On Road Traffic Fatalities Modelling in Nigeria

Osowole, O. I., & Henshaw, B. B., Aghamie, S., Balogun, K.,
1Department of Statistics, University of Ibadan, Nigeria
2Department of Mathematics, College Of Education, Agbor, Delta State, Nigeria
3 Federal School of Statistics, Ajibode, Ibadan, Oyo State, Nigeria
Email: Corresponding Author: academicprofessor2013@gmail.com

Abstract

Road traffic fatalities are on the rise globally and it is a significant cause of violent deaths in Nigeria after Boko Haram insurgency with an estimated loss of N80bn yearly. Prediction of road fatalities with little error and identification of key road crash correlates have always been the concern of stakeholders interested in developing road safety mechanisms. Accordingly, this study considered two regression models as plausible models based on an official road traffic fatality data. The results indicated that the estimated Negative Binomial regression model did well for the data under consideration. Over-speeding, Break failure, Loss of control, Sign light violation, Route violation, Poor weather and Fatigue were the key determinants of road traffic fatality identified. This implies that this model works well under the setting under consideration; especially when over-dispersion is present in the data.

Keywords: Road traffic fatality, Road safety, Accident statistics, Modern transportation system, Assessment criteria, Over-dispersion, Economic growth

1.0 Introduction

Recently in Nigeria, road accidents statistics have become the subject of increased interest among several stakeholders. The objective chiefly herein is to identify succinctly factors contributing to road accidents fatality in Nigeria. Road safety means the totality of measures used in preventing harm to all road users.

Contextually, researchers have shown vast interest on factors significantly affecting road safety; specifically on the monthly changes in accident levels, economic and social impacts of yearly changes in traffic safety and effect of policies (Fridstrøm and Ingebrigtsen (1991); Oppe (1991); Chang and Graham (1993); and Chike and Godwin (2016)).

Lately, the demand for an improved modern transportation system has increased in this age of globalization. Mobility drives economic growth and hence expansion within the transportation sector is a necessity. “The increased importance of transportation does not only demand new standards for efficiency, but also for safety precautions”’ (Trafikverket, 2010, p.1). The notion that injuries are induced by road accidents around the world was evidenced by the resolution of the United Nations’ Assembly in 2010. The assembly declared the years 2011-2020, the ‘Action for Road Safety Decade’. The goal of the resolution is to save approximately 5 million lives during this decade. The resolution followed the first worldwide publication on road safety, which showed,
among other major findings that highly significant deaths were recorded in Africa and in the Mediterranean Region.

Accidents from road traffic are global occurrences with more incidences in developing countries. Annually, road traffic crashes led to approximately 1.24 million deaths globally including adolescents. Less developed nations account for approximately 91% of the world's road crashes though less vehicles are domiciled in these nations globally. Pedestrians, cyclists and motorcyclists are among the approximately fifty percent killed from accidents globally. Without action, fatalities from road traffic are estimated to increase from about 1.3 million to about 1.9 million fatalities by 2020 (WHO, 2013).

Prof Isaac Adewole, the then Nigeria’s Minister of Health, noted that after terror attacks, especially from the Boko Haram’s uprising, violent deaths in Nigeria significantly came from road accidents. In line with the former minister’s words, this paper aims to attempt the identification of an appropriate statistical model that could predict road fatality in Nigeria and also to identify the factors and possible interactions of these factors that could induce road traffic fatality in Nigeria.

2.0 Literature Review

Road accidents in Nigeria are induced by over-loading, unnecessary speeding, wrongful overtaking, lack of proper judgment from drivers, lack of adequate experience, carelessness, machine failure, recklessness, intoxication, tedium, lack of willingness to alight from motion objects (human beings, motor cycles, uncontrolled animals and vehicles), dazzling and defective light, skid and road surface defect, obstruction and improper use of level crossing. Other causes include seat belt misuse, use of mobile phones when driving, and corruption from inadequate application of traffic laws (Federal Road Safety Corps (FRSC, 2010)).

Hendricks et al (1999) conducted a study on risky driving acts in grave traffic crashes to determine the specific driver behavior and risky driving acts that caused the crashes. They reported that 99% of the crashes are induced by improper driver behavior and also identified six causal factors, namely decision errors, driver inattention, vehicle speed, perceptual errors, incapacitation and alcohol impairment to be lead unsafe driving acts in decreasing order of importance.

Gbadamosi (1997) further stressed that significant road traffic accidents and casualties were due to drivers’ errors which include unruliness, over speeding, inappropriate overtaking, lack of attention, inexperience, carelessness and intoxication. Furthermore, according to him nearly six percent were from mechanical vehicular factors while another six percent was due to road construction problems. As for mechanical causes, the incidence is traced to owners’ or drivers’ refusal to practice expected maintenance checks on their vehicles until they degenerate into disastrous conditions. Poor vision of drivers, according to Ocansey (2011) could induce significantly road accidents. The investigation of these factors and causes has not been seen adequately within the Nigerian context statistically to identify the possible factors affecting road fatality in Nigeria. Interestingly, some Nigerians still believe that road accidents are caused by witchcraft and evil forces (Okyere, 2006).

Onakomaiya (1991) sought to find reasons why road traffic accidents had continued to be on the increase and concluded that “safety has been the most neglected aspect of road transportation in
Nigeria; a situation that can be blamed on the failure of the past governments and decision makers to appreciate the magnitude of the problem and attendant costs of road traffic accidents to the national economy” (Gbadamosi, 1997).

Miaou et al (1993) attempted the modeling of road crash data and highlighted the confines of the Poisson Regression (PR) approach. These confines were noted again by Zhong et al (2011). They opined that the Poisson distribution’s “fundamental assumption is that the variance should be equal to its mean and further commented that the “phenomenon of over-dispersion causes what is called over-dispersion and this makes the variances of the estimated parameters to be estimated defectively” (p.4). Zhong et al (2011) noted additionally again that the parameters estimated from the Maximum Likelihood Estimation (MLE) under the PR regression model deviate little from the true parameters; but that the “significance levels of the estimated parameters may be overstated” (p.4).

Zhong et al (2011) noted that to handle over-dispersion in crash data, Negative Binomial Regression (NBR), an alternative to PR, is better for use in accident modeling. Shankar (1995) used the NBR to overcome the over-dispersion problem. He used both PR and NBR to model the effects of road geometry and environmental factors on the number of crashes. He found that NBR modeled the crash data better than the PR under the assumption of over-dispersion.

The parameter of dispersion dictates which model does better between the PR and NBR. When the parameter of dispersion is approximately equal to one, a Poisson model is appropriate. If it is greater than one, this indicates an over-dispersion situation caused by the variations embedded in the highway variables. “These are accident-related factors pertaining to drivers, vehicles, and location not totally included in the highway variables present in the model” (Zhong et al, 2011, p.4).

Caliendo et al, (2007) used both PR and NBR models. Zhong et al (2011) opined that “PR and NBR may be used to examine the relationship between geometric features and accident frequency on multilane roadways in Italy”. They found that the “PR was inappropriate since there was clear evidence that over-dispersion was present. That is, “NBR generalizes the PR by allowing the variance to be over-dispersed”. They further opined that “in the NBR model, the variance equals the mean plus a quadratic term in the mean whose coefficient is called the over-dispersion parameter” (Zhong et al, 2011, p.4).

**3.0 Data Description and Research Methodology**

The study was based on road fatalities data from the Federal Road Safety Corps’ Headquarter Office in Abuja, Nigeria. The data spanned over twenty seven years from 1990-2016. The variables included in the data include the number of people killed (Y) as the dependent variable. Others were over-speeding (SPV), wrongful overtaking (WOT), tyre burst (TBT), Loss of Control (LOC), mechanical deficient vehicle (MDV), motorcycle (MC), overloading (OLV), sign light violation (SLV), dangerous driving (DGD), poor weather (PWR), route violation (RTV), break failure (BFL), sleeping on the steering (SOS), driving under influences (DAD), bad road (BRD), Light Gross Vehicle (LGV), phone use while driving (PWD), dangerous overtaking (DOV), road obstruction violation (ROV), heavy gross vehicle (HGV) and fatigue (FTQ) as independent variables in line with Chike and Godwin (2016). The principal component analysis (PCA) was
first employed to remove the factors that contribute less significantly to traffic fatality in Nigeria. The PCA retained seven factors based on the criterion of eigenvalues greater than or equal to one and these were: over-speeding($X_1$), sign light violation($X_2$), poor weather($X_3$), fatigue($X_4$), route violation($X_5$), loss of control($X_6$) and break failure($X_7$). The R statistical software, version 3.3.3, was used for all the analytical considerations in this study.

3.1 The Poisson Regression Model

By definition, if $Y$ has a Poisson distribution with parameter $\lambda > 0$ then its probability function is

$$P(Y = k) = \frac{\exp(-\lambda) \lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

where

$$E(y_i / x_i) = \lambda = \text{Var} (y_i / x_i)$$

Khishdari and Tafti (2017, p.1) observed that if ‘$y_i$ is the number of road crashes (fatalities), the mean of the crashes (fatalities) can be obtained using

$$\mu_i = \exp(\beta X_j), \quad j = 1, 2, \ldots, n$$

where $\beta$ is a $(1 \times n)$ vector of regression coefficients and $X_j$ is the $j^{\text{th}}$ studied variable.’

Therefore, a more general linear model for (3.0) is

$$E(Y_i) = \mu_i = n_i e^{x_i T \beta}$$

The equalization of the mean with the variance for any given response variance is a major assumption of the Poisson distribution (Hilbe, 2011). This is a limitation for models derived from it. This assumption fails generally in accident data modeling. This is the basis for the recommendation of models derived from the Negative Binomial distribution by researchers.

3.2 The Negative Binomial Regression Model

Cameron and Trivedi (1998) introduced firstly the NB model derived primarily from the PR model. They noted that an over-dispersion parameter ($\alpha$) is needed to handle the inequality between the average value and the variance of the data being considered. The equation connecting the variance to the mean, when $\alpha$ is present is

$$V(y_i) = E(y_i)[1 + \alpha E(y_i)]$$

where $V(y_i)$ and $E(y_i)$ are the variance and expected value of the $i^{\text{th}}$ fatality for the data under consideration.

The probability function representing the NB distribution according to Khishdari and Tafti (2017) is defined as

$$\Pr(Y = y_i | \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1) \Gamma(\alpha^{-1})} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left( \frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i}$$
where the parameters in (6.0) are as defined earlier.

We note that the NB distribution is preferred above the Poisson distribution when there is the presence over-dispersion in the data (Basu and Saha, 2017). Oyedepo and Etu (2016, p.76) noted that both “Poisson and Negative Binomial regression models express the log outcome rate as a linear function of a set of predictors” as shown in (7.0) below

\[
\ln(Y) = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K \\
\]  

(7.0) 

That is,

\[
Y = e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K} \\
\]  

(8.0) 

where Y is the dependent variable (the number of people killed), \( \beta \)'s are the parameters of the model and \( X_1-X_K \) are the selected factors/variables believed to have causal effects on Y. We note that Isa et. al. (2019) considered a regression model not unsimilar to the thought expressed by Oyedepo and Etu (2016).

3.3 The Maximum Likelihood Estimation (MLE)

The estimation procedure used was the maximum likelihood estimation. The goal in MLE is to find the estimates for the parameters of the regression equation to maximize the likelihood function by primarily setting the primary derivative of (9.0) to zero. In most practical cases, finding MLE requires an iterative process. This process usually makes regression models partially complex. Mathematically, let \( X_1, X_2, X_3, \ldots, X_n \) be a random sample from a particular distribution with parameter \( \theta \). Given that we have observed \( X_1=x_1, X_2=x_2, \cdots, X_n=x_n \), a maximum likelihood estimate of \( \theta \), shown by \( \hat{\theta}_{ML} \) is a value of \( \theta \). This value maximizes the likelihood function

\[
L(x_1, x_2, \cdots, x_n; \theta) \\
\]  

(9.0) 

A maximum likelihood estimator (MLE) of the parameter \( \theta \), shown by \( \hat{\theta}_{ML} \) is a random variable

\[
\hat{\theta}_{ML} = \hat{\theta}_{ML}(X_1, X_2, \cdots, X_n) \\
\]  

whose value when \( X_1=x_1, X_2=x_2, \cdots, X_n=x_n \) is given by \( \hat{\theta}_{ML} \)

We sometimes use the transformed likelihood function by taking its natural logarithm because of the simplicity of the process thereafter.

4.0 Results and Discussion

Approximately 2 million road mishaps were recorded in Nigeria from 1990 – 2016 which killed 182,387 people and left 647,014 injured (see Figure 1). This reveals that on the average, about 18,200 people died from road accidents during this period.
Figure 1: Plot of the number of people Killed from 1990-2016 in Nigeria

Source: Researchers’ Analysis

Table 1: The estimates of the Poisson Regression Models

<table>
<thead>
<tr>
<th>Models</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. log(mean of No Killed) = \alpha + \beta_{SPV}</td>
<td>7588.036</td>
<td>7590.627</td>
</tr>
<tr>
<td>2. log(mean of No Killed) = \alpha_1 + \beta_1 SPV + \beta_2 SLV</td>
<td>6981.250</td>
<td>6985.137</td>
</tr>
<tr>
<td>3. log(mean of No Killed) = \alpha + \beta_{SPV} + \beta_{SLV} + \beta_{PWR}</td>
<td>5842.179</td>
<td>5847.362</td>
</tr>
<tr>
<td>4. log(mean of No Killed) = \alpha + \beta_{SPV} + \beta_{SLV} + \beta_{PWR} + \beta_{FTQ}</td>
<td>5225.779</td>
<td>5232.258</td>
</tr>
<tr>
<td>5. log(mean of No Killed) = \alpha + \beta_{SPV} + \beta_{SLV} + \beta_{PWR} + \beta_{FTQ} + \beta_{RTV}</td>
<td>3304.577</td>
<td>3312.352</td>
</tr>
<tr>
<td>6. log(mean of No Killed) = \alpha + \beta_{SPV} + \beta_{SLV} + \beta_{PWR} + \beta_{FTQ} + \beta_{RTV} + \beta_{LOC}</td>
<td>2186.563</td>
<td>2195.634</td>
</tr>
<tr>
<td>7. log(mean of No Killed) = \alpha + \beta_{SPV} + \beta_{SLV} + \beta_{PWR} + \beta_{FTQ} + \beta_{RTV} + \beta_{LOC} + \beta_{BFL}</td>
<td>1107.067</td>
<td>1117.433</td>
</tr>
</tbody>
</table>

Specifically from table 1, model 7 is best suited for the road traffic data in Nigeria from 1990 – 2016 because it had the smallest AIC and BIC estimates. The AIC and BIC estimates of model 7 were 1107.067 and 1117.433. The parameters and their estimates for the selected model 7 including the standard errors are given below in table 2.
Table 2: The Estimates of the Parameters for the PR Model

| Parameter       | Estimate  | Std. Error | Z values | Pr(|Z|) | Exp(B)  |
|-----------------|-----------|------------|----------|--------|---------|
| (Intercept)     | 8.26E+00  | 1.66E-02   | 497.72   | <2e-16 | 3850.826|
| Over-speeding (SPV) | 1.40E-04  | 2.64E-06   | 53.04    | <2e-16 | 1.000   |
| Sign light violation (SLV) | 6.57E-04  | 1.35E-05   | 48.7     | <2e-16 | 1.000   |
| Poor weather (PWR) | -7.56E-04 | 2.80E-05   | -27      | <2e-16 | 0.999   |
| Fatigue (FTQ)   | 4.89E-04  | 2.06E-05   | 23.79    | <2e-16 | 1.000   |
| Route violation (RTV) | 4.64E-05  | 2.80E-05   | 36.03    | <2e-16 | 1.001   |
| Loss of control (LOC) | -9.84E-04 | 1.29E-06   | -35.17   | <2e-16 | 0.999   |
| Break failure (BFL) | -1.04E-03 | 3.15E-05   | -32.97   | <2e-16 | 0.999   |

Table 2 above shows the estimates of the parameters of the PR model for the selected road accident causative factors. The AIC and BIC of this model as stated earlier were 1107.067 and 1117.433. Additional explorative analysis on the data under consideration revealed that the dispersion parameter was 42.316. The Omnibus test also showed the significance of the model at the 5% significance level. However, the expected equality of the mean and variance fails. The parameter of dispersion estimate for the PR model above was 42.316. This represents an over-dispersion situation and that the parameters of the PR model were over estimated. Table 2 also reveals under estimation of the standard errors. To overcome these issues, NBR was used to handle over-dispersion and this led to the results below in table 3.

Table 3: Estimates of the Parameters of the NBR Model

| Parameter       | Estimate  | Std. Error | Z values | Pr(|Z|)  |
|-----------------|-----------|------------|----------|---------|
| (Intercept)     | 8.27E+00  | 9.00E-02   | 91.809   | < 2e-16 |
| Over-speeding (SPV) | 1.41E-04  | 1.51E-05   | 9.347    | < 2e-16 |
| Sign light violation (SLV) | 6.78E-04  | 7.49E-05   | 9.054    | < 2e-16 |
| Poor weather (PWR) | -7.88E-04 | 1.57E-04   | -5.007   | 5.53e-07 |
Table 3 above shows a reduction in the parameter estimates and an increase for the standard errors relative to the estimates given in table 3. This supports the opinion expressed by Basu and Saha (2017) who opined that “over-dispersion is reasonably taken into account by the NBR model” (p.64). Additionally, in table 3, at the 5% significance level, over-speeding, loss of control, route violation, poor weather, fatigue, sign light violation and break failure were all statistically significant since their p-values are all less than 0.05. Over-speeding (SPV) had a positive coefficient; that is an increase in SPV will bring about a rise in road fatality in Nigeria. This is the case with Sign light violation (SLV), Fatigue (FTO), Route violation (RTV) and Loss of control (LOC) while Poor weather (PWR) and Break failure (BFL) had negative coefficients. The implication is that increase in both PWR and BFL may not necessarily bring about increase in road fatality. The assessment criteria analysis to validate the choice of the NBR model vis-a-vis its goodness of fit over Poisson model is shown in table 4 below.

### Table 4: Assessment Criteria for PR and NBR Models

<table>
<thead>
<tr>
<th>Assessment Parameter</th>
<th>Poisson Regression Model</th>
<th>NBR Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Deviance</td>
<td>8855.8</td>
<td>289.660</td>
</tr>
<tr>
<td>Degree of Freedom</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Residual Deviance</td>
<td>804.0</td>
<td>27.013</td>
</tr>
<tr>
<td>Degree of Freedom</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-545.533</td>
<td>-202.691</td>
</tr>
<tr>
<td>Dispersion Parameter</td>
<td>42.316</td>
<td>1.501</td>
</tr>
<tr>
<td>Akaike’s Information</td>
<td>1107.067</td>
<td>423.38</td>
</tr>
<tr>
<td>Bayesian Information (AIC)</td>
<td>435.045</td>
<td></td>
</tr>
<tr>
<td>Bayesian Information (BIC)</td>
<td>1117.433</td>
<td></td>
</tr>
</tbody>
</table>

The results in table 4 shows that the NBR fits well the road traffic data in Nigeria because the parameter of dispersion reduced from 42.316 in the Poisson model to 1.501 in the NB model. The AIC and BIC of the Poisson Regression model further reduced from 1107.067 and 1117.433 to 423.38 and 435.045 in the Negative Binomial model. Lower estimates obtained for the Null Deviance, Residual Deviance, and Log Likelihood under the Negative Binomial model further validated the fact that NBR is a better choice. The fitted NBR model for the road traffic data under consideration is

\[
\log(\text{mean_No.Killed}) = 8.265 + 1.414 \times 10^{-4}(\text{SPV}) + 6.778 \times 10^{-4}(\text{SLV}) - 7.880 \times 10^{-4}(\text{PWR}) + 4.717 \times 10^{-4}(\text{FTQ}) + 4.710 \times 10^{-5}(\text{RTV}) - 9.994 \times 10^{-4}(\text{LOC}) - 1.059 \times 10^{-3}(\text{BFL})
\]

……(10)
Table 5: Ranks of the Estimates of Probability Values for the NBR Model

<table>
<thead>
<tr>
<th>S/n</th>
<th>Coefficients</th>
<th>P-values</th>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Over-speeding (SPV)</td>
<td>&lt; 2e-16</td>
<td>1st</td>
</tr>
<tr>
<td>2</td>
<td>Sign Light Violation (SLV)</td>
<td>&lt; 2e-16</td>
<td>1st</td>
</tr>
<tr>
<td>3</td>
<td>Poor Weather (PWR)</td>
<td>5.53e-07</td>
<td>6th</td>
</tr>
<tr>
<td>4</td>
<td>Fatigue (FTQ)</td>
<td>2.17e-05</td>
<td>7th</td>
</tr>
<tr>
<td>5</td>
<td>Route Violation (RTV)</td>
<td>4.44e-11</td>
<td>3rd</td>
</tr>
<tr>
<td>6</td>
<td>Loss of Control (LOC)</td>
<td>2.32e-10</td>
<td>4rd</td>
</tr>
<tr>
<td>7</td>
<td>Break Failure (BFL)</td>
<td>1.49e-09</td>
<td>5th</td>
</tr>
</tbody>
</table>

Table 5 gives the ranks of the coefficients in NBR model based on their p-values. The ranking indicates that over-speeding and sign light violation were the most likely variables inducing road traffic accidents followed by route violation, loss of control, break failure, poor weather and fatigue in that order.

5.0 Conclusion
This study was aimed at identifying factors that could aid modeling of road traffic fatality in Nigeria. Principal component analysis was first employed to reduce the original 21 independent variables to 7 variables. The seven variables retained were over-speeding, sign light violation, poor weather, fatigue, route violation, loss of control and break failure. These seven variables were used as inputs for the two models considered. The NB model was found to be better. The estimates of the assessment criteria (AIC and BIC) for the NBR model were 423.38 and 435.45; while for the Poisson regression model, the estimates were 1107.067 and 1117.433. The reduction in the estimates of the assessment criteria for the NBR model is an indication of its better performance. This finding is supported by Oyedepo and Etu (2016) and Uwakwasi et al (2018). The negative binomial regression model considered showed that all the seven input variables were all statistically significant. Over-speeding and sign light violation were the most likely variables inducing road traffic accidents followed by route violation, loss of control, break failure, poor weather and fatigue in that order.

6.0 Recommendations
The recommendations of this study are:
1. NBR model is a better model when compared with Poisson regression model in modelling road traffic fatality in Nigeria. Hence, making it a better prediction tool for road traffic rate.

2. According to this research, over-speeding, sign light violation, poor weather, fatigue, route violation, loss of control and break failure are the major contributors to traffic fatality in Nigeria. Therefore, road traffic agencies should intimate road users on the negative effects of the above mentioned factors.

3. Traffic officials including institutes that enforce traffic regulations should be stationed at major roads to checkmate and punish incessant over-speeding and route violation offenders in Nigeria.

4. The accident database of this country should be standardized and also made readily accessible to every interested researcher. Also, qualified and competent professionals should be employed to enhance road accident data handling.

5. Every road user should be sensitized about the effects of the factors identified and others like state of the vehicle, lack of proper sleep and a disturbed mind.

References
Ocansey, (2011). Effect of rapid motorization levels on roads facilities in some rich developing countries. Accident Analysis and Prevention, 17:101-109


