Safety analysis of urban signalized intersections in Kolkata, India using a combined proactive and reactive approach

Submission Date: August 1, 2016
Words in Text: 6015
Number of Tables/Figures: 6
Total word count (including Tables and Figures): 7515

Sudeshna Mitra*
Associate Professor, Department of Civil Engineering,
Indian Institute of Technology Kharagpur,
Kharagpur - 721302, India,
Phone: +91-9475884524; fax:+9103222-282254
Email: sudeshna@civil.iitkgp.ernet.in

Srinivas R. Geedipally, Ph.D., P.E.
Texas A&M Transportation Institute
Texas A&M University System
110 N. Davis Dr., Suite 101
Arlington, Texas 76013
Phone: (817) 462-0519, fax (817) 461-1239
Email: Srinivas-G@tti.tamu.edu

Dominique Lord
Professor, Zachry Department of Civil Engineering
Texas A&M University, 3136 TAMU
College Station, Texas 77843-3136
Phone: (979) 458-3949, fax: (979) 845-6481
Email: d-lord@tamu.edu

(*corresponding author)
ABSTRACT

Urban signalized intersections in Indian cities report high share of fatalities and serious injuries. However, assessment of safety using rigorous statistical technique is still in early stage in India, primarily due to challenges in obtaining crash data, road geometric and traffic data. This study proposed a methodology combining a proactive and reactive approach by utilizing Road Safety Audit (RSA) principles and statistical assessment techniques, to identify elements posing risk at urban signalized intersections in Kolkata, India. A wide range of factors ranging from the built environment, intersection geometric design and operational parameters are identified and recorded from twenty five sites based on their hypothesised effect from RSA and site visits. Subsequently correlation analysis is performed and Safety Performance Functions are developed using count regression models.

Based on the combined proactive and reactive assessments, there is evidence that increased interaction between motorized and non-motorized traffic pose high risk at signalized intersections. Locations, such as markets and commercial areas, bus stops, and locations where footpaths are occupied encourage such interactions and witness higher number of crashes. There is also evidence through statistical analysis that signalized intersections with wide major roads are associated with more number of crashes. Finally, results indicate that higher volume of traffic from minor roads compared to major road traffic has diminishing effect on crash risk at the signalized intersections. This unique finding can help plan improved traffic management specifically at those signalized intersections where high volume carrying major roads meets smaller minor roads, for improved safety.
INTRODUCTION AND BACKGROUND

At-grade intersections are the most complex locations in the transportation network with the maximum numbers of conflicts. Consequently, the identification of elements posing risk of road traffic crashes and the subsequent elimination or relocation of the same are important steps for the safety management of the road network. In doing that, two broad philosophies are often adopted—a) a proactive approach such as Road Safety Audit (RSA) techniques are used to identify potential risky elements and b) a reactive approach where inferences are made by associating the built environment, elements of intersection design and operations with reported crashes. RSA is conducted by an experienced team having knowledge of road design, operation and safety and does not require collection of historical crash data to suggest safety improvements (although they could be collected for specific RSA projects). This approach is widely adopted by practising engineers in many countries around the world. On the other hand, the second category of analysis has been widely used by the safety research community as it involves lesser subjective judgement and utilises scientific methods to identify a set of factors that may influence safety. The Highway Safety Manual (HSM) provides a useful tool in this analysis; even though it requires collection of a sizable data to make a conclusive inference. Safety Performance Function (SPF) is one such method, which has been extensively discussed in HSM, that has been used by researchers for safety analysis for over two decades.

SPFs are mathematical relationships that associate crash frequencies with geometric design elements, traffic volume and regulatory factors; and are used either for better understanding crash risk or for prediction purpose. A significant proportion of past studies (1 - 8) have focused on modeling accidents at signalized intersections, primarily due to high occurrence of crashes at urban signalized intersections. Ideally, if the SPFs are specified correctly, they should help provide insights into safety by identifying factors explaining highest variability in crash occurrence at urban intersections. Thus, in developing SPFs, researchers first identify a set independent variables having high correlation with the response variable, i.e. frequency of crashes in this case. However, correlation coefficients show linear relation between two variables, and it is possible that two variables may have non-linear relationship between them.

As opposed to developed countries, there have been fewer studies for developing countries where road environment and built environment surrounding roads are very different. This is probably due to the fact that obtaining relevant data for safety assessment and the development of SPF is more challenging in developing nations. This is because the data collection and maintenance are often not done systematically. In India, crash data are still collected and recorded manually using First Information Recording (FIR) form except in three southern states. This requires extensive post processing. Furthermore, police reported data consist of significant underreporting of less injury crashes, such as property damage only and minor injury. Obtaining traffic volume data also requires significant resources, which often limits the size of sample. A small sample of network locations also poses bias in model estimation as was shown by Lord (9). Due to such limitation in data availability in general and crash data in particular, few research organizations in India have made efforts to conduct safety assessment of transport network utilising historical crash data (10). However, because of these issues with data, RSA based safety assessment has gained popularity and momentum and faced lesser challenges in execution. While RSA during the design and construction phases probably has no competing alternative, RSA during operation or RSA findings from existing roads should be
cross checked with historical evidence wherever possible. This demands a need to explore a possible alternative way by combining both reactive and proactive approaches using which reasonable assessment of safety, identification of hazardous elements and future prediction of safety are possible.

To that end, the primary objectives of this study are to determine a) if there exists any significant association between elements of built environment such as land use, infrastructure, geometric design, and traffic operation parameters which are obtained from RSA and site visit with occurrence of total crashes at urban signalized intersections, 

b) quantification of the effects of those factors which influence crash occurrence at urban signalized intersections by developing a SPF.

To achieve these objectives, this study utilises concepts of RSA techniques to identify elements of built environment and design that may pose safety hazard, perform a detailed correlation analysis to identify association of various factors with total crashes at signalized intersections, and finally develop an appropriate crash prediction model to quantify the effects of these factors on safety of signalized intersections. The remainder of the paper is organised as follows: a literature review is given first, followed by the methodology, description of data, results and conclusion.

LITERATURE REVIEW

There have been extensive studies in developed and some in developing countries on safety assessment of signalized intersections. The literature in this regard can be divided into two aspects, a) in terms of modelling methodology used and b) in terms of identification of factors influencing intersection crashes.

Count data models such as Poisson and NB regression models are commonly used in crash data modeling and analysis. Although count data modeling always starts with Poisson regression, Poisson formulation is often not suitable in crash modeling due to the over dispersed (i.e. variance > mean) nature of crash data. Negative Binomial models are generally used for such cases as it includes a gamma-distributed error term in Poisson mean to account for a wide range of unobserved heterogeneity such as omission of relevant variables, measurement error, or just the intrinsic randomness in count data. Both Poisson models (11-13) and negative binomial (NB) models (1, 2, 14-17) have been used to investigate relationships between crash occurrence and intersection characteristics. In developing models, one of the major concerns is estimation bias that may arise from model misspecification and unobserved heterogeneity. As shown by Mitra and Washington (18), in the presence of a well-defined mean function, the extra-variance structure of NB generally becomes insignificant, which helps in better inference by reducing standard error of estimation and forming a narrow confidence interval. Other possible sources of bias are omitted variable bias (Mitra and Washington, 2012, (19) bias due to small sample (9) and bias due to unobserved heterogeneity (20-22) which have been addressed by the safety research community. However, traditional negative binomial regression models are used as a starting point to model overdispersed crash data and it’s variation such as random parameter negative binomial models have been widely used by the researchers to capture unobserved heterogeneity present in the data.

Regarding identification of factors affecting the safety of signalized intersections, previous research on intersection related risk factors (e.g., 1-7) has shown that certain variables including traffic volume (e.g., total volume and right-turn volume), geometric design (e.g., intersection type and number of lanes) and traffic control (e.g., phase number and signal control type) can all have significant impacts
on signalized intersection safety. Factors such as lane width (4), presence of median or median width (2,14), presence of turning lane (8), type of signal and number of phases (3), land-use of the site (14), presence of bus stop (2), volume of entering traffic, i.e. major and minor road volume or some variable to capture the effect of these variables (5-8), have been found to be associated with crash occurrence at signalized intersections either in positive or negative manner. There has also been research on built environment and its effect of road safety (4, 23, 24). Built environment refers to the manmade surroundings that provide the setting for human activity, ranging in scale from buildings, offices and parks or green space to neighbourhoods and cities. Thus built environment is represented by land-use characteristics, transport infrastructure and road network characteristics as well as demographics. The literature suggests that the built environment affects crash frequency and severity through the mediators of traffic volume and traffic speed (23). Findings from developed countries suggest that development patterns in urban built environment impact safety primarily through the traffic volumes they generate, and secondarily through the speeds they encourage, which in turn affect frequency and severity of crashes. Further, not only traffic speed but the speed differentials in the traffic stream and the conflicting traffic movement affect crash occurrence. The conflicts are again greatly dependent on land-use, access management, transit and pedestrian infrastructure. In the context of developing countries, role of design built environment is expected to be significant as certain design encourage greater access, resulting in more interaction between motorized and non-motorized traffic. Higher interaction between motorized and non-motorized traffic also affects safety of motorized vehicles, due to their sudden evasive manoeuvre to avoid collisions with non-motorized traffic. As a result, findings from US, UK, or even China may not be extendable for Indian condition due to the very different traffic conditions characterised by higher interaction between slow and high speed vehicles as well as pedestrians.

**METHODOLOGY**

A four step methodology was followed combining proactive and reactive approach to identify factors affecting the safety of signalized intersections in Kolkata, India. The steps are discussed in this section and are illustrated in Figure 1. It includes site visit to conduct an RSA, identification and collection of elements of road design and traffic operations, performing a correlation analysis and finally development of a regression model to identify factors affecting safety of signalized intersections in Kolkata.
Figure 1 Flow chart showing methodological steps followed in safety assessment of signalized intersections in Kolkata.

Road safety audit

A RSA is a formal procedure for assessing crash potential and safety deficiency of an existing or future road by an independent audit. This is viewed as a proactive low-cost measure to improve safety. RSA is a systematic way of checking the road alignment, various cross sectional elements, road side design and appurtenances as well as traffic operational and management measures that will ensure safety of all road users. Hence, a major aim of RSA of a road under operation is to identify elements that may increase risk of road traffic crashes and elimination and/or relocation of the same. In conducting an RSA of existing road network, certain checklists are followed. Since the work involved in this study focuses on safety of signalized intersections, the list shown in Table 1 was used. Another issue with safety reviews of existing roads revolves around the use of crash data. Some auditors like to have access to the Police crash data for the road they are reviewing. They say it helps them understand some of the proven safety issues along the road. Others argue that this can cause the review team to focus too closely on the crash sites, possibly overlooking other high risk locations. Nonetheless, use of crash data, if available, definitely provides some insight about the underlying safety of the location which otherwise would be missed if they are not available.
### Table 1 Checklist for Road safety Audit for Signalized intersection

<table>
<thead>
<tr>
<th>General</th>
<th>Signal-Controlled Junction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Is the general layout of junction cater safely for all road users including disabled road users?</td>
<td>1) Does the signal sequence conform to the requirements of the regulations and standards?</td>
</tr>
<tr>
<td>2) The type of junction (T-type, staggered, signal controlled, roundabout) suitable for the function of the two or more roads, the traffic volume, the traffic movements (pedestrians and vehicular) and the site constraints? Is it safest alternative?</td>
<td>2) Do the signals clearly indicate which movements are allowed at any one time? Are the timings of various phases of signal cycle adequate?</td>
</tr>
<tr>
<td>3) Is the layout of the junction adequate for all permitted vehicular movements and for all types of vehicles?</td>
<td>3) Are the signals for competing phases located in such a way that they are visible only to the traffic for which they are intended?</td>
</tr>
<tr>
<td>4) Is there adequate provision for channelizing the different streams of traffic? (Check the provision for right turn lanes, deceleration lanes and acceleration lanes?)</td>
<td>4) Are all right turning movements protected as far as possible?</td>
</tr>
<tr>
<td>5) Is adequate provision made for pedestrians and non-motorised vehicles?</td>
<td>5) Does the signing; marking and channelization make it clear to drivers what path they should take through the junction?</td>
</tr>
<tr>
<td>6) Is the provision of night-time lighting adequate, if not what are the deficiencies?</td>
<td>6) Are pedestrian crossings places marked, and are pedestrians channelled to these crossings?</td>
</tr>
<tr>
<td>7) Are junction(s) at that stretch having proper markings, signs and studs to avoid accidents?</td>
<td>7) Whether the pedestrian crossing signal controls are provided where appropriate? If so, there is a need for the crossing movements to be fully protected from conflicting traffic movements for example where there will be serious conflicts with turning traffic.</td>
</tr>
</tbody>
</table>

### Data Collection

Crash data for the 25 signalized intersections were obtained from Kolkata Traffic Police. Kolkata police has divided the entire Kolkata urban area in twenty five traffic guards for better management. Out of these traffic guards, this study focused on selected intersections in nine traffic guards where at least one injury crash occurred between 2011 and 2014. The traffic guards were Beliaghata, Howrah Bridge, Shyambazar, Ultadanga, Vidyasagar Setu, Sealdha, Head Quarter, Joradagan, and South. All 25 signalized intersections had a video camera equipped, from which traffic volume data was extracted.

Regarding land-use and design data, a separate site visit was performed after the preliminary audit of the sites. Collected data include information about road type (arterial or minor road), land-use pattern in percentages (residential, commercial, school zone, government office, park/recreational, open space), pavement condition (good, raveling, depressed side drains, pot hole), geometric features, such as straight or curved road, carriageway width, sight distance (clear/obstructed), percentage of carriageway blocked and width of median. Type of traffic operation, such as one way or two way, junction type such as four legged or three legged were collected. Built environment variables, such as presence of pedestrian facilities, zebra crossing, foot over bridge, subway, presence of metro station, bus stop, pedestrian facility at bus stop, central refuse island, zebra crossing at median opening, zebra crossing width foot path width, percentage of foot path occupied, stop marks, road marking visibility, presence of tram, presence of street light were also assembled. The summary statistics of these variables are given in Table 2.
Table 2 Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total_crashes</td>
<td>1</td>
<td>16</td>
<td>4.56</td>
<td>3.990</td>
</tr>
<tr>
<td>Major_road_carrageway_width (m)</td>
<td>7</td>
<td>11</td>
<td>8.96</td>
<td>1.773</td>
</tr>
<tr>
<td>Minor_road_carrageway_width (m)</td>
<td>7</td>
<td>11</td>
<td>7.30</td>
<td>.968</td>
</tr>
<tr>
<td>Major_road_volume (Veh)</td>
<td>4855.0</td>
<td>38086.0</td>
<td>19380.2</td>
<td>7639.8</td>
</tr>
<tr>
<td>Minor_road_volume (Veh)</td>
<td>3393.0</td>
<td>21380.0</td>
<td>9880.0</td>
<td>4394.9</td>
</tr>
<tr>
<td>Ratio_major_minor_volume</td>
<td>0.2</td>
<td>1.0</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>Ped_volume_peak (number)</td>
<td>546.0</td>
<td>6974.0</td>
<td>2666.0</td>
<td>1461.6</td>
</tr>
<tr>
<td>Busstop (0/1)</td>
<td>0</td>
<td>1</td>
<td>0.20</td>
<td>0.408</td>
</tr>
<tr>
<td>Metrostation (0/1)</td>
<td>0</td>
<td>1</td>
<td>0.12</td>
<td>0.332</td>
</tr>
<tr>
<td>Tram_line_location (0/1)</td>
<td>0</td>
<td>1</td>
<td>0.32</td>
<td>0.476</td>
</tr>
<tr>
<td>Presence_road_marking</td>
<td>0</td>
<td>1</td>
<td>0.28</td>
<td>0.46</td>
</tr>
<tr>
<td>Presence_stop_marking</td>
<td>0</td>
<td>1</td>
<td>0.64</td>
<td>0.49</td>
</tr>
<tr>
<td>Footpath_width (m)</td>
<td>1.86</td>
<td>3.88</td>
<td>2.6936</td>
<td>0.45812</td>
</tr>
<tr>
<td>Footpath_occupancy_propriton (%)</td>
<td>0</td>
<td>.5</td>
<td>0.160</td>
<td>0.1683</td>
</tr>
<tr>
<td>Land_use_residential (%)</td>
<td>0</td>
<td>5</td>
<td>0.244</td>
<td>0.1734</td>
</tr>
<tr>
<td>Land_use_commercial (%)</td>
<td>.1</td>
<td>.8</td>
<td>0.476</td>
<td>0.1964</td>
</tr>
<tr>
<td>Land_use_goverment_office (%)</td>
<td>0</td>
<td>1</td>
<td>0.17</td>
<td>0.214</td>
</tr>
<tr>
<td>Land_use_open_space (%)</td>
<td>0</td>
<td>1</td>
<td>0.06</td>
<td>0.138</td>
</tr>
</tbody>
</table>

Correlation analysis

To start with the modelling process, it is important to know the correlation structure present in the data. For this purpose, linear correlation using Pearson’s correlation parameter test is often used. However, if the data are measured on the ordinal or categorical scale, or any data on continuous scale may not satisfy the requirement of approximate normality (25), Spearman rank correlation parameter should be used. Independent variables that are highly correlated with dependent variable (i.e. crash frequency in this case) should be included in the model and two independent variables with high correlation should not be included together in the model. However, two variables with low correlation coefficient may still be correlated in a nonlinear manner.

Once the variables are identified, first of all variables that have high correlation with dependent variable, i.e total crash frequency are included separately to assess their explanatory power. After that, an incremental model building was adopted by adding and dropping variables to arrive at the final model.

Development of statistical regression model

Over the last two decades, there has been significant research into and improvement of count data modelling methodologies for crash frequency data (18, 21, 26, 27, 28).

In a Poisson regression model, the probability of intersection i having $n_i$ accidents (where $n_i$ is a non-negative integer) is given by:

$$P(n_i) = \frac{\mu_i^{n_i} \exp(-\mu_i)}{n_i!}$$ (1)
Where \( P(n_i) \) is the probability of intersection \( i \) having \( n_i \) accidents and \( \mu_i \) is the Poisson parameter for intersection \( i \), which is equal to intersection \( i \)'s expected number of accidents, \( E[n_i] \). In general, Poisson regression models are estimated by specifying the Poisson parameter \( \mu_i \) (the expected number of events per period such as per year) as a function of explanatory variables \( X_1, X_2, \ldots, X_k \) such that

\[
\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}
\]  

(2)

Here \( \beta \) is the vector of estimable coefficients. In this formulation, the expected number of accidents is given by

\[
E[n_i] = \mu_i = \exp(\beta X_i)
\]  

(3)

Further, Poisson model assume mean and variance equality, i.e. \( \mu_i = \text{VAR}[n_i] \). However, for crash data, this assumption often does not hold, i.e the data either under-dispersed (\( \mu_i > \text{VAR}[n_i] \)) or over-dispersed (\( \mu_i < \text{VAR}[n_i] \)), and the coefficient vector will be biased if inequality of variance is not taken into account (Mitra and Washington, 2007).

Cameron and Trivedi (29) have proposed a number of tests to check for under or overdispersion in the Poisson regression model. The testing framework is built around the following hypotheses (30).

**Null hypothesis:** \( H_0 : \text{VAR}[y_i] = \mu_i \)  

**Alternative hypothesis:** \( H_1 : \text{VAR}[y_i] = \mu_i + \alpha g(\mu_i) \)  

(4)

(5)

One of the main assumptions behind this test is that the Poisson model gives consistent estimates of \( E[y_i] = \mu_i \) and result in a model where \( E[y_i] = \text{VAR}[y_i] \). However, in case of overdispersed count data Negative Binomial (NB) model is used, which is derived by adding an additional term in variance as shown in equation 5. Here, \( \alpha \) is the overdispersion parameter and follow a gamma distribution. The test is carried out by testing the significance of the single coefficient in the linear OLS regression of

\[
z_i = \frac{(y_i - \mu_i)^2 - y_i}{\mu_i \sqrt{2}}
\]  

(6)

on

\[
w_i = g(\mu_i)/\mu_i \sqrt{2}
\]  

(7)

Two possibilities are suggested:

\[
g(\mu_i) = \mu_i \quad \text{and} \quad g(\mu_i) = \mu_i^2
\]  

(8)

A significant \( \alpha \), \( g(\mu_i) \) indicates that if \( \alpha \) and \( g(\mu_i) \) are significant, and that \( H_0 \) should be rejected in favor of \( H_1 \). This implies that the count is overdispersed and that negative binominal regression should be adopted.

Finally, to measure the overall goodness of fit, the log-likelihood ratio index is calculated. For count data models, the common practice is to use a “pseudo” \( R^2 \) statistic, which is often known as log-likelihood ratio index \( \rho^2 \). According to Washington et al. (2010), \( \rho^2 \) is given by

\[
\rho^2 = 1 - \frac{L(\beta)}{L(0)}
\]  

(9)

**RESULTS AND DISCUSSION**

Based on the RSA, the major element of risk identified was high interaction of motorized and non-motorized traffic, such as pedestrians, especially near market areas, commercial areas and bus stops. Road conditions were also found to be poor in these locations as opposed to locations where government offices are present. Road
marking was another major problem primarily in residential areas. Signalized intersections in residential areas were also found to have narrower roadway and sidewalk width. A common observation across all locations is poor usage of marked pedestrian crossing. Parking by the side of road was observed only at commercial areas. Major observations from RSA are shown in Figure 2.

Figure 2 Findings from Road Safety Audit across various study intersections.
The statistical assessment in this study started with a correlation analysis using Spearman Rank Correlation test to estimate correlation among the independent variables and that with the dependent variable. The results are shown in Table 3. As can be seen, the total crashes at signalized intersection was highly correlated with presence of bus stop (0.55, p=<0.01) and width of major road (0.418, p<0.05). On the other hand, the correlation with minor road width (-0.09, p>0.2) was weak with total crashes. Traffic volume from major road, minor road, and variables capturing the effect of built environment, such as residential and commercial areas had some correlation with the dependent variables, but none of them were significant at least at 5% level of significance. Further, none of the independent variables had statistically significant high correlation among them, expect the width of minor road and residential land-use (0.5, p<0.01). But these two variables were not found to be significant in the final model. A variable capturing the ratio of minor road volume to major road volume was also created and this variable definitely has high correlation with major road and minor road volume, which dictates that the major road volume and the ratio of major to minor road volume should not be included at the same time in model building. Variables capturing the effect of built environment, such as footpath width and their occupancy showed very interesting results. Footpath widths were found to be statistically significantly wider in areas where government offices are located (0.62, p<0.01), but narrow in residential areas (-0.58, p<0.01). Footpath width is also negatively correlated with commercial areas, but the correlation is not statistically significant. On the other hand, it was interesting to identify that footpath occupancy (occupied illegally by street vendors) is statistically significantly higher in commercial areas (0.6, p<0.01) and significantly lower in areas where parks (-0.5, p<0.01) are located. Narrow sidewalk width and their high encroachment are factors posing risk to non-motorized road users in commercial areas. Presence of road marking was found to be high in areas where government offices are located compared to residential and commercial areas, but none of these correlations were statistically significant. Finally, the variable capturing traffic volume- the ratio of minor road traffic to major road traffic, found to have high correlation with minor road carriageway width (0.41, p<0.05). As a result, these two variables are not included together in regression model.
<table>
<thead>
<tr>
<th></th>
<th>Total crashes</th>
<th>Major_road_carrigeway_width</th>
<th>Major_road_width</th>
<th>Minor_road_carrigeway_width</th>
<th>Minor_road_width</th>
<th>Ratio_major_minor_volume</th>
<th>Ped_volume_peak</th>
<th>Busstop</th>
<th>Metrostation</th>
<th>Presence_stop_marking</th>
<th>Presence_road_marking</th>
<th>Footpath_width</th>
<th>Footpath_occupancy_proportion</th>
<th>Land_use_residential</th>
<th>Land_use_commercial</th>
<th>Land_use_goverment_office</th>
<th>Land_use_park</th>
<th>Land_use_open_space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total crashes</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major road carraigeway width</td>
<td>.418</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minor road carraigeway width</td>
<td>-.09</td>
<td>.327</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major road volume</td>
<td>.203</td>
<td>-15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minor road volume</td>
<td>-.09</td>
<td>.367</td>
<td>.32</td>
<td>.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio major minor volume</td>
<td>-.11</td>
<td>.089</td>
<td>.41</td>
<td>-.5</td>
<td>.6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ped volume peak</td>
<td>.137</td>
<td>.263</td>
<td>-.08</td>
<td>.31</td>
<td>.01</td>
<td>-.32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Busstop</td>
<td>.55</td>
<td>.242</td>
<td>.14</td>
<td>.33</td>
<td>.26</td>
<td>.07</td>
<td>.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrostation</td>
<td>.104</td>
<td>.327</td>
<td>.60</td>
<td>-.08</td>
<td>.38</td>
<td>.29</td>
<td>.31</td>
<td>.12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tram line location</td>
<td>.072</td>
<td>-.23</td>
<td>.30</td>
<td>-.26</td>
<td>-.30</td>
<td>.29</td>
<td>-.12</td>
<td>.30</td>
<td>.27</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence road marking</td>
<td>.106</td>
<td>.014</td>
<td>.05</td>
<td>-.18</td>
<td>-.22</td>
<td>.10</td>
<td>-.12</td>
<td>.13</td>
<td>-.04</td>
<td>.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence stop marking</td>
<td>.257</td>
<td>.17</td>
<td>.04</td>
<td>.12</td>
<td>-.06</td>
<td>.01</td>
<td>-.02</td>
<td>.38</td>
<td>.02</td>
<td>.16</td>
<td>.46</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footpath width</td>
<td>-.24</td>
<td>.10</td>
<td>.49</td>
<td>-.29</td>
<td>.13</td>
<td>.33</td>
<td>-.21</td>
<td>.09</td>
<td>.25</td>
<td>.08</td>
<td>.26</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Footpath occupancy proprtion</td>
<td>-.05</td>
<td>-.14</td>
<td>.09</td>
<td>.14</td>
<td>.21</td>
<td>.05</td>
<td>.33</td>
<td>-.12</td>
<td>.42</td>
<td>.50</td>
<td>-.15</td>
<td>.2</td>
<td>-.3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use residential</td>
<td>.186</td>
<td>.17</td>
<td>-.5</td>
<td>.15</td>
<td>-.26</td>
<td>-.36</td>
<td>.19</td>
<td>.00</td>
<td>-.30</td>
<td>-.18</td>
<td>-.09</td>
<td>.0</td>
<td>-.58</td>
<td>.13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use commercial</td>
<td>.177</td>
<td>-.09</td>
<td>-.06</td>
<td>.12</td>
<td>.12</td>
<td>-.06</td>
<td>.17</td>
<td>-.01</td>
<td>.27</td>
<td>.37</td>
<td>-.21</td>
<td>-.06</td>
<td>-.23</td>
<td>.6</td>
<td>.04</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use government office</td>
<td>-.09</td>
<td>.07</td>
<td>.55</td>
<td>-.12</td>
<td>.10</td>
<td>.31</td>
<td>-.05</td>
<td>.19</td>
<td>.21</td>
<td>.07</td>
<td>.36</td>
<td>.10</td>
<td>.620</td>
<td>-.26</td>
<td>-.06</td>
<td>-.42</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Land use park</td>
<td>-.292</td>
<td>-.16</td>
<td>-.18</td>
<td>-.29</td>
<td>-.24</td>
<td>-.10</td>
<td>-.20</td>
<td>-.25</td>
<td>-.18</td>
<td>-.34</td>
<td>.19</td>
<td>-.20</td>
<td>.26</td>
<td>-.5</td>
<td>-.18</td>
<td>-.51</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Land use open space</td>
<td>.005</td>
<td>-.07</td>
<td>-.20</td>
<td>-.06</td>
<td>-.21</td>
<td>-.01</td>
<td>-.21</td>
<td>.28</td>
<td>-.21</td>
<td>-.38</td>
<td>.00</td>
<td>-.19</td>
<td>-.04</td>
<td>-.35</td>
<td>-.19</td>
<td>-.34</td>
<td>-.03</td>
<td>.14</td>
</tr>
</tbody>
</table>

*significant AT 95% C.I  **significant at 99% C.I
The next step was to test for over-dispersion in crash data for which the test suggested by Cameron and Trivedi [29] was conducted and the result indicated presence of over-dispersion in crash data (Refer Table 4). After this diagnosis, individual variables are added separately to check their statistical significance in explaining the dependent variable. In doing that, the overall model prediction capability by these variables as well as the variable’s ability to correctly specify the model was tested. The results of this exercise are given in Table 4. As can be seen in Table 4, the model explanatory power by adding the variable “Busstop” was good, but the dispersion parameter resulted in an insignificant estimate. Further, the test for over-dispersion indicated absence of over-dispersion in the data, which violates the fact. This is probably due to large variability in the variable bus-stop as it is present only in 20% of locations in the dataset. Further, due to small sample size, it was not possible to drop the locations with bus stop and rerun a model to check the effect of other variables. While bus stop was decided to be excluded for model prediction purpose, the presence of bus stops definitely increases the number of conflicts among the mixed traffic when vehicular speeds are more heterogeneous, as was pointed by Wong et al. [4]. Field observation of the same shows very similar evidence. In Kolkata city, most bus stops are classified as nearside just ahead of signalized intersections as can be seen in Figure 2. In most cases, passengers wait on the road and there is no bus bay present at the site. As a result, there is higher interaction with passengers waiting on the road and with other motorized vehicles, which definitely pose high safety hazard. Hence, even though this variable was not included in the final model due to data limitation; there is evidence that this factor related to design built environment of Kolkata city poses hazard at the signalized intersection. This conclusion was achieved by a combined correlation analysis (Refer Table 3) and by safety audit of the bus stops located at signalized intersections in Kolkata city.

The other variables, such as width of major road, residential and commercial land-use, ratio of minor road and major road traffic volume were all found to be significant when added independently and they all resulted in correct model specification. Once this preliminary investigation was performed, these independent variables are incrementally added or deleted and the final model resulted in a negative binomial model denoted as Model 7 and the results are shown in Table 4. Here it is important to mention that advanced model such as random parameter negative binomial model was not used due to small sample size. Results show that Commercial area ($\beta = 1.19, p < 0.100$) and wider width of major road ($\beta = 0.228, p < 0.006$) increase total crashes whereas a higher ratio of volume of minor road traffic compared to major road traffic volume ($\beta = -0.766, p < 0.118$) decreases total crashes at signalized intersections.
### Table 4 Results from count data models

<table>
<thead>
<tr>
<th>Attributes</th>
<th>NB regression Model-1 (Constants only)</th>
<th>NB regression Model-2 (bus-stops)</th>
<th>NB regression Model-3 (Residential)</th>
<th>NB regression Model-4 (commercial)</th>
<th>NB regression Model-5 (Width)</th>
<th>NB regression Model-6 (Volume Ratio)</th>
<th>NB regression Model-7 (Combined)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.51 (9.45)</td>
<td>1.19 (7.66)</td>
<td>1.1 (3.86)</td>
<td>1.06 (2.51)</td>
<td>-0.45 (-0.55)</td>
<td>1.94 (4.92)</td>
<td>-0.79 (-0.8)</td>
</tr>
<tr>
<td>Bus stops</td>
<td>1.07 (3.74)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial</td>
<td>1.56 (1.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Carriageway Width</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio of minor road traffic to major road traffic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.777 (-1.19)</td>
<td>-0.766 (-1.35)</td>
<td></td>
</tr>
<tr>
<td>Results from overdispersion Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g=μ(i):</td>
<td>2.19</td>
<td>1.49</td>
<td>2.63</td>
<td>2.03</td>
<td>2.48</td>
<td>2.12</td>
<td></td>
</tr>
<tr>
<td>g=μ(i)^2:</td>
<td>2.19</td>
<td>1.33</td>
<td>3.11</td>
<td>2.03</td>
<td>2.88</td>
<td>2.21</td>
<td></td>
</tr>
</tbody>
</table>

**Dispersion parameter for count data model**

| Alpha (α) | 0.426 (2.41) | 0.18 (1.55) | 0.36 (2.25) | 0.39 (2.23) | 0.31 (2.11) | 0.39 (2.33) | 0.23 (1.85) |

**Overall Goodness-of-fit**

| Log-Likelihood function | -63.15 | -57.59 | -61.71 | -62.52 | -60.47 | -62.46 | -58.39 |
| Restricted Log-Likelihood function | -74.89 | -74.89 | -74.89 | -74.89 | -74.89 | -74.89 | -74.89 |
| \( \rho^2 \) | 0.157 | 0.231 | 0.175 | 0.165 | 0.193 | 0.166 | 0.22 |
The variable percentage of commercial area near the signalized intersection is another variable capturing the effect of design built environment where a signalized intersection is located. Generally, commercial areas are characterized by increased access which increases frequent interaction of vehicles. These locations also experience high road side commercial activities. Further, the results from the correlation analysis indicated that commercial areas in the dataset are also associated with high footpath occupancy, which probably forces pedestrians to walk and engage in commercial activities along the road, resulting in frequent interaction between pedestrian and motorized vehicles, leading to higher risk of crashes.

The variable carriageway width of major road or simply width of major road found to be highly significant and positively associated with total crashes. This is not surprising as higher road width encourages higher speed and greater freedom in vehicle maneuver which leads to poor safety. Signalized intersections in Kolkata are found to be no exception and this finding empirically proves that wider roads in urban area pose safety hazard. However, width of minor road was not statistically significant and in the correlation analysis, a wider minor road width was actually found to be negatively associated with total crashes, even though the variable was not statistically significant. Effect of width of road however, found to be mixed in earlier research.

The most interesting finding in this study is the effect of traffic volume. From the correlation analysis, while the traffic volume from major road was somewhat correlated with total crashes in a positive manner, traffic volume from minor roads had very low and negative correlation with total crashes. After incorporating the logarithmic transformation of traffic volume, only major road traffic volume showed some significance and minor road traffic volume was not at all significant. However, in road traffic safety literature, major and minor road traffic volumes are always considered to be the most important variables in intersection crash prediction as they capture exposure. With this in mind, a different form of total entering volume, the ratio of minor road traffic to major road traffic was included in the model and resulted in negative correlation with total crashes. While it may appear to be strange, the width of minor road and the minor road traffic volume were consistently found to be negatively associated with total crashes even though they were not significant. Thus, the effect of minor road traffic compared to the major road traffic was tested and the result indicates that signalized intersections in Kolkata with higher minor road traffic compared to major road volume are safer than intersections where less volume carrying minor road meets with a major road carrying higher volume of traffic. This probably is a surrogate capturing the priority of a signalized intersection from the traffic operations and management point of view, which is quite localized and unique for many Indian cities like Kolkata. Safety Audit and site visit shows that major signalized intersections (often intersections of two roads of equal widths) are better managed with the help of traffic police who enforce and ensure orderly movement of traffic than smaller intersections. As a result, this finding is not against expectation but is a revelation made possible by statistical modeling. While the other two findings indicate safety effects of built environment and design, this finding can directly help Kolkata police to further review safety of such signalized intersections and plan for better traffic management.

Finally, it is important to point out that even though only three variables were found to be statistically significant in the final model, the casual contribution to crashes at intersections may not be fully inferred from a regression analysis, as pointed by Wong et al. (4). This is more so when the sample size is small. However,
due to practical difficulty of data collection and obtaining crash data in a country
where systematic crash data collection is absent, this study is an attempt to study
associations between signalized intersection related variables and total crashes.
Identification of risky elements associated with urban signalized intersection through
the correlation analysis, safety audit by site visit and finally findings from regression
model provide some reasonable directions for improvement and further exploration.

CONCLUSION

Road traffic safety in urban India is a matter of serious concern. This is particularly
true in major cities such as Kolkata. However, due to the absence of systematic crash
data collection and maintenance, there have been limited studies on traffic safety of
urban India. With the exception of only one study (31), existing studies (32) on road
safety in the context of urban India either follow road safety audit techniques or
mostly a descriptive approach where overall trends of safety are documented.
However, no studies to-date focused on spatial location of the intersection by
capturing the elements of built environment, geometric design and traffic operations
on safety in the Indian context. In that sense, this study makes a unique attempt to
collect traffic exposure and a wide range of design built and facility design related
factors to investigate their effects on safety of signalized intersections in Kolkata,
India.

While safety assessment of road network has been widely conducted using
 crash data analysis and development of crash prediction models, this study has
proposed a methodology combining a proactive and reactive approach by utilising
Road Safety Audit principles and statistical assessment techniques involving
correlation and regression analysis. For this purpose, crash data from Kolkata city was
collected and a total of twenty five signalized intersections were visited for safety
audit. Videographic data were also obtained for these intersections from Kolkata
Police.

Based on the combined proactive and reactive analyses it was concluded that
increased interaction between motorized and non-motorized traffic is definitely a
major cause of risk at signalized intersections. Locations, such as markets and
commercial areas, bus stops, and locations where footpaths are occupied encourage
such interactions and evidence from either correlation analysis or statistical model
also show similar results. As was mentioned earlier, presence of bus stop was highly
significant both from correlation analysis and preliminary count model, but this
variable had to be dropped due to its small sample size affecting model specification.
However, commercial built-up areas was found to be significant and included in final
model for prediction purpose. The other important finding through statistical analysis
was width of major road, which positively affected total crashes. Finally, ratio of
traffic from minor and major road provides an interesting insight into safety of the
selected intersections in this study.

Like any other study, this study is not without limitation. A major concern was
small sample size, which could affect estimation. Further, model validation and
transferability for other intersections in the Kolkata have not been tested yet. Thus, an
extended study is required to address these concerns by reinvestigating crash
prediction model for signalized intersections in Kolkata by incorporating more sites
and road users’ behavior. Nonetheless, this study provides important insights into the
safety of signalized intersections in a major urban area in India.

ACKNOWLEDGEMENT
The authors would like to acknowledge the help of Mr. T. Subudhi, Mr. Sanil Jain and Mr. Satyajit Rana for data collection and processing.

REFERENCES


