ISSN:
Print - 2277 - 0593
Online - 2315 - 7461
© FUNAAB 2017

Journal of Natural Science, Engineering and Technology

COMPARISON OF ROAD TRAFFIC ACCIDENT PREDICTION MODELS FOR TWO-LANE HIGHWAY INTEGRATING TRAFFIC AND PAVEMENT CONDITION PARAMETERS

[1]*O. M. POPOOLA, [2]O. S. ABIOLA, [2]S. O. ODUNFA AND [3]S. O. ISMAILA

[1]Department of Civil Engineering, Moshood Abiola Polytechnic, Abeokuta
[2]Department of Civil Engineering, Federal University of Agriculture, Abeokuta
[3]Department of Mechanical Engineering, Federal University of Agriculture, Abeokuta
***Corresponding Author**: oyenolapopoola@yahoo.co.uk, **Tel**. 08033861588

ABSTRACT

In Nigeria, literature on the integration of traffic of pavement condition and traffic characteristics in predicting road traffic accident frequency on 2-lane highways are scanty, hence this article to fill the gap. A comparison of road traffic accident frequency prediction models on Ilesha-Akure-Owo road based on the data observed between 2012 and 2014 is presented. Negative Binomial (NB), Ordered Logistic (OL) and Zero Inflated Negative Binomial (ZINB) models were used to model the frequency of road traffic accident occurrence using road traffic accident data from the Federal Road Safety Commission (FRSC) and pavement conditions parameters from pavement evaluation unit of the Federal Ministry of Works, Kaduna. The explanatory variables were: annual average daily traffic (aadt), shoulder factor (sf), rut depth (rd), pavement condition index (pci), and international roughness index (iri). The explanatory variables that were statistically significant for the three models are aadt, sf and iri with the estimated coefficients having the expected signs. The number of road traffic accident on the road increases with the traffic volume and the international roughness index while it decreases with shoulder factor. The systematic variation explained by the models amounts to 87.7, 78.1 and 74.4% for NB, ZINB and OL respectively. The research findings suggest the accident prediction models that should be integrated into pavement rehabilitation.

Keywords: pavement condition, accident frequency, models, 2-lane highway, traffic

INTRODUCTION

Traffic accident is a major cause of death, injuries and property loss, and as such, a public concern. Nigeria ranked 149th in 2009 out of 178 member states on road traffic accident as reported by World health organization (FRSC, 2012). Highway form the backbone of Nigeria's social-economic activities, however, population and traffic lead to lower riding quality, vehicle damage and road traffic accident.

The relationships between road traffic accidents (RTA) and roadway features, such as horizontal curvature, vertical grade, lane width, and shoulder width, pavement conditions and traffic mix and composition have been studied using multiple linear regression models in numerous previous studies (Chan et al. 2010; Chan et al. 2009; Chandra, 2004;

Al-Masaeid, 1997; Karlaftis and Golias, 2002; Chiou, 2006; Delen *et al.*, 2006).

Mathematical models have been widely used to predict road traffic accident severity and the contributing factors. The most commonly used models are Negative Binomial (NB) and Poisson (P) Regression models due to the nature of RTA data, which consist of both discrete and non-negative integers. Poisson distribution is applicable under the assumption of equidispersion i.e the mean is equal to the variance of the dependent value. NB is an extension of Poisson distribution which is used when the variance is greater than the mean, i.e overdispersion (Abdul-Aty and Essam, 2000; Chan *et al.,* 2010). Milton and Mannering (1997) used logit Regression to estimate the effect of statistically significant factors on accident severity. Logit regression and other related categorical data regression methods have often been used to assess risk factors for various diseases, however, it has been used as well in transportation studies (Al-Ghamdi, 2002). One reason for considering Zero Inflated Negative Binomial (ZINB) regression model is that it can handle the unobserved heterogeneity and excess of zeros. In their study, Zeeger et al., (2000) observed that 55.7% of the road sections studied had no reported vehicle accidents in five-year period, while Miaou and Lum (2003) asserted that over 80% of the road sections had no reported traffic accidents during a one-year period. This suggests that for some years most of the road sections considered would have a high probability of being free from RTA.

Numerous studies on traffic safety have been considered over decades, most of them were focused on various perspectives such as highway geometric, vehicle conditions and human factors. There are limited studies relating the pavement condition to safety. Almost all previous studies were concentrated on pavement/shoulder edge and drop off. Very few focused on relating traffic accident frequency and the compressive pavement condition parameters such as Rut Depth (RD), International Roughness Index (IRI), Annual Average Daily Traffic (AADT), Shoulder Factor (SF), Pavement Condition Index (PCI). In Nigeria literature on the integration of pavement condition and traffic characteristics in prediction RTA frequency on two lane highway are scanty, hence, this paper to fill the gap. Specifically, the objective of this research is to evaluate the robustness of the Negative Binomial (NB) regression model, Ordered Logistic (OL) model and Zero Inflated Negative Binomial (ZINB) model in establishing the relationship between highway traffic accident frequency and pavement condition parameters on Ilesha-Akure-Owo Road.

MATERIALS AND METHODS

Study Area and Data Collection

The study area lies within longitude 02 45' to 05 00' East of Greenwich Meridian and latitude 06 00' to 08 30' North of the equator. The road is Ilesha-Akure-Owo which fall within South-West region of Nigeria, a single carriageway of 107kmans a total of 55 segments (Figure. 1). There years road traffic accident (RTA) data from the FRSC database from 2012-2014 were linked to each of the road by common chainage in the current study. Thus, each segment's accident frequency was obtained through the FRSC database. Pavement condition parameters were obtained from the Pavement Evaluation Unit of the Federal Ministry of Works, Kaduna. Total crashes represent accident frequency, i.e the sum of all the type of accident severity. A total of 442 crashes occurred within the three years of study. Table 1 and Table 2 present the descriptions and summary of the statistics of the independent variables respectively.

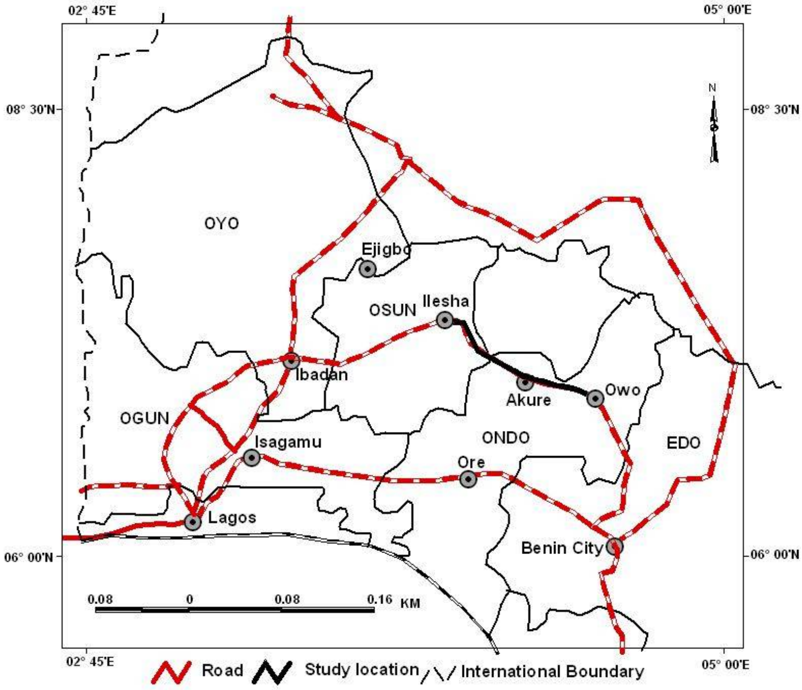

Figure 1: Map of South-Western Nigeria highlighting the road under study

Table 1: Description of Independent variable

S/No	Variable	Symbol	Types	Description
1	Average Annual Daily Traffic	AADT	Continuous	Sectional aadt
2	Shoulder Factor	SF	Continuous	SF scale ranges from 0 (worst) to 5 (best).
3	International Roughness Index	IRI	Continuous	IRI scale ranges from 0 (best) to 15 (worst).
4	Rut Depth	RD	Continuous	Average wheel paths depression measured in millimeters
5	Pavement Condition Index	PCI	Continuous	PCI scale ranges from 0 (worst) to 100 (best).

MODELLING

Data related to RTA and those on pavement condition parameters were fitted into NB, OL, ZINB and Poisson models as earlier defined.

Negative Binomial regression model

A commonly used distribution to deal with over dispersion problem in count data is the NB. The NB regression model considered in this study has the following form:

Table 2: Data summary for the pavement condition parameters on Ilesha-Akure-Owo Single Carriageway

Variables	(Ilesha-Akure-Owo Road)				
	Minimum	Maximum	Mean	Std. Dev.	Variance
aadt	7344	9156	8159.6727	784.14979	614890.89
sf	0.46	3.67	2.242	0.821077	2.674168
iri	2.4	3.6	2.669091	0.342444	30.36498
rd	24	67	31.07273	5.510443	40.36498
pci	33	55	50.03636	6.182303	58.22088

$$p(Y_i = y_i) = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{\Gamma(y_i + 1)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1+\alpha\mu_i}\right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \quad y_i = 0,1,2,.....$$

where

$$\mu_i = E(Y_i) = v_i\left[e^{x_i\beta}\right] = v_i\left[e^{\sum_{j=1}^{k} x_{ij}\beta_j}\right] \quad i = 1,2,3,.........n.$$

and the variance *of* Y_i *is*

$$Var(Y_i) = \mu_i + \alpha\mu_i^2$$

(Agresti, 2000)

where ($\alpha \geq 0$ and is usually referred to as dispersion parameter. one can see that this model allows the variance to exceed the mean. Also, the Poisson regression model can be regarded as a limiting model of the negative binomial regression model as α approaches 0.

The ML estimation of the NB regression model and the calculation of associated statistics are described in detail by (Agresti, 2000). The moment estimation, which was first suggested by Breslow (2004), is also commonly used for estimating the parameters in the NB model.

For comparison, in this study the moment estimation method as described in (Agresti, 2000), was also used for parameter estimation. The method is an iterative procedure which iterates until the estimated dispersion parameter converges.

2.2 Zero-inflated Negative Binomial regression model

The particular ZINB regression model considered in this study has the following form:

$$p(Y_i = y_i) = e^{-\theta r_i} \qquad if\ y_i = 0$$

$$\left(\frac{1-e^{-\theta r_i}}{1-e^{-r_i}}\right)\frac{r_i^{y_i} e^{-r_i}}{y_i!} \qquad if\ y_i = 1,2,3,......$$

and

$$r_i = v_i\left[e^{x_i\beta}\right] = v_i\left[e^{\sum_{j-1}^{k} x_{ij}\beta_j}\right] \qquad i = 1,2,3,......n$$

(Miaou & Lum, 2003)

where $0 < \theta \leq 1$. (Note that for $\theta > 1$, the probability of observing zeros is deflated rather than inflated.)

Under this ZINB regression model, the mean and variance of Y_i can be shown to be

$$\mu_i = E(Y_i) = \left(\frac{1-e^{-\theta r_i}}{1-e^{-r_i}}\right) r_i \quad and$$

$$Var(Y_i) = \mu_i + \left(\frac{1-e^{r_i(\theta-1)}}{e^{\theta r_i}-1}\right)\mu_i^2 = \mu_i + \phi_i \mu_i^2$$

(Miaou & Lum, 2003)

where ϕ_i is a function of r_i and θ. When $\theta = 1$, the ZIP regression model is identical to the Poisson regression model presented. Also, when $0 < \theta < 1$, one can show that the variance of Y_i exceeds its mean. Thus, the ZINB regression model allows overdispersion in the data due to excess zeros when compared to the classic Poisson regression model (Miaou & Lum, 2003).

One reason for considering the ZINB regression model is the potential underreporting of vehicle accidents, especially minor injury and property-damage accidents. If the number of vehicles involved in RTA on road sections follows a Poisson distribution, then, because of underreporting, the "reported" number of accidents cases would follow a ZIP distribution under some underreporting conditions.

Ordered Logit Regression model

It is important to understand that the goal of analysis using logistic regression is the same as that of any model-building technique in statistics: to find the best fit and the most robust one. What distinguishes a logit regression model from a linear regression one is the response variable. In the logistic regression model, the response variable is binary. Once this difference is accounted for, the methods employed in an analysis using logit regression follow the same general principles used in linear regression analysis. In any regression analysis the key quantity is the mean value of the response variable given the values of the independent variable:

where *Y* denotes the response variable, *x* denotes a value of the independent variable,

$$E(Y/x) = \beta_0 + \beta_1 x$$

and the β denote the model parameters. The quantity is called the conditional mean or the expected value of *Y* given the value of *x*. Many distribution functions have been proposed for use in the analysis of a dichotomous response variable (Hosmer and Lemeshow, 1989; Agresti, 1984; Frome *et al*, 2000). The specific form of the logistic regression model is

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

(Frome *et al*, 2000)

Where to simplify the notation,

$$\pi(x) = E(Y/x)$$

The transformation of the π(x) logistic function is known as the logit transformation:

$$g(x) = In\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x$$

RESULTS AND DISCUSSION

The parameter estimation results for models: NB, OL and ZINB at confidence level of 95% to identify statistically significant variables have been described, while comparison is made with Poisson model as presented in Table 5. The pavement parameters variables that were found to be statistically significant for the three models are, aadt, sf and iri, the coefficients of these are positives as logically expected. While PCI and RD are not significantly related to crash frequency.

Regarding accident frequency prediction, logarithm of AADT indicates that the higher the AADT, the higher the accident frequency in agreement with previous research (Chan et al., 2009). Modelling results show that SF was significant variable in all the models. In the current study, if pavement conditions are enhanced by one unit of SF for NB model, the crash frequency would drop by 44% ($e^{-0.4299}$). A higher IRI is associated with a higher crash frequency as observed in all the models considered which indicates that if increasing IRI from 0 to 5 m/km, the crash frequency would increase by 5.65 ($e^{-0.5235}$) times.

While evaluating the performance of the models, overdispersion in the data was considered and this was done by testing the null hypothesis of the overdispersion parameter, Ho: $\alpha = 0$, i.e contrasting the variance-mean equality assumption of the Poisson model against an alternative model in which the variance exceed the mean. The comparison can be made using likelihood ratio statistic, define as $\text{Ţ}_{LR} = -2(\text{ȴ}_p - \text{ȴ}_A)$ (Cameron and Trivedi, 1998), where ȴ_p is the log-likelihood value of the restricted (Poisson) model and ȴ_A is the log-likelihood value estimated of the unrestricted model that consider overdispersion (NB, OL, ZINB). The Ţ_{LR} approximately follow a chi-square distribution, the null hypothesis is rejected if it exceeds a critical value. From Table 3, the log-likelihood value of the Poisson, NB, OL and ZINB distributions are -143.588, -144.566, -110.679 and -108.149 respectively ranking performance of the models from worst to best. Thus Ţ$_{LR}$ is equal 58.44, 55.82 and 60.88, exceeding 5% critical value of x^2 = 9.488. From this, overdispersion was detected with the three alternative models. The Poisson model had the least ranking because it does not allow for overdispersion as earlier explained.

Table 1: Model Calibration Results

Model	Variable	Coefficient	Std. Error	P-Value	Log-likelihood
Poisson	Constant	-58.284	29.675	0.050	-143.588
	lnaadt	4.699	.549	0.000	
	sf	-.554	.0653	0.000	
	rd	.0053	.0059	0.370	
	pci	.166	.296	0.575	
	Iri	3.903	5.353	0.026	
	LR chi2(5)	167.32			
	Prob>chi2	0.000			
	Pseudo R2	0.368			
NB	Constant	-37.302	44.416	0.046	-114.5659
	lnaadt	4.667	.753	0.000	
	sf	-.5799	.0946	0.000	
	rd	.0049	.0103	0.182	
	pci	-.044	.4465	0.066	
	Iri	.123	8.059	0.030	
	/lnalpha	-2.199	.5684		
	Alpha	.111	.0631		
	LR chi2(5)	46.55			
	Prob>chi2	0.000			
	Pseudo R2	0.143			
0L	Constant	-	-	-	-110.6795
	lnaadt	14.632	3.1446	0.000	
	sf	-1.855	.36916	0.000	
	rd	.06459	.0557	0.263	
	pci	-.3425	1.829	0.186	
	Iri	-3.228	33.021	0.042	
	LR chi2(5)	50.65			
	Prob>chi2	0.000			
	Pseudo R2	0.186			
ZINB	Constant	-56.229	35.744	0.038	-108.1496
	lnaadt	3.859	.5906	0.000	
	sf	-.4588	.07339	0.000	Number of obs = 55
	rd	.0062	.0068	0.213	Nonzero obs = 46
	pci	.2285	.3546	0.084	Zero obs = 9
	Iri	4.774	6.4079	0.030	
	/lnalpha	-4.099	1.5069	0.007	
	Alpha	.01657	.02498		
	Inflate const.	22.27	23313.89	0.999	
	LR chi2(5)	51.62			
	Prob>chi2	0.0000			

In each model, the goodness-of-fit statistics parameter were provided for the model comparison, the estimated result shown in Figure 2, 3 and 4 for NB, ZINB and OL respectively. The systematic variation explained by the models amount to 66.44, 58.22 and 64.71% for NB, ZINB and OL respectively. Pearson chi-square revealed that NB has the lowest value, followed by OL, ZINB and Poisson model, while log-likelihood of ZINB is the lowest. In addition to, Akaike Information Criterion-AIC is computed as AIC = -2log-likelihood + 2k, where k = number of estimated parameters included in the model. The model with lowest AIC is preferred. In the present application, k is equal to 5 parameters, and then the AIC values are equal to 297.18, 239.13, 231.36 and 226.3 for Poisson, NB, OL and ZINB models respectively. The NB model outperforms all other models in term od Pearson chi-square and the systematic variation explained while ZINB has the lowest AIC and log-likelihood values, thus ZINB should be considered as a more comprehensive model that integrate traffic and pavement condition parameters in road traffic accident prediction. Therefore, sufficient statistical evidence is established for the relationship between pavement condition parameters and crash frequency.

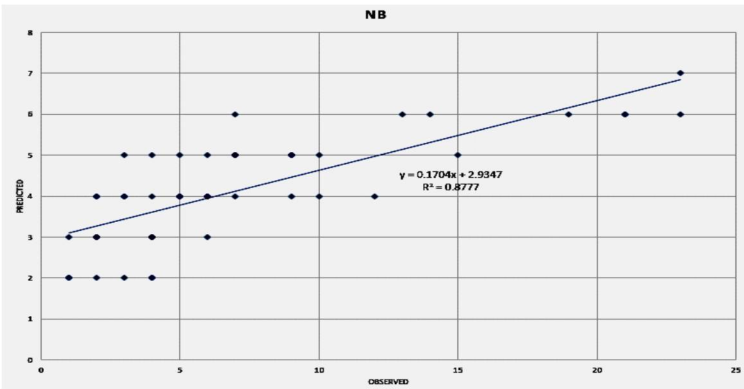

Figure 2: Estimated Crash Frequency with NB Regression Model

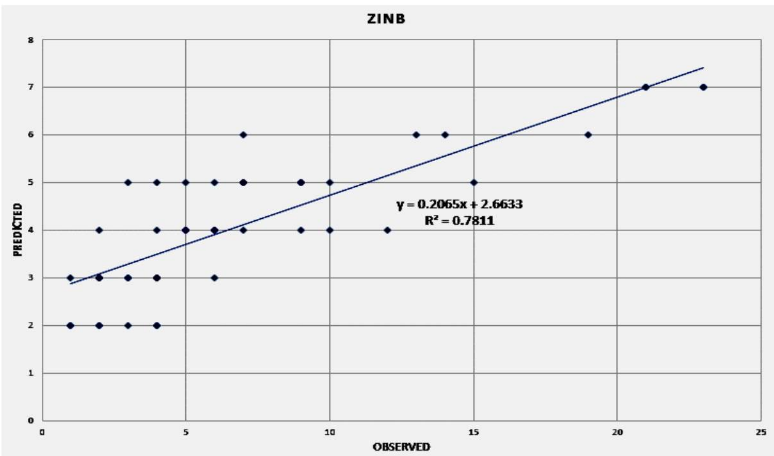

Figure 3: Estimated Crash Frequency with ZINB Regression Model

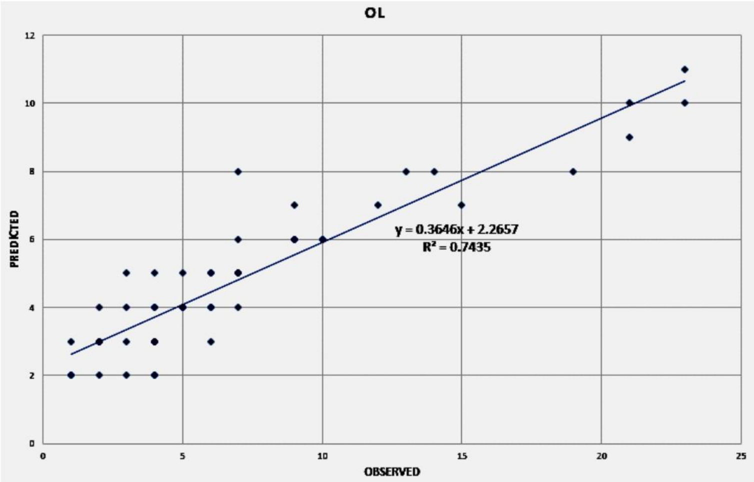

Figure 4: Estimated Crash Frequency with OL Regression Model

CONCLUSION AND RECOMMENDATIONS

There have been numerous efforts to investigate crash frequency in relation to roadway design features, environmental conditions, driver's characteristics and traffic features. However, very few of them have specifically considered the influence of pavement condition and traffic characteristics in predicting road traffic accident (RTA) frequency in Nigeria, hence this paper fill the gap. A comparison of RTA frequency prediction models on Ilesha-Akure-Owo road using data observed from 2012 to 2014 have been presented. NB, OL and ZINB models were used along with the pavement condition parameters from pavement Evaluation Unit of the Federal Ministry of Works, Kaduna. The explanatory variables were annual average daily traffic (AADT), shoulder factor (SF), rut depth (RD), pavement condition index (PCI) and international roughness index (IRI). The model results indicated that AADT, SF and IRI were significant to predict road traffic accident frequency with the calculated coefficient. The number of crashes on the road increases with increase in AADT, and higher IRI will increase the frequency5.65 times, while SF enhanced by one unit, NB model shows that crash frequency reduces by 44%. The NB model is more robust that other models in term of Pearson chi-square, and the systematic variation showed that ZINB has the lowest AIC and log-likelihood values, thus ZINB is considered as a more comprehensive model that integrates traffic and pavement condition parameter. Therefore, sufficient statistical evidence have shown the relationship between pavement condition parameters and crash frequency on the road under study. This research suggests the RTA prediction models that could be integrated into pavement rehabilitation, planning and process.

REFERENCES

Abdel-Aty M.A, Essam Radwan E.A 2000. Modelling traffic accident occurrence and involvement, *Accident Analysis and Prevention 32:* 633-642.

Agresti, A. 1984. Categorical data analysis on effect of pavement on traffic accident, *Transportation Research Record; 11388*: 33-39.

Al-Ghamdi, A. S. 2002. Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention*, (pp. 34, 729-741).

Al-Masaeid H.R 1997, Impact of pavement condition on rural road accidents, *Canadian Journal of Civil engineering* 24 (4): 523-531.

Breslow, N. 2004. Extra-Poisson variation in log-linear models. Applied Statistics: 33: 38-44.

Chan Y.V., Huang B., Yan X., Richards S. 2010. Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system. *Journal of Advanced Transportation* 44: 150-161

Chan Y.V., Huang B., Yan X., Richards S. 2009. Relationship between Highway Pavement Condition, Crash Frequency and Crash type. *Journal of Transportation Safety and Security* 1: 268-281

Chandra S. 2004. Effect of road roughness on capacity of two-lane roads. *Journal of Transportation Engineering* 130 (3): 360-364.

Chiou Y.C. 2006. An artificial network-based expert system for appraisal of two-car crash accidents, Accident Analysis and Prevention 38: 777–785.

Delen D., Sharda R., Besson M. 2006. Identification significant predictors of injury severity in traffic accidents using a series of artificial neural network, Accident Analysis and Prevention 38: 434-444.

Federal Road Safety Commission (FRSC). 2012. *2012 Annual Report.* Abuja: FRSC.

Frome, E. L., Cragle, D. L., McLain, R. W. 2000. Poisson regression analysis of the mortality among a cohort of World War II nucfear industry workers. *Radiation Research*, (pp. 123: I38- 152).

Hosmer, D., Lemeshow, S. 1999. Applied logistic regression on effect of pavement on traffic accident. *Accident Analysis and Prevention 44:* 152-164.

Karlaftis M.G., Golias I. 2002. Effect of roadway geometry and traffic volume on rural roadway accident rates. Accident Analysis and Prevention 34 (3): 357–365.

Miaou, S.P., Lum, H. 2003. A statistical evaluation of the effects of highway geometric design on truck accident involvements. *Transportation Research Record; 1407*: 11-23.

Milton, J., Mannering, F. 1997. Relationship among highway geometric, traffic-related elements, and motor-vehicle accident frequencies. *Presented at the 76th Annual Meeting of the Transportation Research Board.* Washington, DC.

Zegeer, C.V, Hummer, J., Reinfurt, D., Herf, L., Hunter, W. 2000. Safety Effect of cross-section design for two-lane road. Volume I and II. University of North Carolina.

(Manuscript received: 1st January, 2017; accepted: 19th April, 2018).