ARID ZONE JOURNAL OF ENGINEERING, TECHNOLOGY & ENVIRONMENT
AZOJETE December 2020. Vol. 16(4):637-650
Published by the Faculty of Engineering, University of Maiduguri, Maiduguri, Nigeria.
Print ISSN: 1596-2490, Electronic ISSN: 2545-5818
www.azojete.com.ng

ORIGINAL RESEARCH ARTICLE

COMPARISON OF MULTIPLE LINEAR AND QUADRATIC MODELS IN ESTIMATING ROAD CRASHES IN SEMI-URBAN TWO-LANE ROADS: CASE STUDY OF NSUKKA MUNICIPAL COUNCIL, SOUTH EASTERN NIGERIA

***P. A. Ozor[1,2] and C. G. Ozoegwu[2]**

[1.] Department of Quality and Operations Management, Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa
[2.] Department of Mechanical Engineering, University of Nigeria, Nsukka, Nigeria

**Corresponding author's email address:* pozor@uj.ac.za, paul.ozor@unn.edu.ng

ARTICLE INFORMATION

Submitted 2 December, 2019
Revised 7 April, 2020
Accepted 13 April, 2020.

Keywords:
Road crashes
rural communities
developing countries
causal factors
Predictive models.

ABSTRACT

Road crashes are generally characterized by occurrence without prior notice as well as low or moderate to severe losses in terms of lives and properties. In particular, rural communities in developing countries count huge losses to crash exigencies due to lack of basic infrastructure. It is important that such accident trends which possess special and distinct features from those observed for Urban settlement be carefully studied and modeled for effective mitigation strategies. This paper proposes models that quantitatively assess the effects of various causal factors of crashes in rural roads of developing communities. Taking data from Nsukka municipality, South East Nigeria, it provides two reliable, statistically significant techniques of predicting accident rates on such roads. The results show that although important causal variables like "illiteracy, weather condition, alcohol and drugs" affect road accident, "reckless driving, over speeding, annual average daily traffic, poor road and mechanical faults" were variables that demand more assessment and control in road accident management in the studied location. Further, surface comparison of the developed models based on overall statistical indices suggests that multiple linear accident models are more accurate, though the quadratic accident model is statistically significant. A point by point calculation of deviations shows that the quadratic accident model, apart from its ability to take care of non-linearity effects of independent road accident variables, is actually more accurate than the linear model, except where the quadratic model suffered more numerical instability from the combination of numerical parameters.

1.0 Introduction

A road crash can generally be denoted by two important phenomena; 1) it is a random and undesirable event, and 2), it occurs without prior notice (this work). This unforeseen (Hoel et al., 2010) nature underscores the need for an analytical approach to estimating possible indices of its occurrence, to plan safety and precautionary measures. The road version of accident simply means the occurrence of accident on the road. Road traffic accident remains a hazard the World have come to live with. The problem of Road accidents have been reported to be on the increase worldwide, mainly due to the disparity between growth in transportation infrastructure with development investment in key sectors like industry, energy, water and real estate (Prasannakumar et al., 2011). This menace has been credited to be the foremost cause of road crashes and orthopedic hospitalization worldwide (Taimur et al., 2011, Liang et al., 2005). Road safety improvements and deteriorations can be measured in terms of fatality rate (Persia et al., 2015) in an optimally designed and constructed road. The need to ensure road safety has been emphasized in the past (Ona and Garach, 2011).

Schepers and Heinen (2013) suggested the substitution of short car trips with cycling to reduce vehicular crashes.

Vital predicators of road accident include but not limited to human and environmental factors. There are numerous causes of road accidents, attitudinal dimensions and attendant safety repercussions (Shinar, 2019; Konkor et al., 2019; McCartan and Elliott, 2018, Dai and Jaworski, 2016). Although some risk factors are more prevalent in certain regions than others (Papadimitriou et al., 2019; Hordofa, 2018), the modalities of a particular road inferno are not easy to model. Horsman and Conniss (2015) express the investigative difficulty in determining a timeline of proceedings with causal factors in a given road crash. There have been concerted efforts geared towards eliciting detailed list of influential road crash habits. The causative agents might be due to mobile phone usage (Lester et al., 2010, Horsman and Conniss 2015, Musicant et al., 2015), annual daily traffic (Polus and Cohen, 2012), thoroughfare and environmental dependent factors (Abdul Manan et al., 2018; Dai and Jaworski, 2016, Aworemi et al., 2010). Undetected mechanical imperfections, such as faulty tire, bad braking system, or steering malfunctioning can influence crashes in rural roads. Alcohol drinking (Scot-Parker et al., 2014) is another important lead up to road accident. Road users under the influence of psychoactive substances are more likely to experience poor judgment in the presence of some road features like the pavement, length, sharp bends, intersection and the traffic control system. In particular, multiple vehicle collision accident situations can be traced to the action or inaction of the driver of one or both of the vehicles involved, especially in two lane roads. Driver error can increase crash occurrence in many ways (Shim et al., 2015); including going against the traffic, indiscriminate stopping and packing, reckless overtaking, disobedience to both traffic laws, lights and traffic controllers. There are instances where drivers underestimate the speed and overestimate the distance of an approaching vehicle and try to overtake with uncontrollable speed (McCartan and Elliott, 2018; Abegaz et al., 2014). Road accident can also happen due to illiteracy (poor or no knowledge of road codes and signs) on the path of road users or disrespect for existing road rules where knowledge is available. Road accidents are usually characterized by personal, property, vehicle or structural injuries (Konkor et al., 2019; Quistberg et al., 2014).

Road accident prediction and reduction models can provide efficient decision tool in re-designing and maintenance of a given road. It can also guide stake holders, government agencies and road management authorities in formulating road safety policies in the light of proven road accident initiation events, prevalent in specific geographical locations. Though, considerable efforts have been made in the past to model road crashes (Russo et al., 2017; Vilaca et al., 2017; Bagloee and Asadi, 2016; Wu and Jovanis, 2012), development of a generalized prediction model for road accidents can be difficult. This is partly due to the large variety in accident forms and causal factors with respect to specific countries, road types and locations. Past authors have developed algorithms based on the Dodge model to tackle road accident mitigation realities (Xu and Huang 2016). The field had equally seen the evolution of Interactive Highway Safety Design Model (IHSDM) which incorporates a Crash Prediction Module (CPM) for estimating accident frequency on road segments (Aurelio et al., 2012). The IHSDM covers mostly the parameters of interest in crash estimation within the particular regions for which it was designed. A regression analysis model for predicting accidents using exposure and geometry related parameters as explanatory variables is not farfetched (Ona and Garach, 2011). The work demonstrate how speed reduction can affect road safety. The presentation examined consistency of road design based on the consistency criterion developed during the end of the 20th Century (Lamm et al., 1988). Polus and Cohen (2012) proposed a non-canonical Poisson regression methodology for the analysis and prediction of crashes on low-volume rural roads. The database contained accidents in all two-lane highways in Isreal and considered parameters of volume below 3000 annual daily traffic (ADT). Other parameters of concern were length of section as well as degree of fatal and serious crashes, which were expressed as whole numbers. The percentage time-spent-following (PTSF) was used to assess the percentage of time with which a high-speed moving vehicle is impeded by a slow moving one. The authors employ

Poisson distribution to account for non-negative integer value of occurrences and recommended it for counted data. The PTSF does not apply in this study because the peculiarities of the case example do not permit fast movement and 95% of the roads are two-lane.

There were other attempts made towards formulating Poisson regression-based road accident generalized models (Lenguerrand and Martin, 2006, Caliendo et al., 2007). Significant amount of result of such modeling technique show that crash rate increases with either increasing speed or traffic flow. In particular, strong correlation existed between the number of crashes, traffic flow, and infrastructure features (Caliendo et al., 2007). Increase in population and road traffic violation in a given area can lead to road accidents in a large scale (Ateeth et al., 2012). The use of only linear regression models lack competence in adequately describing the nature of crash frequency data (Joshua and Garber, 1990, Polus and Cohen, 2012). There exists negative binomial Poisson regression model for crash prediction (Shankar et al, 1995) and multivariate Poisson lognormal regression model for accident frequency analysis (Ma et al., 2008). The studies introduce geometric variables like horizontal and vertical alignment as well as environmental factors; including number of rainy days, snow and rainfall as well as states of nature. Taking road inventory database from the Indiana Department of Transportation (INDOT), and the Accident Information record from the Indiana State Police, Karlaftis and Golias (2002) present a non-parametric statistical methodology called hierarchical tree-based regression for predicting probability of accident occurrence on rural roads. The work elaborated on the effects of typical roadway geometric indicators on crash rates. Multivariate Bayesian modeling techniques has been added to RA literature, in an attempt to compliment the inadequacy of linear models formulated with Poisson distribution for road accident modelling (Liu and Sharma, 2018; Cai et al., 2017). Joshua and Garber (1990) combine linear regression and Poisson regression technique to model the frequency of road crashes involving Trucks.

One of the aims of this study is to determine the factors that affect accident within the study area and predict how significant the factors contribute to road accident. The literacy level and exposure of traffic officers delineated during six months of opinion surveys and field trips, were relied upon to examine the statistical significance and accuracy of two modelling approaches. The work became necessary due to the difficulty in application of existing sophisticated RA models in the case study area. The basic concern is to make the analytic method as easy to understand and apply as possible, without any loss in quality of its purpose. Use of multiple linear regressions is considered appropriate. This technique was used because it not only predicts how significant the causal factors contribute to accident, but conveniently leads to an in-depth analysis of the data, that is; it explains the relationship that exists between the independent variables and provides adequate explanation for those variables which contribute to the dependent variable. In particular, quadratic accident model (QRM) is explored to offer alternative modeling methodology different from negative binomial, non-canonical Poisson regression or multivariate Bayesian approaches reported widely in literature. The choice of the two models was informed by two reasons. 1) Both models, apart from providing in-depth analysis of the variables that cause accidents in the location, gives an idea of the significance of each variable with reference to the dependent variable. 2) Quadratic accident model can take care of any non-linearity effect of those independent variables with time windows. For instance, during school runs, rainy periods or road blocks. The authors' opinion is that the accuracy of each should be investigated and elicited for policy purposes. There is no existing publication of this sort known to the authors.

2. Materials and Methods

The study used qualitative and quantitative methods. Road accident data was obtained from both primary and secondary sources. Part of the data was sourced by the use of structured questionnaires administered to 1000 respondents drawn from various categories of road users and traffic stake holders in Nsukka, Nigeria. The questionnaire was designed to find out the gender, age, ownership of automobile (including bicycles) and road literacy level (indicating whether the road user attended driving school, has knowledge of road signs, has knowledge of Federal Road Safety Codes and rights of way) of respondents. Other data obtained with the questionnaire included accident involvement or witness index (indicating number of times the respondent was involved in road accident or witnessed one). Respondents that either involved-in or witnessed an accident were requested to specify where it occurred, how it occurred, time of occurrence, accident degree and causal factor (over speeding, recklessness, poor road maintenance, poor vehicle maintenance, road block, weather condition, illiteracy, alcohol and drug, others). Adequate provision was made for respondents to specify the exact cause of accident if the causal factors mentioned are not applicable. Respondents included the motoring public, residents, the Police Traffic Department, Federal Road Safety Commission (FRSC) - RS. 9.11, and Ministry of Works. Hourly traffic data was obtained from three groups of respondents: Police Traffic Unit (PTU), FRSC and Vehicle Inspection Officers (VIO) in the metropolis. The PTU, FRSC and VIO use special instrument (loop detectors) for gathering and recording traffic data. In cases where multiple competing digits or varying answers were obtained for a particular road from all three sources (PTU, FRSC and VIO), an average of the whole values was relied upon. The various data as collected from the sources were screened for both noise and outliers that could result from errors.

2.1 Model I: Multiple Linear Regression Model

Multiple linear regression allows the estimation of a dependent variable from independent variables. In this paper, the basic assumption that any causal factor that does not lead to road accident in up to four years out of the nine years for which data is available, is treated as an outlier, was made. The general form of the model is given by Equations (1) and (2) (Levin, 1998):

$$Y_i = \beta_0 + \beta_1 X_2 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e \tag{1}$$

Where: $i = 1,2\ldots5, Yi = Dependent\ variable,$ $X_{1,2,\cdots5}$ are the Independent variables, βo is The Regression constant, $\beta_{1,2,\cdots5}$ is the Regression Coefficients, e is the Random error associated with Y_i.

While some of the error terms may be small and others large, the variability of the error terms is in no way related to the independent variable used. The constants can be deduced with the following equations:

$$\begin{aligned}
\Sigma Y &= n\beta_0 + \beta_1\Sigma X_1 + \beta_2\Sigma X_2 + \beta_3\Sigma X_3 + \beta_4\Sigma X_4 + \beta_5\Sigma X_5 \\
\Sigma(X_1Y) &= \beta_0\Sigma X_1 + \beta_1\Sigma X_1^2 + \beta_2\Sigma X_1X_2 + \beta_3\Sigma X_1X_3 + \beta_4\Sigma X_1X_4 + \beta_5\Sigma X_1X_5 \\
\Sigma(X_2Y) &= \beta_0\Sigma X_2 + \beta_1\Sigma X_1X_2 + \beta_2\Sigma X_2^2 + \beta_3\Sigma X_2X_3 + \beta_4\Sigma X_2X_4 + \beta_5\Sigma X_2X_5 \\
\Sigma(X_3Y) &= \beta_0\Sigma X_3 + \beta_1\Sigma X_1X_3 + \beta_2\Sigma X_2X_3 + \beta_3\Sigma X_3^2 + \beta_4\Sigma X_3X_4 + \beta_5\Sigma X_3X_5 \\
\Sigma(X_4Y) &= \beta_0\Sigma X_4 + \beta_1\Sigma X_1X_4 + \beta_2\Sigma X_2X_4 + \beta_3\Sigma X_3X_4 + \beta_4\Sigma X_4^2 + \beta_5\Sigma X_4X_5\Sigma(X_5Y) = \beta_0\Sigma X_5 + \beta_1\Sigma X_1X_5 + \\
&\beta_2\Sigma X_2X_5 + \beta_3\Sigma X_3X_5 + \beta_4\Sigma X_4X_5 + \beta_5\Sigma X_5^2
\end{aligned} \tag{2}$$

Where:
For the particular case of the example in this study;
Y= Number of Accidents, X_1= Reckless Driving, X_2 = Over Speeding, X_3= AADT, X_4= Poor Road, X_5 = Mechanical Fault

2.2. Model II: Quadratic Accident Model

The general p-order polynomial least squares model for a d-dimensional multi-variate response can be given as (Levin, 1998):

$$Y(\boldsymbol{x}) = [\boldsymbol{a}(\boldsymbol{x})]^T\left\{\sum_{i=1}^{n} \boldsymbol{a}(\boldsymbol{x}_i)[\boldsymbol{a}(\boldsymbol{x}_i)]^T\right\}^{-1} \sum_{i=1}^{n} \boldsymbol{a}(\boldsymbol{x}_i)Y_i \qquad (3)$$

Where: $\boldsymbol{a}(\boldsymbol{x}) = \{a_1(x)\, a_2(x) ... a_r(x)\}^T$ is the polynomial of basis vector, n is the number of sampling runs. The length r of $\boldsymbol{a}(\boldsymbol{x})$ is generally given as:

$$r = \frac{(d+p)!}{(d!p!)} \qquad (4)$$

where the dimension d is the number of independent variables and p is the order

2.3 Significance of Parameters

On the basis of the data of Table 2, the selected independent variables used in the multi-linear and quadratic linear accident prediction modeling has the following significances:

1. Reckless driving

This factor covers going against the traffic, neglect of right of way, dangerous over taking and any playful act while driving.

2. Over speeding

This describes a situation where the permissible (recommended) speed limit for any road is exceeded by a driver, thereby leading to a road crash.

3. Annual Average Daily Traffic

This index, frequently abbreviated as AADT, is used to denote the total annual volume of vehicular traffic on a particular road divided by 360.

4. Poor road

This takes into account all manner of road failures or encumbrances that resulted in a road crash. It can include potholes, road blocks, damaged vehicles left unattended, etc.

5. Mechanical fault

Mechanical fault refers to vehicles with one or more known problems or failed parts without maintenance that eventually resulted in RA.

2.4. Annual Average Daily Traffic (AADT/Traffic volume)

The daily traffic volume specifications were part of the information obtained through the questionnaire. Only the available data were used to predict AADT in each year. An average growth rate (r) of 5 percent for the succeeding year was assumed, based on opinion surveys and interviews with Vehicle licensing officials (VIO). Documented evidence shows that AADT can be estimated from the identified traffic volume according to the dependence of equation (5) (Singh and Suman, 2012):

$$\mathrm{A} = \mathrm{P}\left(1 \pm \frac{\mathrm{r}}{100}\right)^{\mathrm{n}} \qquad (5)$$

Where $A = AADT$ of the year being predicted, $P = AADT$ of current year, n = number of Years under consideration (size of sample years); a negative sense designates before the current year traffic count while the current year count is indicated by a positive sign.

2.5. Multi-linear Regression Models Specification

The multiple linear regression models attempt to estimate the relationship between two or more explanatory variables (X) as associated with a value of the dependent variable (Y). In this paper, the general form of the model presented in equation (2) can be explored to fully specify the values of the parameters. Application of the values of the parameters in Table 1 through Table 2 to equation (3) results in the dependence relationships depicted in equation (6). Accordingly:

$$
\begin{aligned}
654 &= 9\beta_0 + 415\beta_1 + 180\beta_2 + 22691.05\beta_3 + 19\beta_4 + 13\beta_5 \\
33108 &= 415\beta_0 + 21173\beta_1 + 8955\beta_2 + 914344.33\beta_3 + 969\beta_4 + 613\beta_5 \\
13225 &= 180 + 8955\beta_1 + 3742\beta_2 + 450513.75\beta_3 + 395\beta_4 + 272\beta_5 \\
1631077.64 &= 22691.05\beta_0 + 914344\beta_1 + 450513.75\beta_2 + 59309175.42\beta_3 + 46386.93\beta_4 \\
&\quad + 41586.37\beta_5 \\
1492 &= 19\beta_0 + 969\beta_1 + 395\beta_2 + 46386.93\beta_3 + 55\beta_4 + 23\beta_5 \\
970 &= 13\beta_0 + 613\beta_1 + 272\beta_2 + 41586.37\beta_3 + 23\beta_4 + 27\beta_5
\end{aligned} \quad (6)
$$

A few more steps of Mathematical analysis of equation (6) yields the following values of each parameter.

$\beta_0 = 3.346,\ \beta_1 = 1.106,\ \beta_2 = 1.123,\ \beta_3 = -0.001,\ \beta_4 = -0.663,\ \beta_5 = 0.154$

2.6 Quadratic Regression Models Specification

As presented in equations (4) and (5) for the general p-order polynomial least squares model for a $d-$ dimensional multi-variate response, we consider for illustration the case where $p = 2$ and $d = 2$ (the quadratic bivariate case) to obtain:

$$\boldsymbol{a}(\boldsymbol{x}) = \{1\ X_1\ X_2\ X_1X_2\ X_1^2\ X_2^2\}^T \text{ and } r = 6. \quad (7)$$

Inserting $n = 9$ in equation (4) for the case of multiple $(d = 5)$, linear $(p = 1)$, regression gives the multiple linear accident model as:

$$Y(\boldsymbol{x}) = 3.34587363558554 + 1.10645537492239X_1 + 1.12311165841290X_2 - 0.00133250106231X_1X_2 - 0.66279908020189X_1^2 + 0.15444261162054X_2^2 \quad (8)$$

Also inserting $n = 9$ in equation (4) for the case of multiple $(d = 5)$, Quadratic $(p = 2)$, regression gives the quadratic accident model as:

$$
\begin{aligned}
Y(\boldsymbol{x}) = {} & 29.09080968576436 - 2.20877716640020X_1 - 0.85492873755439X_2 - 0.00361962189169X_3 + \\
& 34.30042009419936X_4 + 1.54128325239977X_5 + 0.09738583088201X_1X_2 + 0.00001863705064X_1X_3 - \\
& 0.28345401496583X_1X_4 + 0.02856208489972X_1X_5 + 0.00023687839939X_2X_3 + \\
& 1.33060383044842X_2X_4 + \\
& 0.61570829132689X_2X_5 + 0.00121973323749X_3X_4 - 0.00084714037295X_3X_5 - \\
& 6.64636921194101X_4X_5 + \\
& 0.01695289120036X_1^2 - 0.13852515708679X_2^2 + 0.00000251868335X_3^2 - 6.58083216092202X_4^2 - \\
& 0.66968073658966X_5^2
\end{aligned} \quad (9)
$$

3. Results and Discussion

This section presents an analysis of the data collected for estimating road crashes in the case study area. The first stage used SPSS for Windows to articulate a regression output for the dependent and independent variables as well as obtain approximate values of other important statistical indices related to the study. Later, the multiple linear accident model result is compared with the field data. Table 1 displays the summary of all the data collected through questionnaire, while Table 2 indicates the dependent variables as a function of the independent variables.

Table 1: Summary of Questionnaire data

Year	NC	MjA	MnA	RD X_1	OS X_2	AADT	PR X_3	MF X_4	Il X_5	WC	AD
2015	112	41	71	43	26	**4125**	20	9	11	0	3
2014	116	47	69	39	29	**3935**	25	11	12	0	0
2013	132	44	88	41	32	**3679**	19	22	15	1	2
2012	115	39	76	37	29	**3579**	17	13	16	2	1
2011	94	41	53	39	21	**3369**	11	3	19	0	1
2010	109	36	73	57	19	**3200**	14	2	17	0	0
2009	101	34	67	43	22	**3198**	13	8	15	0	0
2008	115	38	77	52	17	**2874**	16	5	23	1	1
2007	82	26	56	21	35	**2744**	10	1	15	0	0
2006	135	42	93	61	25	**2607**	21	1	25	1	1
2005	125	44	81	48	30	**2476**	22	1	21	1	2
2003	116	47	69	55	23	**2235**	13	3	22	0	0
2002	96	30	66	34	21	**2123**	15	1	23	1	1
2001	139	42	97	62	24	**2017**	23	1	24	2	2
2000	117	29	88	47	15	**1916**	27	1	27	0	0

Key: NC-Number of crashes, MjA- Major Accident, MnA- Minor Accident, RD- Reckless driving, OS- Over speeding, AADT- Annual Average Daily Traffic, PR- Poor road, MF, Mechanical fault, Il-Illiteracy, WC- Weather condition, AD- Alcohol or drug

Table 2: Values of Dependent and Independent Variable. (Police and Questionnaire)

Year	No. of Accident	Reckless Driving (X_1)	Over speeding (X_2)	AADT(X_3)	Poor road (X_4)	Mechanical fault (X_5)	Error (E_i=e_i)
2011	66	39	21	3369.33	1	3	E_1
2010	82	57	19	3200.86	4	2	E_2
2007	18	12	5	2744.34	0	1	E_3
2006	94	61	25	2607.12	1	1	E_4
2005	84	48	30	2476.78	2	1	E_5
2003	84	55	23	2235.28	3	3	E_6
2002	63	34	21	2123.52	2	1	E_7
2001	97	62	24	2017.34	2	1	E_8
2000	66	47	15	1916.48	4	0	E_9

In the questionnaire, road accident was pre-classified in line with the recommendation of Zakwan (2011). Fatal accident refers to a case where death results from injury sustained in less than 30 days after the accident occurred. In serious accidents, victims are hospitalized due to injury sustained from the accident, but without death within 30 days of its occurrence. Minor or Slight accident results in such injuries as sprain, bruise, cut or even a slight damage on vehicles or properties. The data for year 2004 and 2008 were not available at the time of this study. Figure 1 presents the Annual Average Daily Traffic (AADT) estimated from the collected data. The predicted traffic volume was computed using equation (1). Results of the questionnaire data as statistically analyzed are displayed in Table 1. The specific values of the independent variables as determined from the questionnaire and Police data are presented in Table 2. It is noteworthy that Table 2 shows only the values of the dependent and independent variables and the expected errors for each year. Figure 1 presents the predicted traffic volume.

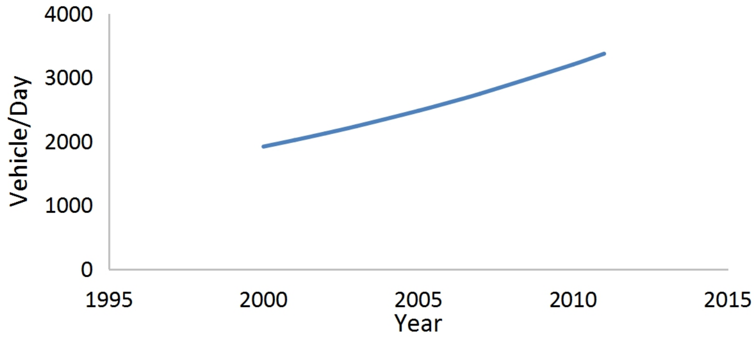

Figure 1: Predicted Traffic Volume

3.1. Regression output from SPSS

Table 3 gives the model summary developed with the statistical software, SPSS and reports the strength of the relationship between the model and the dependent variable (expected number of accidents). The multiple correlation coefficients, R (0.998) indicates a good linear relationship between the observed and predicted values of the dependent variable. The coefficient of determination by the software is $R^2 = 0.996$. This means that 99.6% of the explanatory variables determine the expected number of accident. The coefficient of determination developed by the software ($R^2 = 0.996$) is very close to 1, hence the model holds good fit.

Table 3. Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.998	0.996	0.989	2.48240

The computed variances are displayed in Table 4. Table 5 shows the coefficients of the regression line. Equation (10) became the regression equation.

Table 4: Analysis of Variance (ANOVA)

Model		Sum of Squares	df	Mean Square	F	Significance
	Regression	4543.513	5	908.703	147.46	.001
1	Residual	18.487	3			
	Total	4562.000	8			

Table 5: Coefficients

Model	Un-standardized Coefficients		Standardized Coefficients		Significance
	B	Std. Error	Beta	t	
(Constant)	3.346	6.456		0.518	0.640
Recklessness	1.106	0.1	.739	11.104	0.002
Over speeding	1.123	0.198	.333	5.667	0.011
AADT	-.001	0.002	-.029	-0.586	0.599
Poor Road	-.663	0.836	-.038	-0.793	0.486
Mech. Fault	-.154	1.125	.007	0.137	0.900

$$\mathrm{Y} = 3.349 + 1.106\mathrm{X}_1 + 1.123\mathrm{X}_2 - 0.001\mathrm{X}_3 - 0.663\mathrm{X}_4 + 0.154\mathrm{X}_5 \qquad (10)$$

Figures 2 and 3 indicate that the normality assumption for the model is not violated because violation of normality compromises the estimation of the coefficients. Finally, the plot of residuals by the predicted values shows that the variance of the error increases with increasing predicted number of accidents.

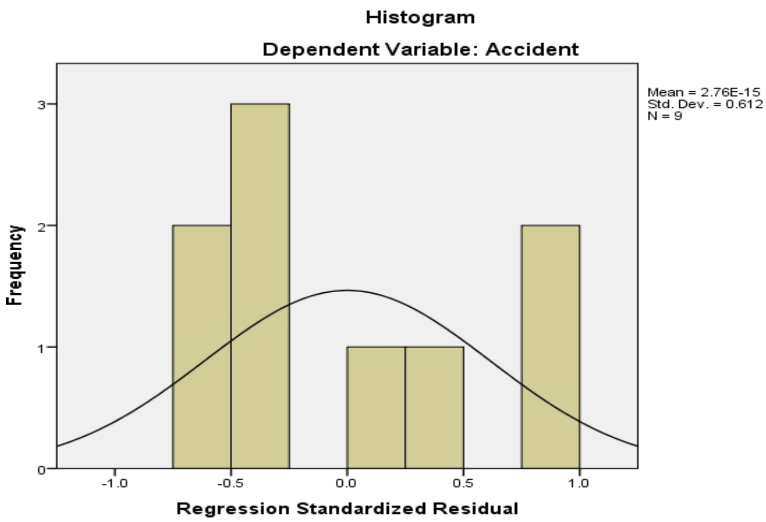

Figure 2. Histogram

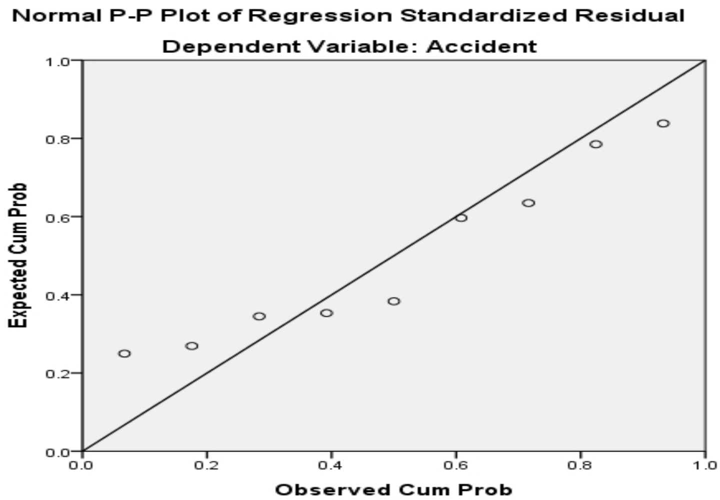

Figure 3 Normal p-p plot of Regression Standardized Residual

3.2 Comparison of Multiple linear Accident Model Result with Field Data

Figure 4 compares the developed accident prediction linear model results with the actual field data. The number of accidents is plotted against year of occurrence. The field data has only a slight observable difference (between 2003 and 2006) with that of the model developed. This means that the model developed can predict the number of accidents with reasonable degree of accuracy.

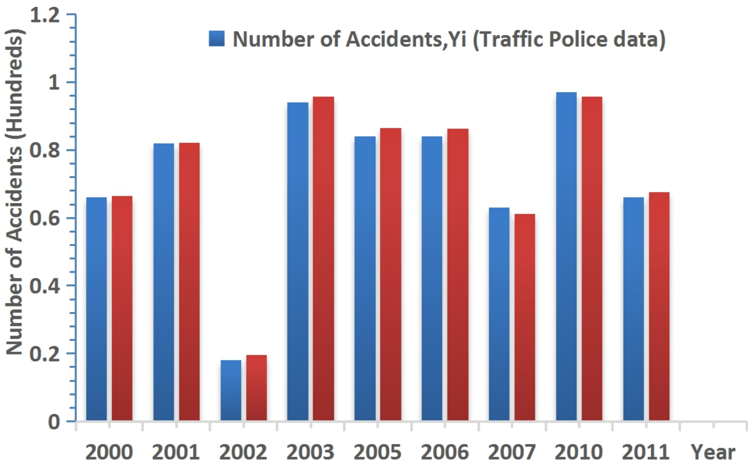

Figure 4. Comparison of number of accidents found by linear model with field data

The overall goodness of fit of the multiple linear and quadratic models were compared on the basis of some statistical error indices. The indices of comparison were the coefficient of determination (R^2), root mean square error (RMSE), mean bias error (MBE), mean absolute bias error (MABE), mean percentage error (MPE), t-Test statistic and correlation coefficient (ρ). More efficient models give R^2 values closer to unity, but there is higher statistical significance when there is lower t-value (Rohani et al., 2011; Levis, 1998). A model is considered statistically significant when the t-value is smaller than the critical t-value of a confidence level. Higher correlation coefficient is an indication of stronger relationship between model and experimental data. The indices are calculated for the linear and quadratic models and results presented in Table 6.

It is seen from values of R^2 in Table 6, that the two models have good fit when judged against 95% confidence level. The R^2 values for the multiple linear and quadratic accident models are respectively 0.9959 and 0.9572, indicating that the former is more reliable in predicting accident occurrence in the studied location. It is seen that each of the indices RMSE, MBE, MABE and MPE is smaller in the linear model than in the quadratic model. This index further supports a better predictive capacity of the multi-linear model. The t-value of each model is less than the critical t-value of 2.201 (at 95% confidence level) indicating that the two models have very good statistical significance. The t-values for the multiple linear and quadratic accident models are respectively 6.0842×10-13 and 0.8127 indicating that the former is more statistically significant than the latter. Even though the correlation coefficients, ρ of each of the models are very good being above 0.98. The linear model performs better on the basis of this index. From the foregoing, it may be said conclusively that the multiple linear accident model of the studied location is highly reliable and that though the multiple quadratic accident model is statistically significant, it does not perform as good as the linear model. This conclusion on superiority of the linear model should be quoted with care as point by point calculation of deviations of the models from the field data would suggest otherwise.

Table 6: Statistical error indices of the multiple linear and quadratic accident models

Model	R^2	RMSE	MBE	MABE	MPE [%]	t-Test	ρ
Multiple Linear Regression Model	0.9959	1.4332	-3.0830×10^{-13}	1.3043	0.2819	6.0842×10^{-13}	0.9980
Multi-Quadratic Regression Model	0.9572	4.6563	-1.2858	1.8007	-1.3172	0.8127	0.9803

Table 7: Statistical error indices of the multiple linear and quadratic accident models

Model	R^2	RMSE	MBE	MABE	MPE [%]	t-Test	ρ
Multiple Linear Regression Model	0.9957	1.4838	-0.1169	1.3504	0.1928	0.2091	0.9979
Multiple Quadratic Regression Model	0.9995	0.5265	0.2896	0.2896	0.3651	1.7425	0.9998

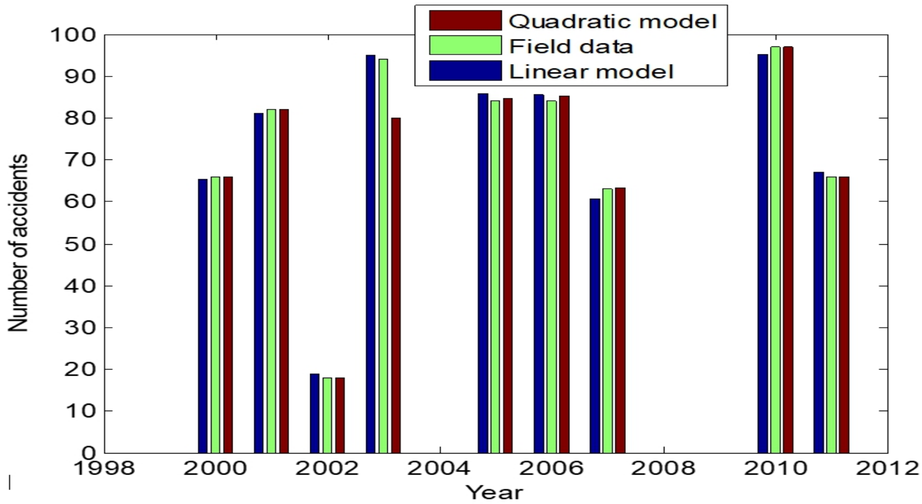

Figure 5. Plot of the field data together with the results of the two models

Figure 5 is a presentation of the field data together with the results of the models. It is seen from comparing the tips of the bars that the quadratic model is more accurate at all data points except for the year 2003. Figure 6 is a plot of magnitude of deviations of the two models from the measured data, to confirm this observation. Therefore, in the studied location, the quadratic model is actually more accurate than the linear model (even though the linear model has better overall statistical indices) except for the year 2003 where it (the quadratic model) seems to have suffered more numerical instability from the combination of numerical parameters. This is verified further by re-computing all the statistical parameters with the exclusion of the year 2003 data. The result is presented in table 7 in which it is seen that the quadratic model perform better in predicting the accident data as judged from the R^2 and ρ values which are almost equal to one. The interaction effects captured in the quadratic model could be the explanation for the better point to point accuracy.

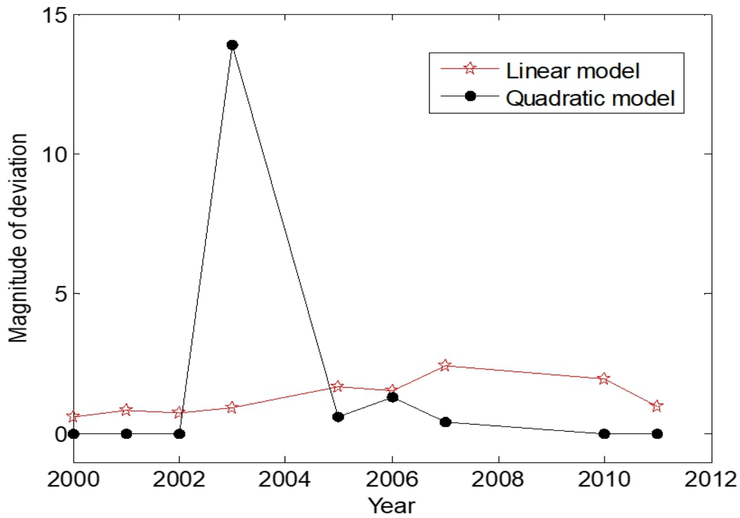

Figure 6. Plot of magnitude of deviations of the two models.

4.0 Conclusion

Based on this study, the following conclusions can be drawn under conditions and causal factors such as obtainable in two-lane roads in the studied rural area:

Both models provide a good means of predicting number of road accidents. The ability of the models is evaluated to reasonably forecast the likely number of rural road crashes as function of the causal parameters using road accident data from the studied location. This study has generated reliable and practical models for the road managers and policy makers using the identified accident causal variables and environmental conditions.

Surface comparison of the multiple linear and quadratic accident model based on overall statistical indices would suggest that the former is highly reliable and that though the latter is statistically significant, it does not perform as good as the linear model. A point-by-point calculation of deviations of the models from the field data contradicts the conclusion on superiority of the linear model over quadratic models. Overall, the quadratic accident model is actually more accurate than the linear model except where it (the quadratic model) seems to have suffered more numerical instability from the combination of numerical parameters.

References

Abdul-Manan, MM., Varhelyi, A. Çelik, AK. and Hashim, HH. 2018. Road characteristics and environment factors associated with motorcycle fatal crashes in Malaysia. IATSS Research, 42: 207–220, https://doi.org/10.1016/j.iatssr.2017.11.001

Abegaz, T. Berhane, Y. Worku, A. and Assefa, A. 2014. Effects of excessive speeding and falling asleep while driving on crash injury severity in Ethiopia: A generalized ordered logit model analysis. Accident Analysis and Prevention, 71: 15 – 21.

Ateeth, KT. Harikrishna, N. and Vigneesh, P. 2012. Intelligent vehicle control based on identification of road and traffic signal operated RFID transponders. Proceedings of the International conference on Advances in Electrical and Electronics Engineering, 11-12 February, Penang, Malaysia, 339-343, Available online, http://psrcentre.org/images/extraimages/4.%2012179.pdf, Accessed on 26th June, 2019

Aurelio, M., Paolo, P. and Nicola, F. 2012. Evaluation of the applicability of IHSDM crash prediction module on Italian two-lane rural roads. In SIIV - 5th International Congress - Sustainability of Road Infrastructures. Procedia - Social and Behavioral Sciences, 53: 933 – 942

Aworemi, JR., Abdul, A., Ibraheem, A., Olabode, FO. and Segun, O. 2010. Analytical study of the causal factors of road traffic crashes in South-West, Nigeria. Educational Research, 1: (4) 118-124.

Bagloee, SA. and Asadi, M. 2016. Crash analysis at intersections in the CBD: A survival analysis model. Transportation Research, Part A, 94: 558–572, http://dx.doi.org/10.1016/j.tra.2016.10.019

Cai, Q., Abdel-Aty, M. and Lee, J. 2017. Macro-level vulnerable road users crash analysis: A Bayesian joint modelling approach of frequency and proportion. Accident Analysis and Prevention 107: 11–19, http://dx.doi.org/10.1016/j.aap.2017.07.020

Dai, D. and Jaworski, D. 2016. Influence of built environment on pedestrian crashes: A network-based GIS analysis. Applied Geography 73: 53-61,

http://dx.doi.org/10.1016/j.apgeog.2016.06.005

Hoel, LA., Garber, NJ. and Sadek, AW. 2008. Transportation Infrastructure Engineering: A Multi-Modal Integration, Second Edition. Chris Carson Publishers, A division of Thomson Canada Limited.

Hordofa, GG., Assegid, S., Girma, A. and Weldemarium, TD. 2018. Prevalence of fatality and associated factors of road traffic accidents among victims reported to Burayu town police stations, between 2010 and 2015. Ethiopian Journal of Transport and Health, 10: 186–193,

https://doi.org/10.1016/j.jth.2018.06.007

Horsman, G. and Conniss, LR. 2015. Investigating evidence of mobile phone usage by drivers in road traffic accidents. Digital Investigation, 12: S30–S37, http://dx.doi.org/10.1016/j.diin.2015.01.008

Joshua, S. and Garber, N. 1990. Estimating truck accident rate and involvement using Linear and Poisson Regression models. Transportation Planning and Technology, 15: 41–58.

Karlaftis, MG. and Golias, I. 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. Accident Analysis and Prevention, 34: 357–365

Konkor, I., Kansanga, M., Sano, Y., Antabe, R. and Luginaah, I. 2019. Community perceptions and misconceptions of motorcycle accident risks in the Upper West Region of Ghana. Travel Behaviour and Society, 15: 157–165, https://doi.org/10.1016/j.tbs.2019.03.001

Lamm, R., Choueiri, EM., Hayward, JC. and Paluri, A. 1988. Possible design procedure to promote design consistency in highway geometric design on two-lane rural roads. Transportation Research Record 11: 111–122.

Lenguerrand, E., Martin, JL. and Laumon, B. 2006. Modelling the hierarchical structure of road crash data—application to severity analysis. Accident Analysis and Prevention, 38: 43–53

Levin, D. 1998. The approximation power of moving least-squares. Mathematics of Computation, 67(24): 1517–1531.

Liang, LY., Hua, LT. and Ma'some, DM. 2005. Traffic accident application using Geographic Information System. Journal of Eastern Asia Society Trans Studies, 6: 3574 – 3589

Liu, C. and Sharma, A. 2018. Using the multivariate Spatio-Temporal Bayesian model to analyze traffic crashes by severity. Analytic Methods in Accident Research, 17: 14–31, https://doi.org/10.1016/j.amar.2018.02.001

Ma, J., Kockelman, K.M. and Damian, P. 2008. A multivariate Poisson-Lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention, 40: 964-975.

McCartan, R. and Elliott, MA. 2018. Bi-dimensional attitudes, attitude accessibility and speeding behaviour. Transportation Research, Part F, 58: 581–593, https://doi.org/10.1016/j.trf.2018.06.036

Musicant, O., Lotan, T. and Albert, G. 2015. Do we really need to use our smartphones while driving? Accident Analysis and Prevention, 85: 13 –21.

Ona, J.D. and Garach, L., 2011. Accidents prediction model based on speed reduction on Spanish two-lane rural highways. Procedia - Social and Behavioral Sciences, 53: 1011 – 1019

Papadimitriou, E., Filtness, A., Theofilatosc, A., Ziakopoulos, A., Quigley, C. and Yannis, G. 2019. Review and ranking of crash risk factors related to the road infrastructure. Accident Analysis and Prevention, 125: 85–97, https://doi.org/10.1016/j.aap.2019.01.002

Persia, L., Gigli, R. and Usami, DS. 2015. Smeed's law and expected road fatality reduction: An assessment of the Italian case. Journal of Safety Research, 55: 121-133

Polus, A. and Cohen, M. 2012. A new non-canonical poisson regression model for the prediction of crashes on low-volume rural roads. International Association of Traffic and Safety Sciences (IATSS) Research, 35: 98-103

Prasannakumar, V. Vijith, H. Charutha, R. Geetha, N. 2011. Spatio-temporal clustering of road accidents: GIS based analysis and assessment. Procedia Social and Behavioral Sciences, 21: 317–325.

Quistberg, DA., Koepsell, TD., Boyle, LN., Miranda, JJ., Johnston, BD. and Ebel, EE. 2014. Pedestrian Signalisation and the risk of Pedestrian – Motor vehicle collisions in Lima, Peru. Accident Analysis and Prevention, 70: 273 – 281.

Rohani, A., Abbaspour-Fard, HM., Abdolahpour, S. 2011. Prediction of tractor repair and maintenance costs using Artificial Neural Network. Expert Systems with Applications, 38: 8999–9007

Russo, F., Di Pace, R., Dell'Acqua, G. and de Luca., S. 2017. Estimating an injury crash rate prediction model based on severity levels evaluation: the case study of single-vehicle run-off-road crashes on rural context. Transportation Research Procedia, 27: 1088–1096

Schepers, JP. and Heinen, E. 2013. How does a modal shift from short car trips to cycling affect road safety? Accident Analysis and Prevention, 50: 1118 – 1127

Scott-Parker, B., Waston, B., King, MJ. and Hyde, MK. 2014. I drove after drinking alcohol and other risky driving behaviours reported by young novice drivers. Accident Analysis and Prevention, 70: 65 – 73.

Shankar, VN., Mannering, FL., Baefield, W. 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention, 27: (3) 371–389.

Shim, J., Park, SH. and Chung, S. 2015. Enforcement avoidance behavior near automated speed enforcement areas in Korea expressways. Accident Analysis and Prevention, 80: 57–66.

Shinar, D. 2019. Crash causes, countermeasures, and safety policy implications, Accident Analysis and Prevention, 125: 224–231, https://doi.org/10.1016/j.aap.2019.02.015

Singh, KS. and Suman, SK., 2012. Accident analysis and prediction of model on national highways, International Journal of Advanced Technology in Civil Engineering, 1: (2) 25-30

Taimur, U., Liping, F. and Luis, F. 2011. Accident prediction models for winter road safety. Journal of the Transportation Research, 22(3): 144-151.

Vilaca, M., Silva, N. and Coelho, MC. 2017. Statistical analysis of the occurrence and severity of crashes involving vulnerable road users. Transportation Research Procedia, 27: 1113–1120,

Wu, K. and Jovanis, P. 2012. Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. Accident Analysis and Prevention, 45: 507-516.

Xu, S. and Huang, H. 2016. Traffic crash liability determination: Danger and Dodge model. Accident Analysis and Prevention, 95: 317–325, http://dx.doi.org/10.1016/j.aap.2016.06.001

Corresponding author's e-mail address: pozor@uj.ac.za