systematic review of the literature. *Int. J. Injury Control Safety Promotion* **24** (2017) 406–422.

[7] K. Bauer and D.W. Harwood, Statistical Models of At-Grade Intersection Accidents. Federal Highway Administration, Addendum, United States (2000).

[8] C. Caliendo, M. Guida and A. Parisi, A crash-prediction model for multilane roads. *Accident Anal. Prev.* **39** (2007) 657–670.

[9] V. Cantillo, P. Garcés and L. Márquez, Factors influencing the occurrence of traffic accidents in urban roads: a combined GIS-Empirical Bayesian approach. *Dyna* **83** (2016) 21–28.

[10] V. Cantillo, L. Márquez and C.J. Díaz, An exploratory analysis of factors associated with traffic crashes severity in Cartagena, Colombia. *Accident Anal. Prev.* **146** (2020) 105749.

[11] Cartagena Comovamos, Encuesta de Percepción Ciudadana Cartagena. [Online]. Available: http://www.cartagenacomovamos.org/nuevo/wp-content/uploads/2019/02/Resultados-Encuesta-de-percepci%C3%B3n-Ciudadana-2018.pdf (2018).

[12] A. Chamber of the Automotive Industry of the National Association of Entrepreneurs of Colombia, Motorcycles in Colombia: allies of the country's development. Available: http://www.andi.com.co/Uploads/LasMotocicletasEnColombia.pdf (2017).

[13] C. Chen, H. Wang, J. Roll, K. Nordback and Y. Wang, Using bicycle app data to develop Safety Performance Functions (SPFs) for bicyclists at intersections: a generic framework. *Transp. Res. Part A: Policy Pract.* **132** (2020) 1034–1052.

[14] W. Cheng and S. Washington, New criteria for evaluating methods of identifying hot spots. *Transp. Res. Record* **2083** (2008) 76–85.

[15] N. Clabaux, T. Brenac, C. Perrin, J. Magnin, B. Canu and P. Van Elslande, Motorcyclists' speed and "looked-but-failed-to-see" accidents (in English). *Accident Anal. Prev.* **49** (2012) 73–77.

[16] M. de Lapparent, Empirical Bayesian analysis of accident severity for motorcyclists in large French urban areas. *Accident Anal. Prev.* **38** (2006) 260–268.

[17] K. El-Basyouny and T. Sayed, Safety performance functions using traffic conflicts. *Safety Sci.* **51** (2013) 160–164.

[18] R. Elvik, State-of-the-art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks. Transportøkonomisk institutt, Oslo (2007).

[19] R. Elvik, A. Høye, T. Vaa and M. Sørensen, Driver Training and Regulation of Professional Drivers. The Handbook of Road Safety Measures. Emerald Group Publishing Limited (2009).

[20] J.D. Febres, F. Mohamadi, M. Mariscal, S. Herrera and S. García-Herrero, The role of journey purpose in road traffic injuries: a Bayesian network approach. *J. Adv. Transp.* **2019** (2019).

[21] L. Fuentes, R. Camargo, J. Arellana, C. Velosa and G. Martinez, Modelling pavement serviceability of urban roads using deterministic and probabilistic approaches. *Int. J. Pavement Eng.* **22** (2021) 77–86.

[22] M.I. Gutierrez and D. Mohan, Safety of motorized two-wheeler riders in the formal and informal transport sector. *Int. J. Injury Control Safety Promotion* **27** (2020) 51–60.

[23] M.M. Haque and H.C. Chin, Right-angle crash vulnerability of motorcycles at signalized intersections: mixed logit analysis. In: Transportation Research Record (2010) 82–90.

[24] M.M. Haque, H.C. Chin and H. Huang, Modeling fault among motorcyclists involved in crashes (in English). *Accident Anal. Prev.* **41** (2009) 327–335.

[25] E. Hauer, Observational before/after studies in road safety. Estimating the effect of highway and traffic engineering measures on road safety (1997).

[26] E. Hauer, D.W. Harwood, F.M. Council and M.S. Griffith, Estimating safety by the empirical Bayes method: a tutorial. *Transp. Res. Record* **1784** (2002) 126–131.

[27] E. Hauer, D.W. Harwood, F. Council and M. Griffith, The empirical Bayes method for estimating safety: a tutorial. *Transp. Res. Record* **1784** (2002) 126–131.

[28] S. Harnen, S. Wong, R.R. Umar and W.W. Hashim, Motorcycle crash prediction model for non-signalized intersections. *IATSS Res.* **27** (2003) 58–65.

[29] S. Harnen, S. Radin, S.V. Wong and W.H. Wan Ibrahim, Development of prediction models for motorcycle crashes at signalized intersections on urban roads in Malaysia. *J. Transp. Stat.* **7** (2004).

[30] S. Harnen, R. Umar, S. Wong and W.H. Wan Ibrahim, Motorcycle accident prediction model for junctions on urban roads in Malaysia. *Adv. Transp. Stud.* **8** (2006).

[31] I.d.E.U.-.IEU, El mototaxismo sigue aumentando en Colombia. Available: http://ieu.unal.edu.co/noticias-del-ieu/item/el-mototaxismo-sigue-aumentando-en-colombia (2019).

[32] A. Jimenez, J.P. Bocarejo, R. Zarama and J. Yerpez, A case study analysis to examine motorcycle crashes in Bogota, Colombia (in English). *J. Safety Res.* **52** (2015) 29–38.

[33] B. Lan, B. Persaud, C. Lyon and R. Bhim, Validation of a full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions. *Accident Anal. Prev.* **41** (2009) 574–580.

[34] J. Lee and F. Mannering, Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accident Anal. Prev.* **34** (2002) 149–161.

[35] M.-R. Lin and J.F. Kraus, A review of risk factors and patterns of motorcycle injuries. *Accident Anal. Prev.* **41** (2009) 710–722.

[36] D. Lord and B.N. Persaud, Estimating the safety performance of urban road transportation networks. *Accident Anal. Prev.* **36** (2004) 609–620.

[37] D. Lord, S.P. Washington and J.N. Ivan, Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Anal. Prev.* **37** (2005) 35–46.

[38] C. Lyon, B. Persaud and S. Himes, Investigating total annual average daily traffic as a surrogate for motorcycle volumes in estimating safety performance functions for motorcycle crashes. *Transp. Res. Record* **2637** (2017) 9–16.

[39] L. Márquez, R. Pico and V. Cantillo, Understanding captive user behavior in the competition between BRT and motorcycle taxis. *Transp. Policy* **61** (2018) 1–9.

[40] F.J. Maza Ávila, M.P. Fals Galezo, L.C. Espinosa Flórez, C.F. Safar Cano and D. Licona Dáger, Percepciones del riesgo asociado a la práctica del mototaxismo en Cartagena, Colombia. *Econ. Reg.* **13** (2019) 57–81.

[41] A. Montella, A comparative analysis of hotspot identification methods. *Accident Anal. Prev.* **42** (2010) 571–581.

[42] A. Montella, M. Aria, A. D'Ambrosio and F. Mauriello, Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accident Anal. Prev.* **49** (2012) 58–72.

[43] ONSV, National Road Safety Observatory. National Automotive Registry. Available: https://ansv.gov.co/observatorio/?op=Contenidos&sec=64 (2019).

[44] ONSV, National Road Safety Observatory. Statistical Bulletins, Deceased Victims, Injured by INMLCF in transit events in Colombia – National, Departmental and Municipal Data 2012–2020. Available: https://ansv.gov.co/observatorio/estad%C3%ADsticas (2019).

[45] H. Ospina-Mateus and L.A. Quintana Jiménez, Understanding the impact of physical fatigue and postural comfort experienced during motorcycling: a systematic review. *J. Transp. Health* **12** (2019) 290–318.

[46] H. Ospina-Mateus, L.A. Quintana Jiménez, F.J. López-Valdés, N. Morales-Londoño and K. Salas-Navarro, Using data-mining techniques for the prediction of the severity of road crashes in Cartagena, Colombia. In: Applied Computer Sciences in Engineering. Springer International Publishing. Cham (2019) 309–320.

[47] H. Ospina-Mateus, L.A. Quintana Jiménez, F.J. Lopez-Valdes and K. Salas-Navarro, Bibliometric analysis in motorcycle accident research: a global overview. *Scientometrics* **121** (2019) 793–815.

[48] H. Ospina-Mateus, L.A. Quintana Jiménez and F.J. Lopez-Valdes, Understanding motorcyclist-related accidents in Colombia. *Int. J. Injury Control Safety Promotion* **27** (2020) 215–231.

[49] H. Ospina-Mateus, L.A. Quintana Jiménez, F.J. Lopez-Valdes, S. Berrio Garcia, L.H. Barrero and S.S. Sana, Extraction of decision rules using genetic algorithms and simulated annealing for prediction of severity of traffic accidents by motorcyclists. *J. Ambient Intell. Humanized Comput.* (2021) 1–22.

[50] K. Ozbay and N. Noyan, Estimation of incident clearance times using Bayesian Networks approach. *Accident Anal. Prev.* **38** (2006) 542–555.

[51] C.W. Pai, Motorcycle right-of-way accidents – A literature review (in English). *Accident Anal. Prev. Rev.* **43** (2011) 971–982.

[52] B. Persaud and C. Lyon, Empirical Bayes before-after safety studies: lessons learned from two decades of experience and future directions. *Accident Anal. Prev.* **39** (2007) 546–555.

[53] E. Polders and T. Brijs, How to Analyse Accident Causation? A Handbook with Focus on Vulnerable Road Users. Hasselt University (2018).

[54] X. Qu and Q. Meng, A note on hotspot identification for urban expressways. *Safety Sci.* **66** (2014) 87–91.

[55] R.S. Radin Umar, M. Mackay and B. Hills, Multivariate analysis of motorcycle accidents and the effects of exclusive motorcycle lanes in Malaysia. *J. Crash Prev. Injury Control* **2** (2000) 11–17.

[56] B.A. Ramírez, F.A. Izquierdo, C.G. Fernández and A.G. Méndez, The influence of heavy goods vehicle traffic on accidents on different types of Spanish interurban roads. *Accident Anal. Prev.* **41** (2009) 15–24.

[57] V. Shankar, F. Mannering and W. Barfield, Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Anal. Prev.* **27** (1995) 371–389.

[58] T. Tang, S. Zhu, Y. Guo, X. Zhou and Y. Cao, Evaluating the safety risk of rural roadsides using a Bayesian network method. *Int. J. Environ. Res. Public Health* **16** (2019) 1166.

[59] R.A. Tegge, J.-H. Jo and Y. Ouyang, Development and application of safety performance functions for Illinois. ICT-10-066 UILU-ENG-2010-2006 (2010).

[60] A. Vogt and J. Bared, Accident models for two-lane rural segments and intersections. *Transp. Res. Record* **1635** (1998) 18–29.

[61] S. Washington, M.G. Karlaftis, F. Mannering and P. Anastasopoulos, Statistical and Econometric Methods for Transportation Data Analysis. CRC Press (2020).

[62] WHO, Powered two- and three-wheelers safety: a road safety manual for decision-makers and practitioners. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/272757/9789243511924-spa.pdf?sequence=1&isAllowed=y (2017).

[63] WHO, Global status report on road safety 2018. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/276462/9789241565684-eng.pdf?ua=1 (2018).

[64] C. Xin, Z. Wang, P.-S. Lin, C. Lee and R. Guo, Safety effects of horizontal curve design on motorcycle crash frequency on rural, two-lane, undivided highways in Florida. *Transp. Res. Record* **2637** (2017) 1–8.

[65] K. Yang, X. Wang and R. Yu, A Bayesian dynamic updating approach for urban expressway real-time crash risk evaluation. *Transp. Res. Part C Emerg. Technol.* **96** (2018) 192–207.