IJTE, Vol. 10, No. 4, Spring 2023, pp. 1217-1236
Research Paper

A New Holistic Crashes Prediction Model based on Zero-Truncated Data for Intercity Four-Lane Highways Curves

Mehran Ghorbani[1], Mahmoud Saffarzadeh[2,*], Ali Naderan[3]

Received: 2022/01/31 *Accepted: 2022/04/30*

Abstract

This study is aimed at exploring the effect of some recognized and new candidate variables of horizontal curves on crash frequency in four-lane highways using zero-truncated crash data. The present study has considered the related variables for 45 curves of four-lane intercity highways during a three-year period (2018-2020). The standard Poisson distribution is a benchmark for modeling Equi-dispersion count data and could not express Under-dispersion zero-truncated data. The modeling was performed using Poisson, Negative Binomial, Zero-Truncated Poisson, Zero-Truncated Negative Binomial, and Conway-Maxwell Poisson (COM-Poisson) regression. The results revealed that the COM-Poisson regression distribution could effectively fit the model Under-dispersion zero-truncated Crashes data. According to the results, using the consistency and self-explaining variables as a useful approach for the estimation of crash frequency in four-lane highway horizontal curves was evaluated.

Keywords: Consistent design, Self-explanatory, Crash model, Intercity, Four-lane, Highway, Horizontal curve, Poisson, Conway-Maxwell

[*] Corresponding author: E-Mail: saffar_m@modares.ac.ir

[1] Ph.D. candidate, Department of Civil Engineering, Faculty of civil and earth resources engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

[2] Professor, Department of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran

[3] Assistant Professor, Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

1. Introduction

Traffic safety has been a growing concern for road safety professionals around the world. Road traffic fatalities and injuries are a major cause of death and disability, with a disproportionate number occurring in Iran as in other developing countries.

The accident health risk index, i.e., the number of deaths per 100,000 inhabitants, is 20.5 in Iran which is higher than that of middle-income countries with an average of 18 [WHO, 2018]. A traffic crash is caused by three general factors: humans, vehicles, and roads. Most traffic crashes are related to human errors [Treat et al. 1997; NHTSA, 2016]. The condition of road elements plays an important role in road safety. Crash analysis has shown that the crash rate in horizontal curves is about 2 to 5 times that of the tangent sections on two-lane rural roads [Lamm et al. 1999]. The safety status of curves on rural highways in developing countries is more important due to the existing inconsistency of the conditions of many roads regarding geometric design standards, including radius, superelevation, and sight distance.

Road design can be confusing and inconsistent with road users' expectations. Generally, the road design is considered inconsistent on a road alignment in which drivers may face high-speed changes and unexpected events and driver's expectations are not met [Cafiso et al. 2007].

A poorly consistent road causes drivers' confusion. In addition, too many changes along different road sections can increase the likelihood of crashes for different drivers. So far, the concept of consistent roads has focused on geometric design aspects such as the radius of curves and the prevention of snap speed reduction on roads.

On the other hand, the self-explaining road presents a more comprehensive definition of the provision of the information required by drivers.

The objective of this paper is to present a consistent and self-explaining based model for crashes in the horizontal curves on four-lane intercity highways in Iran. The main contribution or novelty of this study is identifying and applying a combination of different variables of design consistency an self-explaining in modeling of intercity curves crash prediction on four-lane intercity highways. Another contribution is to find a suitable discrete probability distribution for zero-truncated under-dispersion data. Consistency variables used in this study include performance speed, vehicle stability, lane alignment, and driver workload. Variables related to 6-second logic, the field of view logic, and driver perception logic were considered as the self-explaining criteria. The data from 45 curves on four-lane intercity highways from 15 Iranian provinces were collected and used in modeling.

The most popular distribution for modeling count data is the Poisson distribution, which assumes equi-dispersion of the variables. Since the observed count in this study exhibits under-dispersion, Poisson models become less ideal for modeling. Poisson, Negative Binomial, Zero-Truncated Poisson, Zero-Truncated Negative Binomial, and Conway-Maxwell Poisson (COM-Poisson) regressions were used as an alternative for the regression models.

The remainder of this paper is organized as follows: In Section 2, a review of the background of the consistency and self-explaining measurement indexes is presented. Section 3 describes the research method and data used in the research. Section 4 presents model Development. Finally, Section 5 provides a conclusion and recommendations**.**

2. Literature Review

The following presents a literature review of the potential measures and prediction models.

The term "consistent alignment" is referred to the degree to which a highway complies with the drivers' expectations to avoid crash-leading maneuvers. The importance of a consistent alignment is reflected by its association with safety. Research on design consistency has mainly focused on presenting quantities measures and developing some models for its prediction.

Design consistency is referred to a design in which the geometric components of a road segment are in line with its operational parameters perceived by the drivers. Previous studies have proposed some procedures to measure this consistency quantitatively. Overall, this alignment can be categorized into four main types Operating Speed, Alignment Indices, Vehicle Stability, and Driver Workload [Hassan et al. 2001]. Studies on the consistency of the road geometric design have been related to two-lane roads so far. To the best of our knowledge, no study has assessed the consistency of expressways or four-lane highways. The reason for this research gap is the high standards of expressways construction and the relative certainty of achieving consistent conditions on such roads. In Iran, most four-lane highways have been constructed by widening the previous two-lane roads. As a result, the main geometric safety features of the four-lane road are similar to the two-lane road. However, speed values on such roads are considered the same as the speed limit of expressways (i.e., 110 km/h). In this way, drivers sometimes should lower their speed as low as 60 km/h or less in the curve sites that are unsafe action. In the following, the alignment criteria for the consistency assessment of a geometric design are explained.

Operating speed is the most widely used factor for design consistency assessment. Typically, this criterion is defined as the 85th percentile of driving speed. Many crashes occur due to improper speed adjustment by drivers in areas required for speed adjustments, such as curves and intersections [Al-Masaeed et al. 1994].

In this respect, sudden unexpected changes in roadway alignment may demand snap changes in vehicle operating speed [Lamm et al. 1999]. Changes in vehicle operating speeds can effectively reflect the presence of inconsistencies in the geometric design of road features [Nicholson, 1998]. The speed reduction on a horizontal curve with a preceding curve or tangent may be accompanied by crash frequency [Fitzpatrick et al. 2000a]. Among various methods available for consistency estimation, the most widely used techniques for this purpose are based on the operating speed of vehicles, suggesting the meaningful relationship of this criterion with crashes [Luque and Castro, 2018]. Operating speed is defined as the 85th percentile speed of a vehicle on roads under free-flow conditions.

Operating speed consistency describes the difference in V85 between two successive geometric elements. There are two ways of collecting speed data: 1) using radar, surveillance cameras, etc., and 2) using mathematical equations. Choosing how to collect data depends on the project size and the feasibility of collecting data. Researchers from different countries have proposed several equations for determining the predicted operating speed based on alignment parameters. To the best of our knowledge, most studies employ these models for speed prediction.

Alignment Indices are the second used factor for design consistency assessment that road safety designers should consider [Lamm et al. 1999]. In a consistent alignment, drivers with enough confidence can drive safely at their desired speed throughout the entire alignment. Related indices reflect the general features of an alignment in a road segment.

Some advantages of using these indices are as follows [Anderson et al. 1999]: First, designers and road safety experts can easily use and understand these indices. Second, they can be used to offer a mechanism for the numerical comparison of successive geometric elements from a system-wide perspective. Third, they can quantify the interaction between the vertical and horizontal alignments in a roadway. In a thorough analysis of road accidents in

Washington State, CRR (i.e., the radius of an individual horizontal curve to the average radius of the entire section) was reported as the most determining factor in collision prediction. Moreover, unlike other alignment criteria that are only usable in a relatively long road section, CRR can be applied to assess an individual curve [Fitzpatrick et al., 2000b]. Research has also shown that road safety is sensitive to CRR. According to Anderson and his colleague's study, CRR can be considered an effective criterion to measure design consistency [Anderson et al. 1999]. The reason for the efficiency of this criterion is that when a horizontal curve's radius deviates significantly from the average radius along the road segment, the curve might create inconsistency and violate driver expectancy [Anderson et al. 1999]. Some potential alignment indices recommended by researchers are the average rate of vertical curvature (AVC) by Fitzpatrick et al. [2000a], Curvature change rate (CCR) by Fitzpatrick et al. [2000b], the ratio of the radius of the curve to the mean radius (CRR) by Anderson et al. [1999], and the ratio of maximum radius of curvature to a minimum radius of curvature by Polus [1980].

Vehicle stability is the third used factor for design consistency assessment. In this regard, surface friction is an important and critical characteristic of the pavement surface that provides the driver the ability to accelerate, decelerate and steer the vehicle. It can be reduced as a consequence of surface contamination (from water or pollution) and polished surface aggregate. Vehicle stability is an essential factor in measuring design consistency. Head-on collisions and rollovers of vehicles may be due to huge centripetal forces inserted into a vehicle that moves on a horizontal curve at insufficient friction. Lamm proposed using vehicle stability for design consistency and safety assessment [Lamm et al. 1991]. Studies on vehicle stability analysis are mainly based on determining the safety margin for a vehicle traveling at the operating speed and the safety margin for the difference between side friction demand and side friction supply.

Driver workload is the last used factor for design consistency assessment. The driving workload can be defined as demand tasks applied to a set of undifferentiated mental resources that has a significant effect on the driver's performance. With increasing the complexity of the road's geometric features, the time needed to perform a given driving maneuver increases, as well, resulting in a higher driver workload. In the case of low workload levels, performance will decline due to information missing caused by inattention or making drivers bored or tired. Hence, designers should avoid highway segments with excessively very or low high driver workloads. This workload measures the effort expected by a human operator when performing a task, irrespective of how the task is performed [Senders,1970]. Road safety should be improved by avoiding geometric features inducing very high or low driver workloads.

A visual demand (VD), also known as a direct measure, is the visual information to drive safely in a given roadway path. Both subjective and physiological criteria have been introduced for driver workload measurement [Krammes et al. 1995].

The basic notion of a self-explaining road is a "traffic environment which elicits safe behavior simply by its design" which is considered a new design concept in the roads and their environment. This concept recommends designing roadways in such a way that users can quickly and easily know the expected safe behavior and adapt their expectations for that type of road [Theeuwes and Godthelp, 1995].

A driving error is typically the consequence of poor driving performance where the driver exposes the car to an unsafe position. The drivers may correct their mistakes by changing their driving mode (i.e., acceleration or braking). Hence, these errors may be corrected before drastic consequences such as crashes [PIARC, 2019]. Gestalt psychology can describe several issues associated with the probability of a road accident event. Gestalt is expressed as the perception of certain contents provided from the landscape background. According to Gestalt, regarding immediate perception (i.e., olfactory, auditory, visual, taste, and verbal), the perceiver could detect an "image" that is variable in reality and decide based on it [Stadtler, 1998]. Gestalt laws were soon considered valuable pieces of information by road design experts. Consequently, several experts believe that providing many design conditions may lead to misjudgment of curves, paths, and slopes [PIARC, 2019]. Road environment features can form driving behavior on the road [Bargh and Ferguson, 2000]. For instance, drivers deliberately follow a route they are already familiar with to reach a new destination, regardless of its length. In a modeling study, participants could not recognize a change in an important road safety sign even after driving on the road 24 times [Martens and Fox, 2007]. More than 90% of the information needed by a driver to take correct decisions during driving is in visual forms [Hills, 1980; Sivak, 1996]. Research has also shown that visual impairments are among the major causes of traffic crashes [Charlton and Starkey, 2013]. Based on the human factors' interaction with road design, there are three major principles to consider in the geometric design of roads [PIARC, 2016]. These three principles are explained in the following.

The first principle is the on-time notification of upcoming events (4 to 6 seconds): It takes at least 4 to 6 seconds for a typical driver to adjust from one traffic situation to another or adjust to a new situation. In crashes at curves, the driver most often notices a change in the conditions ahead of the route too late. Therefore, there is not sufficient time for drivers to avoid crashes. Self-explaining and user-friendly design give the driver the required time (4-6 seconds). It takes to adapt, including the advance and warning sections, the encounter section (route prediction), the approach (decision) section, and the maneuver section [PIARC, 2019].

The second principle is the driver's field of view: A safe field of view for road users should be provided. An adequate field of view supports and guides the motor vehicle driver and prevents him from deviating toward the edge of the traffic lane or even leaving the lane. Misleading eye-catching objects on the roadside that do not align with the axis of the road cause unconscious changes in the movement direction. Such objects lead to gross errors in orientation and controllability, such as disturbances in staying on the path. The status of the factors related to the density of the field of view, fixed elements' state in the road environment, and depth of the field of view are factors effective in providing the required conditions of the field of view logic. The absence of boring uniformity of the driver on the road and its surroundings and the absence of cross-section with glare-leading far-distance vision before the curves are considered in the density assessment of the field of view. Elements' condition in maintaining the optimal traffic lane is considered in fixed elements assessment in the lateral road environment. Depth of field of view also requires evaluating the degree of sharpness and the possibility of tracking objects and the absence of ambiguity in detecting distances and objects on the road and its environment [PIARC, 2019].

The third principle is the logic of driver perception: The perceptual logic of road users must be considered in the road environment and along the route. Drivers drive along the road with their anticipation and orientation, which forms through their recent experiences and perceptions, based on the last 5-10 minutes of the driving experience. The driving behavior on the road is adjusted unconsciously. Central vision and side vision make the entire visual field. Unexpected deviations and the presence of unpredictable objects that are far from the driver's pre-planned expectations disrupt the automatic sequence of actions. Also, they may cause the driver to slip and increase the potential of a crash occurrence. The driver's perceptual logic principle is evaluated based on the location of road alignment with the driver's expectations, the lack of a sudden increase in the driver's mental workload, and the absence of deficiencies in traffic control devices [PIARC, 2019]. According to recent studies, road self-explaining variables have not been used in crash prediction modeling due to the novelty of the self-explaining index. Furthermore, the lack of preliminary studies in this context has led to inattention to the role of self-explaining in road safety.

Crash prediction models, known as Safety performance functions (SPFs), are statistical models predicting the probability of crashes on a certain section of roadway that can be used to investigate the effect of various variables on the crash indicator. Research on crash generation models has focused on non-human behavioral parameters such as road geometric design, road environment, and traffic flow factors. Many of the models proposed for road crash prediction employ generalized linear regression (GLM) with the Poisson distribution. These models are established based on variables of traffic volume and road geometry regarding the haphazard nature of crashes and their non-negative values [Anderson et al. 1999; Fitzpatrick et al. 2000b]. In practical problems, e.g., crash prediction, the dependent variable (response) is not normally distributed. Therefore, they cannot be modeled by linear models. There are three conditions for dispersion of count data: equi-dispersion, over-dispersion, and under-dispersion. Equi-dispersion is when the variance is equal to the mean of the crash counts. Over-dispersion is when the variance exceeds the mean of the crash counts. Finally, under-dispersion is a rare phenomenon in which the mean is greater than the variance of the crash counts on-road sections. For equi-dispersion, common Poisson distribution is suitable. However, over-and under-dispersion cases may lead to erroneous crash predictions and inaccurate inferences about the crash factors. The dependent variable(s) with over-dispersion or under-dispersion can violate some of the modeling approaches' basic count-data modeling assumptions. In the case of the dependent variable(s) with over-dispersion, researchers have recommended employing regression modeling with negative binomial (NB) distribution [Ng and Sayed, 2004; Cafiso et al. 2010]. The dependent variable(s) with under-dispersion is more probable when the sample mean value is very low. To this end, researchers have recommended employing regression modeling with Conway-Maxwell Poisson (COM) distribution [Shmueli et al. 2005]. To the best of our knowledge, no previous research has modeled crash data on intercity highway curves with zero-truncated under dispersion status. Numerous prediction and analysis models for intercity horizontal curve crashes have been proposed by various researchers. These models mainly include different types of GLM as 1) Poisson structure, 2) negative binomial, 3) log-normal regression, and 4) zero-inflated negative binomial. As a general statistic, it can be said that about 70% of the surveyed studies have used the negative binomial method [Anderson et al. 1999],[Al-Sahili et al. 2019],[Bird and Hashim, 2006], [Cafiso et al. 2010], [Gooch et al. 2016], [Hamilton et al. 2019], [Khan et al. 2012], [Liopis-Castelló et al. 2018], [Ng and sayed, 2004], [Persaud et al. 2000], 15% have used method poisson structure [De ona and Garach,2013], [Saffarzade et al. 2007] and 15% have used other methods [Schneider et al. 2009], [Schneider et al. 2010], [Dhahir and Hassan, 2017]. Among the methods examined, more than 65% of the researchers used a total number of crashes and some others used other parameters such as Injuries and fatalities crashes [Cafiso et al. 2010], some truck crashes [Schneider et al. 2009], vehicle crashes on motorcycles [Schneider et al. 2010] and rollover crashes [Hamilton et al. 2019].

3. Research Method

The following is a description of the methodology used in data collection, selected distributions for modeling, and criteria for evaluating the fitted models.

3.1. Data Description

The study site in this research is 45 curves with a history of fatal crashes on intercity four-lane highways in 15 Iranian provinces. It is of note that the selection of these curves and the type of their crashes were based on limitations in crash data and their accuracy. Fatal road crashes data collected to identify crash-prone areas have the highest accuracy and reliability in Iran. These data do not have a value of zero. On the other hand, the mean is greater than the variance of crash counts on curve sections. Therefore, the crash data used in this research are under-dispersion zero-truncated. These fatal crash data of the mentioned curves have been collected in cooperation with the National Road Police and the General Directorate of Road Maintenance and Transportation for three years ending in 2020. The required data for the specification of chosen curves were collected from the databases available in the Road Maintenance and Transport Organization (RMTO) of the Islamic Republic of Iran.

The data set includes traffic volume, geometric design, skid resistance, photographs, and video files taken for chosen curves of four-lane intercity highways. It is noteworthy that the conventional methods for collecting speed data using radar, surveillance cameras, etc., were not possible. Hence, in this study, the operating speed was estimated using an equation calibrated in previous studies used in the RMTO for the road rating project. For each curve, three indices related to the three corresponding principles of self-explaining were evaluated.

Then, they were scored by three safety auditors using images and video information based on the PIARC recommended specifications for determining the curves' self-explaining scores [PIARC, 2019].

Also, Table 1 summarizes the characteristics of homogeneous road curves' variables including exposure, geometric and operating, consistency, and self-explaining characteristics considered for model development in the present study.

Table 1. Summary of characteristics of homogeneous curves variables considered for model development

Variable Abbreviation		Description	Mean	Min.	Max.	Standard deviation
Exposure	ADT [veh/day]	Average daily traffic	9337	5142	14546	2088.96
	HV[%]	Percentage of heavy vehicles in traffic	27.51	14.58	71	16.01
Geometric and Operational	Df [deg]	Deflection angle	25.5	6.52	51.59	11.87
	LC [m]	Length of curve	356.7	125	810	151.81
	TL[m]	Length of tangent preceding the curve	890.2	10	5000	1507.66
	R [m]	Curve radius	880.8	216	1648	358.45
	TL/R [m/m]	The ratio of the length of tangent to curve radius	0.992	0.0006	5.81	1.456
	CCR [gon/km]	The curve change rate of the curve	94.81	32.64	220.84	43.19
	Sh [m]	Total curve shoulders width	2.28	1	4.4	1.01
	G [%]]	Longitudinal curve grade	2.741	0	9.91	2.76
	IRI [m/km]	International roughness index	2.604	1.7	4.6	0.563
Consistency	ΔV_{85}[km/h]	The absolute value of the operating speed difference in the tangent-to-curve transition	5.293	0	17.24	5.08
	Δ CCR [gon/km]	Difference between the curve change rate and the average change rate of curves in 3 km before the curve	-88.88	-301	104.42	117.41
	CRR [m/m]	Modified change radius rate: the ratio of the curve radius to the average radius of three previous curves	1.352	0.72	2.356	0.426
	Δf [-]	The difference between existing and demanded side friction (at the 85th-percentile speed)	-0.05	-0.145	0.125	0.059
	Δe [%]	The difference between existing and demanded superelevation (at the 85th-percentile speed)	-1.401	-6.52	12.86	3.207
	VD.lu	Visual demand for unfamiliar driver	0.492	0.41	0.65	0.064
Self-explaining	SE-6sr [num]	6-second rule score of curves self-explaining condition	23.4	14.77	27.28	2.762
	SE-fv [num]	Field of view score of curves self-explaining condition	31.55	17.05	44.32	7.28
	SE-lr [num]	Logic rule score of curves self-explaining condition	18.51	11.36	27.74	3.56
	SEsum [num]	A total self-explaining score of curves condition	73.57	50	94.32	10.94
Crash	CR-fatal [num]	Fatal crashes in curve area (per 3 years)	2.644	1	5	1.264

3.2. Selected Distributions for Modeling

In this study, curve crash variables were estimated using generalized linear modeling (GLM). The GLM is superior to the conventional linear regression because the former can overcome the latter’s limitations [Ng and Sayed, 2004]. In the GLM, Poisson or negative binomial is considered the error structure that best fits the crash occurrence.

Negative binomial and Poisson models have been widely used in numerous studies concerning curve crash modeling. In the case the values of the response variable are under-dispersion, it is not reasonable to apply regression modeling with NB distribution.

However, since the data used for this purpose is non-zero, the following alternative models were used for this purpose: Zero-truncated negative binomial and zero-truncated Poisson. Also, the simulations were performed using the Conway-Maxwell Poisson model regarding the under-dispersed nature of the applied data. Therefore, according to the collected data, which are numerical and non-zero, five models (i.e., Poisson, zero-truncated Poisson, negative binomial, zero-truncated negative binomial, and Conway-Maxwell Poisson) were chosen to meet the research requirements.

The best model is selected among these models through statistical measures. The following is a brief description of these five models and their mathematical equations.

3.2.1. Poisson Regression

The Poisson distribution is based on the independent occurrence of events and discrete probabilities.

This regression is employed in several applications using rare event measurements [Miller and Freund, 1977].

The probability function of this regression is as follows:

$$f(y_i,\mu) = \frac{e^{-\mu}\mu^{y_i}}{y_i!} \quad y_i = 0, 1, 2, 3,.. \tag{1}$$

Where y_i is a counting variable (i.e., the number of occurrences) and $\mu > 0$ is the Poisson parameter.

Independent variables are imported to the model using the function g(μ) = log(μ). Since the mean and the variance are equal in this distribution, they are formulated as $E(y_i) = Var(y_i) = \mu$. The Poisson regression model is expressed by Eq. (2):

$$g(\mu)= \log(\mu)= x.\beta + \varepsilon \tag{2}$$

Where x indicates the independent variables' observation vector and β is a regression coefficient. As $E(y_i) = \mu$, the Poisson regression model is expressed as follows:

$$y= \mu=e^{x\beta} \tag{3}$$

3.2.2. Negative Binomial Regression

The negative binomial distribution is appropriate for modeling count non-negative values with over-dispersion.

Eq. (4) expresses the probability mass function (PMF) of this distribution [Geene, 2008; Cummings, 2009] mathematically:

$$P(y)= \frac{\Gamma(y+\theta)}{y!\Gamma(\theta)} \left(\frac{\mu}{\mu+\theta}\right)^{y} \left(\frac{\theta}{\mu+\theta}\right)^{\theta} \tag{4}$$

Where $\Gamma(.)$ is the Gamma function, θ is the shape parameter, y= 0, 1, 2, …, and μ is the mean. Also, E(y) = μ and Var(y) = μ(1+μθ-1) show this distribution's mean and variance, respectively.

3.2.3. Zero -Truncated Poisson Regression

This distribution, also known as Positive Poisson Distribution is employed only for non-zero integers [Johnson et al. 2005; Wimmer and Altmann, 1999].

The probability mass function of this distribution is expressed as Eq. (5):

$$p(y_i \mid y_i>0)= \frac{p(y_i)}{1-p(y_i=0)} = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i\,!(1-e^{-\mu_i})} = \frac{\mu_i^{y_i}}{y_i\,!(e^{-\mu_i}-1)} \ , \ y_i= 1, 2, 3, \ldots \tag{5}$$

Eq. (6) and Eq. (7) represent the mean and variance of this distribution, respectively:

$$E(y_i)= \frac{\mu_i\, e^{\mu_i}}{e^{\mu_i}-1} \tag{6}$$

$$Var(yi)= \frac{\mu_i e^{\mu_i}}{e^{\mu_i}-1} \left(1- \frac{\mu_i}{e^{\mu_i}-1}\right) \tag{7}$$

Where g (μ_i) = log (μ_i). The regression function of this distribution is similar to that formulated in Eq. (2).

3.2.4. Zero- Truncated Negative Binomial

As a conditional probabilistic distribution for non-zero values, the PMF of the zero negative binomial distribution is expressed by Eq. (8) [Cohen, 1960]:

$$P(y/ \quad \mu,\alpha,y > 0 \quad) = \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{2}\left(\frac{1}{1+\alpha\mu}\right)^{\alpha-1} \times \left(\frac{1}{1-(1+\alpha\mu)^{-\alpha^{-1}}}\right) \quad (8)$$

$$E(y) = \frac{\mu}{1-(1+\mu/\alpha)^{-(1/\alpha)^{-1}}} \quad (9)$$

Where the response variable (y) follows a Poisson distribution with positive numbers, and μ and α show the mean response for all observations and dispersion parameters, respectively.

3.2.5. Conway-Maxwell Poisson

The COM-Poisson distribution is a generalization of the Poisson distribution. Statisticians have re-formulated this distribution to model count data, which are over-dispersed or under-dispersed [Conway and Maxwell, 1962; Shmueli et al. 2005].

This distribution is flexible enough to represent the distribution of a wide range of count data [Sellers and Shmueli.,2010]. The distribution function of this probability is mathematically presented as Eq. (10):

$$P(y, \upsilon) = \frac{\lambda^{y}}{(y!)^{\upsilon}\, z(\lambda,\upsilon)} \quad (10)$$

$$(\lambda, \upsilon)=\sum_{n=0}^{\infty} \frac{\lambda^{n}}{(n!)^{\upsilon}} \quad (11)$$

Where y is a discrete count and λ is a centering parameter almost equal to the mean of the observations in many cases. Also, $\upsilon \geq 0$ is a normalizing constant representing the shape parameter of the COM-Poisson distribution. In Eqs. (10) and (11), υ is the dispersion parameter in a way that $\upsilon > 1$ and $\upsilon < 1$ denotes under-dispersion and over-dispersion, respectively. The COM-Poisson distribution consists of the following three popular distributions: Geometric distribution ($\upsilon = 0, \lambda < 1$), Poisson ($\upsilon = 1$), and Bernoulli distribution ($\upsilon \rightarrow \infty\ with\ \frac{\lambda}{1+\lambda}$) (Shmueli et al., 2005).

Eq. (12) shows a COM-Poisson regression model based on the link function by taking a GLM approach (Sellers and Shmueli, 2010):

$$(Y) = \log \lambda = X'\beta = \beta_0 + \sum_{j=1}^{p} \beta j X_j \quad (12)$$

Eq. (12), mathematically expresses the relationship between $X'\beta$ and (Y) [Sellers and Shmueli, 2013].

3.3. Goodness of Fit Tests

3.3.1. Pearson's Chi-Squared

Pearson's Chi-squared (χ2) is a statistical measure to assess the proximity degree of actual values and prediction values from the proposed models.

This measure is determined using the GOF test. This statistic is mathematically related to the n-v-1 degree of freedom (DOF) expressed by Eq. (13) [Taylor,1982].

$$x^{2}= \sum_{i=1}^{n}\left(\frac{y_i-E(y_i)}{\sqrt{Var(y_i)}}\right)^{2} \quad (13)$$

Where v and n show the number of parameters and observations, respectively.

Next, the $X^{2}_{0.05}/X^{2}$ index is calculated as the relative critical point of the Chi-square distribution with a significant level of 0.05 on the Chi-square distribution of the fitted model. If the expression output is in the range of 0.8 to 1.2, the face and denominator of the fraction are equal or close to each other.

3.3.2. Akaike Information Criterion (AIC)

AIC is one of the well-known criteria for assessing the models' performance based on several likelihood measures. This criterion is a measure of the GOF of an estimated statistical model [Akaike, 1974]. The mathematical formulation of AIC is expressed by Eq. (14):

$$AIC = -2 \log L + 2p \quad (14)$$

Where L shows the maximum likelihood function for the model, and p denotes the number of parameters the statistical model has. A model with a lower AIC has a high best performance and vice versa.

3.3.3. R^2(Cox and Snell)

$R^{2}_{(\text{Cox and Snell})}$ is a criterion to adjust the statistic scale for covering the full range from 0 to 1. This criterion is formulated by Eq. (15), as follows:

$$R^{2}_{C\&S} = 1 - (L_0 / L_M)^{2/n} \quad (15)$$

Where L_0 and L_M are the likelihood function for a model with no predictors and the model to be estimated, respectively, and n indicates the sample size [Cox and Snell, 1989].

3.3.4. Mean Prediction Bias (MPB)

MPB can be used as a criterion to measure the bias direction and magnitude of the average model [Oh and Lyon, 2003].

In a positive MPB, the model over-predicts car accidents. On the other hand, a model with a negative MPB underestimates the crashes. MPB is expressed using Eq. (16):

$$\text{MPB} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i) \quad (16)$$

Where, y_i and $\hat{y_i}$ are observed and predicted crashes, respectively, and n is the sample size.

3.3.5. Mean Absolute Deviance (MAD)

MAD is a criterion to estimate the average predictive error of the model [Oh and Lyon, 2003].

This criterion is calculated using Eq. (17):

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y_i} - y_i| \quad (17)$$

3.3.6. Mean Squared Predictive Error (MSPE)

MSPE is a criterion to evaluate the error of an external data set or the validation error [Oh and Lyon, 2003]. MSPE is calculated using Eq. (18), as follows:

$$\text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2 \quad (18)$$

4. Model Development

The GLM was used to estimate the parameters of the crash prediction models. The GLM has the advantage over random linear regression by overcoming the limitations associated with random, discrete, and non-negative data [Ng and Sayed, 2004]. For the GLM approach, the error structure which best fits the crash occurrence is usually assumed Poisson or negative binomial. Every regression technique selection should always be based on analyzing the data under study. The model development phase included variable selection, fitting models, and comparison of the models to determine the best model. The data in this study included zero-truncated count data with under-dispersion. The best model and the most significant variables were identified by fitting five types of regression models called Poisson, negative binomial (NB), Zero-truncated Poisson(ZT-Poisson), Zero-truncated negative binomial(ZT-NB), and Conway-Maxwell Poisson. There is limited software for estimating Poisson models, including zero-truncated and Conway-Maxwell Poisson. The statistical analyses and modeling were performed in the statistical R (Ver.4.1.2) software using its functional package.

4.1. Variable Selection

Selecting the Candidate variables among the potential variables, especially consistency and self-explaining variables, is an important stage in modeling. The decision for keeping a variable in the model was based on all three following criteria [Swalha and Sayed, 2006]:

(1) The logic (i.e., +/-) of the estimated parameter should be associated with crashes, (2) The P-value for the t (z for n> 30 sample data) statistic for each parameter should be significant at the 95% confidence level, and (3) The added variable should have a minimal correlation (i.e., < 0.3) with any other independent variables in the same model. The most common correlation measure for data with normal distribution is Pearson Correlation. For abnormal distribution data, the Spearman correlation is a common measure. Multi-collinearity would greatly increase sampling variations in coefficients. Some variables (e.g., CRR, CCR, VDlu, and ΔV85) are functions of curve radius(R).

Also, some self-explaining measures are related together and have a strong correlation. In this study, the normality of variables was investigated using the Shapiro-Wilk test. The test results showed that the variables ADT, CRR, Df, R, SE.fv, and SE.sum have a normal distribution, and the normality assumption for other variables was not obtained. The Pearson correlation coefficient was used to examine the correlation between variables for the case that both have a normal distribution, and the Spearman correlation coefficient was used for other cases. A correlation matrix was developed to check the correlation between variables (Table 2). Highly correlated variables are those with a correlation value of 0.7 or higher [Sawalha and Sayed, 2006]. Variable pairs that are strongly correlated (i.e., Cp or Ep higher than ± 0.7) were considered for discarding. Specifically, there were high correlations between R and Df, TL and TL/R, LC and VD.lu, R and Δf, and SE.lr and SE.sum. After the preliminary screening of alternative consistency and self-explaining measures based on Correlation analysis and logical considerations, the following twelve main variables were selected for potential inclusion in the final model: ADT, HV, SH, G, ΔV_{85}, Δe, CRR, IRI, SE.6sr, SE.fv, and SE.lr (or only SE.sum instead of the last three variables).

4.2. Models Fitting

The study is based on using the forward stepwise procedure by considering both AIC and p-value measures (drop a covariate if it is not statistically significant) such that the candidate variable is being added to the model one at a time.

In the forward stepwise procedure, the leading exposure variable is recommended as the first variable to be added because of its dominating prediction effect [Swalha and Sayed, 2006]. Therefore, the model included exposure variables and a set of variables followed by step-by-step nominating the leading variables which are consequently ADT and the percentage of heavy vehicles in traffic (HV). Two-tailed estimated parameter coefficients with significant levels of (<10%) would suggest remaining the corresponding variables in the modeling based on what is additionally discussed in section 4.1. The result of the processing of these models outlined by using the well-known statistical software of "R" have been tabulated briefly in Tables 3, 4, 5, 6, and 7 for curves crashes.

Table 2. Correlation coefficients matrix for potential variables

	CR.fatal	ADT	HV	Df	R	LC	TL	SH	G	ΔV85	Δf	Δe	TL/R	CRR	ΔCCR	CCR	IRI	SE.6sr	S.fv	SE.lr	Se.sum	VD.lu
CR.fatal	1	0.39	0.28	-0.02	0.05	0.17	-0.16	0.02	-0.23	0.36	0.03	-0.46	-0.11	-0.41	-0.08	-0.07	0.12	-0.25	-0.48	-0.15	-0.44	-0.17
ADT	0.39	1	-0.01	0.16	-0.03	0.1	0.06	0.05	-0.21	0.12	0.12	-0.16	0.08	-0.32	-0.11	0.18	0.14	-0.24	-0.16	0.02	-0.18	-0.1
HV	0.28	-0.01	1	-0.21	0.2	0.2	0.18	0.11	-0.17	0.25	-0.08	-0.3	0.16	-0.2	-0.09	-0.11	-0.13	0.32	-0.03	0.21	0.14	-0.18
Df	-0.02	0.16	-0.21	1	-0.69	0.22	-0.04	-0.26	-0.03	0.07	0.65	0.1	0.12	-0.11	-0.22	0.5	-0.09	-0.21	-0.04	0.14	-0.01	-0.23
R	0.05	-0.03	0.2	-0.69	1	-0.08	0.01	0.3	0.22	-0.14	-0.71	-0.14	-0.23	0.03	0.37	-0.77	-0.05	0.25	0.27	0.08	0.27	0.09
LC	0.17	0.1	0.2	0.22	-0.08	1	0.26	0.03	-0.24	-0.02	0.05	-0.25	0.26	-0.07	-0.13	0	0.1	-0.13	-0.21	-0.14	-0.19	-0.99
TL	-0.16	0.06	0.18	-0.04	0.01	0.26	1	0.05	0.05	0.04	-0.01	-0.1	0.95	-0.02	0.08	-0.06	-0.16	0.01	-0.09	0.01	-0.03	-0.24
Sh	0.02	0.05	0.11	-0.26	0.3	0.03	0.05	1	-0.1	-0.09	-0.41	0	0	0.09	0.09	-0.28	-0.07	0.1	0.2	0	0.12	-0.05
G	-0.23	-0.21	-0.17	-0.03	0.22	-0.24	0.05	-0.1	1	-0.24	-0.11	0.11	-0.04	-0.01	0.3	-0.31	-0.16	-0.03	-0.02	0	0.02	0.25
ΔV85	0.36	0.12	0.25	0.07	-0.14	-0.02	0.04	-0.09	-0.24	1	0	-0.22	0.15	-0.03	-0.06	0.16	-0.03	-0.16	-0.31	-0.08	-0.27	0.02
Δf	0.03	0.12	-0.08	0.65	-0.71	0.05	-0.01	-0.41	-0.11	0.01	1	0.08	0.15	0.02	-0.42	0.61	0.1	-0.14	-0.19	0.11	-0.08	-0.06
Δe	-0.46	-0.16	-0.3	0.1	-0.14	-0.25	-0.1	0	0.11	-0.22	0.08	1	-0.13	0.18	0.22	0.1	-0.03	-0.15	0.18	-0.07	0.06	0.23
TL/R	-0.11	0.08	0.16	0.12	-0.23	0.26	0.95	0.01	-0.04	0.15	0.15	-0.13	1	-0.02	-0.03	0.11	-0.17	-0.06	-0.2	0.01	-0.12	-0.25
CRR	-0.41	-0.32	-0.2	-0.11	0.03	-0.07	-0.02	0.09	-0.01	-0.03	0.02	0.18	-0.02	1	-0.03	0.03	0.22	0.15	0.16	0	0.15	0.06
ΔCCR	-0.08	-0.11	-0.09	-0.22	0.37	-0.13	0.08	0.09	0.3	-0.06	-0.42	0.22	-0.03	-0.03	1	-0.47	-0.22	0.09	-0.09	-0.03	-0.04	0.12
CCR	-0.07	0.18	-0.11	0.5	-0.77	0	-0.06	-0.28	-0.31	0.16	0.61	0.1	0.11	0.03	-0.47	1	0.14	-0.16	-0.08	0.06	-0.09	-0.01
IRI	0.12	0.14	-0.13	-0.09	-0.05	0.1	-0.16	-0.07	-0.16	-0.03	0.1	-0.03	-0.17	0.22	-0.22	0.14	1	0.06	0.06	-0.3	-0.08	-0.09
SE.6sr	-0.25	-0.24	0.32	-0.21	0.25	-0.13	0.01	0.1	-0.03	-0.16	-0.14	-0.15	-0.06	0.15	0.09	-0.16	0.06	1	0.43	0.52	0.69	0.15
S.fv	-0.48	-0.16	-0.03	-0.04	0.27	-0.21	-0.09	0.2	-0.02	-0.31	-0.19	0.18	-0.2	0.16	-0.09	-0.08	0.06	0.43	1	0.46	0.9	0.21
SE.lr	-0.15	0.02	0.21	0.14	0.08	-0.14	0.01	0	0	-0.08	0.11	-0.07	0.01	0	-0.03	0.06	-0.3	0.52	0.46	1	0.75	0.15
Se.sum	-0.44	-0.18	0.14	-0.01	0.27	-0.19	-0.03	0.12	0.02	-0.27	-0.08	0.06	-0.12	0.15	-0.04	-0.09	-0.08	0.69	0.9	0.75	1	0.19
VD.lu	-0.17	-0.1	-0.18	-0.23	0.09	-0.99	-0.24	-0.05	0.25	0.02	-0.06	0.23	-0.25	0.06	0.12	-0.01	-0.09	0.15	0.21	0.15	0.19	1

Table 3. Poisson model for curves fatality crashes

Variable	Estimate	Standard Error	Z-value	P-value
(Intercept)	-1.97	4.43	-.44	0.66
Lin (ADT)	0.41	0.44	0.90	0.37
Lin(HV)	0.20	0.16	1.26	0.21
CRR	-0.26	0.24	-1.08	0.28
Δe	-0.06	0.04	-1.56	0.12
SE.sum	-0.01	0.009	-1.61	0.11
$X^2_{0.05}/X^2$	1.11			
AIC	148.27			
$R^2_{C\&S}$	0.310			

** Significant at the 0.05 level and * Significant at the 0.10 level

Table 4. Negative binomial model for curves fatality crashes

Variable	Estimate	Standard Error	Z-value	P-value
(Intercept)	-1.97	4.43	-.44	0.657
Lin (ADT)	0.40	0.44	0.00	0.369
Lin(HV)	0.20	0.16	1.26	0.209
CRR	-0.26	0.24	-1.08	0.280
Δe	-0.06	0.04	-1.56	0.119
SE.sum	-0.01	0.009	-1.61	0.107
$X^2_{0.05}/X^2$	1.11			
AIC	150.27			
$R^2_{C\&S}$	0.310			

** Significant at the 0.05 level and * Significant at the 0.10 level

Table 5. Zero-truncated poisson model for curves fatality crashes

Variable	Estimate	Standard Error	Z-value	P-value
(Intercept)	--2.36	5.36	-0.44	0.660
Lin (ADT)	0.46	0.54	0.86	0.392
Lin(HV)	0.27	0.20	1.37	0.170
CRR	-0.40	0.29	-1.37	0.169
Δe	-0.10	0.05	-2.004	0.045^{**}
SE.sum	-0.01	0.01	-1.82	0.068^{*}
$X^2_{0.05}/X^2$	0.68			
AIC	134.58			
$R^2_{C\&S}$	0.40			

** Significant at the 0.05 level and * Significant at the 0.10 level

Table 6. Zero-truncated negative binomial model for curves fatality crashes

Variable	Estimate	Standard Error	Z-value	P-value
(Intercept)	--2.36	5.36	-0.44	0.659
Lin (ADT)	0.46	0.54	0.86	0.393
Lin(HV)	0.27	0.20	1.37	0.169
CRR	-0.40	0.29	-1.37	0.170
Δe	-0.10	0.05	-2.003	0.045^{**}
SE.sum	-0.01	0.01	-1.82	0.068^{*}
$X^2_{0.05}/X^2$	0.68			
AIC	136.58			

Variable	Estimate	Standard Error	Z-value	P-value
$R^2_{C\&S}$	0.31			

** Significant at the 0.05 level and * Significant at the 0.10 level

Table 7.COM-poisson model for curves fatality crashes

Variable	Estimate	Standard Error	Z-value	P-value
(Intercept)	-1.92	2.30	-0.84	0.404
Lin (ADT)	0.398	0.23	1.72	0.085^{*}
Lin(HV)	0.20	0.08	2.40	0.016^{**}
CRR	-0.27	0.13	-2.13	0.033^{**}
Δe	-0.06	0.02	-2.90	0.004^{**}
SE.sum	-0.01	0.004	-3.13	0.002^{**}
$X^2_{0.05}/X^2$	0.42			
AIC	121.41			
$R^2_{C\&S}$	0.677			

** Significant at the 0.05 level and * Significant at the 0.10 level

It was found that Poisson, negative binomial, zero-truncated Poisson, and zero-truncated negative binomial distributions could not significantly fit the model for non-zero and under-dispersion data. In Poisson and negative binomial models, no exposure, consistency, and self-explaining variables appeared significantly in the model. In the zero-truncated Poisson model and the zero-truncated negative binomial model, only the variable Δe was significant in the fitted models, while even the exposure variables (i.e., ADT and HV) have not been proven to be significant. However, this study shows that the COM-Poisson regression model based on whole goodness of fit indicators can model under-dispersed zero-truncated crash data. Two consistency variables (i.e., CRR and Δe) and a self-explaining variable (i.e., SE.sum) in addition to the exposure variables (i.e., ADT and HV) appeared significantly in the model. A comparison of the fitted models based on goodness-of-fit indicators is presented in Table 8. All parameters used in the selected model were explained earlier by the authors in Table 1. Some of the variables were excluded from the modeling process for reasons such as lacking the estimated parameter’s logic (i.e., +/-) that could be declared that the variables may give inverse correlation, not because of the lack of a safety relationship, but because of limitations in the accuracy of the data obtained. This model provided good validation results. Model validity is concerned with the ability of crash models to explain the underlying phenomenon and focus on logical defensibility. In GLM models, it is not possible to use R^2 index. The validity of this model was calculated and evaluated by the Cox-Snell generalization index ($R^2_{C\&S}$) to assess fitness. The large value of this index may suggest a better fit for the model. The value of this index for the Com-Poisson and Zero-truncated Poisson models is larger than the rest. Therefore, based on this index, the Com-Poisson model has a better fit, around 0.677, which means a moderate association between predicted and observed crash frequencies Due to the absence of many components expressing the influenced factor in crashes (e.g., weather, vehicle, and driver conditions), the value of $R^2_{C\&S}$ is considered acceptable. The model with negative MPB indicates an under-prediction of the dependent variable. A value close to 0 in this index indicates that the predicted values are very close to the actual value.

Table 8. A comparison of the fitted models based on goodness-of-fit indicators

** Significant at the 0.05 level and * Significant at the 0.10 level

Evaluation Criteria	Type of Model				
	Poisson	**NB**	**ZT-Poisson**	**ZT-NB**	**COM-Poisson**
Significant variables	-	-	Δe^{**} , $SE.sum^{*}$	Δe^{**} , $SE.sum^{*}$	ADT*, HV** , Δe** ,CRR** and SE.sum**
X^2 / $X^2_{0.05}$	1.11	1.11	0.68	0.68	0.42
AIC	148.27	150.27	134.58	136.58	121.41
$\mathrm{R}^2{}_{C\&S}$	0.310	0.310	0.40	0.31	0.677
MPB	9.64 * 10^{-12}	3.98*10^{-8}	8.78*10^{-12}	3.98*10^{-8}	-0.0003
MAD	0.643	0.642	0.663	0.642	0.462
MSPE	0.656	0.660	0.690	0.660	0.560

The MAD index considers the difference between the predicted and actual values of the dependent variable as an absolute value.

A value close to 0 of this index is desirable. Also, the MSPE value close to 0 indicates that the predicted values are very close to the actual value. Fatal crash frequency was positively correlated with the model's exposure variables (ADT and HV) and was negatively correlated with consistency and self-explaining variables (Δe, CRR, and SE.sum). By increasing the difference between existing and demanded superelevation (Δe), crashes are expected to decrease.

For the alignment index CRR, as the radius of a curve is higher than the average radius, the crash frequency is expected to decrease, and vice versa. Additionally, sharp curves would give a small average radius of curvature. These curves are expected to experience higher crashes than milder curves and vice versa. A large value for this index indicates a decrease in crashes occurring.

For the self-explaining index SE.sum, as the total self-explaining score of the curves condition is increased, the crash frequency is expected to decrease. The calibrated model was analyzed for sensitivity to evaluate the effect of each mentioned factor on the overall performance.

The results of sensitivity analysis show that the maximum sensitivity curves' crashes are related to the ADT, with the HV percentage being in the second rank. A 10% reduction in ADT reduces curve crashes by 32%.

Also, a 10% reduction in HV reduces crashes by 7.4%. This result is because most heavy vehicles, especially trucks, are without any control and monitoring system for speed and maximum driving time. Regarding consistency characteristics, a 10% increase in Δe causes a 3.7% decrease in on-road curves 'crashes.

Also, regarding self-explaining characteristics, a 10% increase in SEsum causes an 8.6% in reduction in on-road curves' crashes.

5. Conclusions and Recommendations

This paper presented the effect of some recognized and new candidate variables of horizontal curves on crash frequency in four-lane intercity highways for 45 curves with fatal crash history among the 15 provinces of Iran using zero-truncated crash data.

Up to now, studies about road design consistency have focused only on rural two-lane highways since these highways have higher crash rates with considerable inconsistencies. The variables of road self-explaining have not been used in road safety modeling. Based on the obtained results, the main conclusions of this research are as follows:

- Crash data for this research included zero-truncated and under-dispersion count data.

Generalized linear regression modeling (GLM) approach was adopted for model development. The application of various generalized linear regression (i.e., Poisson, Negative binomial, Zero-truncated Poisson, Zero-truncated Negative binomial, and Conway-Maxwell Poisson techniques) was investigated. Overall, Conway-Maxwell Poisson regression was considerably more suitable for under-dispersion zero-truncated dependent data.

- A relationship exists between geometric design consistency and curve self-explaining and safety on four-lane intercity highways. Validation of the model indicates that the fitted model experienced an acceptable level of goodness of fit and can be used in identifying the potential of four-lane highway curves' crashes and prioritizing them. The value $R^2_{C\&S}$ of this model was 0.677, which is relatively acceptable, especially considering the complicated nature of the crash occurrence. This result suggests that this model accounts for a large proportion of the variability in crash frequencies on road curves.

- The ultimately developed model showed that Δe and CRR, as the main design consistency variables, have significant impacts on four-lane intercity highways safety for alignment indices and vehicle stability index, respectively.

- It was found that four-lane intercity highways constructed by widening the previous two-lane roads in Iran need to be modified to ensure vehicle stability consistency in terms of adjusting the superelevation of curves.

- The variable of the total self-explaining index (SE.sum) showed statistically significant modeling results. Overall, increasing road self-explaining is an effective approach for improving curves' safety on four-lane intercity highways. Therefore, a low-cost approach using self-explaining measures is recommended to improve traffic safety on four-lane intercity highway curves.

- Highway designers should pay special attention to inconsistent and non-self-explaining designs of four-lane intercity highways during the widening of two-lane highways to reduce crash frequencies. This model can be used in redesigning and improving existing four-lane intercity highways, designing new highways, and identifying accident hotspots.

The final recommended model is still in the preliminary stage, and more work is needed to modify and develop for its consistency and self-explanatory criteria.

For future studies, it is suggested to explore and model a combination of curve self-explaining and consistency in modeling intercity two-lane highway crashes. More work and research should be done on highway sections with other geometric features such as vertical curves, inverse, and compound curves, access, and sight distances.

6. Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

7. Acknowledgments

The authors would like to acknowledge IRAN's Road Maintenance and Transportation Organization (RMTO) for supporting data acquisition.

8. References

- Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control. 19(6):716–772.

- Al-Masaeid, H. R., Hamed, M., Aboul-Ela, M., and Ghannam, A. G. (1994). Consistency of horizontal alignment for different vehicle classes. Transp. Res. Rec. 1500, Transportation Research Board, Washington, D.C., 178–183.

- Al-Sahili, K., and Dwaikat, M. (2019). Modeling geometric design consistency and road safety for two-lane rural highway in the West Bank, Palestine. Arabian Journal for Science and Engineering, 44(10), 4895-4909.

- Anderson, I.B., Bauer, K.M., Harwood, D.W., and Fritzpatrick, L.(1999). Relationship to the safety of geometric design consistency measures for rural two-lane highways.J. Transp. Res. Board 1658, 43–51.

- Bargh, J. A., and Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. Psychological Bulletin, 126(6), 925–945.

- Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., and Persaud, B. (2010). Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency, and context variables. Accid. Anal. Prev. 42, 1072–1079.

- Cafiso, S., La Cava, G., and Montella, A.(2007). Safety index for evaluation of two-lane rural highways. Transportation Research Record, 2019(1), 136-145.

- Charlton, S. G., and Starkey, N. J. (2013). Driving on familiar roads: Automaticity and inattention blindness. Transportation Research Part F: Traffic Psychology and Behaviour, 19, 121–133.

- Cohen, A.C. (1960). Estimating the parameter in a conditional Poisson distribution. Biometrics, vol. 16, no. 2, 203–211.

- Conway, R.W., and Maxwell, W.L.A. (1962). Queuing model with state-dependent service rates. Journal of Industrial Engineering. 12, 132-136.

- Cox, D.R. and E.J. Snell.(1989). Analysis of Binary Data. Second Edition. Chapman & Hall.

- Cummings, P. (2009). Methods for estimating adjusted risk ratios, The Stata Journal. Vol. 9, No. 2, 2009, 175–196.

- De Ona, J.D., and Garach, L.,Calvo ,F., and Garcia-Munoz,T. (2013).Relationship between predicted speed reduction on horizontal curves and safety on two-lane rural roads in Spain. Journal of Transportation Engineering,140(3),04013015.

- Dhahir, B., and Hassan, Y. (2017). Relationship between traffic safety and speed differential on horizontal curves based on naturalistic driving studies, In Proc. Road Safety & Simulation International Conference, Paper (Vol. 58).

- Eenink, R., Reurings, M., Elvik, R., Cardoso, J., Wichert, S., and Stefan, C. (2007).Accident prediction models and road safety impact assessment: recommendations for using these tools. Report RI-SWOV-WP2-D2-F. SWOV Institute for Road Safety Research, Leidschendam,4p.

- Fitzpatrick, K., Elefteriadou, L., Harwood., D.W, Collins, J.M., McFadden.J., Anderson, I.B., Krammes, R.A., Irizarry, N., Parma, K.D., Bauer, K.M., and Passetti, K . (2000a). Speed prediction for two-lane rural highways. Federal Highway Administration, Report FHWA-RD-99-171, Springfield, Virginia.

- Fitzpatrick, K., M. Wooldridge, O. Tsimoni, J. Collins, P. Green, K.Bauer, K. Parma, R. Koppa, D. Harwood, I. Anderson, R. Krammes, and B. Poggioli. (2000b). Alternative design consistency rating methods for two-lane rural highways. FHWA Report FHWA-RD-99-172. Springfield,Va.

- Global status report on road safety.(2018).World Health Organization.

- Gooch, J. P., Gayah, V. V., and Donnell, E. T. (2016).Quantifying the safety effects of horizontal curves on two-way, two-lane rural roads. Accident Analysis and Prevention, 92, 71-81.

- Greene, W. (2008). Functional forms for the Negative Binomial model for count data, Economics Letters.Vol. 99, 585–590.

- Hamilton, I., Himes, S., Porter, R.J., and Donnell. (2019).Safety evaluation of horizontal alignment design consistency on rural two-lane highways. Transportation Research Record,1-9.

- Hassan, Y., T. Sayed, and P. Tabernero. (2001). Establishing a practical approach for design consistency evaluation. Journal of Transportation Engineering, Vol. 127, No. 4, 295–302.

- Hills, B. L. (1980). Vision, visibility, and perception in driving. Perception, 9,183–216.

- Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). Univariate discrete distributions.Wiley.

- Khan, G., Bill, A. R., Chitturi, M., and Noyce, D. A. (2012). Horizontal curves, signs, and safety. Transportation research record, 2279(1), 124-131.

- Krammes, R. A., Rao, K. S., and Oh, H. (1995). Highway geometric design consistency evaluation software. Transp. Res. Rec. 1500, Transportation Research Board, Washington, D.C., 19–24.

- Lamm, R., Choueiri, E.M., and Mailaender, T.(1991).Side friction demand versus side friction assumed for curve design on two-lane rural highways. Transportation Research Record, 1303: 11-21.

- Lamm, R., Psarianos, B., and Mailaender, T. (1999). Highway design and traffic safety engineering handbook, McGraw-Hill, New York.

- Liopis-Castello, D., Bella, F., Camacho-Torregrosa, F.J., and Garcia, A.(2018). New consistency model based on inertial operating speed profiles for road safety evaluation. Journal of Transportation Engineering, Part A: Systems, 144(4).

- Luque, R., and Castro,M. (2018). Highway geometric design consistency speed models and local or global assessment.Int.J.Civ.Eng. 14(6),347-355.

- Martens, M. H., and Fox, M. R. J. (2007). Do familiarity and expectations change perception? Drivers glances and responses to changes. Transportation Research Part F, 10, 476–492.

- Miller, I., and Freund, J.E. (1977). Probability and statistics for engineers, 2nd ed. Prentice-Hall, Englewood Cliffs, N.J.

- Ng, J. and Sayed, T. (2004). Effect of geometric design consistency on road safety. Canadian Journal of Civil Engineering, 31(2), 218-227.

- NHTSA. (2016). Traffic safety facts. National highway traffic safety administration, National Center for Statistics and Analysis, U.S. Department of Transportation, Washington, DC 20590.

- Nicholson, A. (1998). Superelevation, side friction, and roadway consistency. J.Transp.Engrg.,ASCE,124(5),411- 418.

- Oh, J., Lyon, C., Washington,SP., Persaud ,BN., and Bared, J. (2003). Validation of the FHWA crash models for rural intersections: Lessons learned. Transportation Research Record 1840,41–49.

- PIARC. (2016). Human factors guidelines for a safer man-road interface. Technical Committee C.2 Design and Operation of Safer Road Infrastructure.

- PIARC. (2019). Road safety evaluation based on human factors method. Technical

Committee C.2 Design and Operation of Safer Road Infrastructure.

- Polus, A. (1980). The relationship of overall geometric characteristics to the safety level of rural highways. Eno Transportation Foundation, Traffic Quarterly, 34(4): 575-585.

- Saffarzadeh,M., Shabani,S., and Azarmi,A. (2007). Accident prediction model in two-way two-lane highway curves, Bulletin of Transportation (Pajouheshnameh Haml va Naghl, IR), 4(3),213-221.

- Sawalha, Z. and Sayed, T.(2006). Traffic accident modeling: some statistical issues. Can. J. Civ. Eng. 33(9), 1115–1123.

- Schneider, W. H. , Savolainen, P. T. and Moore, D. N. (2010). Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways. Transportation Research Record: Journal of the Transportation Research Board, vol. 2194, no. 1, 91-98.

- Schneider, W. H. , Zimmerman, K. , Van Boxel, D. and Vavilikolanu, S. (2009). Bayesian analysis of the effect of horizontal curvature on truck crashes using training and validation data sets. Transportation Research Record: Journal of the Transportation Research Board, vol. 2096, no. 1, 41-46.

- Sellers, K.F., and Shmueli,G. (2010). A flexible regression model for count data. The Annals of Applied Statistics. 4, Issue 2, 943-961.

- Sellers,K.F., and Shmueli, G. (2013). Data dispersion: Now you see it...Now you don't, Communication in Statistics: Theory and Methods. 42, Issue 17, 3134-47.

- Senders, J. W. The estimation of operator workload in complex systems. In Systems Psychology, McGraw-Hill, New York. (1970). 207–216.

- Shmueli,G., Minka, T.P., Kadane, J.B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: Revival of the Conway–Maxwell–Poisson distribution. Journal of The Royal Statistical Society. Series C (Applied Statistics). 54, Issue 1, 127-142.

- Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? Perception, 25, 1081–1089.

- Stadtler, T. (1998). Lexikon der psychologie. Alkfred Kroner Verlag, Stuttgart.

- Taylor,j.R. (1982). An introduction to error analysis university science books.

- Theeuwes, J., and Godthelp, H. (1995). Self-explaining roads. Safety Science, Vol. 19, No. 2–3, 217–225.

- Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hume, R. D., Mayer, R. E.,Stanisfer, R. L., and Castellan, N. J. (1977). Tri-level study of the causes of traffic accidents. Report No. DOT-HS-034-3-535-77(TAC).

- Wimmer, G. and Altmann, G. (1999). Thesaurus of univariate discrete probability distributions. Mathematica Slovaca, Vol. 49, No. 5, 599-600.