Exploratory Assessment of Road Traffic Crashes on the Intercity Expressway in India

Laxman Singh Bisht[1], Geetam Tiwari[2], Kalaga Ramachandra Rao[3], Kumar Neeraj Jha[4]

[1]PhD student, Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India-110016, Email: laxmanlsbisht@gmail.com
[2]Professor, Transportation Research and Injury Prevention Centre, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India-110016, Email: geetamt@gmail.com
[3]Professor, Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India-110016, Email: rrkalaga@civil.iitd.ac.in
[4]Professor, Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India-110016, Email: knjha@civil.iitd.ac.in

Abstract

The study aims to evaluate the risk of RTC on the selected intercity expressway in India. An exploratory data analysis technique was employed to ascertain crash characteristics. Random parameter negative binomial modelling approaches were used to account for segment-specific unobserved heterogeneity on the 168 km intercity six-lane expressway. According to the revealed crash characteristics, rear-end crashes had the highest proportion of total crashes. Pedestrians' fatality was also significant, a unique characteristic of expressway crashes in low-and-middle-income countries (LMICs) like India. The crashes involving median and guardrails also substantially contributed to the total crashes. Cars and trucks were involved in most of the crashes striking vehicles. Cars as striking vehicles caused the highest pedestrian fatalities. Single-vehicle crashes (SVC) also had a substantial proportion of the total crashes. Model results showed that segments with attributes such as the presence of hazards, presence of access location and underpasses, vertical alignment length, horizontal alignment radius, horizontal curve length and high AADT have a high risk of fatal crashes. In contrast, the segments with the attributes such as the presence of a village or settlement, vertical alignment gradient, vertical curve length and speed are negatively associated with the fatal crashes. The effect of these variables was consistent across all the developed models. The statistically significant random parameters are speed, AADT and vertical curve length. The findings from this study are likely to help the decision-makers and engineers alter the expressway design, mainly when it passes through the villages or settlements.

Keywords: crash characteristics; geometric design; highway safety; LMICs; safety analytics; unobserved heterogeneity

1. Introduction

India has the second-largest road network; however, most highways are two-lane [1]. The growth of private vehicles has also increased by 10.1% from 2007 to 2017. To boost the economy Indian government is targetting to reduce logistic costs as it accounts for 14% of GDP [2]. Therefore, lately, India has witnessed rapid development in the network length highways, especially expressways. The benefit of rapid expansion of the expressway network has been offset by the fatalities and injury burden due to road traffic crashes (RTCs) on these expressways [1]. RTC is a rare event, as the location and its occurrence on the highway network are random [3]. An RTC can be thought of as a complex event due to interaction among road, vehicle, and road users characteristics. A recent report on the status of road safety across the globe shows that RTCs are the 8th leading cause of death for all age groups [4]. It also revealed that the risk of death due to RTCs is three times higher in low-and-middle-income countries (LMICs) than in high-income countries (HICs). In India, national highways (NHs), including expressways, totalled 36% of the fatalities in 2019 [5].
The road owing agency or transportation engineers are interested in knowing the effect of road geometric elements on the number and frequency of RTCs [6]. Consequently, one of the objectives is to improve highway safety with the help of various safety evaluation methods [7]. These methods are built on multiple analytical approaches that identify the risk factors that affect the number and frequency of crashes on the expressway. The crash number and frequency are count data and non-negative integers, indicating that regression-based count data models are suitable for application [8]. Lord & Mannering [8] listed the issues with the count data. Significant issues are over-dispersion, temporal variation, temporal and spatial correlation, low mean, small sample size, and omitted variable

[1] * Corresponding author. Tel.: +91-956-0407-507.
E-mail address: laxmanlsbisht@gmail.com

bias. Poisson regression is the conventional method to determine the risk factor effect on crashes [9]. However, due failure of the Poisson regression model to account for overdispersion, its variant s negative binomial (NB) became a popular method. Various other count-data models and variants have also been applied over the years, including the Gamma, Conway–Maxwell–Poisson, the negative binomial-Lindley, and other models [8]. Researchers have also modelled crashes using duration models considering the duration between crashes [11] and the generalized ordered-response method [10].
One of the safety evaluation studies aims to identify the crash contributory risk factors and consequently suggest effective safety interventions to reduce the RTC frequency and severity [11]. In addition, various risk factors of highway crashes remain unobserved or omitted while developing the regression models due to various reasons [8]. Therefore, lately, advanced analytical methods have accounted for unobserved heterogeneity [12]. Due to variations across roadway segments, unobserved factors are likely to influence the risk factors' effect on crashes. Random parameter models are widely used techniques to capture segment-specific unobserved heterogeneity effect on RTC [13]. In this modelling technique, the parameters are allowed to vary across the highway observation, such as expressway segments [9].
Nonetheless, the literature review suggests the crash characteristics of expressways are not studied compared to the NH category in India. Against this backdrop, the study aims to evaluate the risk due to RTC on the selected intercity expressway in India. The first aim is to ascertain the crash characteristics on the selected expressway. Second aim is to identify the fatal crash contributory factors of the expressway. This study utilizes crash data of 168 km intercity six-lane expressway from August 2012 to October 2018.

2. Methodology

2.1 Study Area and Data Description

The selected highway is a 168 km high speed and semi access-controlled intercity six-lane expressway in Uttar Pradesh, India. It has a continuous 6 m wide raised median and no central guardrail. The study utilized speed, traffic volume, geometrics, and crash data. RTC data includes the information related to the fatal and non-fatal crash data from August 2012 to October 2018. The speed data was collected with the help of speed cameras installed on the expressway in ten different locations, five in each travel direction. In contrast, traffic volume data were extracted using toll transaction details data of all three toll plazas. The processed crash data was coded for the exploratory analysis to understand crash characteristics and build the segment-specific explanatory model.

2.1 Exploratory Data Analysis and Crash Contributory Factors Estimation

The study incorporates exploratory data analysis techniques to assess the crash characteristics of the selected expressway. Conventionally, the statistical analysis of crashes begins with a count data model such as Poisson regression. However, the crash data is generally over-dispersed [12]. Preponderant zero crash segments or segments with large crash counts can present challenges during the statistical analysis of crash count data. Thus, this property of crash data implies to use negative binomial (NB) model. However, the highway segments with identical attributes often exhibit significantly different crash counts due to unobserved reasons known as unobserved heterogeneity [14]. The unobserved heterogeneity is due to various reasons in the crash data [12]. Mannering et al. [12] explained that the unobserved heterogeneity could also be due to the unavailability of many factors.
Further, the authors suggest the perils of ignoring unobserved heterogeneity. Hence, the traditional negative binomial model with fixed parameters (FPNB) does not adequately handle highly over-dispersed crash data, spatially correlated or unobserved heterogeneity [15]. There is a need to address it with the help of advanced modelling techniques. Various researchers have used the random parameters negative binomial (RPNB) model to improve the inference of estimated coefficients by explicitly accounting for extra variation in crash data [15].
Therefore, this study proposes a negative binomial specification with random parameters to model fatal crash counts of expressway segments of 100 m length. The unobserved heterogeneity is addressed using a random parameters specification that relaxes any distributional assumptions of parameters. The model extracts the expressway segments with fatal crash counts that are equally sensitive to the road attributes on an average. Therefore, the estimable parameters of RPNB are allowed to vary across the expressway segment. An estimated coefficient is a random parameter if its estimated standard deviation (σ) is statistically different than zero. In this study, the normal distribution of the random parameters was statistically significant.
Further, to account for the inter-correlation of the distribution of the random parameter, a correlated random parameter (CRPNB) is used. In the case of uncorrelated random parameters (RPNB), the off-diagonal elements in the variance-covariance matrix (C) are zero. However, in the case of the correlated random parameters model (CRPNB), the off-diagonal elements in the C are statistically significant non-zero values that suggest the potential correlation amongst random parameters. Thus, the covariance matrix, M of the random parameters, can be expressed as $M = CC'$ [16]. The variance of the random parameters is the diagonal elements of the matrix M. The

statistical significance of the standard deviations is computed with the help of their standard errors and z-statistics. In this study, model estimation was done with the help of the stimulated maximum likelihood method was employed using 200 Halton draws in the simulated likelihood functions using NLOGIT6 software [17].
At last, statistical evaluation of the competing models was done with the help of several established statistical tests and statistics. The log-likelihood at convergence, Akaike information criteria (AIC), Bayesian information criterion (BIC) and the likelihood ratio test statistics were computed [18]. The model with the smallest AIC, smallest BIC and high log-likelihood magnitude at convergence was reported as the final model. In this study, the effect of segment-specific attributes on fatal crashes was studied with the help of the explanatory models.

3. Results and Discussion

3.1 Crash Characteristics on the Expressway

Crash data revealed 552 fatal and 5445 non-fatal crashes occurred from August 2012 to October 2018. Crash characteristics revealed that rear-end crashes were predominant in both types of crashes. They accounted for 49% of the fatal and 35 % of the non-fatal crashes. A few reasons for the high occurrence of rear-end crashes could be the poor conspicuity of parked vehicles on the shoulders, lack of visibility in case of foggy weather crashes, and inadequate information to the user of the entry and exit locations. In addition, in the case of fatal crashes, hit pedestrian crashes were the second-highest, followed by the hit median. In the case of non-fatal crashes, rollover crashes (22%) had the second-highest share, followed by the hit median (13%). It was also revealed that the median crossed crashes also contributed significantly to fatal and non-fatal crashes with a 5% and 6% share, respectively. Moreover, it was also found that a significant number of crashes occurred with the guardrail and other roadside furniture on the expressway. Cumulatively hit-guardrail and hit-culvert/bridge parapet wall collision types had 9% of fatal and 12% of the non-fatal crashes.
Details of hit-median crashes and median crossed crashes suggested that the design of median is inappropriate for high-speed roads such as expressways. Even though the expressway was access-controlled, hit-pedestrian fatalities were quite substantial (11%). This characteristic is unique and is expected on the expressways of the LMICs like India [19]. Besides, geolocations of hit-pedestrian crashes show that they were scattered all along the expressway. However, these crashes were high in the vicinity of the underpasses and entry and exit locations. It was observed that passengers were waiting for buses at these locations, which were, over time, developed as informal bus stops. It was also observed that young men were using the expressway for physical exercise and jogging, leading to pedestrian fatality in the evening. This unique issue exists on the studied expressways and could have been persistent on the other Indian expressways. However, further studies need to be carried out to reach a concrete conclusion and highlight this issue. In addition, it was also revealed that cars and trucks had contributed significantly to striking vehicles in both crash types. Compellingly, motorised two-wheelers (MTW) were also involved as striking vehicles in case of car user crashes and fatalities. In the case of fatal crashes, hit-run cases had a significant contribution as in 19% of such cases, and striking vehicle type remained unknown.
It was also shown that the characteristics of striking vehicle types posing a risk to road users as the occupant of the impacted vehicle was consistent in both crash types. However, slight variation in the striking vehicle types involved in car user-related non-fatal crashes were observed. Further, in the case of non-fatal car crashes, the prominent striking vehicle types were cars, MTWs and trucks**.** Thus, one of the salient features of crash characteristics on Indian expressway is the high pedestrian fatalities. In most cases of hit-pedestrian crashes, the striking vehicle types remain unknown. In addition, MTW crashes are also high compared to their low proportion in the total volume on the expressway. Another important finding from crash characteristics revealed that 38% and 22% of the total fatal crashes are related to truck and car users. Whereas in the case of non-fatal crashes, car crashes were disproportionately high. Further, Single-vehicle crashes (SVC) were 48% and 62% of the fatal and non-fatal crashes, respectively. In SVC, the highest proportion of crashes was due to cars, trucks, and MLS. The overall characteristics revealed that nighttime crashes were slightly high compared to daytime crashes. However, the overall volume remains low in this duration, promoting high speeds and consequently more fatalities.
At last, crashes were also analysed based on the weekday and month; however, there was not much variation, and characteristics were consistent between both crash types. Further, it was observed that the first half of the expressway experienced a higher number of fatalities and crashes. The high-risk sections were the initial section on the right-hand side (i.e., exit direction towards Delhi), the entry and exit sections, interchange weaving sections, at grade midblock sections, underpass locations and horizontal curves sections.

3.2 Estimates of the risk factors of fatal crashes

The final model's specification includes statistically significant variables, and Table 1 illustrates the estimates of all three models developed in this study. The goodness-of-fit measures and log-likelihood value results were also presented in this table. It was observed that the direction of the effect and statistical significance of the variables was consistent across the models. In all models, the variables such as the presence of hazards, presence of access location and underpasses, vertical alignment length, horizontal alignment radius, horizontal curve length and AADT were positively associated with the fatal crashes. The variable vertical alignment length captures the effect of the length of rising and fall along the longitudinal direction of the expressway on crashes.

Table 1: Risk factors effect estimation and comparison of models for fatal crashes

Variables	FPNB		RPNB		CRPNB	
	Coefficient	Prob. z|>Z*	Coefficient	Prob. z|>Z*	Coefficient	Prob. z|>Z*
			Nonrandom parameters			
Constant	-1.06043	0.3688	-1.05192	0.3589	-0.7495	0.513
Presence of village/settlement	-0.24376***	0.0074	-0.24127***	0.0081	-0.23072***	0.0092
Presence of hazards	0.11879	0.4835	0.11669	0.4892	0.13777	0.3958
Presence of access location and underpass	0.60648***	0.000	0.62760***	0.000	0.61151***	0.000
			Random parameters			
Vertical alignment gradient	-0.06487	0.4108	-0.06719	0.4032	-0.06565	0.3454
Standard deviation of parameter distribution			0.00042	0.9958	0.0874	0.148
Vertical alignment length	0.03371	0.9257	0.05379	0.8829	0.20269	0.5819
Standard deviation of parameter distribution			0.25419	0.2447	0.52606	0.1033
Vertical curve length	-0.42858	0.1214	-0.44692	0.1089	-0.37342	0.2073
Standard deviation of parameter distribution			0.23055	0.1748	0.56459**	0.0126
Horizontal alignment radius	0.05053**	0.0345	0.04982**	0.0383	0.06729***	0.0054
Standard deviation of parameter distribution			0.00246	0.9145	0.00877	0.6945
Horizontal curve length	0.12257**	0.0476	0.11409*	0.0694	0.14712**	0.0205
Standard deviation of parameter distribution			0.00288	0.9536	0.08105	0.1754
Speed	-0.00548	0.6157	-0.00581	0.5825	-0.01084	0.3095
Standard deviation of parameter distribution			0.00289***	0.000	0.00311***	0.0018
AADT	0.13596***	0.0002	0.12266***	0.0009	0.13387***	0.0004
Standard deviation of parameter distribution			0.06203***	0.000	0.00305	0.7824
Scale parameter/Alpha	0.12988***	0.0097	5.37077	0.111	77494.9	0.9999
K	13		20		41	
Log likelihood function	-1211.23		-1209.63		-1205.48	
Restricted log likelihod	-1216.56		-1309.68		-1309.68	
Chi squared	10.67		200.1		208.42	
McFadden Pseudo R-sqaured	0.004		0.076		0.08	

*note: ***, **, * ==> Significance at 1%,5%,10% level*

The marginal effects estimate of the RPNB model shows that the segments with hazards have higher chances of the average fatal crashes, and the probability marginally increases by 0.03%. For the segments with access locations and the presence of underpasses, the probability of fatal crashes increases significantly by 0.18%. The increase in fatal crash probability was also significant for the segments with high horizontal curves at 0.03%. Likewise, the segments with high AADT have a 0.03% high probability of fatal crashes. However, the segments with vertical alignment radius have higher chances of fatal crashes, and the fatal crash risk increases significantly by 0.14%. Further, results show that the significant risk factors of fatal crashes are the presence of access location and underpass, horizontal curve length, horizontal alignment radius and AADT. Thus, expressway segments dominated by these attributes have a high risk of fatal crashes compared to other segments. Pearson's correlation estimates show that speed and AADT were highly correlated.

Further, the marginal effect estimate shows that segments with villages or settlements have less risk of fatal crashes and the decrease in fatal crash probability is significant at 0.07%. Speed, AADT and vertical curve length variables were statistically significant random parameters. Of which, only speed and vertical curve variables were statistically significant as correlated random parameters. It was also observed that the magnitude of the random-parameters estimates had changed when the correlation was accounted for in the CRPNB model. The change in magnitude of random parameter estimates suggests that considering correlation is vital to remove bias and improve the accuracy of the estimates. Hence, doing this will help understand the effect of the expressway segment attributes on fatal crash risk. The relative statistical performance of the CRPNB and RPNB models is based on the log-likelihood value (shown in Table 2). The CRPNB model statistically outperforms uncorrelated RPNB and FPNB models. The estimated effects of risk factors using CRPNB were less dispersed when compared to the RPNB model, according to the estimated standard deviation of random parameters. Thus, this study suggests that the CRPNB modelling approach is statistically superior to its counterpart in explaining unobserved heterogeneity.

4. Conclusions

This study assessed the crash characteristics of the 168 km selected intercity expressway. Results showed that the rear-end crashes on this expressway were dominant. In the case of fatal crashes, hit-pedestrian crashes were the second-highest, followed by the hit median. In the case of non-fatal crashes, rollover crashes had the second-highest share, followed by the hit-median. Median related crashes significantly contribute to the total crashes. Cars and trucks contributed significantly to striking vehicles in both crash types. In the case of fatal crashes, hit-run cases had a significant contribution (19%), and striking vehicle type remained unknown in these crashes. Cars caused the highest pedestrian fatalities on the expressway. SVC were 48% and 62% of the fatal and non-fatal crashes, respectively. The highest share in the SVCs was due to cars, trucks, and MTW. Temporal crash characteristics revealed that nighttime crashes were slightly high compared to daytime crashes. This study proposed the application of NB model variants for analyzing fatal crash data by implementing RPNB models to account for segment-specific unobserved heterogeneity. The models have been applied to segment-specific datasets with 1674 observations. The model results were compared to the traditional FPNB, uncorrelated RPNB, and CRPNB models. Consequently, results showed that both the RPNB and CRPNB performed better than the FPNB model. The estimated effects of risk factors using CRPNB were less dispersed when compared to the RPNB model, according to the estimated standard deviation of random parameters. Model results also showed that the variables such as the presence of hazards, presence of access location and underpasses, vertical alignment length, horizontal alignment radius, horizontal curve length and AADT were risk factors of the fatal crashes. The effect of these variables was consistent across all the developed models. Speed, AADT and vertical curve length variables were statistically significant random parameters.

To sum up, this study suggests and makes a strong case for the calibration of the geometric design of expressways in India. It is to be acknowledged that solutions for many of these unique issues are not readily available. Hence, dedicated research needs to be carried out to calibrate the geometric designs of expressways. This study contributes by indicating that the design calibrations require to incorporate the safety of the local people accessing expressways due to other needs. The findings from this study are likely to help the decision-makers and engineers alter the expressway design, mainly when it passes through the villages or settlements. This study has also contributed by identifying fatal crash risk factors in the dearth of studies on the expressway in India. Hence, the estimates suggest that expressway design needs to be reviewed, especially the median design. This study did not assess heterogeneity in means and variance of random parameters. Thus, the underlying nature of unobserved heterogeneity in the dataset was not explored. In a future study, a modelling approach needs to be implemented to account for heterogeneity in the mean and variance of the random parameters.

References

[1] MoRTH, "Annual Report 2020-2021," 2021.

[2] NITI Aayog, RMI, and R. India, "Fast Tracking Freight in India," 2021.

[3] Z. Xie and J. Yan, "Kernel Density Estimation of traffic accidents in a network space," *Comput. Environ. Urban Syst.*, vol. 32, no. 5, pp. 396–406, 2008.

[4] WHO, "Global Status Report on Road Safety 2018," 2018.

[5] MoRTH, "Road Accidents in India 2019," 2020.

[6] M. G. Karlaftis and I. Golias, "Effects of road geometry and traffic volumes on rural roadway accident rates," *Accid. Anal. Prev.*, vol. 34, no. 34, pp. 357–365, 2002.

[7] F. Gross, B. Persaud, and C. Lyon, "A Guide to Developing Quality Crash Modification Factors," 2010.

[8] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transp. Res. Part A Policy Pract.*, vol. 44, no. 5, pp. 291–305, Jun. 2010.

[9] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: Methodological frontier and future directions," *Anal. Methods Accid. Res.*, vol. 1, pp. 1–22, Jan. 2014.

[10] M. Castro, R. Paleti, and C. R. Bhat, "A spatial generalized ordered response model to examine highway crash injury severity," *Accid. Anal. Prev.*, vol. 52, pp. 188–203, 2013.

[11] C. Wang, M. A. Quddus, and S. G. Ison, "The effect of traffic and road characteristics on road safety: A review and future research direction," *Saf. Sci.*, vol. 57, pp. 264–275, 2013.

[12] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Anal. Methods Accid. Res.*, vol. 11, pp. 1–16, 2016.

[13] P. C. Anastasopoulos and F. L. Mannering, "A note on modeling vehicle accident frequencies with random-parameters count models," *Accid. Anal. Prev.*, vol. 41, no. 1, pp. 153–159, 2009.

[14] P. Buddhavarapu, J. G. Scott, and J. A. Prozzi, "Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data," *Transp. Res. Part B Methodol.*, vol. 91, pp. 492–510, 2016.

[15] M. R. R. Shaon, X. Qin, M. Shirazi, D. Lord, and S. R. Geedipally, "Developing a Random Parameters Negative Binomial-Lindley Model to analyze highly over-dispersed crash count data," *Anal. Methods Accid. Res.*, vol. 18, pp. 33–44, 2018.

[16] W. H. Greene, "LIMDEP Version 8." 2002.

[17] J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numer. Math.*, vol. 2, no. 1, pp. 84–90, 1960.

[18] S. Washington, M. Karlaftis, F. Mannering, and P. Anastasopoulos, *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC, 2020.

[19] G. Tiwari, "Principles for Development of Safer Rural Highway Systems for Conditions Prevailing in Low and Middle-Income Countries," in *Transport and Safety*, Springer, 2021, pp. 105–127.