Investigation on Risk Factors Influencing Crash Occurrence in Urban Context: A Case Study of Hyderabad, India

Pratik AGARWAL[a], Sudeshna MITRA[b], Bandhan B. MAJUMDAR[c], Priyatham RISHWA[d], Prasanta K. SAHU[e]

[a]*Undergraduate student, Department of Civil Engineering, Birla Institute of Technology and Science Pilani, Hyderabad - 500 078, Telangana, India*
[a]*E-mail: f20150589@hyderabad.bits-pilani.ac.in*
[b]*Associate Professor, Department of Civil Engineering, Indian Institute of Technology Kharagpur, Pin-721302, West Bengal, India*
[b]*E-mail: sudeshna@civil.iitkgp.ernet.in*
[c,e]*Assistant Professor, Department of Civil Engineering, Birla Institute of Technology and Science Pilani, Hyderabad - 500 078, Telangana, India*
[c]*E-mail: bandykolkata@gmail.com*
[e]*E-mail: prasantsahu222@gmail.com*
[d]*Former M. Tech scholar, Department of Civil Engineering, Indian Institute of Technology Kharagpur, Pin-721302, West Bengal, India*
[d]*E-mail:- appriyatham@gmail.com*

Abstract: Road crash statistics for the last decade indicate a steep-rise in both number of crashes and number of fatalities in India. In order to reduce the crashes on a long-term basis and suggest necessary mitigation-measures, identification of causative factors associated with different crash types is important. In this regard, this paper aims to explore the risk factors affecting crash-occurrence and severity in Indian urban-context; Hyderabad, an Indian metropolitan city with significant road crash history is selected as a case study city. In this study, location wise road crash data (crash severity, accused, victim vehicle details, personal information of victims), traffic and geometric characteristics of the locations (classified vehicle volume, geometric, land-use and road related information) were collected and a detailed database was developed. Based on the data, separate safety performance function models were developed for signalized, un-signalized and midblock-sections using negative-binomial-regression models for Hyderabad city. Based on the results, non-motorized traffic volumes, median width, presence of T-junction were found to be key determinants of fatal and non-fatal crashes in typical Indian urban context.

Keywords: Risk-factors; Crash Severity; Safety Performance Function; Negative binomial Regression; Signalized, un-signalized, mid-block locations; Hyderabad

1.INTRODUCTION

India is overwhelmed by a worrying number of road traffic related crashes and injuries every year. The number of road crashes per 100,000 population has been increasing since 1970s; 84% increase could be observed from 1980 to 1990. Mohan et al. (2017) reported that a total of 150,785 persons were killed and 494,624 were injured in road traffic related crashes in India in 2016. Mohan et al. (2017) also mentioned that the road traffic-related fatalities were the 11th leading cause of death in 1990, however, it has become the 8th leading cause of death in India in 2016. The authors also mentioned that the total number of vulnerable road user (pedestrians, bicyclists, and motorised two-wheeler users) deaths in all the metropolitan cities range between 84% and 93%; this phenomenon could be clearly attributed to the lack of adequate infrastructure for vulnerable and non-motorized transport (NMT) users in big cities including Hyderabad requiring interventions.

Road safety has thus become one of the most critical urban issues demanding significant priority for developing necessary intervention measures in typical Indian context. In order to address this issue, identification of the contributing risk factors for different geometric features such as signalized, un-signalized and mid-block locations is necessary. However, a brief review of existing research literature suggests a significant lack of scientific road safety investigations in Indian context in general, lack of studies exploring the risk factors associated with crash occurrence in urban Indian settings in specific. In this regard, development of a safety performance function (SPF) for different locations is necessary, as it will be key for both crash prediction at a particular location, as well as identification of elements of traffic, geometric or land-use characteristics contributing to different types crashes across various locations. Development of safety performance functions for different road categories and predicting future crashes has already become a global practice across developed nations, however development of SPFs for roads in developing countries with mixed traffic operations, where data availability is a major challenge has been rather limited. Moreover, due to difference in traffic and road-geometric characteristics, the developed crash prediction models cannot be directly transferred to Indian settings. In order to address this research gap, this study has attempted to develop SPF's for different locations for Fatal, no-fatal and total (including both fatal and non-fatal) crashes in a typical urban setting. For demonstration of the proposed method, Hyderabad, an Indian metropolitan city has been selected.

Hyderabad is capital of Telangana, it lies in south of central India and has a territory of 625 sq./km and 6852 sq.km of metropolitan region is the fifth biggest city in India, with a population of over 8 million. The expanding weight of the prospering populace is putting Hyderabad's transport framework under consistent pressure. Due to its prominence as a major high-tech center, it is regarded as one of the fastest growing city with a population density of more than 17,000 persons per sq.km. As per the official statistics on the Hyderabad police website, 2,585 road crashes occurred in 2014 in which 358 people died and 2,540 got injured. Hyderabad police, claimed 53 more crash deaths. According to the Cyberabad police website, 3,296 road crashes had occurred in 2014, in which 1,036 persons died and 2,260 suffered injuries. Hence, development of crash prediction models and identification of causative factors will be a key input to Hyderabad city officials for curbing the rate of crashes in the city and the similar methodology could be followed by other Indian cities of similar size and characteristics.

The paper proceeds as follows. Section 2 reviews the existing research literature highlighting the major work already done in this area and simultaneously identifies the research gap and presents the specific scopes for this work. Section 3 details about the methodology adopted for the analysis. The following section 4 presents the database development, preliminary analysis, crash severity modelling through formulation of SPFs and also presents interpretation of the models. Section 5 includes the major findings and conclusions of the work.

2. LITERATURE REVIEW

In this section, a brief review of existing research literature related to road traffic safety analysis and investigation in general, formulation of SPF in specific is carried out to understand the existing research status on this topic and identify the present research gaps in this context. Several of the existing studies have carried out detailed research on crash data modelling and formulated SPF for crash prediction across developed (Hadi et al. 1995; Shankar et al. 1995; Maher and Summersgill, 1996; Mountain et al. 1996; Yan et al. 2005; El-Basyouny and Sayed, 2010; Choi et al. 2019; Farid et al. 2019) as well as developing countries (Robert, et al. 2006; Dinu and Veeraragavan, 2011; ChikkaKrishna et al. 2016; Mitra et al. 2017). Hence, a brief review of existing research would be key for understanding the existing research gaps and identifying the specific scopes for this study. In this regard, a set of existing researches along with their key findings are presented in the following section.

The safety performance function for a segment or junction with respect to its traffic and geometric characteristics give the expected crashes. Researchers used both Poisson and more extensively, Negative Binomial (NB) models for developing SPF for establishing empirical relation between crash occurrence and other geometric and traffic characteristics. Several researches in in middle income countries have developed SPF's for crash prediction (Hadi et al. 1995; Shankar et al. 1995; Maher and Summersgill, 1996; Mountain et al. 1996; El-Basyouny and Sayed, 2010; Choi et al. 2019; Farid et al. 2019), however, development of SPFs for roadways in developing countries with mixed traffic operations has been rather limited. Only a few studies (Robert, et al. 2006; Dinu and Veeraragavan, 2011; ChikkaKrishna et al. 2016; Mitra et al. 2017) have developed SPF's for highways, but none for urban roads. In one of such key research study, Yan et al. (2005) conducted an in-depth investigation to understand the characteristics of rear-end crashes at signalized intersections using multiple logistic regression model in Florida, USA. Based on the study results, authors concluded that crash rates are significantly higher for six-lanes compared to four-lanes. They also observed that traffic speed is one of the key attributes contributing to increased rate of crashes. Age was also found to be a significant factor influencing crash occurrence in the state of Florida. In another similar study, Wang and Abdel-Aty (2006) developed generalized estimating equations along with negative binomial link functions to model crash frequencies at signalized intersections and concluded that traffic characteristics, geometric design features, traffic control, location type and corridor level factors to be five significant variables. Based on their research results, they suggested for reduced provision of free turns to reduce the number of crashes. They also recommended for coordination across intersections along the corridors for reducing crash numbers in signalized intersections. In another relevant study, Wong et al. (2007) developed SPF's through negative binomial and poison and negative binomial regression models to identify the contributory factors to traffic crashes at signalized intersections in Hong Kong. They found increase in curvature, presence of tram stops and traffic volume to be key factors influencing crash occurrence. Similarly, several other researches have been conducted in the international context with respect to crash factor identification, however, as it is difficult to transfer the models in Indian context, it is necessary to develop such models in Indian urban settings. In this context, a brief review of related researches conducted in India is presented in the following section.

Among such relevant studies, Raju and Apparao (2013) conducted a temporal analysis of crashes in Meerut and Muzaffarnagar, India. They found that weekends, evening hours and months between August and December observed highest number of crashes. In another relevant research study, Bandyopadhyaya (2015) studied the reasons behind occurrence of crashes in Howrah, India along NH-6 through developing fuzzy-clustering and SPF models.

Based on both negative binomial and poison regression model, she concluded that collision partners, time of crash and condition of road to be the three main factors influencing crash occurrence along a typical highway in India. There were some related studies conducted on the city of Hyderabad as well. Among such studies, Murthy and Rao (2015) found that number of intersections, major traffic, unpaved shoulder, speed and turning radius have a positive relation with crash rate. Akkelagunta (2017) conducted a study on spatio-temporal distribution of crash risk in Hyderabad city. He concluded that along with signal characteristics, road markings, footpaths, pavement conditions, car parking, presence of median, purpose of use of road affect crash occurrence significantly in Hyderabad.

Based on the literature review, the following research gaps could be identified. Firstly, it could be observed that, studies investigating the identification of causative factors associated with crash occurrence through crash severity model are rather limited in typical Indian context. Such analysis is of relevance when availability of data is limited, which is a rather common instance in Indian settings. Secondly, majority of the research literature reported in Indian context mentions about lack of suitable database for analysis and evaluation. In order to address this issue, this study develops a comprehensive database require for a detailed crash study analysis. Thirdly, it could also be seen that establishment of empirical relationships between crash occurrence and geometric and traffic characteristics of roadways separately for signalized, un-signalized intersections and mid-block locations are rather limited in Indian context in general, development of Safety Performance Function in particular. This specific issue needs to be addressed through judicious study to identify the reasons of crashes and predict the influence of built environment on crash. In order to address these research gaps, the following scopes were set to be attained through the study:

- Development of a comprehensive crash database for Hyderabad city
- Formulation of Safety Performance Functions for fatal, non-fatal and total (fatal and non-fatal) crashes across signalized, un-signalized and midblock locations.
- Identification of critical factors contributing to the crash occurrence in typical Indian urban settings

3. DATABASE DEVELOPMENT

This section presents the data collection process followed by the database development process. The data used in this study comprises of three components: crash data, road geometric data and traffic volume data.

The past crash data is obtained from Hyderabad Police, the road geometric data is taken from Greater Hyderabad Municipality Corporation (GHMC) and traffic volume data were collected from Hyderabad Police. Crash Information Data has information about date and time of the crash occurrence, details of the victims including name of the person involved and gender, type of the crash as fatal or non-fatal, detailed location of the crash, details of the vehicles involved etc.

Road Geometric Data includes information about classification of location as signalized, unsignalised or mid-block, information of the location of crash in form of the road in which crash took place, road width, pavement condition, type of road, type of traffic, information about markings and parking etc. Videos from surveillance cameras of selected locations have been obtained from the Hyderabad Traffic Police. Traffic volume data is being extracted to calculate right turning, left-turning and through traffic movement. For the study purpose, peak 15 minute volume was used during the occurrence of fatal crash at that particular location. Peak volume is being extracted for the duration of fatal crash occurrence only; a classified volume count was carried out. Traffic count of seven major types of vehicles namely motorized two-wheelers, bicycles, cars or SUVs, buses, auto, NMTs, light commercial vehicles and heavy commercial vehicles are considered as independent causative factors. The traffic volume data is obtained through video clips of the locations chosen for analysis. It is assumed that the traffic volume remains constant throughout the year. Also it should be mentioned that obtaining traffic volume is a very difficult task in Indian context and even national level road safety records failed to incorporate traffic volume while explaining historical trends of road traffic crashes.

Data related a total of fifty-six variables were collected through inventory survey of the locations. While identifying locations for the study, locations with at-least one fatal crash were selected. In this study, a total of 30 signalized, 14 un-signalized and 11 midblock locations were considered and relevant geometric and traffic data were collected through both primary as well as secondary sources. The initial set of data was collected from the selected signalized, midblock and un-signalized locations with at-least one fatal crash during 2011 to 2016. The data regarding the following independent variables were collected and a comprehensive database is prepared.

- Location type (Signalized/un-signalized/mid-block)
- Sight distance blocked or clear, and if blocked whether removable/irremovable.
- Road characteristics like carriageway width, median width
- Presence of car parking and road side market
- Percentage carriageway blocked
- Land use pattern (Commercial, residential, school zone, parks, etc.)
- Footpath width
- Percentage of footpath encroached
- Presence of certain features like metro station, bus bay, bus tops etc.
- Presence of road markings
- Zebra crossing width
- Signal control variables like cycle time, free left turn, all Red time, number of phases, protected right turn, etc.
- Type of zebra crossing
- Pedestrian facility
- Presence of stop mark

A preliminary analysis of the crash data collected from Hyderabad Traffic Police during 2011 to 2016 of all crashes in the corridors of Hyderabad is conducted. For ease of analysis, this collected data is further sorted with respect to type of crashes, temporal characteristics, details of accused and victim vehicles, personal details of the victim. The major observations from the preliminary descriptive analysis are as follows: Firstly, 16 % and 84% of all crashes were found to be fatal and non-fatal in nature respectively. Secondly, crashes were found to follow no specific trend across year or month of occurrence of crashes. Thirdly, majority of the crashes were observed to take place in the peak period due to presence of heavy traffic volume. Fourthly, it was observed that two wheelers, pedestrians and the three wheelers are the major victims of the crashes. On the other hand, two wheelers, three wheelers, four wheelers and the buses are found to be major accused vehicles associated with fatal crashes in Hyderabad. The ratio of men is to women involved in crashes were found to be 4:1. Figure 1 presents the overall distribution of crashes with respect to severity and Figure 2 presents the month-wise crash occurrence across Hyderabad during 2011-2016. Table 1 presents the distribution of fatal, non-fatal (major and minor crashes) and total crashes across three types of road geometry namely, signalized, un-signalized and midblock locations spread across Hyderabad. The locations with at-least one fatal crash was selected in the study for crash factor investigation. The dependent variables for formulation of SPF models are total, fatal and non-fatal crashes, which is the total number of each type of crashes occurred at a particular location during 2011-2016.

Table 1: Crash type distribution across signalized, un-signalized and midblock locations

Crash types	**Signalized**		**Un-signalized**		**Midblock**	
	Max	**Min.**	**Max**	**Min.**	**Max**	**Min.**
Total crash	156	4	70	6	90	11
Fatal Crash	29	1	11	1	20	2
Non-fatal crash	127	3	59	5	70	8

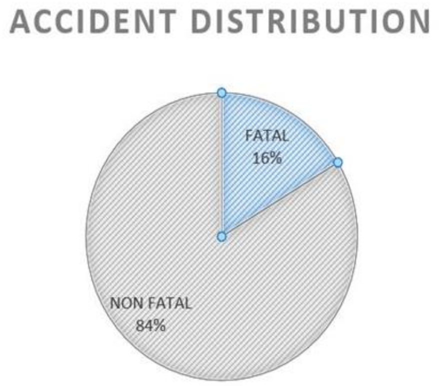

Figure 1 Type of crashes

Figure 2 Month wise Distribution of occurrence of crashes in Hyderabad

4. CRASH FACTOR INVESTIGATION

In this section, the key factors influencing fatal or total crashes are investigated. The occurrence of fatal and non-fatal crashes depends on a number of risk factors which combine and those complex combinations are responsible for these crashes to occur. The crash can be called fatal in case of loss of life and non-fatal (major or minor) in case where the person survives. This crash severity is the most important input for the safety performance function generation. The crash occurrence at any place can only be predicted by the model if the contributing factors play a role more often than not. These occurrences may/may not be completely random.

The crash injury severity is ordered and completely random for all the locations. The ordered discrete model provides two different models: Poisson and Negative Binomial model. Poisson Model can only be used if the mean and variance of the observations are equal and Negative Binomial model can be used otherwise too. The Poisson assumption of crash data restricts mean and variance to be equal (Washington et al. 2011). This equality may not hold in actual situation and the parameter estimate is biased if corrective measures are not taken. The data is over-dispersed when variance is greater than the mean and under dispersed when variance is lesser than the mean (Bandyopadhyaya, 2015). The dispersion in crash data can result from a variety of reasons. The primary reason in many studies is that variables influencing the Poisson rate may have been omitted from the regression. This can be addressed by including a gamma distributed error term in the Poisson parameter. The relationship between explanatory variables and the expected number of crashes per year at the i^{th} segment, can be written as in equation 1.

$$E[n_i] = \lambda_i = e^{([\beta][X] + \varepsilon_i)} \tag{1}$$

` The gamma distributed error term ε_i has a mean 1 and variance α^2, where α is called the over-dispersion parameter. The addition of this term allows variance to differ from the mean and the variance can be written as in equation 2.

$$Var[n_i] = E[n_i] + \alpha\, E[n_i]^2 \tag{2}$$

With this assumption the probability of the i^{th} segment experiencing n_i crashes per year can be said to follow Negative Binomial (NB) distribution. The Poisson regression model is a limiting model of NB model when α approaches 0. The NB distribution of probability of the i^{th} segment experiencing n_i crashes per year can be written as in equation 3.

$$P(n_i) = \frac{\Gamma\left(\left(1/\alpha\right) + n_i\right)}{\Gamma\left(1/\alpha\right) n_i!}\left(\frac{1/\alpha}{\left(1/\alpha\right) + \lambda_i}\right)^{1/\alpha}\left(\frac{\lambda_i}{\left(1/\alpha\right) + \lambda_i}\right)^{n_i} \tag{3}$$

Where, Γ(.) is the gamma function. The model is estimated using maximum likelihood estimation (MLE) technique. The technique attempts to maximise the likelihood that the observed data comes from the hypothesised probability distribution by adjusting the model coefficients (Greene, 2007). The likelihood function is the joint probability of the function at the observation data points and can be written as in equation 4.

$$L(\beta, X, \alpha) = \prod_i \frac{\Gamma\left(\left(1/_{\alpha}\right) + n_i\right)}{\Gamma\left(1/_{\alpha}\right) n_i!} \left(\frac{1/_{\alpha}}{\left(1/_{\alpha}\right) + e^{([\beta][X] + \varepsilon_i)}}\right)^{1/_{\alpha}} \left(\frac{\lambda_i}{\left(1/_{\alpha}\right) + \lambda_i}\right)^{e^{([\beta][X] + \varepsilon_i)}} \quad (4)$$

Given the crash data and the explanatory variables, the parameters β and α can be obtained by maximizing the likelihood function given in equation 4. The parameters are estimated using the software LIMDEP 4.0 (Greene, 2007). Once the estimates of the parameters β and α are obtained, expected crashes λi at a site can be determined. The following section presents the formulation of different safety performance functions for various crash severity types across signalized, un-signalized and mid-block locations.

4.1. Formulation of SPF

Safety Performance Functions (SPFs) would predict the crashes for the crash-prone sites in the future which would help us prevent those crashes by suggesting suitable measures. In this section, the details of the development of SPFs for different types of roads and different crash severities are provided. In this study, SPF's are developed to predict three specific types of crashes, namely, fatal, non-fatal and total (includes both fatal and non-fatal type) crashes at signalized, un-signalized and midblock locations across Hyderabad. As crash occurrence and severity characteristics varies significantly with respect to location of occurrence, different crash prediction models are developed to account for the aforementioned difference. While developing SPF's, an exhaustive set of around fifty variables were used to develop the preliminary negative binomial regression model, however only the statistically significant variables at 95% confidence interval were retained in the model, whereas other variables were excluded.

The negative binomial regression model parameters are estimated using the software LIMDEP 4.0 (Greene, 2007).To generate the SPF, the t-test is useful to determine whether the variables are to be included in the model. The explanatory variables, which are found to be statistically significant at 10% level of significance, are included in the final model. All the individual models are tested for their overall goodness-of-fit. The overall goodness-of-fit of the model is tested with the chi-square test. The test is used to assess whether the inclusion of explanatory variables has improved the performance of the model. The proposed model with explanatory variables is compared against a null model with no explanatory variables. The ρ^2 statistic is used to assess the goodness-of-fit. In the following sections, different SPF's with respect to signalized, un-signalized and midblock locations are presented.

4.1.1. SPF- signalized intersections

In this section, three types of SPF's are developed for signalized intersections and are presented.

SPF for Total crash- Signalized Intersections

In the SPF for total crashes for signalized intersections, average daily volume of NMT and presence of median width greater than 2.5m are found to be statistically significant variables, all other variables were excluded from the initial model. The expected numbers of crashes tend to increase with increase in average daily traffic volume of NMT but decrease with increase in median width. Such findings clearly indicate that inadequate infrastructure for NMT is one of the main reasons behind crashes at signalized junctions in Indian urban context. The results indicate the model ρ^2 statistic to be 0.204 and the χ^2 value of the total crash model is 42.3 with degree of freedom as 2. The ρ^2 statistic of 0.204 indicates that the crash explanatory variables in the model explain the change in the expected crashes reasonably well. Table 2 presents the model estimates for total crashes at signalized intersections.

Table 2: SPF for total crashes at signalized intersection

Attribute	Negative Binomial Regression model		
	Coefficient estimates	t-statistics	p-Value
Constant	1.71	13.121	.00
Log(Volume of NMT)	.616	5.235	.00
Median Width>2.5m	-1.24	-4.613	.00
Over-dispersion parameter α	1.23		
ρ^2	0.204		

SPF for fatal crash- Signalized Intersections

The crash prediction model for fatal crashes for signalized intersections is presented in Table 3. The variables found to be significant in predicting total crashes at intersections are average daily volume of NMT and median width greater than 2.5m at 5% significance level. The expected number of crashes tend to increase with increase in average daily traffic volume of NMT and decreases as median width increases. The model ρ^2 statistic is 0.202 of the formulated model. The ρ^2 statistic of 0.202 indicates that the crash explanatory variables in the model explain the change in the expected crashes reasonably.

Table 3: SPF for fatal crashes at signalized intersection

Attributes	Negative Binomial Regression model		
	Coefficient estimates	t-statistics	p-Value
Constant	.50	1.297	.0194
Log(Volume of NMT)	.78	2.325	.0201
Median Width>2.5m	-2.09	-2.001	.0454
Over-dispersion parameter α	1.21		
ρ^2	0.202		

SPF for non-fatal crash- Signalized Intersections

The SPF model for nonfatal crashes for signalized intersections is presented in Table 4. The variables found to be significant in predicting total crashes at intersections are average daily volume of NMT and median width greater than 2.5m. Co-efficient estimates clearly indicate that NMT volume plays a relatively less significant role towards influencing The expected number of crashes tend to increase with increase in average daily traffic volume of NMT and decreases as median width increases. The model ρ^2 statistic is 0.320 and the χ^2 value of the total crash model is 2.57 with degree of freedom as 1. The ρ^2 statistic of 0.320 indicates that the crash explanatory variables in the model explain the change in the expected crashes reasonably.

Table 4: SPF for non-fatal crashes at signalized intersection

Attributes	Negative Binomial Regression model		
	Coefficient estimates	t-statistics	p-Value
Constant	1.596	6.282	.000
Log(Volume of NMT)	.598	2.098	.0359
Median Width>2.5m	-1.14	-3.858	.0001
Over-dispersion parameter α	1.261		
ρ^2	0.320		

4.1.2. SPF for midblock locations

In this section, three types of SPF's are developed for midblock locations and the results are interpreted for better policy implications

SPF for Total crash- Midblock locations

The estimated SPF for total crashes for midblock locations is presented in Table 5. Average volume of NMT and presence of government offices were found to be negatively influencing the crash occurrence. It could be attributed to the fact that, at midblock locations, presence of NMT modes make drivers more aware of the situation, and they tend to reduce the speed, as a result, probability of crash occurrence reduces. Similarly, presence of government offices near a particular mid-block locations reduce the probability of crashes due to better enforcement, road marking, traffic signs and traffic management system near government offices. The model ρ^2 statistic is 0.302 shows that the crash explanatory variables in the model explain the change in the expected crashes reasonably well.

Table 5: SPF for total crashes at Mid-block

Attributes	Negative Binomial Regression model		
	Coefficient estimates	t-statistics	p-Value
Constant	2.279	8.732	.0000
Log(Volume of NMT)	-.003	-1.344	.0791
Presence of Govt. Office	-.9955	-4.376	.0000
Over-dispersion parameter α	1.21		
ρ^2	0.302		

SPF for Fatal crash- Midblock locations

The crash prediction model for fatal crashes for midblock intersections is shown in Table 6. The variables found to be significant in predicting total crashes at intersections are presence of commercial zones and government offices. The expected number of crashes tend to increase with increase in presence of commercial zones due to increased traffic volume and relatively lower traffic enforcement but decrease due to presence of government offices. The model ρ^2 statistic is 0.303, which is reasonably satisfactory.

Table 6: SPF for Fatal crash- Midblock locations

Attributes	Negative Binomial Regression model		
	Coefficient estimates	t-statistics	p-Value
Constant	0.001	.000	.0000
Presence of Commercial Zone	1.386	2.567	.0103
Presence of Govt. Office	-1.098	-1.017	.0091
Over-dispersion parameter α	1.21		
ρ^2	0.303		

SPF for Non-Fatal crash- Midblock locations

In this particular section, the SPF for non-fatal crashes at mid-block locations were developed. The variables found to be significant in predicting total crashes at intersections are presence of car parking and commercial zones. The expected number of crashes tend to increase with increase in car parking and commercial zones, which are quite expected outcome. The model ρ^2 statistic is 0.318 and the χ^2 value of the total crash model is 24.20 with degree of freedom as 2. The ρ^2 statistic of 0.318 indicates that the independent variables could explain the crash occurrence reasonably satisfactorily. Table 7 presents the co-efficient estimated for the SPF for non-fatal crashes at midblock locations.

Table 7: SPF for non-Fatal crash- Midblock locations

Attributes	Negative Binomial Regression model		
	Coefficient estimates	t-statistics	p-Value
Constant	1.098	3.296	.0010
Presence of Car Parking	.773	2.091	.0365
Presence of commercial zone	.869	3.612	.0003
Over-dispersion parameter α	1.212		
ρ^2	0.318		

Results of SPF for mid-block locations in Hyderabad clearly indicate the effect of land-use such as a government or commercial or a residential zone on crash occurence is quite interesting and unique. A government area being more protected and traffic regulated would have less propensity of being involved in a fatal crash, on the other hand results clearly indicate that presence of a commercial zone, where the heterogeneous traffic movement will be significantly higher has a high chance of being involved in a fatal crash. Such results provide clear indication towards better safety related interventions in commercial zones for reducing number of crashes in mid-block locations.

4.1.2. SPF for un-signalized intersections

This following section presents the Safety Performance functions developed for total crashes at un-signalized intersections across Hyderabad as the models for fatal and non-fatal crashes were not significant. The crash prediction model for total crashes for un-signalized intersections is presented in Table 8. The variables found to be significant in predicting total crashes at intersections are presence of T junction and main carriageway width. The expected number of crashes tend to increase with presence of T junction and as carriage way width increases more than 20.23m. Results indicate that, T-junctions would be more prone to fatal and non-fatal crashes compared to other types of un-signalized junctions. This observation could be attributed to inadequate sight distance at T-junctions in Indian context. The model ρ^2 statistic is 0 .208. The ρ^2 statistic of 0.208 indicates that the crash explanatory variables in the model explain the change in the expected crashes reasonably.

Table 8: SPF for total crash- un-signalized intersections

Attributes	Negative Binomial Regression model		
	Coefficient estimates	t-statistics	p-Value
Constant	1.138	3.161	.0016
Presence of T Junction	.573	1.855	.0636
Main carriageway width	1.068	2.887	.0039
Over-dispersion parameter α	1.232		
Number of observations	14		
ρ^2	0.208		

Based on the various safety performance functions developed for different types of crashes across various locations, a combined table showing various statistically significant risk factors associated with different types of crashes could be presented in Table 9. Based on the findings, the risk factors associated with different types of crash severity could be identified for signalized, un-signalized and midblock locations. For example, the results clearly indicate that with increase in main carriageway width, the crash occurrence severity will increase at un-signalized locations. Hence, better traffic management at wider roads, adequate arrangement for safe NMT movement needs to be ensured to improve the overall safety scenario at city level.

Table 9: Risk factors for different locations across Hyderabad

Risk Factors	Negative Binomial Regression model		
	Signalized junction	Midblock locations	Un-signalized locations
NMT Volume	+	-	NS
Presence of Govt. offices	NS	-	NS
Presence of commercial zones	NS	+	NS
Median width more than 2.5 m	-	NS	NS
Presence of car parking	NS	+	NS
Presence of T Junction	NS	NS	+
Main carriageway width	NS	NS	+
+ indicates that the attribute increase the probability of crash occurrence; - indicates that the attribute would reduce the probability of crash occurrence; NS indicates that the attribute was not found statistically significant at 95% confidence interval			

5. CONCLUSION

Based on the key findings derived from the study we firstly find that the results provide an in-depth understanding on different types of crashes on different locations. This study also explains the reason behind these crashes and how different factors affect the severity of the crashes. Such findings would be useful input for safe planning of urban roadway sections in typical Indian context. Based on the results and findings, a set of concluding remarks.

- Firstly, it could be inferred that for crashes occurring in signalized intersections, the total crashes, fatal crashes and the non-fatal crashes significantly depend on the volume of NMT vehicles and the median width. For mid blocks the total crashes, fatal and non-fatal type crashes depend on different factors such as present of government offices or commercial areas or residential areas etc. For unsignalized intersections, presence of T junctions increase the probability of being involved in a crash. This results clearly indicate the areas to strengthen for improving overall urban safety level.
- Secondly, based on the preliminary investigation of the raw crash data derived from Hyderabad Police it could be clearly concluded that factors such as time of crashes or human errors which caused crashes or socioeconomic profile of accused and victims are not statistically correlated with the crash severity at different locations. However, the authors feel that there is causal relationship between these variables and crash severity data, which requires further investigations.
- Thirdly, it can also be concluded that statistical models such as negative binomial regression models can be successfully applied to assess the safety performance of a particular location and identify the key factors influencing different types of crashes in typical Indian context. However, for better model prediction and inclusion of more number of variables in the model, more accurate and real-time data collection is necessary.
- Finally, based on this study, specific mitigation measures can be suggested to reduce different types of crashes at several locations across Hyderabad city. For example, results clearly indicate that presence of NMT traffic (bicyclists and pedestrians) in heterogeneous

traffic would increase the probability of both fatal and non-fatal crashes. Hence, provision of dedicated bicycle lane, adequate pedestrian crossing, pedestrian refuge, grade separated pedestrian crossing arrangements can be recommended at key locations. Similarly, it was observed that locations with wider median has relatively lower probability of crash occurrence, hence provision of wider medians, glare blockers in narrow medians can be recommended to address this issue.

Likewise, other studies, this study is also not without limitations. One of the major limitations of this study is the limited number of locations in this study, more number of data points are required for improved model prediction. This study has omitted the variables which are not significant at 95% confidence interval, however these could lead to omitted variable bias. In the further study, these limitations would be adequately addressed.

Before closing, we would like to state that even though this thesis has sought to develop a methodological framework for identifying the factors influencing different types of crashes in Indian cities, the results are not tested for transferability across similar cities. Therefore, similar studies could be taken up for a more generalized methodology. Nonetheless, the methodology developed could be used with findings serving as basis for comparison across other cities with similar size and characteristics.

REFERENCES

Abdel-Aty, M., and Wang, X. (2006). Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. *Transportation Research Record: Journal of the Transportation Research Board*, (1953), 98-111.

Bandyopadhyaya, R. (2015). A model Based Method for Hotspot Identification Using Fuzzy Clustering Technique *(Doctoral dissertation, IIT, Kharagpur)*.

Cheng, W., and Washington, S. (2008). New criteria for evaluating methods of identifying hot spots. *Transportation Research Record: Journal of the Transportation Research Board*, (2083), 76-85.

ChikkaKrishna, N. K., Parida, M., and Jain, S. S. (2016). Identifying safety factors associated with crash frequency and severity on non-urban four lane highway stretch in India. Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2016.1150927.

Choi, Y. Y., Kim, D. K., Kho, S. Y., & Park, P. Y. (2019). Safety evaluation of four safety countermeasures on freeways in South Korea. Journal of Transportation Safety & Security, 1-14.

Dinu, R. R., and Veeraragavan, A. (2011). Random parameter models for accident prediction on two-lane undivided highways in India. Journal of Safety Research, 42, 39–42

El-Basyouny, K and Sayed, T. (2010) Safety Performance Function with Measurement Errors in Traffic Volume. Safety Science, Vol. 48, pp. 1339-1344.

Farid, A., Abdel-Aty, M., and Lee, J. (2019). Comparative analysis of multiple techniques for developing and transferring safety performance functions. Accident Analysis & Prevention, 122, 85-98.

Greene, W. (2007). NLOGIT student reference guide table of contents. *Econometric Software, Plainview, NY*.

Hadi, M A, Aruldhas, J, Chow, L S and Wa J A. (1995). Estimating safety effects of cross-section design for various highway types using negative binomial regression. 1500, *Transportation Research Record*, pp. 169-177.

Maher, M J and Summersgill, I. (1996) A comprehensive methodology for the fitting of predictive accident models. 3, *Accident Analysis & Prevention*, Vol. 28, pp. 281–296.

Mitra S, Haque M and King MJ (2017) Effects of access, geometric design, and heterogeneous traffic on safety performance of divided multilane highways in India, *Journal of Transportation Safety & Security*, 9:sup1, 216-235, DOI: 10.1080/19439962.2016.1237600

Mohan, D., Tiwari, G., and Bhalla, K. (2016). Road safety in India: status report. 2015. *Transportation Research & Injury Prevention Programme (TRIPP)*.

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

Murthy, N. S. S. R., and Rao, R. S. (2015). Development of model for road crashes based on intersection parameters using regression models. *International Journal of Scientific and Research Publications*, *5*(1), 1-8.

Raju, P. M. D. S. G. (2013). Identification of crash black spots for national highway using GIS. *International Journal of Scientific & Technology Research*, *2*(2), 154-157.

Robert, R. V., Veeraaragavan, A., and Murthy, K. (2006). Safety relationships for highway segments in developing countries. In Transportation Research Board 85th Annual Meeting (No. 06–0508).

Shankar, V, Mannering, F and Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. 3, *Accident Analysis & Prevention*, Vol. 27, pp. 371-389.

Washington, S.P., Karlaftis, M.G., and Mannering, F.L. (2011). Statistical and econometric methods for transportation data analysis, second ed. *Chapman & Hall/ CRC*, Boca Raton, FL, USA.

Wong, S. C., Sze, N. N., & Li, Y. C. (2007). Contributory factors to traffic crashes at signalized intersections in Hong Kong. *Crash Analysis & Prevention*, *39*(6), 1107-1113.

Yan, X., Radwan, E., & Abdel-Aty, M. (2005). Characteristics of rear-end crashes at signalized intersections using multiple logistic regression model. *Crash Analysis & Prevention*, *37*(6), 983-995.