

Article

Accident Frequency Prediction Model for Flat Rural Roads in Serbia

Spasoje Mičić ¹, Radoje Vujadinović ², Goran Amidžić ³, Milanko Damjanović ² and Boško Matović ^{2,*}

¹ Faculty of Transportation, Pan-European University APEIRON, 78000 Banja Luka, Bosnia and Herzegovina; spasoje.n.micic@apeiron-edu.eu

² Faculty of Mechanical Engineering, University of Montenegro, 81000 Podgorica, Montenegro; radojev@ucg.ac.me (R.V.); milanko@ucg.ac.me (M.D.)

³ Faculty of Security Science, University of Banja Luka, 78000 Banja Luka, Bosnia and Herzegovina; goran.amidzic@fbn.unibl.org

* Correspondence: boskom@ucg.ac.me

Abstract: Traffic accidents, by their nature, are random events; therefore, it is difficult to estimate the exact places and times of their occurrences and the true nature of their impacts. Although they are hard to precisely predict, preventative actions can be taken and their numbers (in a certain period) can be approximately predicted. In this study, we investigated the relationship between accident frequency and factors that affect accident frequency; we used accident data for events that occurred on a flat rural state road in Serbia. The analysis was conducted using five statistical models, i.e., Poisson, negative binomial, random effect negative binomial, zero-inflated Poisson, and zero-inflated negative binomial models. The results indicated that the random effect negative binomial model outperformed the other models in terms of goodness-of-fit measures; it was chosen as the accident prediction model for flat rural roads. Four explanatory variables—annual average daily traffic, segment length, number of horizontal curves, and access road density—were found to significantly affect accident frequency. The results of this research can help road authorities make decisions about interventions and investments in road networks, designing new roads, and reconstructing existing roads.

Keywords: accident frequency; flat rural roads; prediction; road safety



Citation: Mičić, S.; Vujadinović, R.; Amidžić, G.; Damjanović, M.; Matović, B. Accident Frequency Prediction Model for Flat Rural Roads in Serbia. *Sustainability* **2022**, *14*, 7704. <https://doi.org/10.3390/su14137704>

Academic Editors:
Elżbieta Macioszek,
Margarida Coelho, Anna Granà and
Raffaele Mauro

Received: 20 May 2022

Accepted: 21 June 2022

Published: 24 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Road accidents are becoming global, social, public health, and economic issues; as such, broad measures are required to prevent them, including having a better understanding of the impacts of the roadway and environmental factors on traffic crashes. Great attention has been placed on the influences of urban and rural environments on road accidents. The risks of road deaths on rural roads (per kilometer traveled) are generally higher than on urban roads, and four to six times higher than on motorways [1]. Of the traffic fatalities that took place in the US in 2016—18,590 occurred in rural areas (or 50% of all traffic fatalities for that year) [2]. In 2015, 8% of fatalities occurred on European motorways, 55% on rural roads, and 37% on urban roads [3]. According to research carried out by the Road Traffic Safety Agency of the Republic of Serbia, in 2015, the majority of traffic fatalities occurred on rural state roads (48.7%), municipal roads (25.9%), and motorways (17.7%) [4].

Traffic accidents, by their nature, are random events [5]; therefore, it is difficult to estimate the exact places and times of their occurrences and the true nature of their impacts. Although they are hard to precisely predict, preventative actions can be taken and their numbers (in a certain period) can be approximately predicted. All key stakeholders in road safety systems should take appropriate steps and measures to increase road safety measures and reduce the number of accidents and casualties. One task for researchers, road safety specialists, and engineers is to maintain the highest level of safety on road networks [6]. Much effort has been devoted to improving road safety, through modeling

safety issues and examining the factors related to accident frequency and severity. The authors of [7–9] have conducted meta-analyses regarding accident frequencies and severity. They emphasized that there are methodological issues that influence the selection of appropriate models for predicting accidents. These statistical models have been considered for different spatial units of observation (e.g., sections, intersections, etc.) in the functions of their characteristics [10].

Predictive models have been developed (depending on the type of terrain through which a road passes). The Highway Safety Manual (HSM) identified three road terrain types (flat, rolling, and mountainous) that were used to develop safety performance functions (SPFs) [11]. Regarding accident frequency analysis, there is a research gap in modeling the accident frequencies of flat rural roads. Flat rural roads are mainly observed through the influence of exploratory terrain variables. Persaud et al. [12] developed a predictive model for horizontal curves based on the road environment. They found that the accident frequency is lower if the road is in a flat environment compared to a rolling environment. Choi et al. [13] found that including the terrain type (flat and mountainous) in their analysis of accidents helped with formulating better models. Moreover, they found that long horizontal curved elements in a flat terrain contributed to reducing accident costs. It was also found that there are no similar studies in the Western Balkans region. Flat rural roads in the Republic of Serbia are characterized by mild shoulders, long horizontal curved elements, long directions, the absence of vertical curves, and a large number of access roads (private driveways, industrial–commercial access roads, and agricultural roads).

1.1. Models in Accident Frequency Analysis

In prior research, accident frequencies and determining the relationships between accidents and the factors that contributed to the occurrences of accidents were examined using conventional multiple linear regression techniques. However, Joshua and Garber [14] have shown that conventional linear regression models are not suitable for modeling events such as traffic accidents, because they have non-negative values and (approximately) follow the Poisson distribution. The greatest advantage of the Poisson model is its simplicity in the calculation—the characteristics of the model involve the mean and variance being equal [7]. This also presents the limitations of the Poisson model (because this assumption is rarely verified in the actual data). In addition, it cannot cope with overdispersion and underdispersion of data [15], small sample sizes, or low sample mean values [16,17]. Numerous extensions of the Poisson model have been conducted in previous studies [18–21]. For example, Kronprasert et al. [19] developed an accident frequency predictive model based on the Poisson regression model—a useful tool when the observation numbers are very small, such as with the fatal accident frequency. To resolve the problem of overdispersion, a negative binomial (NB) model was introduced, which assumes that the Poisson parameter follows the gamma distribution [22] and relaxes the assumption of the equality of the mean and variance. The NB model has been widely used in accident modeling [23]. The NB model is shown to be good at resolving overdispersion [24], but in certain data sets, where there are temporal [25] and spatial [26] correlations, it has not been the best solution, and the use of a random effect negative binomial (RENB) model has been suggested [27,28]. Another problem that occurs in accident frequency modeling is excess zeros in the observations, due to the fact that traffic accidents are rare events. The Poisson and NB models tend to produce erroneous estimates for over-dispersed accident data when there are large numbers of zeros in a data set; thus, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models are more adequate [29]. However, the problem with the inflated models is in the assumption of a dual state. Lord et al. [30] argue that the safe state is not theoretically possible with a data set involving traffic accidents. In general, count data models allow for analyses when the dependent variables of interest are numerical counts. Count data represent types of data in which observations only have non-negative integer values. Variables of countable types have common properties: (a) their values are always integers, (b) the smallest possible value is 0 (and can never be negative), and (c) they are mostly positively flattened, with more data having small values (and small

numbers—higher values) [31]. Count data models can be used to estimate the effects of policy intervention or accident prediction.

1.2. Factors Affecting the Accident Frequency

Many factors that affect accident frequency have been analyzed. They could be grouped into five main groups: traffic flow characteristics, road characteristics, environment characteristics, vehicle characteristics, and human-related factors [32,33].

Traffic volume is the most commonly used independent variable for accident frequency modeling. Studies have shown that accidents increase with the increase in the annual average daily traffic (AADT) [19,32,34,35] or the average daily traffic (ADT) [36,37]. Conversely, Milton and Mannering [38] determined that an increase in traffic volume during peak hours reduces accident frequency (due to traffic congestion). The authors of [25,39] claimed that a larger percentage of heavy vehicles in traffic flow contributes to an increase in the accident frequency.

Many studies have found that accident frequency is affected by the segment length [21,24,29,32], lane width [29,40], lane numbers [27,32,38], carriageway width [21,29,41], and the shoulder width and type [29,42]. Significant attention has been dedicated to horizontal and vertical road geometric elements, which are found to affect the accident frequency [28,37,43,44]. Shankar et al. [45] revealed that an increase in the number of horizontal curves at a road section leads to an increase in accident frequency. Moreover, they determined that horizontal curves with higher design speeds have larger effects on the accident frequency than curves with lower design speeds, due to risk compensation by the drivers. The effects of horizontal curves have been investigated through various elements; it was found that a larger horizontal curvature [27], a smaller radius of a horizontal curve [29,46], a greater angle of a curvature [28,47], and a shorter horizontal curve [46], increase the accident frequency. Huo et al. [48] claimed that the number of accidents increases with the increase in the longitudinal grade. Similarly, Miaou [49] found that the truck accident involvement rate increases with the increase in the vertical grade. Montella and Imbriani [50] established that with the increase in the longitudinal grade, the accident frequency rises by 4% more at the downgrade than at the upgrade. The length of a vertical curve is negatively related to the accident frequency, and with the increase in length, the accident frequency declines [23]. The carriageway condition can have a significant impact on the accident frequency. The authors [21,51] determined that a greater friction coefficient tends to reduce the accident frequency. The rutting depth is positively related to accidents, and adverse safety effects are produced with the increase of the rutting depth [27]. The authors of [27,52] have shown that an increase in the value of the International Roughness Index (IRI) contributes to an increase in the number of accidents.

Great efforts have been made to determine the relationship between accidents and speed. Milton and Mannering [38] determined that a high posted speed limit is associated with a lower accident frequency. Hosseinpour et al. [39] suggested that this fact is the consequence of low-speed limits at unsafe locations and on roads with poor geometric characteristics. On the other hand, Donnell and Mason [53] found that on interstate highways, with the increase in the speed limit, there is an increase in the accident frequency. Imprialou et al. [54] have shown that higher speeds cause accidents with more severe consequences. The greater the difference between the posted speed limit and the average speed of the traffic flow—the so-called “speed gap”—causes an increase in the accident rate [55].

Access roads may also have a significant role in the occurrence of traffic accidents on rural roads. Research conducted by Vaiana et al. [56] has shown that the increase in driveway density increases the accident frequency. Similarly, Saccomanno et al. [18] claimed that the increase in the number of personal access points to sections increases the number of accidents.

1.3. Aims of the Research

Given the level of knowledge and the state of the practice concerning road safety on flat rural roads, this research attempted to fill in some of the knowledge gaps in this area. In particular, we focused on two-way rural roads. The main goal of this paper was to develop a predictive model for flat rural roads using accident data over three years. The predictive model was developed based on the comparisons of the performances of several selected statistical models. In addition, the influences of several independent variables on the accident frequency were analyzed—primarily traffic and geometric variables and then the variables that define the pavement conditions and the influence of access roads. Finally, an elasticity analysis was performed to assess the relative importance of the selected independent variables on the accident frequency.

2. Materials and Methods

2.1. Data

To carry out this research, data were collected for state road IB-12. The length of the road is 276.70 km, divided into 23 sections. This road is primarily a two-lane, two-way main road (approximately 90.85% of the total length, i.e., 251.38 km). The road passes through an urban area, at a length of 70.28 km (25.4%), while the remaining part of the road (length: 206.42 km (74.6%)) is located in a rural area. State road IB-12 is a flatland road with a longitudinal slope of $\pm 2.5\%$, connecting the eastern city of Subotica to the Romanian border (Srpska Crnja) in the west; its entire length is in the area of the Autonomous Province of Vojvodina. The starting and ending points are at nodes 1102 (X: 393,302.85 m E; Y: 5,102,918.84 m N) and 1222 (X: 476,708.20 m E; Y: 5,066,386.42 m N), respectively.

For research purposes, data were collected from 3 databases:

- Accident data, collected and maintained by the Ministry of Internal Affairs of the Republic of Serbia.
- Traffic data, collected and maintained by the Public Enterprise “Roads of Serbia”.
- Roadway geometrics, cross-sectional elements, and traffic signalization data; collected and maintained by the Public Enterprise “Roads of Serbia”.

Three years of data (from 2015 to 2017) were observed for state road IB-12. During this period, data on accidents, traffic flow, section lengths, posted speed limit, horizontal curves, access roads, and IRI were collected. Accident data contained a large number of details (place and time of an accident, type of accident, consequences, number of vehicles involved in the accident, number of people involved in the accident, carriageway conditions, etc.). In this study, only the total accident frequency was observed; in the observed period, a total of 1216 accidents occurred. Traffic volume data were collected from automatic counters, occasional automatic traffic counting, and interpolation of data, containing information about the total traffic flow, passenger vehicles, buses, and heavy vehicles. From the reference system of state roads, data on traffic nodes, section lengths, horizontal curves (number of horizontal curves, length of the arc, angle of curvature, etc.), posted speed limits (with GPS coordinates), and IRI were obtained. The data in the reference system are updated annually, except for traffic signalization data, which are updated on a monthly basis. Access road information was collected from the IB-12 road map available on Google Earth.

After data collection, the process of dynamic segmentation was carried out using ArcGIS 10.8. The four criteria for the dynamic segmentation were (1) traffic nodes, (2) traffic volume, (3) posted speed limit, and (4) number of traffic lanes. Initially, in the segmentation process, 136 homogeneous segments were obtained. Then, segments for which there were no complete data and segments that were located in the urban area were excluded. Segments that were shorter than 100 m were associated with neighboring segments with similar characteristics. During the segmentation process, accidents related to each segment were summarized. As a result of the dynamic segmentation, 59 homogeneous segments in rural sections were obtained.

Table 1 shows information about the minimums and maximums, the means, and the standard deviations of the selected variables from the database. The segment lengths ranged from 0.23 to 13.31 km, with posted speed limits ranging from 40 to 80 km/h. A

low AADT was identified along the road, with a mean of 3384 veh/day. The number of horizontal curves in the segments ranged from 0 to 7. The variable “Access roads density” is a relative variable that measures the number of roads per km of a homogeneous segment, ranging from 0 to 45.52. Although it is a rural road, a large number of access roads are evident, which can be justified by the fact that it is an agricultural area, and that in Vojvodina, the problems of linear settlements and uncontrolled access to the IB state roads are expressed. The pavement conditions are represented by the IRI, ranging from 0 to 4.53. To analyze the temporal effect and solve the problem of temporal correlations, the dummy variable “Year”, which takes values (0, 1), was introduced in accordance with previous research conducted by Caliendo et al. [51].

Table 1. Summary statistics of the considered variables.

Variables	Description	Min	Max	Mean	Std. Dev
Dependent variable					
RTA	Number of traffic accidents	0	20	2.16	3.00
Independent variables					
L	Segment length (km)	0.23	13.31	3.38	3.11
AADT	Annual average daily traffic (veh/day)	760	15195	3785.72	2391.52
SPEEDLIMIT	Posted speed limit (km/h)	40	80	74.07	10.94
NRCURVE	Number of horizontal curves	0	7	1.64	2.04
DAR	Access road density (no. roads/km)	0	42.52	7.46	8.11
IRI	International roughness index	0	4.53	2.47	1.07
Year 2015	Year of occurrence indicator (1—if 2015, 0—otherwise)				
Year 2016	Year of occurrence indicator (1—if 2016, 0—otherwise)				
Year 2017	Year of occurrence indicator (1—if 2017, 0—otherwise)				

After data collection and the dynamic segmentation process, a unique database was created, containing 177 independent observations ($n = 177 = 59 * 3$). The database only included segments for which there were accidents and traffic volume data. The base included 59 homogeneous segments, at a total length of 199.63 km. In the period 2015–2017, regarding these segments, 382 accidents occurred. After the database was created, the same was imported into statistical software STATA 13.0 where further processing and data analysis was performed.

2.2. Statistical Models

The Poisson regression model is based on the assumption that the number of accidents that occur on a road section i during an observed time period has a Poisson distribution with a mean value μ_i . The probability distribution function of the Poisson regression model is defined by equation [31]:

$$p(Y_i = y_i) = p(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}; \quad i = 1, 2, 3, \dots, n, \quad (1)$$

where $p(Y_i = y_i)$ is the probability of road section i having y_i accidents during the observed time period; Y_i —is a random variable that represents the number of accidents on a road section i during the observed time period; y_i —represents a real observation of Y_i on a road section i during the observed time period; μ_i —Poisson parameter for road section i , which is equal to the expected number of accidents (i.e., $E(Y_i)$) on a road section i during the observed time period. The basic assumption of the Poisson model is that mean and variance are equal, $E(Y_i) = Var(Y_i) = \mu_i$. If $E(Y_i) > Var(Y_i)$ then there is underdispersion, otherwise, if $E(Y_i) < Var(Y_i)$, then there is overdispersion.

The NB model relaxes the assumption of the equality of the mean and the variance by adding a random error, which has a gamma distribution. It is actually an extension of the Poisson regression model, which allows the variance of the predicted coefficients to be different from the mean and to overcome the possible overdispersion in the data. The NB

model $E(Y_i) < Var(Y_i)$. The NB distribution contains two parameters: the mean— μ_i and the dispersion parameter— α or its inversion ($\theta = 1/\alpha$). The dispersion parameter is used to “capture” an extra variation in the accident data. The probability distribution function of the NB model is defined by equation [31]:

$$p(Y_i = y_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta)y_i!} \left(\frac{\theta}{\theta + \mu_i}\right)^\theta \left(\frac{\mu_i}{\theta + \mu_i}\right)^{y_i}; i = 1, 2, 3, \dots, n, \quad (2)$$

where μ_i is the expected number of accidents on a road section i during the observed time period; θ is inversely proportional to the dispersion parameter α ($\theta = 1/\alpha$); $\Gamma(\cdot)$ is a gamma function.

The choice between the Poisson and the NB models can be largely determined by the statistical significance of the estimated coefficient α . If α is not significantly different from 0 (as measured by t —statistics), the NB model simply reduces to a Poisson regression model with $Var(Y_i) = E(Y_i)$. If α is significantly different from 0, the NB model is the right choice, and the Poisson regression model is inadequate [57].

If there is unobserved heterogeneity of the individual effects (assumed to be fixed, while, in fact, they vary through observations), the model parameter estimates will be inaccurate and lead to wrong conclusions. Solving the problem of unobserved heterogeneity is achieved using the RENB model. It allows regression parameters to vary through observations, counting on spatial and temporal correlations. The probability distribution function of the RENB model is defined by equation [28]:

$$p(Y_i = y_i) = \frac{\Gamma(a+b)\Gamma(a+\sum_T \mu_{it})\Gamma(b+\sum_T \mu_{it})}{\Gamma(a)\Gamma(b)\Gamma(a+b+\sum_T \mu_{it})} \prod_T \frac{\Gamma(\mu_{it}+y_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)}; i = 1, 2, 3, \dots, n, \quad (3)$$

where a and b are distribution parameters for θ .

The ZIP model is used to count data sets (such as accident data), when there is an excess zero in the data set. The problem with large numbers of zeros is solved by fitting a mixed model, which combines different distributions. The ZIP model consists of two parts: the “zero-inflated” and the Poisson part. The distribution of the dependent variable in the ZIP model is approximated by mixing two models and two distributions. The ZIP model produces two sets of coefficients. The first coefficient set estimates the probability of zero (the zero-inflated part), and the second coefficient set estimates the mean μ_i (the part “count”—Poisson model). The ZIP model assumes that events $Y_i = Y_1, Y_2, \dots, Y_n$ are independent. The probability distribution function of the ZIP model is defined by equation [58]:

$$p(Y_i = y_i) = p(y_i) = \begin{cases} p_i + (1 - p_i)e^{-\mu_i}; & \text{if } Y_i = 0; i = 1, 2, 3, \dots, n \\ (1 - p_i) \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}; & \text{if } > 0; i = 1, 2, 3, \dots, n \end{cases} \quad (4)$$

where p_i is the probability of the existence of the condition with the excess zero on-road section i (the logistic link function is given by the term $p_i = \lambda_i / (1 + \lambda_i)$, where λ_i is the parameter of the logit link representing covariates); $1 - p_i$ is the probability that accidents follow the Poisson distribution. If $p_i = 0$, then there is no excess zero and the mean of the ZIP model is equal to the mean of the Poisson model.

The ZINB model simultaneously solves the problem of excess zeros and overdispersion. The model consists of two parts: the “zero-inflated” and the NB part. As with the ZIP models, the distribution of the dependent variable in the ZINB model is approximated by mixing two models and two distributions. The first model uses logistic distribution to predict the non-occurrence of accidents on the road section. The second model uses a negative binomial distribution to predict how often accidents occur on a road section. The

ZINB model also assumes that events $Y_i = Y_1, Y_2, \dots, Y_n$ are independent. The probability distribution function of the ZINB model is defined by equation [31]:

$$p(Y_i = y_i) = p(y_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{\theta}{\theta + \mu_i} \right)^\theta; & \text{if } Y_i = 0; i = 1, 2, 3, \dots, n \\ (1 - p_i) \frac{\Gamma(\theta + y_i)}{\Gamma(\theta) y_i!} \left(\frac{\theta}{\theta + \mu_i} \right)^\theta \left(\frac{\mu_i}{\theta + \mu_i} \right)^{y_i}; & \text{if } Y_i > 0; i = 1, 2, 3, \dots, n \end{cases} \quad (5)$$

The estimation of the parameters in the models was conducted using the maximum likelihood estimation (MLE) method; by maximizing the likelihood function, estimates were obtained for the β , α , a and b .

There were several measures used to test the model, i.e., to verify that the developed model fit with the actual data. They summarized the differences between the observed and predicted values of accidents. In this study, the log-likelihood ratio expressed through McFadden ρ^2 (Equation (6)), the Akaike information criterion (AIC) (Equation (7)), and the Bayesian information criterion (BIC) (Equation (8)), were used to estimate the overall goodness-of-fit statistics. In addition, these measures were used to compare competing models.

$$\rho^2 = 1 - \ln L / \ln 0 \quad (6)$$

$$AIC = -2 \ln L + 2k \quad (7)$$

$$BIC = -2 \ln L + k \ln n \quad (8)$$

where $\ln L$ is the log-likelihood at the converging of the “unrestricted” model; $\ln 0$ is the log-likelihood at the converging of the “restricted” model, which contains only a constant term (with all parameters set to zero); k is the number of unknown parameters in the model i.e., the number of parameters to be estimated; n is the sample size.

Additional testing was performed using the Vuong test, i.e., the choice between the Poisson and the NB models with the ZIP and the ZINB models. More precisely, if it is $V > 1.96$, the advantage is given to the ZIP and the ZINB models, and if it is $V \leq 1.96$, the advantage is given to the Poisson and the NB models [59].

In this study, two prediction-based model selection criteria were used, including (1) mean absolute deviation (MAD) and mean squared predictive error (MSPE), which are represented by the formula:

$$MAD = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (9)$$

$$MSPE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (10)$$

where y_i is the observed number of accidents, \hat{y}_i is the expected number of accidents, and n is the number of observations. A model with a lower value of MAD and MSPE was chosen as the superior model in terms of predictive ability.

Interpretation of the effects of the estimated parameters of the exploratory variables on the accident frequency was conducted by expressing the elasticity. The elasticity (Equation (11)) measures the effect of a 1% change in the exploratory variable on the accident frequency. The elasticity is defined as the proportionate change in the accident frequency resulting from a proportionate change in a given attribute and it is obtained by taking the derivative of the accident frequency with respect to a given attribute, as follows:

$$E_{x_{itj}}^{\lambda_i} = \frac{\partial \lambda_{it}}{\partial x_{itj}} \frac{x_{itj}}{\lambda_{it}} \quad (11)$$

where $E_{x_{itj}}^{\lambda_i}$ is the elasticity coefficient of the j^{th} explanatory variable, x_{itj} is the value of the j^{th} explanatory variable for road segment i in year t .

3. Results

Table 2 presents the modeling results for five selected models. The values of the coefficients of independent variables with standard errors are shown. Moreover, the values of the goodness-of-fit statistics for all models are presented. Candidates for exploratory variables are: L, AADT, SPEEDLIMIT, NRCURVE, DAR, and IRI. In addition to these variables, dummy variables were made to take into account the annual changes in the independent variables encoded by values of 1 (if the observed road segment was in the year) and 0 (if the observed road segment was not in a certain year). The log-linear regression model $\mu_i(x_i; \beta) = \exp(\sum_{i=1}^n \beta_i x_i)$ was assumed for the expected number of accidents on the road segment i . The analysis was conducted using five statistical models, i.e., Poisson, NB, RENB, ZIP, and ZINB models. When applying Poisson, NB, ZIP, and ZINB models, it was assumed that there were 177 independent observations of the accident frequency. When we applied the RENB model, it was assumed to have 59 independent clusters, with 3 observations in each cluster. In all models, almost all the coefficients of independent variables had the expected signs. In addition, in the ZIP model, the independent variables NRCURVE and DAR had opposite signs. In addition to these variables, in the ZINB model, the AADT had an opposite sign than expected. The dummy variable “Year2017” was omitted because of the multicollinearity in all models, while the dummy variables “Year2015” and “Year2016” were not statistically significant, indicating the absence of temporal correlations in this data set. In accident frequency modeling, the Poisson regression model was first applied, in which there were (statistically significant) six independent variables: L, AADT, SPEEDLIMIT, NRCURVE, DAR, and IRI (Table 2). After the Poisson model, the NB model was applied, for which there were also (statistically significant) six independent variables. Overdispersion in data ($\alpha = 0.12$) indicates that the NB model is more appropriate than the Poisson model. Zero-inflated models were tested, i.e., ZIP and ZINB models. In the inflated part of ZIP and ZINB models, no independent variable was statistically significant, indicating the inadequacy of the zero-inflated model for this data set. This is the consequence of a small number of zeros in the data set (31.07%). The last was applied to the RENB model. It was found that there were (statistically significant) four independent variables: L ($\beta = 0.1151$; $p < 0.05$), AADT ($\beta = 0.0001$; $p < 0.05$), NRCURVE ($\beta = 0.1039$; $p < 0.05$), and DAR ($\beta = 0.0315$; $p < 0.05$). Unlike the NB model in which the dispersion parameter was constant and did not change between different road segments over time, and had a value of 0.12, the RENB model determined that the dispersion parameter was not constant between different road segments over time and that it followed the beta distribution with the parameters a (61.17) and b (9.02).

The results indicate that the RENB model had the best overall goodness-of-fit statistics. McFadden ρ^2 had the lowest value in the zero-inflated models, i.e., ZIP (0.255) and ZINB (0.277) models, which increased to 0.336 and 0.338 in the Poisson and NB models, respectively. The highest value of McFadden ρ^2 had a RENB model (0.348), which indicated the best fit of the model with the actual data. The Vuong test showed that the data set (with the Poisson and NB models) was superior to the ZIP and ZINB model, with the values of the tests being $V = 1.71$ and $V = 1.60$, respectively. To compare the competition models based on information criteria, the same set of independent variables had to be applied to all models. The lowest value of the information criteria had the RENB model, i.e., AIC (594.50) and BIC (629.44); then, the NB (AIC = 604.78; BIC = 635.54) and the Poisson model (AIC = 607.61; BIC = 636.20). The highest value of AIC and BIC had zero-inflated models. Based on the above tests, it has been determined that the RENB model has the best performance and the best goodness-of-fit statistics.

A comparison between the observed and predicted accident frequencies for the five statistical models is depicted in Figure 1. Based on the values of MAD and MSPE, it has been established that the RENB model has the greatest predictive power compared to the other tested models. The value of the MAD in the RENB model was the smallest (0.16), followed by the NB model (0.20), the Poisson model (0.22), and the zero-inflated models (ZIP = 3.87 and ZINB = 3.97). Similarly, with MSPE, the RENB model has the lowest value

(4.41), followed by the NB model (7.23) and the Poisson model (8.34). Zero-inflated models were far more inaccurate compared to the previous three models, so the ZIP model had a value of MSPE at 2657.42, and the ZINB model was 2793.77. Figure 1 shows the apparent adequacy of the RENB model in terms of predictive accuracy compared to the other models. Although the results indicate that the RENB model has the best performance compared to the other tested models, it should be noted that the performance difference between the RENB and NB models is very low. The performance of the RENB model was slightly improved compared to the NB model, e.g., at McFadden ρ^2 by only 1%.

Table 2. Model estimation results and goodness-of-fit statistics.

Variables	Poisson (S.E)	NB (S.E)	RENB (S.E)	ZIP (S.E.) ²	ZINB (S.E.) ²
Intercept	−3.0234 (0.7328) ¹	−2.8188 (0.7770) ¹	−0.7516 (0.5065) ¹	4.7542 (4.9658)	8.3044 (4.6790)
L	0.0877 (0.0221) ¹	0.1014 (0.0268) ¹	0.1151 (0.0348) ¹	−1.4961 (1.4624)	−2.3008 (1.7572)
AADT	0.0001 (0.00002) ¹	0.0001 (0.00003) ¹	0.0001(0.00003) ¹	−0.00002 (0.0002)	0.00003 (0.0001)
SPEEDLIMIT	0.0236 (0.0089) ¹	0.0216 (0.0095) ¹	0.0231 (0.0119)	−0.0693 (0.0616)	−0.1138 (0.0652)
NRCURVE	0.1332 (0.0300) ¹	0.1171 (0.0358) ¹	0.1039 (0.0465) ¹	0.5901 (0.7187)	0.2767 (0.9915)
DAR	0.0336 (0.0060) ¹	0.0319 (0.0070) ¹	0.0315 (0.0090) ¹	0.0304 (0.0727)	0.0751 (0.0751)
IRI	0.1589 (0.0613) ¹	0.1502 (0.0671) ¹	0.1339 (0.0822)	−0.1986 (0.5081)	−0.5073 (0.4828)
Year2015	0.0927 (0.1295)	0.0997 (0.1545)	0.0811 (0.1402)	0.2768 (1.1022)	0.0939 (1.2694)
Year2016	0.2126 (0.1253)	0.1797 (0.1510)	0.2175 (0.1345)	0.1720 (1.1126)	−0.2815 (1.2653)
Dispersion parameter		0.122 ¹			0.060
a			61.16802		
b			9.02539		
Summary statistics					
McFadden ρ^2	0.336	0.338	0.348	0.255	0.277
MAD	0.22	0.20	0.16	3.87	3.97
MSPE	8.34	7.23	4.41	2657.42	2793.77
Vuong test	1.71	1.60	-	1.71	1.60
AIC	607.61	604.78	594.50	609.48	608.69
BIC	636.20	635.54	629.44	666.65	660.04
No of observations	177	177	177	177	177
No of parameters	8	8	8	8	8

¹ means significant at a 0.05 significance level, ² inflated part.

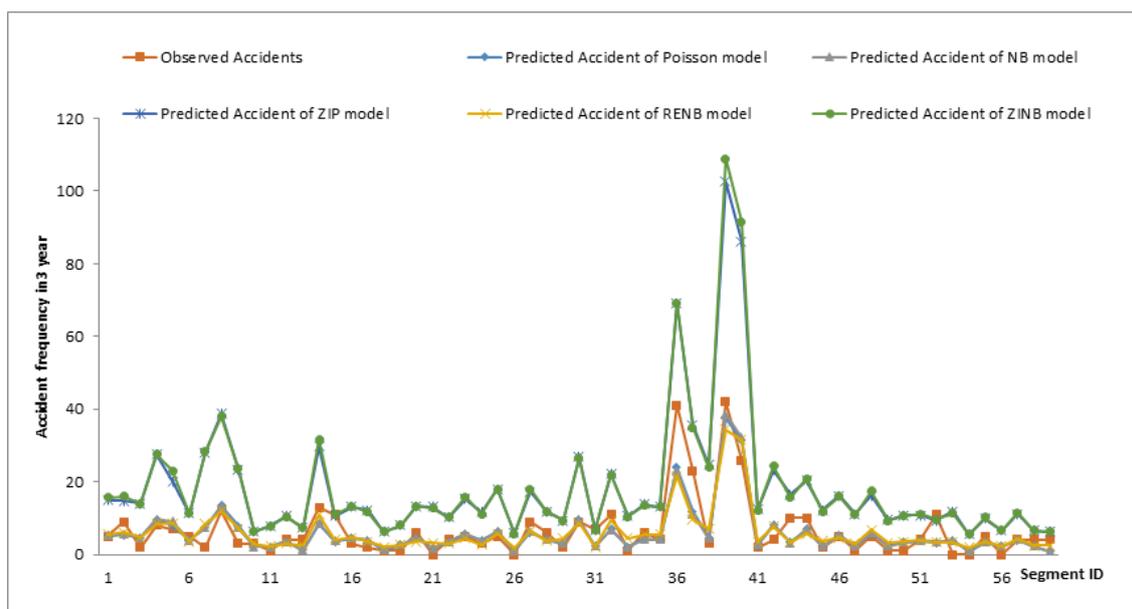


Figure 1. Comparison between the five competing models regarding the observed and predicted accident frequencies.

Table 3 shows the results of the elasticity analysis based on the RENB model. Four independent variables were statistically significant in the RENB model. The elasticity coefficient of the independent variable AADT was 0.18, indicating that an increase of 10% in AADT led to an increase of 1.8% in the accident frequency. L was also positively related to the accident frequency, indicating that as L increased by 10%, accidents increased by about 1.5% in the accident frequency. Elasticity estimates suggest that an increase in DAR by 10% was associated with an increase in accident frequency by 1%. The elasticity coefficient of the independent variable NRCURVE was 0.06, suggesting that an increase of 10% in the NRCURVE led to an increase of 0.6% in the accident frequency.

Table 3. Elasticity analysis (RENB model).

Independent Variable	Mean	Coefficient	Elasticity Coefficient
L	3.384	0.0927	0.15
AADT	3785.72	0.0001	0.18
NRCURVE	1.644	0.1278	0.06
DAR	7.462	0.0336	0.10

4. Discussion and Conclusions

Traffic accidents on flat rural roads are linked to a wide range of factors, including traffic volume, road geometric elements, and the road's environment. In this paper, an accident frequency predictive model of flat rural roads was developed, based on the selected set of independent variables, using appropriate statistical regression tools. In addition, the relative influences of selected independent variables on the accident frequencies on flat rural roads were investigated.

Selecting the most suitable statistical tool is challenging for researchers because researchers need to select (from a large number of statistical models) the one that best describes the given data set. In this study, we started with the Poisson regression model, which proved inadequate for this data set due to overdispersion. The results confirmed the findings of earlier research [15,17] that the existence of overdispersion in the data, in the application of the Poisson regression model, causes inaccurate parameter estimates and leads to the erroneous interpretation of the conclusions. The application of the NB model has shown that the existence of overdispersion in the data ($\alpha = 0.12$) justifies the use of this model [24], and—compared with the Poisson model—has better overall goodness-of-fit statistics, which is in agreement with previous research [22,49]. The poorest performances in the accident frequency modeling of flat rural roads were from the zero-inflated models (i.e., ZIP and ZINB). There was a particularly large deviation of the MSPE compared to the other models. It should be emphasized that this outcome was expected—given a small number of zeros in the observed data set (31.07%), contrary to the earlier research in which the percentage of zeros was greater than 80% [29] and, therefore, the zero-inflated models were superior compared to the Poisson and NB models. However, Lord et al. [30] claimed that the excess zeros in a data set can be the consequence of (1) the existence of sites with a combination of low exposures and high heterogeneities, (2) analyses conducted with too small spatial or time scales, (3) data with relatively high percentages of missing or unreported accidents, or (4) the application of a model with important predictive variables omitted. The conclusion was made that zero-inflated models are not suitable for modeling events such as accidents. The last model tested was the RENB model, which improved the performance of the NB model. The RENB model had the best value of overall goodness-of-fit statistics, and the best predictive power—i.e., the best fit with the actual accident frequency, which is in line with previous research [27,28]. The RENB model was chosen for the accident frequency predictive model on flat rural roads, which had the highest McFadden value of ρ^2 (0.35), and the smallest information criteria value (AIC = 594.50; BIC = 629.44). The developed model took into account the intuitive and credible traffic and geometrical factors that affected the accident frequency.

In all, six independent variables were statistically significant in the Poisson and NB models; with the introduction of the RENB model, this number was reduced to four independent variables. As expected, the segment length was statistically significant in all three specified models and was positively correlated to the accident frequency. These results are supported and confirmed by numerous previous studies [21,24,28,32,37]. Hou et al. [27] revealed that the effect of the segment length on the accident frequency, regarding the function of exposure, was also positive, and that the relationship was almost linear to the homogeneous segment lengths. Moreover, for the AADT variable, it has been found that it is consistently statistically significant and has a positive effect on the accident frequency. Rusli et al. [37] observed the effect of the ADT in the function of exposure on the accident frequency for a particular type of accident (single-vehicle accidents); they determined that this connection was positive and nonlinear. The results of this study confirm these findings and the findings of other research [28,32,35], which claim that the increase in the AADT leads to an increase in accident frequencies. The factor NRCURVE presents the effect of horizontal curves on the accident frequency. It has been found that with the increase in horizontal curves, the probability of the total number of traffic accidents is higher. The positive sign followed by this variable has also been reported by other authors. Shankar et al. [60] argued that the number of horizontal curves with a design speed of up to 112.60 kph positively affects the accident frequency. Venkataraman et al. [61] reported that the number of horizontal curves in a segment is associated with the increase in property damage accident frequency and accident frequency with possible or evident injury. On the other hand, Milton et al. [62] argued that the higher the horizontal curve density decreases, the more likelihood of injury accidents. These results can be the consequences of risk compensation (on the part of the drivers). Namely, if there are a large number of horizontal curves on a road segment, especially with a smaller radius, it is expected that a lower speed limit will be posted on the road segment and that drivers will drive at a lower speed, resulting in a reduction in the accident severity rather than accident frequency. A positive sign of the coefficient of the independent variable DAR is that the accident frequency will increase if there is a higher density of access roads on a road segment. These results confirm the findings of earlier studies that a higher density of access roads [55], number of access roads [18], or access points [41] increase the accident frequency. The number of access roads on a road segment increases the number of conflict points, and it is expected that the accident frequency will be higher. The independent variables SPEEDLIMIT and IRI were statistically significant in the Poisson and NB models. However, the RENB model showed that these variables were not statistically significant predictors in this data set. The influence of the posted speed limit on the accident frequency is complex. The authors [53] showed that a higher speed limit leads to a higher number of accidents, while another group of researchers [29,38,39] claimed it led to a reduction in the accident frequency. Differences in the results obtained can be interpreted as a consequence of unobserved heterogeneity. However, in this study, similar to the study by Ahmed et al. [63], it has been revealed that the posted speed limit is not significantly related to the accident frequency. Although the independent variable SPEEDLIMIT is not statistically significant, the positive sign of its coefficient indicates that it tends to affect the increase in the accident frequency. In contrast, research by [27,52], concerning the IRI coefficient, indicates that it tends to influence the increase in the accident frequency.

It is necessary to emphasize certain limitations in this study. A number of property damage accidents were not reported and they were not included in the accident database. Yamamoto et al. [64] found that unreported accidents lead to an inaccurate assessment of the parameters of independent variables. In addition, some accidents were incorrectly entered into the Unique Information System, which could not be included in the analysis. On the other hand, the statistical models themselves produced certain limitations in the modeling of accidents. Lord and Mannering [7] systematized the potential problems and limitations that modelers can encounter when modeling and analyzing the accident frequency.

The results of this research can help road authorities make decisions about interventions and investments in road networks, designing new roads, and reconstructing existing ones. The developed model is a particularly useful tool in the implementation of the Road Safety Impact Assessment because it represents a new method for predicting accident frequency, which can help when choosing the design variants. In addition, it has been established that it is necessary to reduce the number of horizontal curves whenever possible. However, if the number of horizontal curves cannot be reduced, it is necessary to design curves with more favorable geometric elements or to design additional traffic signs. Moreover, the results of the impact of access road density on accident frequency support the implementation of modern procedures, i.e., road safety audits and road safety inspections. More precisely, they support the reduction of the number of access roads to the main roads when possible, to reduce the number of conflict points.

Further research should be directed to developing road accident prediction models by taking into account different terrain characteristics besides flat environments. In addition, due to the problems that often exist in urban areas, such as traffic management, road geometry, etc., it is necessary to develop a suitable evaluation method for urban roads. Contemporary tools and techniques, such as machine learning or alternative regression models for count data (e.g., the Conway–Maxwell–Poisson model), offer the potential for accident frequency modeling, and an opportunity for further research in this area.

Author Contributions: Conceptualization, M.D. and S.M.; methodology, B.M.; software, B.M.; validation, G.A., S.M. and R.V.; formal analysis, M.D.; investigation, G.A.; resources, S.M.; data curation, S.M.; writing—original draft preparation, R.V.; writing—review and editing, R.V.; visualization, B.M.; supervision, M.D.; project administration, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. The Organisation for Economic Co-operation and Development (OECD). *Safety Strategies for Rural Roads*; OECD Publishing: Paris, France, 1999.
2. National Center for Statistics and Analysis. *Rural/Urban Comparison of Traffic Fatalities: 2015 Data*; Report No. DOT HS 812 521; National Highway Traffic Safety Administration: Washington, DC, USA, 2018.
3. European Commission. *Annual Accident Report*; European Commission: Directorate General for Transport: Brussels, Belgium, 2017.
4. Road Traffic Safety Agency. *Review Report: The Impact of the Road on Traffic Safety*; Road Traffic Safety Agency: Belgrade, Serbia, 2017.
5. Elvik, R.; Høye, A.; Vaa, T.; Sørensen, M. *The Handbook of Road Safety Measures*; Emerald Group Publishing Limited: Bingley, UK, 2009.
6. Butsick, A.J.; Wood, J.S.; Jovanis, P.P. Using Network Screening Methods to Determine Locations with Specific Safety Issues: A Design Consistency Case Study. *Accid. Anal. Prev.* **2017**, *106*, 223–233. [[CrossRef](#)] [[PubMed](#)]
7. Lord, D.; Mannering, F. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transp. Res. Part Policy Pract.* **2010**, *44*, 291–305. [[CrossRef](#)]
8. Mannering, F.L.; Bhat, C.R. Analytic Methods in Accident Research: Methodological Frontier and Future Directions. *Anal. Methods Accid. Res.* **2014**, *1*, 1–22. [[CrossRef](#)]
9. Savolainen, P.T.; Mannering, F.L.; Lord, D.; Quddus, M.A. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accid. Anal. Prev.* **2011**, *43*, 1666–1676. [[CrossRef](#)]
10. Hauer, E. *The Art of Regression Modeling in Road Safety*; Springer International Publishing: Cham, Switzerland, 2015. [[CrossRef](#)]
11. The American Association of State Highway and Transportation Officials (AASHTO). *Highway Safety Manual*; American Association of State Highways and Transportation Officials, Transportation Research Board of the National Academies: Washington, DC, USA, 2010.

12. Persaud, B.; Retting, R.; Lyon, C. Guidelines for identification of hazardous highway curves. *Transp. Res. Rec.* **2000**, *1717*, 14–18. [[CrossRef](#)]
13. Choi, J.; Kim, S.; Heo, T.-Y.; Lee, J. Safety Effects of Highway Terrain Types in Vehicle Crash Model of Major Rural Roads. *KSCE J. Civ. Eng.* **2011**, *15*, 405–412. [[CrossRef](#)]
14. Joshua, S.C.; Garber, N.J. Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression Models. *Transp. Plan. Technol.* **1990**, *15*, 41–58. [[CrossRef](#)]
15. Miaou, S.-P.; Lum, H. Modeling Vehicle Accidents and Highway Geometric Design Relationships. *Accid. Anal. Prev.* **1993**, *25*, 689–709. [[CrossRef](#)]
16. Lord, D. Modeling Motor Vehicle Crashes Using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter. *Accid. Anal. Prev.* **2006**, *38*, 751–766. [[CrossRef](#)]
17. Maher, M.J.; Summersgill, I. A Comprehensive Methodology for the Fitting of Predictive Accident Models. *Accid. Anal. Prev.* **1996**, *28*, 281–296. [[CrossRef](#)]
18. Saccomanno, F.F.; Grossi, R.; Greco, D.; Mehmood, A. Identifying Black Spots along Highway SS107 in Southern Italy Using Two Models. *J. Transp. Eng.* **2001**, *127*, 515–522. [[CrossRef](#)]
19. Kronprasert, N.; Boontan, K.; Kanha, P. Crash Prediction Models for Horizontal Curve Segments on Two-Lane Rural Roads in Thailand. *Sustainability* **2021**, *13*, 9011. [[CrossRef](#)]
20. Jamal, A.; Mahmood, T.; Riaz, M.; Al-Ahmadi, H.M. GLM-Based Flexible Monitoring Methods: An Application to Real-Time Highway Safety Surveillance. *Symmetry* **2021**, *13*, 362. [[CrossRef](#)]
21. Ye, X.; Pendyala, R.M.; Shankar, V.; Konduri, K.C. A Simultaneous Equations Model of Crash Frequency by Severity Level for Freeway Sections. *Accid. Anal. Prev.* **2013**, *57*, 140–149. [[CrossRef](#)] [[PubMed](#)]
22. Bonneson, J.A.; McCoy, P.T. Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways. *Transp. Res. Rec.* **1993**, *1401*, 83–89.
23. Anarkooli, A.J.; Persaud, B.; Hosseinpour, M.; Saleem, T. Comparison of Univariate and Two-Stage Approaches for Estimating Crash Frequency by Severity—Case Study for Horizontal Curves on Two-Lane Rural Roads. *Accid. Anal. Prev.* **2019**, *129*, 382–389. [[CrossRef](#)]
24. Abdel-Aty, M.A.; Radwan, A.E. Modeling Traffic Accident Occurrence and Involvement. *Accid. Anal. Prev.* **2000**, *32*, 633–642. [[CrossRef](#)]
25. Caliendo, C.; De Guglielmo, M.L.; Guida, M. A Crash-Prediction Model for Road Tunnels. *Accid. Anal. Prev.* **2013**, *55*, 107–115. [[CrossRef](#)]
26. Xu, X.; Teng, H.; Kwigizile, V.; Mulokozi, E. Modeling signalized-intersection safety with corner clearance. *J. Transp. Eng.* **2014**, *140*, 04014016. [[CrossRef](#)]
27. Hou, Q.; Tarko, A.P.; Meng, X. Investigating Factors of Crash Frequency with Random Effects and Random Parameters Models: New Insights from Chinese Freeway Study. *Accid. Anal. Prev.* **2018**, *120*, 1–12. [[CrossRef](#)]
28. Ma, Z.; Zhang, H.; Chien, S.I.-J.; Wang, J.; Dong, C. Predicting Expressway Crash Frequency Using a Random Effect Negative Binomial Model: A Case Study in China. *Accid. Anal. Prev.* **2017**, *98*, 214–222. [[CrossRef](#)] [[PubMed](#)]
29. Champahom, T.; Jomnonkwao, S.; Banyong, C.; Nambulee, W.; Karoonsoontawong, A.; Ratanavaraha, V. Analysis of Crash Frequency and Crash Severity in Thailand: Hierarchical Structure Models Approach. *Sustainability* **2021**, *13*, 10086. [[CrossRef](#)]
30. Lord, D.; Washington, S.P.; Ivan, J.N. Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accid. Anal. Prev.* **2005**, *37*, 35–46. [[CrossRef](#)]
31. Washington, S.; Karlaftis, M.; Mannering, F.; Anastasopoulos, P. *Statistical and Econometric Methods for Transportation Data Analysis*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2020. [[CrossRef](#)]
32. La Torre, F.; Meocci, M.; Domenichini, L.; Branzi, V.; Tanzi, N.; Paliotto, A. Development of an Accident Prediction Model for Italian Freeways. *Accid. Anal. Prev.* **2019**, *124*, 1–11. [[CrossRef](#)]
33. Mannering, F.L.; Shankar, V.; Bhat, C.R. Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data. *Anal. Methods Accid. Res.* **2016**, *11*, 1–16. [[CrossRef](#)]
34. Hou, Q.; Tarko, A.P.; Meng, X. Analyzing Crash Frequency in Freeway Tunnels: A Correlated Random Parameters Approach. *Accid. Anal. Prev.* **2018**, *111*, 94–100. [[CrossRef](#)]
35. Vangala, P.; Lord, D.; Geedipally, S.R. Exploring the Application of the Negative Binomial-Generalized Exponential Model for Analyzing Traffic Crash Data with Excess Zeros. *Anal. Methods Accid. Res.* **2015**, *7*, 29–36. [[CrossRef](#)]
36. Alghafli, A.; Mohamad, E.; Ahmed, A.Z. The Effect of Geometric Road Conditions on Safety Performance of Abu Dhabi Road Intersections. *Safety* **2021**, *7*, 73. [[CrossRef](#)]
37. Rusli, R.; Haque, M.M.; King, M.; Voon, W.S. Single-Vehicle Crashes along Rural Mountainous Highways in Malaysia: An Application of Random Parameters Negative Binomial Model. *Accid. Anal. Prev.* **2017**, *102*, 153–164. [[CrossRef](#)]
38. Milton, J.; Mannering, F. The Relationship among Highway Geometrics, Traffic-Related Elements and Motor-Vehicle Accident Frequencies. *Transportation* **1998**, *25*, 395–413. [[CrossRef](#)]
39. Hosseinpour, M.; Yahaya, A.S.; Sadullah, A.F. Exploring the Effects of Roadway Characteristics on the Frequency and Severity of Head-on Crashes: Case Studies from Malaysian Federal Roads. *Accid. Anal. Prev.* **2014**, *62*, 209–222. [[CrossRef](#)] [[PubMed](#)]
40. Karlaftis, M.G.; Golias, I. Effects of Road Geometry and Traffic Volumes on Rural Roadway Accident Rates. *Accid. Anal. Prev.* **2002**, *34*, 357–365. [[CrossRef](#)]

41. Greibe, P. Accident Prediction Models for Urban Roads. *Accid. Anal. Prev.* **2003**, *35*, 273–285. [[CrossRef](#)]
42. Hadi, M.A.; Aruldas, J.; Chow, L.F.; Wattleworth, J.A. Estimating safety effects of cross-section design for various highway types using negative binomial regression. *Transp. Res. Rec.* **1995**, *1500*, 169–177.
43. Shi, L.; Huseynova, N.; Yang, B.; Li, C.; Gao, L. A Cask Evaluation Model to Assess Safety in Chinese Rural Roads. *Sustainability* **2018**, *10*, 3864. [[CrossRef](#)]
44. Zhu, L.; Lu, L.; Zhang, W.; Zhao, Y.; Song, M. Analysis of Accident Severity for Curved Roadways Based on Bayesian Networks. *Sustainability* **2019**, *11*, 2223. [[CrossRef](#)]
45. Shankar, V.N.; Albin, R.B.; Milton, J.C.; Mannering, F.L. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model. *Transp. Res. Rec.* **1998**, *1635*, 44–48. [[CrossRef](#)]
46. Bauer, K.M.; Harwood, D.W. Safety Effects of Horizontal Curve and Grade Combinations on Rural Two-Lane Highways. *Transp. Res. Rec.* **2013**, *2398*, 37–49. [[CrossRef](#)]
47. Shalkamy, A.; El-Basyouny, K. Multivariate Models to Investigate the Relationship between Collision Risk and Reliability Outcomes on Horizontal Curves. *Accid. Anal. Prev.* **2020**, *147*, 105745. [[CrossRef](#)]
48. Huo, X.; Leng, J.; Luo, L.; Wang, D. A Mixed Logit Model with Mean-Variance Heterogeneity to Investigate Factors of Crash Occurrence. *Int. J. Injury Control Saf. Promot.* **2021**, *28*, 301–308. [[CrossRef](#)]
49. Miaou, S.-P. The Relationship between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions. *Accid. Anal. Prev.* **1994**, *26*, 471–482. [[CrossRef](#)]
50. Montella, A.; Imbriani, L.L. Safety Performance Functions Incorporating Design Consistency Variables. *Accid. Anal. Prev.* **2015**, *74*, 133–144. [[CrossRef](#)] [[PubMed](#)]
51. Caliendo, C.; Guida, M.; Parisi, A. A Crash-Prediction Model for Multilane Roads. *Accid. Anal. Prev.* **2007**, *39*, 657–670. [[CrossRef](#)]
52. Anastasopoulos, P.C.; Mannering, F.L. A Note on Modeling Vehicle Accident Frequencies with Random-Parameters Count Models. *Accid. Anal. Prev.* **2009**, *41*, 153–159. [[CrossRef](#)] [[PubMed](#)]
53. Donnell, E.T.; Mason, J.M. Predicting the Frequency of Median Barrier Crashes on Pennsylvania Interstate Highways. *Accid. Anal. Prev.* **2006**, *38*, 590–599. [[CrossRef](#)] [[PubMed](#)]
54. Imprialou, M.-I.M.; Quddus, M.; Pitfield, D.E.; Lord, D. Re-Visiting Crash–Speed Relationships: A New Perspective in Crash Modelling. *Accid. Anal. Prev.* **2016**, *86*, 173–185. [[CrossRef](#)]
55. Ma, X.; Chen, F.; Chen, S. Modeling Crash Rates for a Mountainous Highway by Using Refined-Scale Panel Data. *Transp. Res. Rec.* **2015**, *2515*, 10–16. [[CrossRef](#)]
56. Vaiana, R.; Perri, G.; Iuele, T.; Gallelli, V. A Comprehensive Approach Combining Regulatory Procedures and Accident Data Analysis for Road Safety Management Based on the European Directive 2019/1936/EC. *Safety* **2021**, *7*, 6. [[CrossRef](#)]
57. Poch, M.; Mannering, F. Negative Binomial Analysis of Intersection-Accident Frequencies. *J. Transp. Eng.* **1996**, *122*, 105–113. [[CrossRef](#)]
58. Lambert, D. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* **1992**, *34*, 1–14. [[CrossRef](#)]
59. Vuong, Q.H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **1989**, *57*, 307–333. [[CrossRef](#)]
60. Shankar, V.; Mannering, F.; Barfield, W. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accid. Anal. Prev.* **1995**, *27*, 371–389. [[CrossRef](#)]
61. Venkataraman, N.; Ulfarsson, G.F.; Shankar, V.N. Random Parameter Models of Interstate Crash Frequencies by Severity, Number of Vehicles Involved, Collision and Location Type. *Accid. Anal. Prev.* **2013**, *59*, 309–318. [[CrossRef](#)] [[PubMed](#)]
62. Milton, J.C.; Shankar, V.N.; Mannering, F.L. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accid. Anal. Prev.* **2008**, *40*, 260–266. [[CrossRef](#)] [[PubMed](#)]
63. Ahmed, M.; Huang, H.; Abdel-Aty, M.; Guevara, B. Exploring a Bayesian Hierarchical Approach for Developing Safety Performance Functions for a Mountainous Freeway. *Accid. Anal. Prev.* **2011**, *43*, 1581–1589. [[CrossRef](#)]
64. Yamamoto, T.; Hashiji, J.; Shankar, V.N. Underreporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models. *Accid. Anal. Prev.* **2008**, *40*, 1320–1329. [[CrossRef](#)]