Accident Prediction by Using Poisson Regression for Unsignalised Junction in Khulna Metropolitan City, Bangladesh

M. E. SHAIK[1], Q. S. HOSSAIN[2]

[1]Department of Civil Engineering, KUET, Bangladesh (ebrahimkuet82@gmail.com)
[2]Department of Civil Engineering, KUET, Bangladesh (sazzad@ce.kuet.ac.bd)

Abstract

The Poisson regression models were used in the large area of application and likely to improve the quality and important for the engineering aspects of accident prevention in Khulna Metropolitan City. A wide range of accident data at each junction for the fifteen-year period (2000-2015) were used in the model development process. In this model, numbers of accident were selected as an outcome variable and indicate the number of accidents occurred at different types of junction in a year. Also total accident and junction types were selected as a continuous predictor variable and categorical predictor variable respectively. It was found that the models evaluated 2.19 times more variation in accidents at Not junctions and 1.80 times more variation at crossing junctions than at T-junctions respectively. In absence of available dedicated left turning lane and traffic spot speeds vehicle on the roads generally related with accident severity in each junction.

Keywords: *Poisson Regression, Accident Prediction, Road, Unsignalised Junction, Variable.*

1 Introduction

Road traffic accident is one of the most undesirable situations to occur to a road user resulting harmful to the people and damages to the valuable properties. Now road traffic accident is the global phenomenon to the almost all countries of the world. They are seriously concerned about the increasing number of people killed and injured on their highways. According to the World Health Organization (WHO), more than 1.25 million people lives are cut short for the result of road traffic accident every year. More people between 20 and 50 million suffer simple or non- fatal injuries, with many incurring a disability as a result of their injury. For Bangladesh, accident severity and fatality in road accident is now an alarming issue. In Bangladesh Khulna metropolitan city is the 3rd largest city with an area of 45.65 square kilometers and more than 1.022 million people live here. In this city, Total numbers of roads are 1215 with a total length of 356.65 kilometers (Khulna City Corporation, 2015). Recently significant numbers of killed and injured people are increasing enormously in Khulna metropolitan city (Ebrahim and Hossain, 2018).

For transportation safety, road traffic accident prediction models are very necessary tools given their effective for evaluating both the frequency of accident frequency and the contributing factors that could then be determined by transportation methods (Azad, 2017). To determine the accident frequency, accident severity level and different factors which are generally responsible for road traffic accident, transportation authority and many research institute are often want to accident data and factor to identify the most vulnerable road environment site. The different regression model including Poisson regression model helps to explain the relationship between accident occurrence and responsible risk factor. Though multiple linear regression models are widely used for prediction, it has been found that the Poisson regression model can often be better fitted for prediction accident occurrence. The development of generalized theories concerning highway safety, road traffic accident prediction models can also keep important contribution (Azad, 2017). Poisson regression has been often applied for count road traffic and different transportation policies including frequency.

Generally Poisson regression has more advantages than other regression also related to different discrete distribution and constraint to predicted value to non-negative integer number (Glenberg, 1996). In accident frequency analysis for different platform, almost all the researcher has normally applied Poisson regression as a beginning point for analysis, they have often found that the model data exhibit over and under dispersion that make the application of Poisson regression model problematic (Lord and Mannering, 2010).As accidents are not occurred frequently, categorical style of accident data is more efficient for expression of accident reaction and using this categorized accident data Poisson regression model make the taste significance as fixed value (Sonia, 2012).

The situation of road accident in Khulna metropolitan city is actually frightful and the death of lives and infrastructure damages are desired to continue if necessary corrective measures are not taken accordingly by applying proper engineering measures through proper research. These situations for all metropolitan cities in Bangladesh are very dangerous. About 20 percent of road accident occurred in metropolitan cities viz. Dhaka, Chittagong, Khulna and Rajshahi (Hoque, M.M., 1991). So, it is more important for Khulna city to predict road accident considering the basic factors including different types of junctions.

2 Methodology

2.1 Data Collection

Five police stations were selected as study area namely Khulna sadar, Daulatpur, Khalishpur, Khanjahan Ali and Sonadanga. The accident data from 2000 to 2015 were collected from the Khulna Metropolitan Police (KMP) head quarter and Accident Research Institute (ARI), BUET.

2.2 Accidents Distribution

Figure 1 shows the yearly accident distribution.

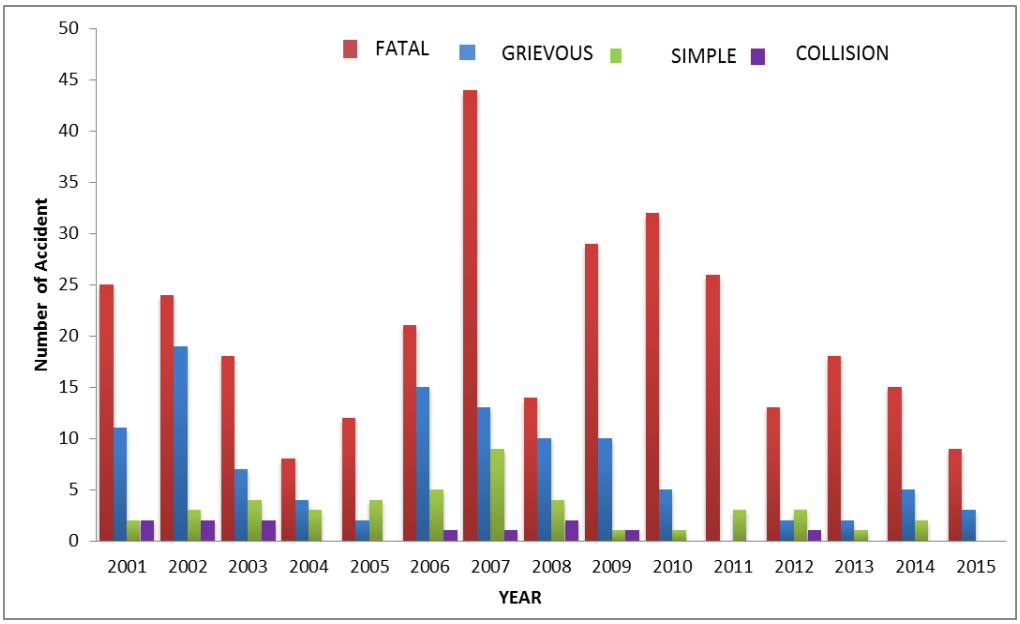

Figure 1. Yearly accidents distribution.

Total 475 accidents were recorded within the year 2001 to 2015. Maximum 67 accidents and minimum 15 accidents were occurred in the year 2007 and 2004 respectively. Figure1 also showed that the accident rate decreases in the last three years.

2.3 Model Development

IBM SPSS statistics 22 software was used for the development of Poisson regression model. In this study, number of accidents, total number of accident and junction types were selected as model parameter where number of accident was a outcome variable, total number of accident was a continuous predictor variable and junction types is a categorical predictor variable.

The Poisson regression model is expressed as (Lord and Mannering, 2010):

$$P(k_n) = \frac{\exp(-\lambda_n)\lambda_n^{k_n}}{k_n!} \quad (1)$$

Where,
$n = k_n$ accidents per some time period,
k_n = non- negative integer,
$P(k_n)$ = roadway entity probability, n having k_n accidents per some time period,

λ_n = Poisson parameter for roadway entity n,
Which is equal to roadway entity n's expected number of accident per year, e [k_n],

Poisson regression models were estimated by specifying the Poisson parameter λ_n (the expected number of accidents per period) as a function of explanatory variables, the most common functional form being
$\lambda_n = \exp(\beta x_n)$,
Where x_n is a vector of explanatory variables and β is a vector of estimable parameters.

2.4 Akaike's Information Criterion (AIC)
The performance of different statistical model for given data is measured by Akaike's Information Criterion (Akaike, 1973). Akaike's Information Criterion (AIC) compares the models quality relative to each other for their collective data. Therefore, AIC is widely used for the selection of better model. Smallest value of AIC indicates the best model (Omari-Sasu et al, 2016).

2.5 Bayesian Information Criterions (BIC)
Bayesian Information Criterion (BIC) is generally used for model selection among finite data sets of different model. BIC is the exponential function and small parts of a likelihood function are closely related to Akaike's Information Criterion (AIC). Smallest value of Bayesian Information Criterions (BIC) indicates the best model (Omari-Sasu et al, 2016).

3 Results and Discussions

The total accidents 231, 40 and 60 were recorded at the sites of Not junctions, T junctions and crossing junction respectively for the study period 2000 to 2015. Only junction type is used and most of the parameter and variable are not included for developing this model. Therefore, this model can be used only for approximate determinates of accident severity in Khulna Metropolitan City.

Table 1 shows the narrative statistics of model data. Table 1 also shows that each variable contains 16 legal observations and from their distributions it indicates quite reasonable. The outcome variable for unconditional mean and variance are not immensely different. These values and predictor variable for different condition will be equal by the model assumption.

Table 1. Narrative Statistics

	N	Min.	Max.	Mean	Standard Dev.
No. of accident	16	1.80	0	32	4.56
Total .accident	16	1.85	12	67	32.19
Valid N (list wise)	16	1.93			

Table 2 shows that the qualities fit of the model and the model output starts by this table. The statistics lists of Table 2 indicate the model is quite fit. The first row of the Table 2 indicates the qualities of fit Chi- Squared test. It was also found that the Akaike's Information Criterion (AIC) of this model 55.669 and Bayesian Information Criterions (BIC) equal to 58.760. Model also evaluates the deviance 20.20 as Chi-square distributed with the 12 degrees of freedom. Lowest value of both AIC and BIC indicates the developed model is the best model.

Table 2. Qualities of Fit

	Value	DF	Value/DF
Deviance	20.200	12	1.683
Scaled deviance	20.200	12	
Chi- Square (Pearson)	14.301	12	1.192
Chi- Square (Scaled-Pearson)	14.301	12	
Log-likelihood	-23.835		
AIC	55.669		
Corrected AIC (AICC)	59.306		
BIC	58.760		
Consistent AIC (CAIC)	62.760		

Dependent Variable: Number of Accident, Model: (Intercept), Parameter, Total Accident.

Table 3 shows the Estimate of Accident Parameters. These comprise the Poisson regression coefficients for each variable including standard error, 95% confidence interval due to the coefficients. The regression coefficient for total accident was found 0.082. This indicates that the expected increase in the log count for one unit increase in total accident is 0.082. It was observed from the model effects that the output of this study is statistically significant.

Table 3. Estimate of Accident Parameters

Parameter	B	Standard error	95% Wald interval confidence level		Hypothesis test		
			Lower	Upper	Wald Chi-square	DF	Significance
Intercept	28.395	.4423	-29.262	-27.528	4121.570	1	.000
Parameter=1	26.548	.4027	25.758	27.337	4346.635	1	.000
Parameter=2	26.350	.					
Parameter=2	0	.					
Total accident	.082	.0087	.065	.099			
Scale	1						

Dependent variable: Number of accident, Model: (intercept), parameter, total accident

Table 4 shows the parameter estimates of different types of junction. The average predicted value for Not junctions was found 2.19. It indicates that the developed model evaluated 2.19 times more variation in road accident at Not junctions and 1.80 times more variation at Crossing junctions than at T junctions respectively. So, Not Junction is more severe for occurring accident among these junctions in Khulna Metropolitan City.

Table 4. Different types of parameter estimates

Types of parameter	Mean	Standard error	95% Wald interval confidence level	
			Lower	Upper
Not Junction	2.19	.582	1.30	3.69
Crossing Junction	1.80	.577	.96	3.37
T Junction	.00	.000	.00	.00

This Poisson regression model is the similar as that used in ordinary regression model except that the random component is the Poisson distribution. The Poisson Regression model prediction is sometimes mentioned as a Poisson Log linear model. Most of the regression model provides better for over dispersed data while Poisson regression model provides better for equal- dispersion data.

4 Conclusions

The study of this research work was carried out the ability of Poisson regression model for prediction of accident in Khulna Metropolitan City, Bangladesh. This paper developed the traffic accident prediction models to gather a better knowledge by using this useful systems for predict the road traffic accidents including their hazard factors. From the results it was found that the developed model evaluated 2.19 times more variation in road accident at Not junctions and 1.80 times more variation at Crossing junctions than at T junctions respectively. So, Not Junction is more severe for occurring accident among these junctions in Khulna Metropolitan City. It was found from the Tests of Model Effects output that all the junctions overall, is statistically significant. It can be concluded that the model fits reasonably well because the goodness-of-fit chi-squared test is not statistically significant with 12 degrees of freedom. This model can able to predict the road accident at unsignalised junction of Khulna Metropolitan City, Bangladesh.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In the second international symposium on information theory, edited B.V Petrov and B.F Csaki, Academical Kiado.

A.Y. Omari- Sasu, A. M. Isaac and R. K Boadi (2016). Statistical Models for Count Data with Applications to Road Accidents in Ghana. International Journal of Statistics and Applications. vol. 6(3), 123-137.

Azad, A. (2017). Road Crash Prediction Models: Different Statistical Modeling Approaches. Journal of Transportation Technologies. vol. 7, 190-205.

Ebrahim, S. and Hossain, Q. S. (2018). An Artificial Neural Network Model for Road Accident Prediction: A Case Study of Khulna Metropolitan City. ICCESD 2018, 5193, 1-8.

Glenberg, A. (1996). Learning from Data: An Introduction to Statistical Reasoning. 2nd Edition, Lawrence Erlbaum Associates, Mahwah.

Hoque. M.M. (1991). Accident investigation for the safety improvement of Dhaka- Aricha highway: A section of Asian highway. Final report, Department of Civil Engineering, Bangladesh University of Engineering & Technology, Dhaka, Bangladesh.

Khulna City Corporation (2015), Basic Statistics, Retrieved from: http://www.khulnacity.org/Content/index.php?page=About_KCC&Z2Y&pid=30

Lord, D. and Mannering. F (2010). The statistical Analysis of Crash Frequency Data: A Review and Assessment of Methodological Alternatives. Accident Analysis and Prevention. vol. 44, 292-303.

Sonia, R. (2012). Development of an Accident Prediction Model for Intersections of Dhaka City, Bangladesh. International Journal of Computer Applications. vol. 47(16), 10-16.