Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Safety Performance Functions for Fatal Crashes on National Highways in India

Hasan Mehdi Naqvi[a], Geetam Tiwari[b]

[a]Chief General Manager(Tech.), National Highways Authority of India, G-5 & 6, Sector-10, Dwarka, New Delhi-110075, India
[b]Professor, Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi-110016, India

Abstract

This paper investigates factors contributing to fatal crashes on three National Highways (NHs) sections in India by developing safety performance functions. The generalized linear model technique is used for analyzing linear and non-linear effect of continuous and categorical predictor variables on discreet dependent variable (fatal crashes) separately for each NH segments. In India, NHs are not usually access controlled, and heterogeneous vehicles travel on highways. The probable explanatory variables are short-listed after thorough literature review, and availability of data. These variables comprise of vehicular traffic (ADT), highway elements, and roadside land use. The fatal crash data for the past five years (2009-2013), traffic and highway inventory data have been collected for three NHs having varying lane configuration: two-lane NH-8, four-lane NH-24 and six-lane NH-1. The study results revealed negative binomial regression model fit the data statistically, and also identified number of statistically significant variables (‘segment length’, ‘roadside land use’, ‘presence of service road (SR)’ and ‘terrain type’) to estimate fatal crashes at NHs segments. The results of the safety performance functions (SPFs) showed that out of seven explanatory variables examined for each NH (segments), the significant explanatory variable is found to be ‘segment length’ in km for all three models of NHs (segments). Other significant variable is ‘land use’ along NHs for both two-lane NH-8 and four-lane NH-24. Similarly, the explanatory variables ‘presence of SR’ and ‘terrain type’ are found significant for four-lane NH-24 and two-lane NH-8 respectively.

Keywords: Fatal crashes; National Highways; negative binomial regression; land use; access road(s);

1. Introduction

In India, road crashes led to fatalities of 150,785 persons (MoRTH, 2017). National Highways (NHs), accounted for a share of 35% (52,075) in total number of persons killed in road crashes during 2016 (MoRTH, 2017). NHs constitute about 2% (101,011 km as on March 2016) of total road network (5.603 million km) in India (MoRTH,

2352-1465

2017b), but cater to about 40% of road traffic as cited by Mohan et al. (2009). The share of two-lane NHs or less than two-lane NHs is 75% of total NHs length in India (MoRTH, 2017b).

NHs are primary arteries running through length and breadth of the country, and provide connectivity to state capitals/major ports/district centres and so on. NHs in India are not access controlled and passes through towns and villages. These high-speed highways cater to travel needs of both short and long distance trips by villagers and intercity passengers/goods traffic respectively owing to absence of full-length service road (SR) on both side of highways (Naqvi and Tiwari, 2018).

To improve highway infrastructure in India, upgradation of NHs from single- to two-lane, two- to four-lane and so on have been taking place for the past few years. Despite improvement in accessibility and mobility of passengers and freight through construction of new NHs and upgradation in existing highway infrastructure (MoRTH, 2017) in terms of highway capacity, riding quality, vehicles speed, road connectivity, road furniture, wayside amenities for road users in the past, there is no reduction in number of both fatal road crashes and fatalities on NHs.

With the increase in road crashes in particular fatal, it becomes necessary to carry out diagnostic studies to identify causes of crashes preferably with respect to elements of roads (highways) as these are the most controllable than other elements such as, vehicle and driver (Vogt and Bared, 1998). Therefore, road safety modelling or development of safety performance functions (SPFs) is gaining importance for couple of decades mainly because of its ability to forecast traffic crashes, and thereby assist the concerned agencies/highway engineers to take suitable remedial measures at identified crash spots.

The study objectives are to examine fatal crash pattern on studied NHs (two-, four- and six-lane); and, also to develop SPFs for fatal crashes for studied NHs sections suitable for prevailing heterogeneous vehicular traffic (including pedestrians) and roadside land use in India.

Subsequent sections of the paper present summary of previous research on crash prediction model for rural roads. An overview of generalised linear model (GLM) with Poisson and negative binomial (NB) regression modelling approach is presented to establish relationship between dependent variable and possible explanatory factors. Thereafter, descriptions of the database, fatal crash data analysis and selected variables are discussed. Subsequently, results of the SPFs developed using NB regression for two-, four- and six-lane NHs sections are discussed. Finally, study inferences and limitations are discussed.

2. Literature review

Numbers of research papers in the past have discussed various approaches to develop SPFs for rural roads. However, literature review revealed that majority of researchers (Miaou and Lum, 1993; Milton and Mannering, 1998; Abdel-aty and Radwan, 2000) employed three statistical approaches: multiple linear regression, Poisson and NB regression. The multiple linear regression models lack the distributional property to describe adequately random, discreet, non-negativity, sporadic vehicle crash events (Miaou and Lum, 1993) and makes assumption of a normal distribution for the dependent variable (Abdel-aty and Radwan, 2000; Cafiso et al., 2010) and violation of the homoscedasticity assumption (Milton and Mannering, 1998).

Milton and Mannering (1998) study found that an increase in section length, vertical grade, average annual daily traffic (AADT) and more lanes tend to increase crash frequency. Whereas higher peak hour percentages, an increase in the percentage of all trucks, and high posted speed limits have potential to decrease crash frequency. Abdel-Aty and Radwan (2000) study results showed that heavy traffic, speeding, narrow lane width, larger number of lanes, urban roadways, narrow shoulder width, and reduced median width increase the likelihood of crash involvement.

Persaud and Mucsi (1995) developed crash prediction models for two-lane rural roads for different combinations of time periods (24 hours, day hours and night hours) and geometric (roadway and shoulder width) characteristics. Authors concluded that for single-vehicle crashes, the crash potential was higher during night, whereas for multi-vehicles crashes the opposite was true.

Hadi et al. (1995) quantified the effects of cross-section design elements on total, fatality, and injury crash rates for various types of rural and urban highways of Florida, USA. The study results pertaining to rural highways revealed that i) crash frequency increases with higher AADT for all highway types investigated, ii) The two-way, two-lane rural highways with same AADT have the highest crash rates, iii) increasing unpaved shoulder width to decrease crash rates on two-lane rural highways and iv) significant reduction in crash rates could be expected from greater median width for all highway types. Cafiso et al. (2010) study identified parameters for safety were AADT, driveway density, curve ratio, roadside hazard, and speed differentials density.

Miaou (1994) evaluated the performance of the Poisson and NB regression models in establishing relationship between truck crashes and geometric design of road sections. Author recommended that NB, zero inflated Poisson (ZIP) and Poisson models performed the best in estimating the frequency of road sections with zero, one to three and four or more truck accident involvement respectively.

Noland (2003) has examined effect of changes in highway infrastructure on traffic fatalities and injuries in US states. The study results showed that infrastructure improvements were not effective at reducing total fatalities and injuries.

Mountain et al. (1996) used generalized linear model to predict crashes on main roads with minor junctions where traffic counts on minor approaches were not available. Their study results showed that crashes on highway link were not proportional to exposure (traffic volume and link length). In addition, authors stated that the presence of minor junctions had an important influence on link crash frequencies.

Vangala et al. (2015) discussed the application of the NB-Generalized Exponential (GE) distribution model with excess zero counts (over-dispersed data). The study results showed that NB–GE and NB–Lindley distributions fit well with excess zeros and over-dispersed datasets better than the NB distribution.

Hauer (2004) discussed issues pertaining to statistical road safety modelling (SRSM) and suggested few ways for improving the SRSM process. Author has suggested consistency of results as a necessary condition for making inferences about cause and effect. Author concluded that the effect of variables that influence the probability of crash occurrence along significant portions of a highway segment was better represented by multiplicative than additive factors and most phenomena related to safety are non-linear. Author also mentioned that the count of crashes obeys NB distribution.

From the review of literature, it is established that count regression models are appropriate to predict the crash frequency for given explanatory (road geometry and traffic) variables. However, most of these crash prediction models have been developed for countries where traffic characteristics and roadside land use, and traffic enforcement measures are observed to be different than observed in India or developing countries. In India, heterogenous vehicles travel on NHs (Naqvi and Tiwari, 2018), which includes pedestrians, non-motorized/motorized vehicles (slow/fast, and light/heavy). The presence of heavy roadside land use (villages/industries/towns/educational institutes) along NHs at regular interval (say, 3-5 km), heterogeneous traffic, high vehicle speed (above 80 km/hr) and partial access controlled NHs in India have led to large share of crashes especially fatal on highways, which include pedestrians and two-wheelers occupants. Moreover, it is noticed through literature review that SPFs or crash prediction models were developed generally either for two-lane or multi-lane highways. In this study, an attempt is made to develop SPFs for fatal crashes for three NHs (segments) with varying lane configuration to understand variation in fatal crash characteristics and significant variables for each NH lane configuration.

The study is limited to fatal crashes due to under-reporting of non-fatal crashes in India (Mohan et al., 2009). Further, fatal traffic crashes have severe implications with regard to human loss to victims' family and society. Hence, it is important to develop SPFs for fatal crashes on NHs to assist highway engineers/planners to take suitable remedial measure to improve safety. Secondary data pertaining to vehicles speed on cited NHs was not available, and hence, the same is not used in this study.

3. Data

Data comprise of sections of three NHs having varying lanes configuration, namely, NH-8: two-lane undivided carriageway (c/w); NH-24: four-lane divided c/w; and, NH-1: six-lane divided c/w in three states in India. The division of c/w for each travel direction for NHs is generally done through raised median with width 2.5 m or more. It is mentioned that these three NHs are short-listed for the study after ensuring that no geometric changes (change in number of lane/horizontal curvature/vertical gradient/availability of raised median/paved shoulder and change in width of lane/raised median/paved shoulder) were made on these highways during the study period.

In India, the respective state police department reports crashes in First Information Reports (FIRs) (Bhalla et al., 2017). Crash FIR data fields include crash date, day, crash time, crash location, brief description of crash, and so on. The inputs for data fields are obtained from the fatal crash victim's relative or friend or eyewitness. Fatal crash data, namely, FIRs for the sections of NHs are collected for the past five years from the respective police stations. Highway inventory and roadside land use data are collected during field visits. Average daily traffic (ADT) for the studied NHs are collected from the concerned department. Total 1,534 fatal crash FIRs are examined.

The breakup of FIRs for each NH (FIRs number and period), NH details (NH number, length, number of lanes, and section chainage-start and end kilometre), and ADT in vehicles are presented in table 1. The highest ADT is recorded as 51,085 at km 52.65 on six-lane NH-1 during 2012.

Table 1. Summary details of national highways, traffic and fatal crash first information reports.

National Highway number, Lanes (type of c/w, traffic direction)	Section Length Km (mile)	Section Chainage$	Total NH length in km* (mile)	Average Daily Traffic	Fatal Crash FIRs	
				Vehicles; chainage; year	period	number
NH-8 Two-lane (undivided, two-way)	98 (60.9)	km 64 to km 162	1375 (854)	2,934^; at km 84.0; 2013**	2009-2012	355
NH-24 Four-lane (divided, two-way)	30 (18.6)	km 7 to km 37	438 (272)	46,836, at km 29.3; 2014!	2010-2013	290
NH-1 Six-lane (divided, two-way)	57 (35.4)	km 29 to km 86	456 (283)	51,085 at km 52.65; 2012!	2009-2013	889

*Source: http://morth.nic.in/showfile.asp?lid=365 accessed on 11/07/2015; **Source: Ministry of Road Transport &Highways, New Delhi. !Source: National Highways Authority of India;
^ includes non-toll-able and toll exempted traffic, such as army vehicle, motorcycle, bicycle, bullock-cart, three-wheeler rickshaw, etc.
$ Section chainage mentioned against each NH presents portion of the National Highway selected for the study. Chainage (start and end) in kilometres are given by the respective Highways Department.

4. Data analysis

4.1. Fatal crashes per km (NHs length) per year

The fatal crashes per km (NHs length) per year for three NHs sections are computed. The results revealed that fatal crashes/km/year is found to be highest (3.08) on six-lane NH-1, followed by 2.42 on four-lane NH-24, and 0.72 on two-lane NH-8 during the study period (Naqvi and Tiwari, 2018).

4.2. Distribution of fatal crashes by collision type

Fig. 1 shows share of fatal crashes (in per cent) by collision type on each of the studied NHs during the study period (Naqvi and Tiwari, 2018).

The share of collisions involving pedestrians (hit pedestrians) is observed highest (45%) on six-lane NH-1, followed by 34% on four-lane NH-24 and 19% on two-lane NH-8 (Fig. 1). High share of pedestrians in fatal crashes on four- and six-lane NHs could be attributed to close proximity to Delhi for part sections of these two NHs, having sizable residential (existing and upcoming projects) and industrial development along these highways (Naqvi and Tiwari, 2018).

The share of ‘head-on’ collision type is found highest (33%) on two-lane (undivided c/w), followed by 9% on six-lane NH-1(divided c/w) and 6% on four-lane NH-24 (divided c/w) NH-24 (Fig. 1). The high percentage share of ‘head-on’ collision type on six-lane NH-1 (9%) and on four-lane NH-24 (6%) could be inter-alia attributed to plying of some vehicles in opposite lane owing to absence of SR on either side of NHs or all along the studied NHs. It is also witnessed during the field visits of cited NHs that drivers often violate traffic rules by travelling vehicles in the opposite direction to reach nearby destination (roadside eateries, vehicle garage, fuel station and so on) through shorter path (Naqvi and Tiwari, 2018). The other reason for ‘head-on’ collisions on divided c/w (four-lane NH-24 and six-lane NH-1) could be due to high-speed vehicles crossing the raised median and colliding with vehicles travelling in the opposite direction.

Futhermore, fig. 1 depicts the percentage share of ‘rear-end’ collisions is observed to be varying from 23% (two-lane NH-8) to 52% (four-lane NH-24) of fatal crashes. The high percentage of 'rear-end' collisions could be inter-alia due to unauthorized parking of trucks and buses along roadside eateries or to carry out minor repair of vehicle(s) say, change of tyre on roadside without switching on parking light indicators of stationary vehicles (Naqvi and Tiwari, 2018).

The share of collision involving 'sideswipe' is observed to be to the tune of 6% on two-lane NH-8, 2% on four-lane NH-24 and 9% on six-lane NH-1 (fig. 1). Majority of these crashes occur at or near intersections or at median openings (Naqvi and Tiwari, 2018).

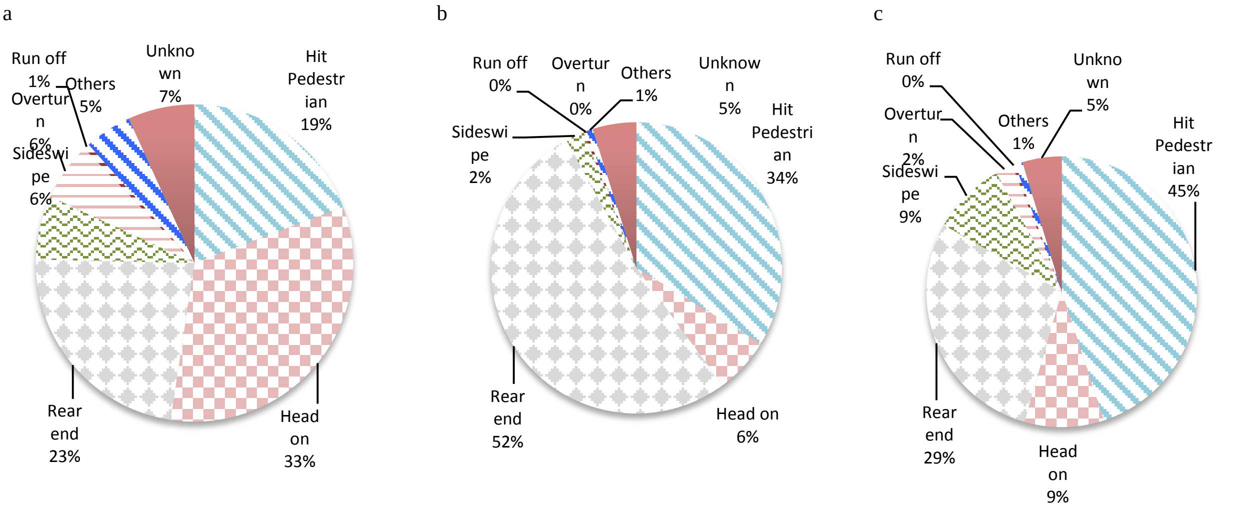

Fig. 1. Distribution of fatal crashes (in per cent) by collision type (a) two-lane NH-8; (b) four-lane NH-24; (c) six-lane NH-1. (Naqvi and Tiwari 2018).

5. Methodology

To develop relationship between fatal crash frequency and explanatory variables (highway elements, geometric features, ADT and land use), the modeling techniques, namely, Poisson and NB regression are proposed to be used. These models are suitable for non-negative and discrete count data, such as crash frequency than linear regression (Persaud and Mucsi, 1995; Mountain et al., 1996; Milton and Mannering, 1998; Washington et al., 2011). To begin with, Poisson regression modeling is used for developing relationship between fatal crash frequency for a chosen period of time and the number of explanatory variables. A characteristic of the Poisson distribution is that its mean is equal to its variance. It is also limitation of the model. Poisson model (Milton and Mannering 1998) is defined as below:

$$P(n_i)=[\lambda_i^{ni} \exp(\lambda_i)] / n_i! \quad (1)$$

Where $P(n_i)$ represents probability of n fatal crashes on a highway section i, and λ_i is the expected fatal crash frequency for highway section i. For applying Poisson regression model, the expected crash frequency is assumed to be a function of independent variables such that:

$$\lambda_i = \exp(\beta X_i) \quad (2)$$

Where λ_i is the expected fatal crash frequency for a chosen period of time for highway section i,

Xi is a vector of explanatory variables which include highway elements, geometric features, traffic characteristics and land use of highway section i, and β is a vector coefficient to be estimated by maximum likelihood methods.

In case of the NB regression model (Milton and Mannering, 1998), the expected fatal crash frequency for a chosen period of time for highway segment i is assumed to be function of explanatory variables such that:

$$\lambda_i = \exp(\beta X_i + \epsilon) \quad (3)$$

Where λ_i is the expected fatal crash frequency for a chosen period of time at highway section i, say E [ni] where n refers to number of crashes,

Xi is a vector of explanatory variables,

ϵ is an error term, and,

$\exp(\epsilon_i)$ is a gamma-distributed error term with mean one and variance α.

X_i may include the highway elements, geometric features, traffic characteristics and land use of highway section i which is used to determine fatal crash frequency and β is vector of estimate coefficients. Using the form of λ_i in equation (3), the coefficient vector β can be estimated by standard maximum likelihood methods with the likelihood function $L(\beta)$. The cited model permits the mean to differ with the variance such that:

$$\text{Var}\,[n_i] = E\,[n_i]\,[1 + a\,E\,[n_i]] \quad (4)$$

Generalized linear model is used to estimate the model coefficient using SPSS Statistics 19 software. Firstly, Poisson distribution is assumed for developing model. Models are fitted by specifying the dependent variable, explanatory variables and the error structure Poisson.

Subsequently, NB error distribution is assumed. Models are fitted by specifying the dependent variable, explanatory variables and the error structure NB with the link function (log). Model parameters are estimated using maximum likelihood approach. Explanatory variables are added to the model one by one using the forward procedure (Hauer 2004). The results of both Poisson regression and NB regression models with respect to goodness-of-fit statistics are reviewed before adding new variable to ensure that the model fits reasonably well with the data.

6. Model estimation

6.1 Division of NHs into segments

The division of the road into number of segments is required to assess impact of road elements/traffic/land use on fatal crashes. In the past, researchers (Okamoto and Koshi, 1989) developed crash prediction models using both fixed length and homogenous road segments (variable segment length) mainly with respect to geometric elements and traffic characteristics. The segment length in homogenous segments of road may vary. Okamoto and Koshi (1989) explained how to divide a road stretch into segments for crash modeling, and suggested to ensure that the random errors of the segments to be as much as possible equal to each other and small enough compared with accident rates.

Based on the literature review, it is observed that to study fatal crash pattern and characteristics as well as developing SPFs for fatal crashes, majority of researchers have divided portion of the identified highway/road corridor into homogenous segments based on number of factors. Literature review revealed that Hadi et al. (1995) considered minimum segment length of 80 m (0.05 mile) to exclude short sections for developing SPFs using NB regression. Similarly, Chen and Wang (1999) used variable segment length (ranges from 0.08 km to 4.43 km) for developing relationship between accident types and highway geometric features for an access controlled San Yat-Sen National freeway facility in Taiwan. Knuiman et al. (1993) considered homogenous highway segments (average section length 1.6 km) for developing a log-linear regression model to examine the effect of median width on the frequency and severity of accidents. Cafiso et al. (2010) divided the roads to segments with its length varying from 0.50 km to 4.29 km for developing models. Garber and Ehrhart (2000) considered road segments between two major junctions to ensure homogenous traffic flow characteristics. Donnell and Mason Jr. (2006) developed model median barrier crash frequency model, and considered homogenous segments (each nominally half mile long) with respect to traffic volume and geometric design characteristics.

Having reviewed literature on SPFs, it is observed that number of researchers (Abdel-Aty and Radwan, 2000; Chen and Wang, 1999) have considered homogenous road segments for developing crash prediction models. Homogenous segments (variable segments length) are more appropriate for crash prediction modeling as it ensures uniformity in variable values within a given segment. Hence in this study, homogenous segments excluding intersections are considered with respect to road elements (number of lanes, presence of raised median, horizontal curvature, number of access roads, presence of SR and terrain), traffic and roadside land use characteristics for two-lane NH-8, four-lane NH-24 and six-lane NH-1. Fig. 2 illustrates graphically division of three segments (segment 1, segment 2, and segment 3) in a curved portion of road along with change in lane-width (number of lanes). Accordingly, in this study each NH segment is a divided portion of the studied NHs sections.

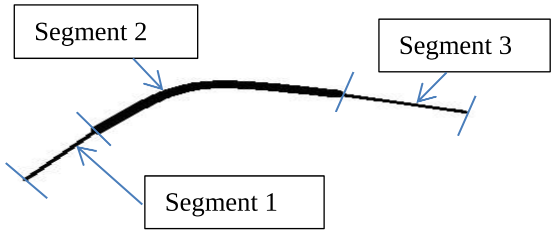

Fig. 2. Illustration of division of segments in a curved portion and number of lanes.

6.2 Determination of models variables

To identify the independent variables influencing fatal crashes, variables are identified through literature review, available information in the fatal crash FIRs for the studied NHs and other data. Based the identified explanatory variables, SPFs for fatal crashes on NHs segments are developed.

6.2.1 Segment length

In this study, NHs segments are defined by any change in road elements, roadside land use or other variables (for example, a new segment is identified when raised median is introduced in undivided two-lane NH or number of lanes changes from two/four to four/six or major bridge approach starts/ends). Therefore, each NHs segment is uniform with respect to all studied features of road, traffic and roadside land use recording during highway inventory and as per secondary database.

Elevated NHs segment(s) (flyover/rail over bridge) has/have not been considered in this study owing to significant variation in characteristics and owing to non-clarity about fatal crash locations (either on SR or on elevated highway portion) in fatal crash FIRs. Three highway sections on NH-1 are found to have elevated roads, and these sections have not been included in the study for the cited reasons.

The studied NHs segments are divided: 77 segments on two-lane NH-8, 33 segments on four-lane NH-24 and 52 segments on six-lane NH-1. Segment length of the three studied NHs ranges: two-lane NH-8 (from 0.15 m to 2.4 m), four-lane NH-24 (from 0.15 m to 1.25 m) and six-lane NH-1 (from 0.3 m to 1.85 m). The minimum segment length of studied NHs is kept 1.0 km (0.621 mile), which is in line with the previous research (Chen and Wang, 1999; Cafiso et al., 2010). Researchers (Abdel-Aty and Radwan, 2000; Cafiso et al., 2010; Ackaah and Salifu, 2011) have studied impact of segment length as exposure variable for crash modeling. Accordingly, ‘segment length in km’ is considered as an independent variable in this study.

6.2.2 Average daily traffic

Vehicular traffic including pedestrians’ movement on highways influences fatal crashes. Hence, to assess the impact of traffic on fatal crashes, average daily traffic in vehicles are considered for developing SPFs. Traffic count data obtained at toll plaza locations on each studied NHs are considered. Therefore, ‘ADT (in vehicles)’ is considered as an independent variable in this study. This is in line with other researchers (Abdel-Aty and Radwan 2000; Cafiso et al. 2010) who have also used ADT or AADT (vehicle/day) as exposure variable for fatal crash modeling. ADT (in vehicles) for studied NHs are collected at respective toll plaza locations of NHs for a particular period.

6.2.3 Number of access roads

The access roads or crossroads or driveways are common features of highways, and they provide direct access to nearby district/town/village road(s) to highways. The crossroads mentioned here are those roads where SR(s) with proper acceleration (merging) and deceleration (demerging) lanes with NHs are not available. Owing to absence of SR(s) along highway(s), the presence of access roads affects highway safety significantly. At crossroads locations, the absence of raised median openings for four- and six-lane (divided c/w) NHs are ensured to identify them. In case at crossroad locations, raised median openings at NHs are found, then these locations become road junctions. Depending on the land use along highways and presence of village(s)/town(s), number of crossroads along highways varies. Hence, in this study, variable 'presence of access road' along NHs is considered as an independent variable. Numbers of researchers (Cafiso et al., 2010; Ackaah and Salifu, 2011) have studied impact of number of driveways or access density on crashes.

6.2.4 Presence of service road

SRs of 5.5 to 7.5 m width are usually provided along NHs especially for four- and six-lane NHs to cater to local traffic safe movement near urban areas/towns, and it also allow safe merging/demerging to local vehicular traffic from NHs. Therefore, SRs for short length (2-3 km) are intermittently provided on NHs. The provision of SRs along NHs for short-distance changes road features. In order to capture influence of SRs on fatal crashes, 'presence of SR' along a given highway segment has been considered as variable for developing fatal crash prediction model in this study.

6.2.5 Presence of horizontal curve

The presence of horizontal curve on highways is considered to have influence on fatal crashes. Accordingly, this data was captured from the design plans (strip plans) of the respective NHs. The strip plan of each NH shows highway chainage-wise geometric details such as number of lane, median width, curvature: horizontal and vertical, and so on. The horizontal curvature values at the studied NHs are observed varying: two-lane NH-8 (115 – 2,500 m) and four-lane NH-24 (650-10,000 m). The data pertaining to horizontal curvature for six-lane NH-1 is not found in the design drawings. Researchers (Milton and Mannering, 1998; Ackaah and Salifu, 2011) have also used number of horizontal curves per mile or within segment. Few researchers (Abdel-Aty and Radwan, 2000) have used degree of horizontal curves (degree/100 m arc) for developing models. Researchers (Cafiso et al., 2010; Ackaah and Salifu, 2011) have studied impact of presence of horizontal curves on crashes. Hence, 'presence of horizontal curve' is included as independent variable in this study.

6.2.6 Land use

'Land use' along the highways is considered an important variable in Indian context, as it generates local traffic including pedestrian traffic. NHs also cater to local traffic (short-distance) in the absence of SRs. As in India, villages/towns/industries along NHs are usually located within 2-5 kms. Owing to the above and non-availability of full-length access controlled NHs in India lead to large number of crashes and majority of crash victims are vulnerable road users (pedestrians/bicyclists/two-wheelers).

In order to capture impact of the 'land use' along NHs on fatal crashes, villages/towns/industries located within 100 m lateral distance of the studied NHs (50 m on either side of the centerline of NHs) are considered for the study. The variable 'land use' along NHs is categorized into two categories: sparse (agriculture/forest/barren land) and heavy (residential/industrial/commercial) settlement(s). Furthermore, to capture influence of variable 'roadside land use along NHs' within 100 m of the studied NHs on fatal crashes is captured through google earth maps, and the cited data is also confirmed during field visits. 'Roadside land use' along NHs within 100 m of the highway is considered as right of way (RoW) of NHs usually varies between 25 m (two-lane NHs) to 60 m (for six-lane NHs) with maximum 100 m RoW. The RoW of NH represents the extent of land boundary upto that it pertain to NHs, and may be used for expansion of highway in future. Furthermore, it is mentioned that the roadway (paved surface) width in six-lane NHs is to the tune of 30 m (typical cross section of six-lane NH is: 10.5 m c/w on either side, 4.5 m wide raised median, 2.5 m wide paved shoulder on either side). The right of way (boundary) of NHs usually varies from 30 m to 60 m (maximum 100 m) depending upon number of lanes. Numbers of researchers (Ackaah and Salifu 2011) have studied impact of number of 'land use' or presence of village settlement within road segment or driveway density on crashes. (Ivan et al., 2000) have also studied land use effects by the number of driveways observed on each highway segement and classified into categories (private residence; apartment building with more than four units; gas station; and so on).

6.2.7 Terrain type

For two-lane NH-8, 'terrain type' is also considered as explanatory variable owing to presence of both flat and rolling in the studied NH section. The other two studied NHs (four- and six-lane NHs) have only flat terrain throughout the NHs sections, hence 'terrain type' as variable is not considered. Ackaah and Salifu (2011) have studied the impact of terrain type on highway crashes.

Table 2 provides data summary for NHs segments for SPFs for fatal crashes.Table 3 presents short-listed variables for development of SPFs for fatal crashes for NHs (segments) alongwith their descriptive summary statistics. The dependent variable for the model is fatal crash frequency per segment for the study period. Based on thorough literature review, availability of highway geometric details from highway strip plans and data captured during field visits, explanatory variables have been identified.

Furthermore, it is mentioned that changes in highway infrastructure on the studied three NHs segments are not observed during the study period (2008-2013) and hence explanatory variables values, such as number of access roads, number of lanes and number of access roads per km remain constant. Using random number generation through web (www.random.org/integers), ninety per cent of random numbers of data are used for model development and balance ten per cent random numbers from each NH segment are kept for model validation.

Table 2. Data summary of NHs (segments) for SPFs for fatal crashes.

Data description	Code	Two-lane NH-8	Four-lane NH-24	Six-lane NH-1
No. of segments	-	69	30	45
Access road per segment	-	0-9	0-11	0-4
Presence of SR at segments	0-No	N.A.	19	34
	1-Yes	N.A.	11	11
Presence of horizontal curvature at segments	0-No	2	22	2
	1-Yes	67	8	43
Terrain type at segments	0-Rolling	54	0	0
	1-Plain	15	30	45
Roadside land use at segments	0-sparse	51	21	23
	1-heavy	18	9	22

N.A.-Not applicable

Table 3. Variable description and summary statistics of data (2008-2013): models using segments for each NH (excluding intersections).

Variable Description	Unit	Data type	Range of values			Mean			Standard deviation		
			Two-lane NH-8	Four-lane NH-24	Six NH-1	Two -lane NH-8	Four -lane NH-24	Six NH-1	Two-lane NH-8	Four -lane NH-24	Six NH-1

(A) Dependent											
Fatal crashes at segments	Number	Count	0-19	0-9	0-25	2.68	2.9	8.57	4.09	2.9	7.59
(B) Independent											
Segment length	Km	Conti-nuous	0.15-2.4	0.15-1.25	0.3-1.85	1.49	0.66	0.90	0.74	0.29	0.38
Roadside land use for a road segment	Heavy/ sparse	Categorical	0-sparse(agriculture /forest /barren land) 1-heavy (residential/industrial/commercial)				-			-	
Number access roads per km within a road segment	Number/ km	Continuous	0.0-12.5		0-6.82	2.49		1.96	2.07		1.82
ADT in vehicles within a road segment*	-	Continuous	2605-2705	46436	48028-65535	-	-	-	-	-	-
Presence of raised median	-	categorical	0-not present in segment 1-present in segment								
Presence of SR	-	categorical	N.A.	0-not present in segment 1-present in segment		N.A.	-		N.A	-	
Presence of horizontal curve	-	categorical	0-not present in segment 1-present in segment								
Terrain type	-	Categorical	0-plain or level 1- rolling								

N.A.-Not applicable **data from toll plaza(s) at each NH*

To develop statistical models, general shape of fatal crash frequency distribution at segments for each of the studied NHs are examined. Fig. 3 depicts fatal crash frequency distribution at segments for each three NHs based on fatal crash data for 2008-2013. It can be observed from fig. 3.(a); (b); (c). that the data for two-lane NH-8, four-lane NH-24 and six-lane NH-1 are skewed to the right and hence, clearly ordinary least square (OLS) regression would be inappropriate and it suggest to explore generalized regression models, such as Poisson or NB regression. The mean and variance for the dependent variable (fatal crashes per segment) presented in table 3 are: 2.68 and 16.73 for two-lane NH-8, 3.03 and 9.20 for four-lane NH-24 and 8.57 and 58.93 for six-lane NH-1 respectively. These figures suggest that overdispersion may exist in the data. However, to further investigate the overdispersion for Poisson regression models, the ratios of both deviance and its degrees of freedom, and Pearson chi-square statistics and its degrees of freedom are obtained as: 1.25 and 1.20 for two-lane NH-8, for four-lane NH-24, for six-lane NH-1 respectively through statistics software. This analysis also indicates that the data are overdispersed. Based on the analysis discussed above, it can be concluded that NB regression approach seems to be better for developing SPF for fatal crashes for NHs (segments) than Poisson regression.

This study has examined both Poisson regression and NB regression models. However, the results of NB regression are presented in table 4.

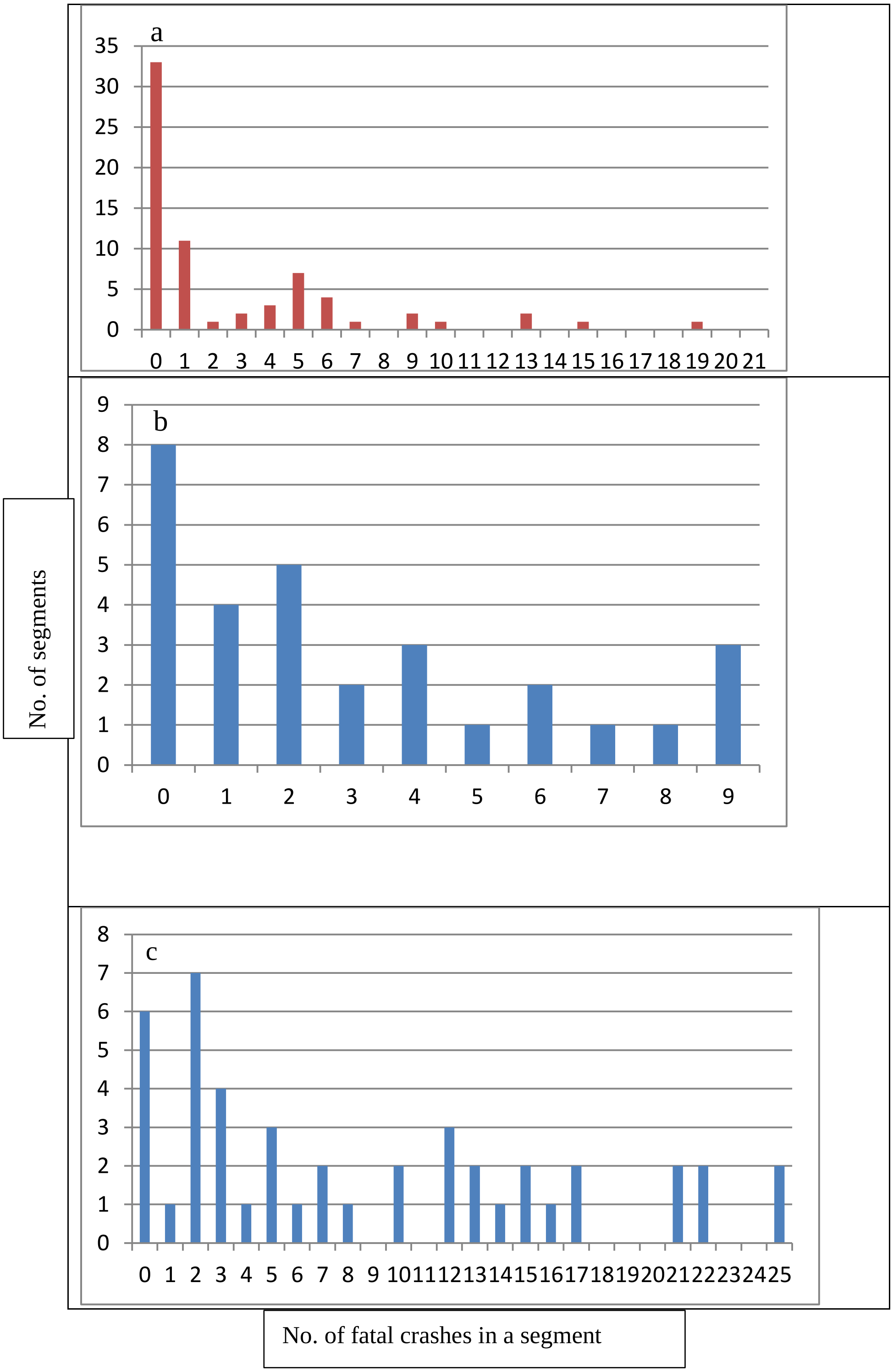

Fig. 3. Fatal crash frequency distribution (a) two-lane NH-8; (b) four-lane NH-24; (c) six-lane NH-1.

7. Estimation results

The statistical models are developed for the data presented in table 2 for three NHs segments to estimate fatal crash frequency. Due to significant overdispersion in data (variance > mean) and violation of Poisson regression assumption (mean = variance) as discussed under 'model estimation' paragraph, parameter estimates for the fatal crash prediction models for three studied NHs employing NB regression method as presented in table 4 are discussed. In this study software namely, SPSS Statistics (version 19) is used to run NB regression models. The statistical models are developed for the data presented in table 2 for three NHs segments to estimate fatal crash frequency. Due to significant overdispersion in data (variance > mean) and violation of Poisson regression assumption (mean = variance) as discussed under 'model estimation' paragraph, parameter estimates for the fatal crash prediction models for three studied NHs employing NB regression method as presented in table 4 are discussed. In this study software namely, SPSS Statistics (version 19) is used to run NB regression models.

7.1 Model-1: Two-lane NH-8

Table 4 shows that for two-lane NH-8, the NB regression model results are found to be significant at 95% confidence interval. The goodness-of-fit statistics for the model shows that model fits reasonably well with three explanatory variables: 'segment length' (km), 'roadside land use': (heavy/sparse), and 'terrain': (plain/rolling). These variables are found to be significant at 95% confidence interval ($p<0.05$) with positive estimated model parameters of the model. The recommended fatal crash frequency model for two-lane NH-8 is presented below.

Fatal crashes at two-lane NH-8 segment (in number)

$$= e^{1.802} \exp(0.625X1) \exp(-1.239X2) \exp(-1.428X3) \quad (5)$$

Where, X1= segment length (km) and X2 = roadside land use on either/both sides of NHs (0 for sparse, 1-heavy), and X3=terrain (0-plain, 1- rolling).

7.2 Model-2: Four-lane NH-24

Table 4 presents NB regression model for four-lane NH-24. The goodness-of-fit statistics for the model shows that model fits reasonably well with two explanatory variables: 'segment length' (km), and 'presence of SR' on either side or both sides of carriageway. These variables are found to be significant at 95% confidence interval ($p<0.05$) with positive estimated model parameters of the model. The recommended fatal crash frequency model for four-lane NH-24 is given below.

Fatal crashes at four-lane NH-24 segment (in number)

$$=e^{-1.402} \exp(2.804X1) \exp(0.437X2) \quad (6)$$

Where, X1= segment length (km) and X2 = presence of SR (0-No, 1-Yes).

7.3 Model-3: Six-lane NH-1

Table 4 presents NB regression model results for six-lane NH-1. The goodness-of-fit statistics for the model shows that model fits reasonably well with two explanatory variables, namely, 'segment length' (km), and 'roadside land use': (heavy/sparse). These variables are found to be significant at 95% confidence interval ($p<0.05$) with positive estimated model parameters of the model. The recommended fatal crash frequency model for six-lane NH-1 is presented as follows.

Fatal crashes at six-lane NH-1 segment (in number)

$$= e^{-0.197} \exp(0.087X1) \exp(-0.056X2) \quad (7)$$

Where, X1= segment length (km) and X2 = roadside land use on both sides of NHs (0 for sparse, 1 for heavy).

7.4 Inferences

In the above-mentioned three models (two-, four- and six-lane NHs segments) presented in Table 4, the exposure variable 'segment length (km)' is observed significant for fatal crashes on the cited three NHs. The positive sign of coefficients in all three models shows that the fatal crash frequency increases with an increase in the 'segment length' while other parameter is/are held constant. For variable 'segment length in km' the parameter estimate is found to be the highest [exp(2.8)=16.4] in four-lane NH-24 and lowest [exp(0.087)=1.09] in six-lane NH-1. The results of model for explanatory variable are as per expectation. It also shows that the longer distance traveled generate more likelihood for fatal crash frequency on NHs. The study results of the model are found to be consistent with the previous studies (Milton and Mannering 1998; Ackaah and Salifu 2011) that as the length of section increases, the probability to crashes also increases.

For two-lane NH-8 and six-lane NH-1 models (NHs segments) presented in table 4, the variable 'roadside land use' is observed significant for fatal crashes at NHs. In case of variable 'roadside land use' (either side or both sides) along NHs, it is noticed that for two-lane NH-8 and six-lane NH-1 models that 'roadside' variable value 'sparse' lowers fatal crash frequency [exp (-1.239) = 0.29] by 63% and [exp (-0.056) = 0.94] by 43% for two-lane NH-8 and six-lane NH-1 respectively in comparison with variable value 'heavy'. The study result for the variable 'segment length' is found consistent with previous study (Milton and Mannering 1998; Ackaah and Salifu 2011). Furthermore, it is mentioned that for six-lane NH-1 the explanatory variable 'roadside land use' along NHs is not found statistically significant at 95% CI and hence, it is not considered in the model.

The parameter estimates for NB regression model with variables such as, 'presence of raised median', 'presence of horizontal curve', 'number of access roads/km' in a segment, 'ADT' in vehicles are not found to be statistically significant, and hence their results are not presented in table 4. The reason(s) for cited explanatory variables found not statistically fit with the cited model is/are not known, except in case of ADT (in vehicles) data is collected at the respective toll plaza(s) location(s) for the studied NHs. Toll plaza location(s) for collecting ADT data for two-lane NH-8, four-lane NH-24 and six-lane NH-1 are two, one and one respectively. Hence, variations in ADT values at intersections along each NH are not noticed.

Table **Error! No text of specified style in document.**. Estimation results of the SPFs for fatal crashes at NHs (segments): Two-lane NH-8, Four-lane NH-24 and Six-lane NH-1.

Explanatory Variables	Model-1 Two-lane NH-8			Model-2 Four-lane NH-24			Model-3 Six-lane NH-1		
	Coefficient (B)	Standard Error	Significant at 95% CI	Coefficient (B)	Standard Error	Significant at 95% CI	Coefficient (B)	Standard Error	Significant at 95% CI
Constant	1.802	0.557	0.001	-1.402	0.561	0.013	-0.197	0.065	0.003
Exposure factor									
segment length (km)	0.625	0.318	0.050	2.804	0.599	0.000	0.087	0.0388	0.024
Land use related factors									
Land use (either side or both sides of NH)									
0-Sparse	-1.239	0.522	0.018	-	-	-	-0.056	0.037	0.131
1-Heavy	0^b	-	-	-	-	-	0^b	-	-
Presence of SR on either side or both side of c/w									
0-No	-	-	-	0.437	0.324	0.178	-	-	-
1-Yes	-	-	-	0^b	-	-	-	-	-
Topography related factor									
Terrain									
0-Plain	-1.428	0.457	0.002	-	-	-	-	-	-
1- rolling	0^b	-	-	-	-	-	-	-	-

	Omnibus Test: Likelihood ratio Chi-square= 14.419; df = 3; P-value = 0.002	Omnibus Test: Likelihood ratio Chi-square=21.789.; df = 2; P-value = 0.000	Omnibus Test: Likelihood ratio Chi-square= 6.015; df = 2; P-value = 0.049
	Goodness of fit measures: Deviance (value/d.f.)=66.339 (1.037), Pearson chi-square(value/d.f.)=66.808 (1.044), log-likelihood = -132.265 Number of observations= 69 Negative binomial B=1.943	Goodness of fit measures: Deviance (value/d.f.) =31.219 (1.249), Pearson chi-square (value/d.f.) =29.551 (1.182), log-likelihood = -53.449 Number of observations= 30 Negative binomial B=0.205	Goodness of fit measures: Deviance (value/d.f.) =53.174 (1.297), Pearson chi-square (value/d.f.) =34.369 (0.83), log-likelihood = -141.227 Number of observations= 45 Negative binomial B=0.82

b-set to zero because this parameter is redundant

8. Summary and conclusion

The results of the developed models showed that out of seven explanatory variables examined for each NH (segments), the significant explanatory variables are found to be 'segment length' in km for all three models of NHs (segments). The other significant variable is 'roadside land use' for two-lane NH-8 and four-lane NH-24. These study results are in line with previous studies (Milton and Mannering, 1998; Ackaah and Salifu, 2011). Furthermore, the variables, namely, 'presence of SR' and 'terrain type' are found significant for four-lane NH-24 and for two-lane NH-8 respectively.

The study results clearly highlight need to focus on NHs segments, which are passing through heavy land use to reduce fatal crashes by developing bypasses or access controlled highways with SRs at either travel direction for local traffic. The provision of SRs and pedestrians crossing facilities are important to cater to local traffic safe maneuvering needs especially at segments having heavy land use.

Acknowledgements

The authors convey thanks to the NHAI and MoRTH for sharing traffic data, and Police departments of Haryana, Rajasthan and Uttar Pradesh for providing fatal crash data. The opinions, findings and conclusions expressed here are those of authors.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention, 32, pp. 633-642.

Ackaah, W., Salifu, M., 2011. Crash prediction model for two-lane rural highways in the Ashanti region of Ghana. IATSS Research, 35(1), pp. 34-40.

Bhalla, K., Khurana, N., Bose, D., Navaratne, K.V., Tiwari, G., Mohan, D., 2017. Official government statistics of road traffic deaths in India under-represent pedestrians and motorised two wheeler riders. Injury Prevention, 23, pp. 1-7.

Cafiso, S., Graziano, A.D., Silvestro, G.D., Cava, G.L., Persaud, B., 2010. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. Accident Analysis and Prevention, 42, pp. 1072-1079.

Chen, J.-S., Wang, S.-C., 1999. Statistically modelling relationships between accident types and highway features. Civil Engineering and Environmental Systems, 16(1), pp. 51-65.

Donnell, E.T., and Mason Jr., J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accident Analysis and Prevention, 38(3), pp. 590-599.

Garber, N., Ehrhart, A.A., 2000. Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. Transportation Research Record, 1717 (January), pp. 76-83.

Hadi, M.A., Aruldhas, J., Chow, L-F., Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative Binomial regression. Transportation Research Record, 1500, pp. 169-177.

Hauer, E., 2004. Statistical road safety modeling. Transportation Research Record, 1897, pp. 81-87.

Haynes, R., Lake, I.R., Kinghom, S., Sabel, C.E., Pearce, J., Barnett, R., 2008. The influence of road curvature on fatal crashes in New Zealand. Accident Analysis and Prevention, 40, pp. 843-850.

Ivan, J.N., Wang, C., and Bernardo, N.R., 2000. Explaining two-lane highway crash rates using land use and hourly exposure. Accident Analysis and Prevention, 32, pp. 787-785.

Knuiman, M.W., Council, F.M., Reinfurt, D.W., 1993. Association of median width and highway accident rates. Transportation Research Record, 1401, pp. 70-80.

Persaud, B.N., Mucsi, K., 1995. Microscopic accident potential models for two-lane rural roads. Transportation Research Record, 1485, pp. 134-139.

Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis and Prevention, 26(4), pp. 471-482.

Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention, 25(6), 689-709.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation, 25, 395-413.

Ministry of Road Transport & Highways (MoRTH)., 2017. Road accidents in India 2016. Ministry of Road Transport & Highways, Govt. of India, New Delhi, pp. 4-15.

Ministry of Road Transport & Highways (MoRTH)., 2017b. Basic road statistics of India 2015-16. Ministry of Road Transport & Highways, Govt. of India, New Delhi, pp. 3.

Mohan, D., Tsimhoni, O., Sivak, M., Flannagan, M.J., 2009. Road safety in India: Challenges and opportunities. Transportation Research Institute, University of Michigan, USA, pp. 4.

Mountain, L., Fawaz, B., Jarrett, D., 1996. Accident prediction models for roads with minor junctions. Accident Analysis and Prevention, 28(6), pp. 695-707.

Naqvi, H.M., Tiwari, G., 2018. Factors contributing to motorcycle fatal crashes on National Highways in India. International Journal of Injury Control and Safety Promotion, pp. 1-10.

Noland, R. B., 2003. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. Accident Analysis and Prevention, 35, pp. 599-611.

Okamoto, H., and Koshi, M., 1989. A method to cope with the random errors of observed accident rates in regression analysis. Accident Analysis and Prevention, 21(4), pp. 317-332.

Vangala, P., Lord, D., Geedipally S.R., 2015. Exploring the application of the negative Binomial– generalized exponential model for analyzing traffic crash data with excess zeros. Analytic Methods in Accident Research, 7, pp. 29-36.

Vogt, A., Bared, J., 1998. Accident models for two-lane rural segments and intersections. Transportation Research Record, 1635, pp. 18-29.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Count data models. Statistical and econometric methods for trasportation data analysis, CRC Press, Taylor & Francis Group, London, pp. 283-300.