Crash Prediction for Multilane Highway Stretch in India

Naveen Kumar CHIKKAKRISHNA [a], Manoranjan PARIDA [b], Sukhvir Singh JAIN [c]

[a,b, c] *Indian Institute of Technology Roorkee, Roorkee, Uttarakhand - 247667, India.*
[a] *E-mail: cnaveenkh@gmail.com*
[b] *E-mail: mparida@gmail.com*
[c] *E-mail: profssjain@gmail.com*

Abstract: This paper documents the application of Bayesian modeling techniques for road traffic crash analysis on a sample of Indian National Highways. Poisson-Gamma Hierarchical Bayes and Poisson-Weibull Bayesian models were applied to the collected crash data.

Explanatory variables were Geometric Characteristics like Median Opening (MedOpn), Access Roads to main highway (AcsRds) and Traffic Characteristics like Average Daily Traffic (ADT) and road-side developments like Industrial (Ind), Commercial (Com), Residential (Resi) and School (School) were analyzed against dependent variable as crash count per two hundred meter per year. The results of this study show that Poisson-Gamma hierarchical model best predicts the crashes with higher accuracy compared to other technique implemented. Traffic volume, Access Roads and Median opening emphasizes on increase in the probability of occurrence of crashes.

Keywords: Crash Prediction Models, Bayesian Technique, Poisson - Gamma and Poisson - Weibull Models.

1 INTRODUCTION

Road safety is a socio-economic problem, leading to tremendous life and property loss. To improve road traffic safety, comprehensive understanding of traffic system safety is always emphasized. Road traffic is a complicated system which may be affected by a diversity of risk factors representing environmental, road geometric, traffic, driver, and vehicle characteristics. The understanding of traffic system safety may be further obscured since crash occurrences are discrete events, often irregular and random events. Hence, obtaining unbiased and relatively accurate estimation and prediction of traffic system safety has become a major concern in road safety management.

Crash prediction models are one of the most important techniques in investigating the relationship between crash occurrence and risk factors associated with various traffic entities. These risk factors are assumed to provide information on the behavior of crash occurrence, which is commonly measured by crash frequency with various degrees of crash severity. Appropriate probabilistic forms and statistically significant factors are identified based on the examination of crash occurrence mechanism and model fitting performance to the historical crash data. As a result, the safety variability associated with traffic entities is modeled by risk factors identified and noise/error terms used to account for hidden or unobserved safety-related features. So better the capture of these errors, better will the model performance and consequently better the safety estimation and prediction. Hence, apart from exploring risk factors, to address modeling errors is a major challenge for safety modelers. Unfortunately in current road safety research, effort is significantly insufficient towards a better understanding of these errors.

The mixed traffic on Indian highways comes with a lot of variability within, ranging from difference in vehicle types to variability in driver behavior. This could result in variability in the effect of explanatory variables on crashes across locations. To properly model the potential heterogeneities due to the multilevel data structure, Bayesian hierarchical approach can be used. This approach explicitly specifies multilevel structure and reliably yields parameter estimates. A Bayesian calculation combines prior information and current information to derive an estimate for the expected safety of a site that is being evaluated. In the context of accident analysis, the prior information is the expected accident frequency from a group of similar sites and the current information is the site-specific observed accident frequency.

1.1 Global Scenario of Road Safety

As per the Commission for Global Road Safety (2009), road traffic accidents kill an estimated 1.3 million people and injure 50 million people per year globally. Further, the global road fatalities are forecasted to reach 1.9 million by 2020. It is estimated that the number of deaths from road accidents in Asia is about 700,000 per year, accounting for more than half of the world’s road fatalities even though Asia accounted for only 43% of the global vehicle population in 2007. In view of above, the importance of road safety studies are being felt all around the world.

1.2 Road Safety in Indian Context

With rapid increase in industrialization, motorization and multilane highways in India; the casualties due to accidents on the roads are increasing alarmingly year by year. The dominance of road transport will continue in India as it has in the rest of the world. The share of the movement of both passengers and goods is expected to increase further in the coming years with the full implementation of the current road development programs being undertaken in the country. Most of these high-speed road development programs are expected to be completed by 2015.

Studies of the relationship between gross domestic product (GDP) per capita, growth of motor vehicles and road fatalities have shown that fatality rate increases as GDP increases at relatively low levels of GDP per capita, but then start to decline with continued GDP growth. The peak position on this inverted U-shaped curve is not, however, immutable.

At present more than 600,000 accidents occur annually with about 147,000 people losing their life in these accidents. Besides fatalities, nearly 600,000 people suffer injuries in road accidents which lead to lifelong misery for the victims and their families. Road Safety Study can ensure that various safety deficiencies in road are reviewed so that these can be taken care at appropriate stage of road design or operation in a cost effective way.

1.3 Literature Survey

There has been a significant amount of research conducted on the safety performance of rural highways. Most of this research has been concentrated on rural two-lane highways (Vogt and Bared, 1998; Harwood et al., 2000; Qin et al., 2004). This focus is not surprising given the fact that the rural network is composed mainly of two-lane highways. Consequently, there has been very little research conducted on the safety performance of multilane highways in rural areas.

There have been few studies that examined the safety performance of multilane rural highways in North America. Persaud (1991) investigated the safety performance of rural and urban multilane highway segments in Ontario, Canada. To accomplish this task, Persaud developed several statistical models and used the Empirical Bayes (EB) method to refine (or improve) the estimated long-term safety performance of these facilities. He found that statistical models predicted more collisions for rural multilane highways than urban segments for the same level of exposure and number of lanes. As expected, divided rural segments performed better than undivided segments.

Council and Stewart (1999) developed statistical models for rural four-lane undivided and divided highways. The models were developed as part of a cross-sectional study for comparing the safety performance between rural two-lane and four-lane highways. Council and Stewart used data collected in California, Minnesota, North Carolina, and Washington. The models were developed using traffic flow and shoulder width as input variables. Larger shoulder widths were associated with fewer crashes for divided four-lane highways.

Persaud and Bahar (2000) investigated the use of statistical models for screening high-risk sections of rural highways in Ontario, Canada. They developed a Potential for Safety Improvement Index (PSI) for identifying these sites. Several statistical models were produced, including models for rural divided and undivided four+-lane highways. Similar to the study conducted by Persaud (1991), the models predicted more crashes for undivided than divided rural multilane highways.

In a study conducted in Italy, Caliendo et al. (2007) developed a series of statistical models to estimate the effects of geometric design features on the safety performance of multilane rural highways; tangents and curves were modeled separately. Three model types were estimated: the Poisson model, the NB model, and the NB multinomial, which is a model where the dispersion parameter is modeled as a function of the segment length.

2 MODELLING METHODOLOGY

2.1 Poisson - Gamma Model

The Poisson-Gamma model has the following model structure by Lord (2006): the number of crashes 'Y_{it}' for a particular i^{th} site and time period t when conditional on its mean μ_{it} is Poisson distributed and independent over all sites and time periods.

$$Y_{it} \mid \mu_{it} \sim Po(\mu_{it}) \quad \text{i} = 1,2,.....,\text{I} \text{ and } \text{t} = 1,2,.....,\text{T} \tag{1}$$

The mean of the Poisson is structured as:

$$\mu_{it} = f(X;\beta)\exp(e_{it})$$

where,

$f\ (.)$ is a function of the covariates (X);

β is a vector of unknown coefficients; and,

e_{it} is the model error independent of all the covariates.

With this characteristic, it can be shown that Y_{it}, conditional on μ_{it} and φ, is distributed as a Poisson-Gamma random variable with a mean μ_{it} and a variance $\mu_{it} + \frac{\mu_{it}^2}{\varphi}$, respectively.

The probability density function (PDF) of the Poisson-Gamma structure described above is given by the following equation:

$$f\left(y_{it};\varphi,\mu_{it}\right)=\begin{pmatrix} y_{it}+\varphi-1 \\ \varphi-1 \end{pmatrix}\left(\frac{\varphi}{\mu_{it}+\varphi}\right)^{\varphi}\left(\frac{\mu_{it}}{\mu_{it}+\varphi}\right)^{y_{it}} \tag{2}$$

where,

y_{it} = response variable for observation i and time period t;

μ_{it} = mean response for observation i and time period t; and,

φ = inverse dispersion parameter of the Poisson-Gamma distribution.

Note that if $\varphi \to \infty$, the crash variance equals the crash mean and this model reverts back to the standard Poisson regression model.

The term φ is usually defined as the "inverse dispersion parameter" of the Poisson-Gamma distribution. (Note: in the statistical and econometric literature, $\alpha = 1/\varphi$ is usually defined as the dispersion parameter; in some published documents, the variable α has also been defined as the "over-dispersion parameter."). This term has traditionally been assumed to be fixed and a unique value applied to the entire dataset in the study. As discussed above, recent research in highway safety has shown that the dispersion parameter can potentially be dependent upon the covariates of the model and could vary from one observation to another (Lord *et al.*, 2005; Hauer, 1997).

2.2 Poisson – Weibull (PW) Model

The PW distribution is a mixture of Poisson and Weibull distribution, as the name implies. Similar to most Poisson-based distributions (e.g., Poisson-gamma, Poisson-lognormal, etc.), the PW model is also designed to accommodate the over-dispersion (Raghavachari *et al.*,1997; Lord *et al.*, 2005; Maher *et al.*, 2009). First, the Poisson and Weibull distributions need to be defined, respectively. Crash data can be characterized as the product of Bernoulli trials with unequal probability of events (also known as Poisson trials). As the number of trials increases the distribution may approximately follow a Poisson process and the amount of dispersion is governed by the characteristics of this process. Thus, the number of crashes at i^{th} entity Y_i is assumed to be Poisson distributed with mean and independent over all entities:

$$Y_i \mid \mu_i \sim Poisson \quad (\mu_i) \qquad i = 1,2,3...\mathrm{I}$$

The Poisson mean μ_i is structured as:

$$\mu_i = \hat{\mu}_i \varepsilon_i = f(X;\beta).\varepsilon_i$$

And,

$$\hat{\mu}_i = f(X;\beta) = \exp(\beta_0 + \textstyle\sum_{j=1}^{q} \beta_j X_j),\ j = 1,2,3...q \tag{3}$$

Where, Xs are the independent variables; represents the total number of independent variables; βs are the regression coefficients; and ε_i is the model error independent of all covariates (7-8).

For the PW model, it is assumed that ε_i is independent and Weibull distributed. This is used for capturing the extra variation that the traditional Poisson cannot fully handle. The Weibull probability density function (p.d.f) is given as follows:

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{x}{\lambda}\right)^{k}\right] \qquad x > 0, k > 0, \lambda > 0 \tag{4}$$

Where λ and k are scale and shape parameters, respectively. The p.d.f. of the Weibull distribution has a wide variety of shapes depending on the k values and the shape can be similar to that of the gamma, gamma-like, exponential or approximate normal distributions. This characteristic gives the model a lot of flexibility to fit different kinds of data.

The mean and variance of the Weibull distribution are:

$$E(\varepsilon) = \lambda\left(1+\frac{1}{k}\right)$$

$$Var(\varepsilon) = \lambda^2\Gamma\left(1+\frac{2}{k}\right) - \left[\lambda\Gamma\left(1+\frac{1}{k}\right)\right]^2$$

Given the statistical characteristics of Poisson and Weibull distributions, the PW distribution is defined as the mixture of those two distributions such that:

$$P(Y=y;\mu,\lambda,k) = \int Poisson(y;\hat{\mu}\varepsilon)Weibull(\varepsilon;\lambda,k)d\varepsilon \tag{5}$$

The mean or expected value of the PW distribution is given as:

$$E(Y) = \hat{\mu}E(\varepsilon) = \hat{\mu}\times\lambda\Gamma\left(1+\frac{1}{k}\right)$$

and the variance is given by:

$$Var(Y) = \hat{\mu}\times\lambda\Gamma\left(1+\frac{1}{k}\right) + \hat{\mu}^2\times\lambda^2\Gamma\left(1+\frac{2}{k}\right) - \hat{\mu}^2\times\left[\lambda\Gamma\left(1+\frac{1}{k}\right)\right]^2$$

2.3 Goodness of Fit Statistics

Likelihood Statistics

There are many measures that can be used for estimating how well the model fits the data. There are statistics for indicating the likelihood level of a model, that is, how well the model maximizes the likelihood function. Among these statistics are:

2.3.1 Pearson chi-square

Another useful likelihood statistic is the Pearson Chi-square and is defined as

$$\text{Pearson} - \chi^2 = \sum_{i=1}^{N}\frac{(y_i-\hat{\mu}_i)^2}{VAR(y_i)} \tag{6}$$

2.3.2 Bayesian model selection - Deviance Information Criterion (DIC)

The DIC (9) set in WinBUGS was used as the criterion for comparing the different Bayesian hierarchical models, since the DIC includes a penalty for the complexity of the model. The DIC for the *jth* model is given by:

$$DIC_j = D(\bar{\theta}_j) + 2pD_j = \bar{D} + pD_j \tag{7}$$

where $D(\bar{\theta}_j)$= the deviance $D(\bar{\theta}_j|y)$ at the posterior mean $\bar{\theta}_j$ of the parameters for model j, called Dhat in WinBUGS, $\bar{D}$= the expected devaiance $\bar{D} = E(D|y,\theta)$, given by the mean $\bar{D}$ of the sampled deviances D(t) from Markov Chain Monte Carlo simulations, also called Dbar in WinBUGS, and pDj = the effective number of parameters of the model, computed as the difference between $\bar{D}_j$ and $D(\bar{\theta}_j)$, namely, pDj = $\bar{D}_j$ - $D(\bar{\theta}_j)$.

In comparing two models, a difference of more than 10 in the value of the DIC might rule out the model with the higher DIC (Spiegelhalter *et al.*, 2003). Where the difference is less than 10, the models are competitive.

2.4 Model Error Estimates

There are statistics for estimating how well the model fit the data and the converse, how much error was in the model. Two error statistics are particularly useful.

2.4.1 Mean Absolute Deviation (MAD)

This criterion has been proposed by Oh et al. (Oh *et al.*, 2003) to evaluate the fit of models. The Mean Absolute Deviance (MAD) calculates the absolute difference between the estimated and observed values.

$$\mathrm{MAD} = \frac{1}{n}\sum_{i=1}^{n}|\hat{\mu}_i - y_i| \tag{8}$$

The model closer to zero value is considered to be best among all the available models.

2.4.2 Mean Squared Prediction Error (MSPE)

The Mean Squared Prediction Error (MSPE) is a traditional indicator of error and calculates the difference between the estimated and observed values squared.

$$\mathrm{MSPE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_i - y_i)^2 \tag{9}$$

A value closer to 1 means the model fits the data better.

3 DATA DESCRIPTION

The National Highway 58 traverses mainly through a plain terrain of mostly agricultural and industrial areas. Most of the highway study segment falls in rural areas (approximately 85%). The National Highway-58 originates from national capital New Delhi and goes up to Mana, near China border in Uttrakhand state. It serves as a life line to the hilly part of the state. The road is strategically important as being the shortest route from Delhi to international China border. The highway has length of 536 Km of which 230 Km length in plain and rest in the hilly terrain. The highway connects important religious destinations which attract tourists from all over the country and world throughout the year.

The highway has two-lane and four-lane stretches. Traffic on the highway is mixed in nature and comprises of heavy and light vehicles. The study road section is an upgraded four-lane divided National Highway. This road stretch is the best suited to conduct post opening road safety study for divided four-lane National Highway. The main settlements along the NH are Siwaya, Daurala, Sakauti, Jarouda, Khatauli, Bhainsi, Mansurpur, Bengrajpur, Janshath Bypass, Bhopa, and Sisauna. Earlier the road was passing through three major cities Meerut, Khatauli and Muzaffarnagar but these settlements are now bypassed during reconstruction. The road level has been raised with respect to nearby area. Maximum access roads have level difference. Route map of study section of National Highway-58 is shown in Figure 1.

3.1 Site Selection

The stretch from Km 76.00 to Km 130.00 of National Highway 58 has been selected for candidate analysis. The selected highway stretch has been newly reconstructed and upgraded to four lanes. The two important obligatory points on the study area are Meerut and Muzaffarnagar of the highway in the state of Uttar-Pradesh, India.

The road stretch traverses through a flat and rolling terrain of mostly agricultural and urban settlement land. This national highway is maintained and operated by National Highway Authority of India (NHAI).

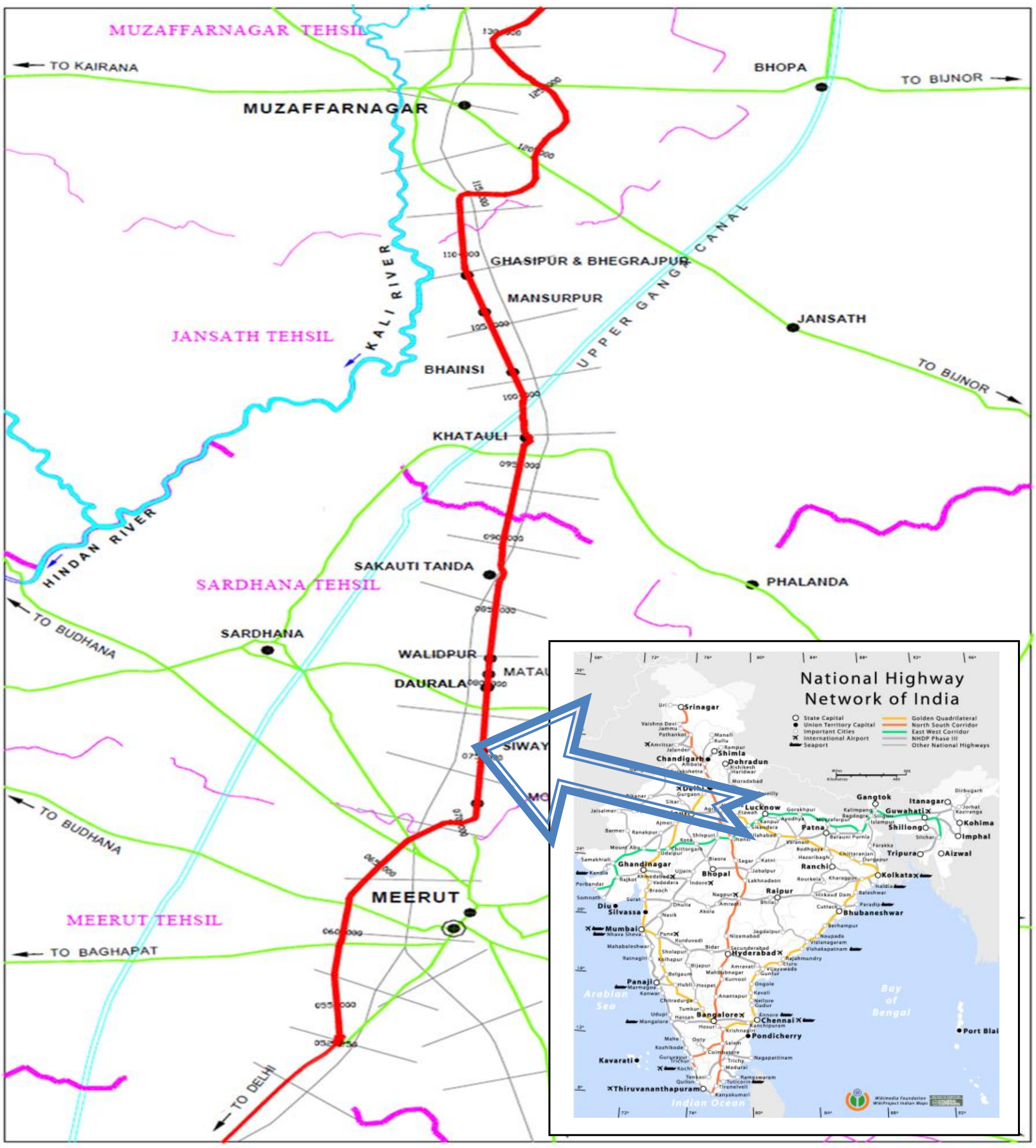

Figure 1 Study Area Route Map of National Highway - 58.

3.2 Details of Road Geometrics

Main Carriageway is 7.00m wide with 1.50m paved and 2.00m earthen shoulder on either side of the roadway. The median width is 5m and raised type in general. Table 1 shows the road infrastructure details for the study area.

4 DATA ANALYSIS AND RESULTS

4.1 Data Description

With consideration given to variables applied in past models and data availability, data were obtained for estimating the crash prediction models. To carry out extensive safety assessment of highway environment, two hundred meter sections were considered. All identified safety parameters and their conditions were noted in the survey Performa. Road safety data was collected for both direction of the divided four-lane National Highway 58 in between Km 76.00 and Km 130.00.

Table 1 Road Infrastructures Details

Highway Features	Description/ Quantity
Major Junctions	07 numbers
Major Bridges	02 numbers (Over Major Ganga Canal & Khatauli Escape Canal)
Minor Bridges	03 numbers at Km. 109.260, 115.250 & 117.700.
R O B	02 numbers at Km. 87.583 & 114.289.
Underpasses	05 numbers at Km 78.815, 87.400, 102.896, 118.550 & 122.175.
Culverts	186 numbers
Truck lay byes	02 numbers
Bus lay bays	07 numbers
Toll Plaza	1 number at Km.75.990

Accident records for two years, 913 and 1268 crashes in 2011 and 2012 respectively were collected from Western Uttar Pradesh Toll-Way Limited and police stations. Crash count (CC) per two hundred meter per year was taken as dependent variable in the prediction models. The Annual Average Daily Traffic (AADT) for 2011 was 20050 and for year 2012 is 23616 in PCU. Project details were also obtained from the highway agency.

Study area comprised of a total 19 intersections – 8 Four-legged and 11 Three-legged intersections. Crashes occurring on the highway segments were considered for the analysis. Crashes occurring within a circle of 76 meters (250 feet) (Lord et al., 2008a) were considered as intersection crash and were excluded from the analysis data. Hence there were thirty segments which were further divided into two hundred meter stretches. Traffic volume data was collected at each of the intersections and the main highway data was extracted for modelling.

The safety parameters included for safety study were Geometric Characteristics like Median Type (MedTyp), Median Opening (MedOpn), Access Roads to main highway (AcsRds), Road Markings (RdMrkgs) and Traffic Characteristics like Average Daily Traffic (ADT) and road-side developments like Industrial (Ind), Commercial (Com), Residential (Resi) and School (School).

For convenient calculations natural log of ADT was considered in analysis. Median openings were coded as per number of opening per two hundred meter stretch like code1 for one opening, code2 for two openings and code0 for no opening.

4.2 Analysis of Results

We have used the following generalized model functional form in both hierarchical bayesian analyses:

$$\mu_i = \hat{\mu}_i \varepsilon_i = \exp\ (B0 + B2*(Ln[ADTi]) + B3*MedOpni + B4*AcsRdsi + b5*Indi + b6*Comi + B7*Resi + B8*Schooli)*\varepsilon i \tag{10}$$

Parameters like Median Type (MedTyp) and Road Markings (RdMrkgs) showed less correlation to dependent variable Crash Count and were more inter correlated, leading to the omission of these insignificant variables in the modeling.

The parameter estimation and inference can be obtained by using Markov Chain Monte Carlo (MCMC) and software such as WinBUGS. Since the Bayesian formulation requires priors for all unknown parameters, non-informative normal priors for β's and gamma priors for error terms.

A total of two Markov chains were used in the coefficient estimation process. The G-R convergence statistic is generally used to verify that the simulation runs converged properly. For model comparison, it was suggested that convergence was achieved when the G-R statistic was less than 1.2 (Miaou et al., 2003). The first half of iterations (60,000) was used as burn-in samples and was discarded. Thus, the remaining half of the iterations was used for estimating the coefficients.

Table 2 Statistical Summary of the Parameters used for Analysis

	CrashCount	LnADT	AcsRds	MedOpn	Ind	Com	Resi	School	Offset(mts)
N(Sample Size)	1103	1103	1103	1103	1103	1103	1103	1103	1103
Mean	1.936	9.846	0.050	0.029	0.112	0.560	0.334	0.007	189.422
Variance	5.99	0.079	0.063	0.028	0.350	5.147	3.812	0.007	1272.373
Std. Deviation	2.448	0.28170	0.250	0.168	0.592	2.269	1.953	0.085	35.670
Minimum	0.00	9.22	0.00	0.00	0.00	0.00	0.00	0.00	12.00
Maximum	17.00	10.46	2.00	1.00	7.00	28.00	27.00	1.00	200.00
Log of Average Daily Traffic (LnADT), Median Opening (MedOpn), Access Roads (AcsRds), Industrial (Ind), Commercial (Com), Residential (Resi) and School (School).									

Table 2 gives the statistical summary of the variables selected to build the final models. The results in Table 2 revealed that for any subset of the independent variables, the crash count exhibits overdispersion.

As the nature of the data collected were overdispersed, Poisson-Gamma Hierarchical and Poisson-Weibull Bayesian models were appropriate models to analyze the data.

Table 3 Bayesian Estimate for Poisson – Gamma Hierarchical Model

Node	**Mean**	**SD**	**MC Error**
Intercept (B0)	-2.992	1.22	6.05%
LnADT (B2)	0.342	0.124	0.006
MedOpn (B3)	0.972	0.207	0.002
AcsRds (B4)	0.468	0.150	0.001
Ind (B5)	0.033	0.071	7.81E-04
Com (B6)	-0.025	1.98E-02	2.12E-04
Resi (B7)	0.011	1.93E-02	1.90E-04
School (B9)	1.521	4.38E-01	0.006
$\alpha = 1/\Phi$	1.091	0.086	8.851E-4
Pearson Chi Square	**357.146**		
MAD	**0.562**		
MSPE	**0.573**		
DIC	**3518.32**		
Log of Average Daily Traffic (LnADT), Median Opening (MedOpn), Access Roads (AcsRds), Industrial (Ind), Commercial (Com), Residential (Resi) and School (School).			

Poisson-Gamma hierarchical and Poisson-Weibull Bayesian models output are summarized in Table 3 and Table 4 respectively. Median openings have major effect on crash probability which is also practically true. Movements at median openings affect the speed and smooth traffic movement of the main highway. Traffic volume also has direct impact on crash occurrence. Access Roads from nearby developments also hinder the speed of the main highway traffic leading to occurrence of crashes. Most of the commercial developments have better access and exits along with sufficient lateral clearance which shows inverse effect on the probability of crash occurrence. Crashes are relatively high nearer to Schools as there are inadequate access and exit roads for school vehicles to maneuver.

Table 4 Bayesian Estimate for Poisson – Weibull Model

Node	**Mean**	**SD**	**MC Error**
Intercept (B0)	-16.74	7.13	0.381
LnADT (B2)	2.227	0.786	0.042
MedOpn (B3)	11.18	6.342	0.267
AcsRds (B4)	4.639	2.604	0.099
Ind (B5)	0.382	0.654	0.016
Com (B6)	-0.364	0.27	0.008
Resi (B7)	0.267	0.312	0.008
School (B9)	27.36	25.07	1.066
v	0.867	0.059	0.001
lambda	8.824	2.592	0.126
Pearson Chi Square	**831.865**		
MAD	**1.042**		
MSPE	**1.302**		
DIC	**3933.43**		
Log of Average Daily Traffic (LnADT), Median Opening (MedOpn), Access Roads (AcsRds), Industrial (Ind), Commercial (Com), Residential (Resi) and School (School).			

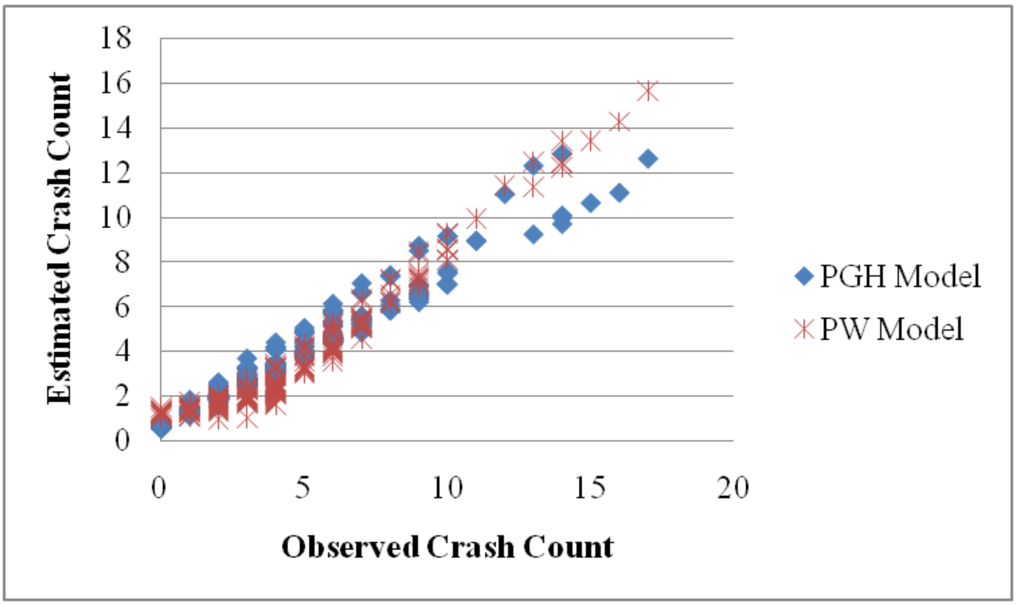

Figure 2 Predicted Values vs Observed Values for PGH and PW Model.

From goodness of fit measures Pearson chi-square (PGH = 357.15, PW = 831.86) and critical value is 1074.68 for degree of freedom 1095 at 95% confidence, MAD (PGH = 0.562, PW = 1.042) and MSPE (PGH = 0.57, PW = 1.30) values suggested the superiority of model fit by hierarchical Poisson-Gamma model on Poisson-Weibull technique. Figure 2 plots the observed Crash Count versus estimated values from each of these models. By graph it is easy to justify the superiority of the PGH modelling technique as PW model miss-predicts the values at lower range.

5 CONCLUSIONS

This paper presents two approaches to analyze road traffic crashes on a stretch of National Highway 58 in India. As the obtained data was overdispersed, Poisson-Gamma and Poisson-Weibull modelling techniques were efficiently used for the analysis of the data.
After careful application and assessment of statistical model, accompanied by detailed examination of the road crash model, the following general conclusions are drawn.

1. Median opening has major influence on the occurrence of crashes.
2. The traffic flow is also showing direct impact on occurrence of crashes as justified practically.
3. From the analysis, it has been observed that as access roads to the main highway increases the chances of crashes on highways will be more which is as per realistic experience.
4. Road side developments also increase the movement and hinder the smooth traffic movement which is also justified. Whereas the commercial activities is showing negative impact as there is enough lateral clearance from the highway shoulder for ingress and egress of the vehicles.

The results of this study lead to support the superior data fit by Poisson-Gamma hierarchical bayes model. There are scarce safety studies adopting this technique for crash analysis on Indian Highways.

Examination of the modeling results suggest that explanatory variables like traffic volume, access roads, median openings and road side developments led to the occurrence of crash.

ACKNOWLEDGEMENTS
The author would like to thank Ministry of Human Resource Development (MHRD) for funding the research work and my friends for carrying out the data extraction.

REFERENCES

Bolstad, W. M. Introduction to Bayesian Statistics. John Wiley & Sons, Ltd, 2007.

Cheng, L., Geedipally, S, R., Lord, D. Examining the Poisson-Weibull Generalized Linear Model for Analyzing Crash Data. Transportation Research Board Annual Meeting 2012, Paper No. 12-2793, 2012.

Congdon, P. Bayesian Statistical Modeling. John Wiley & Sons, Ltd, 2006.

El-Basyouny, K. and Sayed, T. Statistical Analysis of Crash Data: Alternative Modeling Methods. LAMBERT Academic Publishing, 2012.

Hauer, E. (1997) Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Elsevier

Science Ltd, Oxford.
Jain, S.S. and Parida, M. Final Report on Development of a Scientific Approach to Road Safety for Indian Environment. Submitted to AICTE, New Delhi, 2001.
Lord, D. (2006) Modeling Motor Vehicle Crashes Using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter. Accident Analysis & Prevention, Vol. 38, No. 4, pp. 751-766.
Lord, D., S.P. Washington, and J.N. Ivan. (2005) Poisson, Poisson-Gamma and Zero Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. Accident Analysis & Prevention, Vol. 37, No. 1, pp. 35-46.
Lord, D., S.P. Washington and J.N Ivan. Poisson, Poisson-Gamma and Zero-inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory, Accident Analysis & Prevention Vol. 37, No. 1, 2005, pp. 35-46.
Lord, D. and Park, B. J. Appendix D-Negative Binomial Regression Models and Estimation Methods. In Crime Stat Version 3.3 Update Notes: Part 2: Regression Modeling, The National Institute of Justice, Washington, D.C. 2010.
Lord, D., Geedipally, S.R., Persaud, B.N., Washington, S.P., van Schalkwyk, I.,Ivan, J.N., Lyon, C., Jonsson, T., 2008a. Methodology to Predict the Safety Performance of Rural Multilane Highways. NCHRP Web-Only Document 126, National Cooperation Highway Research Program, Washington, D.C. (**http://onlinepubs.trb.org /onlinepubs/** nchrp/ nch rp_w 126.pdf, accessed on Feb 11 2013).
Maher, M and L. Mountain. The Sensitivity of Estimates of Regression to the Mean. Accident Analysis & Prevention, Vol. 41, No. 4, 2009, pp. 861-868.
Miaou, S. P., and Lord, D. Modeling Traffic-Flow Relationships at Signalized Intersection: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes. Transportation Research Record, No. 1840, 2003, pp. 31-40.
Mitra, S. and Washington, S. On The Nature of Over-Dispersion in Motor Vehicle Crash Prediction Models. Accident Analysis and Prevention 39 (3), 2007, pp. 459-468.
Oh, J., C. Lyon, S. Washington, B. Persaud and J. Bared. Validation of FHWA Crash Models for Rural Intersections: Lessons Learned. Transportation Research Record: 1840, 2003, pp. 41-49.
Parida, M., Jain, S.S., and Naveen Kumar, C., 2012. Road Traffic Crash Prediction on National Highways. Indian Highways Journal, Indian Road Congress, Vol. 40 (6), pp. 93-103.
Raghavachari, M., A. Srinivasan and P. Sullo. (1997) Poisson Mixture Yield Models for Integrated Circuits: A Critical Review. Microelectronics Reliability, Vol. 37, No. 4, pp. 565-580.
Spiegelhalter, D., Thomas, A., Best, N., Lunn, D. WinBUGS Version 1.4 User Manual. MRC Biostatistics Unit, Cambridge, **http://www.mrc-cam.ac.uk/bugs**. 2003.
Washington , S.P., Karlaftis, M.G., Mannering, F.L. Statistical and Econometric Methods for Transportation Data Analysis, Second Edition. CRC Press, 2011.