A Logistic Regression model of Road Traffic Fatalities in Benue State: Implication to Public Health

*Apeagee, B.B.[1] and Haaor, S.A[2]
[1] Benue State School of Nursing, Makurdi
[2] Benue State Health Insurance Agency
Corresponding Author: Apeagee, B.B

Abstract
Road accident fatalities remain one of the major causes of deaths in Nigeria. These accidents sometimes appear to occur at some flashpoints where there are sharp bends or curves, Potholes due to bad roads, and bad sections of the highways. The study aimed to model road traffic fatalities in Benue State with implications for public health. The data used for the study was sourced from the department of the rescue unit, Federal Road Safety Corps, Benue State Command from July 2011-December 2019. To better comprehend the impact of sex on successfully having a non-fatal accident, we employed a logistic regression analysis. The results indicated that for the 3,545 RTAs used in the model, the model correctly predicted whether or not someone will have a fatal accident 81.7% of the time. The results further showed that females have .970 times the odds of not having a fatal accident compared to males with statistically significant value (P =.015, 95% C.I .652, .956). The results also stated that males demonstrated a greater likelihood to be involved in a fatal accident than the female with statistically significant difference (P =.015, 95% C.I .652, .956). The study, therefore concluded that sex or gender is an influential predictor and as such, has a significant effect on the status of an accident. The implication of the results to public health is that people are likely to spend more money on hospital bills with an increased number of disabilities caused as a result of these accidents which may result to a reduction in physical activity in the next decade if the rate of accidents is not controlled.

Keywords: Logistic, Regression, Model, Road, Traffic, Fatalities

Introduction

Road transportation is one of the types of transportation system used by Nigerians. It provides benefits both to the nation and to individuals by facilitating the easy movement of goods and people within the country. Nigeria has the largest road network in Sub‑Saharan Africa with a total road network of 194,394 km (Onyemaechi and Ofoma, 2016).

Road accident fatalities remain one of the major causes of deaths in Nigeria. These accidents sometimes appear to occur at some flashpoints where there are sharp bends or curves, Potholes due to bad roads, and bad sections of the highways. However, reckless and over-speeding drivers usually find it difficult to control their vehicles which possibly leads to vehicle somersaulting, and in turn results in fatal traffic accidents, recorded over the past years (Ncube, Cheteni, and Sindiyandiya, 2016). World Health Organization (WHO, 2004) stated that with a high amount of motorized road transport, poor economic settings, unemployment of youths in the labour market, most of them ends up as road drivers. Many of which are not qualified and/or not roadworthy, and as such contribute substantially to the majority of the road traffic crashes.

According to the world health organization (WHO, 2009), Over 1.2 million people die each year on the world's roads, and between 20 and 50 million suffer non-fatal injuries. Over 90% of the world's fatalities on the roads occur in low-income and middle-income countries (LMIC), which have only 48% of the world's registered vehicles (Ncube, *et al.*, 2016). They predicted road accidents as the fifth leading cause of death by 2030. Similarly, the mortality rate due to Road Traffic Accidents (RTAs) in LMICs is about 20.0 per 100,000 population, nearly twice as high as the rate in High-Income Countries (HICs) (Schlottmann, Tyson, Caims, varela, and Charles, 2017). Specifically in Africa, the number of road traffic injuries and deaths has been increasing over the last three decades, and the African region continues to have the highest road traffic death rates (26.6 per 100,000 population)

(Schlottmann *et al.*, 2017).The rising incidence of RTA in Nigeria and other developing countries with adverse physical and socioeconomic implications cannot be over-emphasized. This further explained that injuries and deaths resulting from RTA are on the rise and are Nigeria's third‑leading cause of overall deaths, the leading cause of trauma‑related deaths and the most common cause of disability (Onyemaechi and Ofoma, 2016; Ncube *et al.*, 2016). The study correlates with WHO (2013) report which proscribed Nigeria as the most dangerous country in Africa with 33.7 deaths per 100,000 populations every year. The report further noted that one in every four deaths in Africa occurs in Nigeria with the remaining 64% in the Democratic Republic of Congo, Ethiopia, Kenya, South Africa, Tanzania, and Uganda (WHO, 2013). Road Traffic Accidents (RTA) are an important public health problem with considerable morbidity, mortality, and disability especially in low-income countries (Olusoji, Olusola, Kolawole, Gbenayon, and Albert, 2017). The study was conducted to explore the epidemiology of road traffic accidents in Nigeria using spatial analytical tools. They utilize secondary data on road traffic accidents and mortality between 2007 and 2015 from the Federal Road Safety Commission in Nigeria. The results showed that a total of 83,548 road traffic accidents and 76,822 deaths were reported in Nigeria from 2007-2015. The total road crashes in 2007 were 5.7/100,000 population and this increased gradually to a peak of 8.7/100,000 population in 2009 and then declined to 2.9/100,000 in 2011 and to another peak of 7.8/100,000 in 2011. The study concluded that there is a decline of RTA over the years under study but however noted significant clustering of RTA occurrence and death in the Federal Capital Territory and Nasarawa State. The study recommended that further research is required to explore the determinants for the high rates of RTA in the identified clusters (Olusoji *et al.*, 2017). Further studies also reviewed that Nigeria's situation has reached such an alarming proportion even to the point of sheer frustration and near helplessness. Nigeria continues to feature in the bottom half of the World Health Organisation country rankings of road traffic accidents (Sumaila, 2013). This is following the results obtained from the study carried by (Sumaila, 2013). The results showed the general high rate incidence of RTAs in Nigeria with the driver as the main culprit, the functional limitations of Federal Road Safety Commission (FRSC) as the lead agency for road safety matters, the practical difficulties of implementing the driver license and vehicle registration schemes, and poor driving culture of Nigerians arising from weak traffic education, public awareness, and enforcement programs. The amount of accident data stored in Nigeria's FRSC database grows every twelve months at a rate of 100%, which shows that a lot of data are obtained and still more data are still being collected (Ogwueleka and Ogwueleka, 2012).

Materials and Methods

Study population/sample size

A total of 3, 545 accidents recorded were collected from the rescue registry of the Federal Road Safety Commission (FRSC), Benue State Command Headquarters, Makurdi.

Data collection

The data for the study was sourced from the Federal Road Safety Commission (FRSC). Data on accidents and sex of victims in Benue State from July 2011 to December 2019 were collected.

Methods of Data Analysis

A binary logistic regression was used to model the collected data

The logistic regression model

Logistic regression is a category of the generalized linear model (GLM) that follows a Bernoulli (binary) distribution. The dependent variable is usually dichotomous, that is, the dependent variable can take value 1 with a probability of success or the value 0, with a probability of failure. The basic mathematical concept that underlay logistic regression is the natural logarithm of an odds ratio (Logit) (Park, 2013). The Logit model analyses the relationship between multiple independent variables and a categorical dependent variable and then estimates the probability of occurrence of an event by fitting data to a logistic curve (Park, 2013). The mean of the response variable Y in terms of an explanatory variable X is modeled relating Y and X through the equation $Y = a + \beta X$ (Chao-Ying, *et al.*, 2002). Relationships between $\pi(x)$ and x are usually non-linear rather than linear (Agresti, 2007). For a binary response variable Y, and an explanatory variable X, let

$\pi(x) = P(Y = 1/X = x)$

The logistic regression model is

$$\pi(x) = \frac{exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} \qquad \{1\}$$

The probability of death under the logistic model is

$$\pi(x) = \frac{exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)} \qquad \{2\}$$

Hence, $1 - \pi(x) = probability\ of\ survival$

$$= \frac{1+exp(\alpha+\beta x)-exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)} \qquad \{3\}$$

$$= \frac{1}{1+\exp(\alpha+\beta x)} \qquad \{4\}$$

The Logit function

For any number π between 0 and 1, the Logit function is defined by

$$Logit\ [\pi(x)] = log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \qquad \{5\}$$

This implies that $\pi(x)$ increases or decreases as an s-shaped function of x (Agresti, 2007)

Odds and Odds Ratio (OR)

Odds of an event are the ratio of the probability that an outcome will occur to the probability that it will not occur. If the probability of an event occurring is P the probability of the same event not occurring is$(1 - P)$. Hence, the corresponding odd is a value which according to Chao-Ying, *et al.* (2002), is given by

$$odds\ of\ [event] = \frac{p}{1-p} \qquad \{6\}$$

Evaluation of the Logistic regression model
The likelihood ratio test

The overall fit of the model shows the strong relationship between all the independent variables, taken together, and the dependent variable. It can be assessed by comparing the fit of the two models with and without the independent variables. A logistic regression model with the *k* independent variables (L1) is said to provide a better fit to the data if it demonstrates an improvement over the model with no independent variables (L0)(Park, 2013). The likelihood-ratio test statistic equals:

$$= -2\log\left(\frac{L0}{L1}\right) = -2[\log(L0) - \log(L1)] = -2(L0 - L1) \qquad \{7\}$$

This log transformation of the likelihood functions yields a chi-squared statistic. If the *p*-value for the overall model fit statistic is less than the conventional 0.05, then H_0 is rejected with the conclusion that there is evidence that at least one of the independent variables contributes to the prediction of the outcome (Park, 2013).

Statistical significance of individual regression coefficients
Wald statistic

The Wald statistic is the ratio of the square of the regression coefficient to the square of the standard error of the coefficient. The Wald statistic is asymptotically distributed as a Chi-square distribution.

$$W_{stat} = \frac{Bj^2}{SEBj^2} \qquad \{8\}$$

Each Wald statistic is compared with a Chi-square with 1 degree of freedom

Odds ratios with 95% Confidence Interval (CI)

The 95% CI was used to estimate the precision of the OR. A large CI indicates a low level of precision of the OR, whereas a small CI indicates a higher precision of the OR. An approximate confidence interval for the population log odds ratio is 95% CI for the In (OR) =In (OR) ±1.96× {SE In (OR)}, where SE is the standard error.

Use of Statistical Software

A statistical Package for Social Sciences (SPSS, version 23) was used for computing the descriptive statistics of socio-demographic data collected for the study. Also, the data collected for the study were fit to the model using a logistic regression model

Results and Discussion

A total of 3,545 road traffic accidents were recorded during the period under review. However, 647 road fatalities were accounting for 18.3% and 2898 non-fatalities (81.7%) as shown in the frequency table (Table 1a). Table 1b showed the sex distribution of road traffic accidents recorded and 69.3% of the accidents recorded were male and 30.7% were female.

Table 1a: Frequency Table for road accidents

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fatality	647	18.3	18.3	18.3
	Non-Fatality	2898	81.7	81.7	100.0
	Total	**3545**	**100.0**	**100.0**	

Table 1b: Sex distribution of traffic accidents

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	2456	69.3	69.3	69.3
	Female	1089	30.7	30.7	100.0
	Total	3545	100.0	100.0	

Model of road traffic fatalities

A binary logistic modelling was carried out to access the impact of sex on the likelihood of having a road traffic accident. The response variable is the accident status concerning the predictor variable sex. Both the response and predictor variables are categorical, and stated as follows;

$$Accident\ Status\ = \begin{cases} 0, & if\ died\ or\ fatal \\ 1, & if\ survived\ or\ nonfatal \end{cases}$$

Evaluation of road traffic fatalities using the logistic model
The likelihood ratio test

The model containing the predictor variable is statistically significant with a chi-square value of 6.013 ($p = .014$) for Omnibus Tests of Model Coefficient as shown in table 2a below. The Cox & Snell R^2 of .002 and the Nageikerke R^2 of .003 indicated that between 0.2% and 0.3% of the variance in the response variable is explained by the predictor variable as shown in table 2b.

Table 2a: Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	6.013	1	.014
	Block	6.013	1	.014
	Model	6.013	1	.014

Table 2b: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	3363.007[a]	.002	.003

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Odds ratios with 95% Confidence Interval (CI)

Table 3: Odds ratios with 95% Confidence Interval (CI)

Variables in the Equation

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	Sex(1)	-.236	.097	5.875	1	.015	.790	.652	.956
	Constant	1.667	.083	404.242	1	.000	5.295		

a. Variable(s) entered on step 1: Sex.

Table 4: Model correct outcome classification

Classification Table[a]

			Predicted		
			Accident Status		Percentage
	Observed		Fatality	Non-Fatality	Correct
Step 1	Accident Status	Fatality	0	647	.0
		Non-Fatality	0	2898	100.0
	Overall Percentage				81.7

a. The cut value is .500

Notice that the confidence interval [95% C.I for $Exp(\beta)$] is included in Table 3 above. The first row (Sex) indicated a lower C.I of .652 and an upper C.I of .956, relating to the $Exp(\beta)$ of .790. This implies that for the odds ratio relating to Sex, 95% of the values are expected to be between .652 and .956 respectively. This shows that sex is a statistically significant predictor of the model (P=*.015*). Hence, females have 0.790 times the odds of not having a fatal accident compared to males (95% C.I: .652, .956) with statistically significant value (P=*.015*).The negative coefficient (B = -.236) indicated that males were demonstrating a greater likelihood to be involved in a fatal accident than the female and the difference is statistically significant (P=*.015)*.However, when $Exp(\beta)$ less than 1 for a categorical predictor (Sex) was observed from the results, the reciprocal of $Exp(\beta)$ was calculated as; $\frac{1}{Exp(\beta)}$which is $\frac{1}{0.790} = 1.266$. Thus, it was stated that males have 1.266 times the odds of not having a fatal accident compared to females.

Similarly, the model predictive accuracy was also shown in Table 4. The results showed the prediction-accuracy produced by the logistic regression. Thus, the percentage of correct predictions was given as 81.7%. This results indicated that for the 3,545 RTAs used in the model, the model correctly predicted whether or not someone will have a fatal accident 81.7% of the time.

Study Implication to Public health

The results of the study implied that people are likely to spend more money on hospital bills with an increased number of disabilities caused as a result of these accidents which may result to a reduction in physical activity in the next decade if the rate of accidents is not controlled.

Conclusions

The study, therefore, concluded that sex as a predictor variable had a significant effect on the status of an accident. They were 0.2% and 0.3% variation in the response variable explained by the predictor variable. However, females have .970 times the odds of not having a fatal accident compared to males with statistically significant value (P=*.015*, 95% C.I: 652, .956). Also, male demonstrated a greater likelihood to be involved in a fatal accident than the female with statistically significant difference (P =.015, 95% C.I: 652, .956). The model correctly predicted whether or not someone will have a fatal accident 81.7% of the time.

References

World Health Organization (2009). Global Status Report on Road Safety: Time for Action

Ncube, P.Z., Cheteni, P., & Sindiyandiya, K.P. (2016). Road accidents fatalities trends and safety management in South-Africa. *Problems and perspectives in Management, 14*(3): 627-633

Schlottmann, F., Tyson, A.F., Caims, B.A., & Verela, C. (2014). Road traffic collisions in Malawi: Trends and Patterns of Mortality on Scene. *Malawi Medical Journal, 29*(4):301-305

Onyemaechi, N.O.C., & Ofoma, U.R. (2016). The public health threat of road traffic accidents in Nigeria: A call for Action. *Annale of Medical and Health Science Research, 6*: 199-204

WHO (2013). World Health Statistics

WHO (2004). World Report on Road Traffic Injury Prevention

Olusoji, J.D., Olusola, A.A., Kolawole, S.O., Gbenayon, T.M., & Albert, A. S. (2017). Spatial Epidemiology of road traffic crashes and mortality in Nigeria. *British Journals of Applied Science & Technology, 20*(5): 1-11

Sumaila, A.F. (2013). Road crashes trends and safety management in Nigeria. *Journal of Geography and Regional Planning, 6*(3): 53-62

Ogwueleka, F.N., & Ogwueleka, T.C. (2012). Traffic accident data profiling and clustering with the data mining process. *IOSR Journal of Computer Engineering, 6*(2): 14-22

Park, H. (2013). An introduction to logistic regression from basic concepts to interpretation with particular attention to the nursing domain. *J Korean AcadNurs, 43*(2): 154-164

Chao-Ying, J.P., Kuk, L.L., & Gary, M.I. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research, Indiana University-Bloomington, 96*(1): 4-14

Agresti, A. (2007). *An introduction to categorical data analysis*, 2nd Ed. John Wiley and Sons; Hoboken, New Jersey.