

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/325694859>

Modelling Vehicular Crash Mortalities in Ghana

Article in *Model Assisted Statistics and Applications* · June 2017

DOI: 10.3233/MAS-180433

CITATIONS

0

READS

119

4 authors, including:



Ezekiel Nortey

University of Ghana, College of Basic and Applied Sciences

51 PUBLICATIONS 112 CITATIONS

[SEE PROFILE](#)



Samuel Iddi

University of Ghana

22 PUBLICATIONS 129 CITATIONS

[SEE PROFILE](#)



Abeku Atta Asare-Kumi

University of Ghana

7 PUBLICATIONS 4 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Report Writing of the 2010 Population and Housing Census of Ghana [View project](#)



A Student Loan Trust Fund Means Testing Project of which I was the lead consultant. [View project](#)

Modelling Vehicular Crash Mortalities in Ghana

Ernest Somua-Wiafe, Abeku Asare-Kumi, Ezekiel N. N. Nortey, and Samuel Iddi

University of Ghana, Department of Statistics, P. O. Box LG115, Legon, Ghana

*Corresponding author. Email: siddi@ug.edu.gh

Abstract

Deaths due to road accidents are major concern to many stakeholders in Ghana especially because road accidents only come second behind malaria for cause of deaths. Statistical models can be helpful in evaluating the effect of factors responsible for mortality and morbidity during vehicular accidents. There is often a spoilt for choice on the type of models that may be used to explain a particular phenomenon. Picking a model can be based on the researcher's knowledge or experience and the simplicity of the model. However, in common applications, the models applied are often not adequate to accurately and efficiently explain underlying phenomenon particularly when it fails to address certain characteristics of the data. In this paper, an appropriate statistical model on the number of vehicular deaths in Ghana is fitted. The Poisson, Negative Binomial (NB), Zero-Inflation Poisson (ZIP) and Zero-Inflation Negative Binomial (ZINB) models, estimated by the method of maximum likelihood, are compared to determine the most appropriate model for the data at hand. In addition, due to the large number of explanatory variables, the backward model selection procedure was adopted to select the most significant factors associated with crash fatalities. After a careful model building process, the ZINB model was identified as the most appropriate for modelling road crash mortality. The model also identified factors such as shoulder type, time of crash, driver's sex, road environment landmarks, among others as having significant effect on the fatalities during vehicular accidents in Ghana. It is recommended that authorities focus on installing reflective markings on the shoulders of roads and increase education of drivers in adhering to road regulations while also paying keen attention to road environmental landmarks.

Keywords: Accident; Maximum likelihood estimation; Negative binomial; Overdispersion; Poisson model; Zero-Inflation.

1 Introduction

Safety on the road is of paramount interest to various cities, regional and country administrators. The major objective of the departments responsible for safety on roads is often to reduce accidents to the barest minimum while ensuring that the number of people killed and injured victims of accidents on the road is as low as possible. To keep accidents on the road to a minimum, city administrators have often proposed and amended traffic regulations and invested in roads and road signs to avert the impacts of crashes.

The World Health Organization (WHO) in 2004 estimated the death and damage effects caused by roadway crashes to exceed 1.2 million with about 50 million persons injured. The WHO stipulates that, roadway traffic crashes which affects the young people of our societies are reported globally as a second cause of death followed by HIV/AIDS and tuberculosis. WHO in 2013, estimated that

roadway traffic crashes costs most countries 1% to 3% of their Gross National Product. In Ghana, roadway crashes causes more death (as measured in human-years) than almost any other causes of death - falling behind malaria (National Roadway and Safety Commission (NRSC), 2012). It is reported that, an average of 1800 human lives are lost annually as result of roadway crashes. In Ghana alone, the number of roadway traffic crashes reported between the years 2001 to 2011 was 125,657 (NRSC, 2012). In these reported crash cases, 21,265 people lost their lives in motor-vehicle crashes. Also during this year interval, 63,867 victims sustained disabling-injury and 96,078 people had non-disabling injuries. In the year 2012, the reported number of roadway traffic crashes exceeded 1100 recording over 1800 number of deaths (NRSC, 2012).

The Directorate of Planning and Programmes of the NRSC has indicated that, driver error contributes about 80% of all roadway traffic crashes in Ghana with driver inattentiveness representing 28.1%, over-speeding 24.8% and lost-steering-control 23.4% (Ghana News Agency (GNA), 2013). The summary statistics by Motor Traffic and Transport Unit (MTTU) for the year 2013 reported the number of roadway crashes, the crash severity, number of person injured and the number of victims who lost their lives through those crashes for the various regions for the year. The MTTU has also over the past years identified certain critical crash black spots across the country. These sites includes Wulungu highway, Bolga-Sunbrungu-Navrongo highway, Bolga-Zuarungu- Bawku highway, Bolga-Tamale highway, Bolga - Navrongo Highway, Mpaha -highway, Chorgu-Highway, Kushegu highway, Kpong-Tamale Highway, Agric-Gumani-Sankpala Highway, Josonawili town, Zuo Highway, Dormaa Berekum highway, Techiman - Nkuranza road, Agbozume - Aflao highway, Tema motorway, Kanda Highway, George Walker Bush highway, Secondi - Tarkwa Highway, Apemanim-Elubo-Highway, Abuakwa highway, Abofour Junction, Kumasi - Obuasi Highway, Kumasi - Offinso Highway, and Nsawam-Accra highway (MTTU, 2013). The Building and Roadway Research Institute (BRRI, 2006) of the Council of Scientific and Industrial Research outlines the location of crash, the driver's age and sex, the type of vehicle involved in the crash, the nature of the roadway, the weather conditions and also the day and time of crash as major causes of road traffic crashes.

Zhang (2011) reiterates that various factors such as traffic volumes, alcohol- impaired driving and no restraint use affect significantly crash occurrence with other variables such as the lane width, shoulder width, the light conditions, the weather and environmental conditions detected to have some level of influence on the occurrence of roadway crashes. Salifu (2004) and Oppong (2012), among other research conducted in Ghana, have focused on the frequency of road crashes on major roads in Ghana. The frequency of accidents have been attributed to the lack of traffic lights at major junctions, vehicle type (car or bus), the day and time of accident and age of victims amongst others. The number of fatalities associated with traffic crashes in Ghana is increasing with time (Oppong, 2012) and in a nation where currently roadway crashes claim an average of 1800 people annually; it seems highly unlikely for Ghana to satisfy the Decade of Action programme which aims at drastically reducing the number of road fatalities by 2020 as stipulated by the World Health Organization (WHO,

2015).

The Poisson and Negative Binomial (NB) models are two basic generalized linear model (GLM) specifications, which are usually employed to analyze count data (Agresti, 2002). The Poisson regression is characterized by equal mean and variance, and fits well for equidispersed data whilst the NB is utilized in cases where there is suspicion of overdispersion in the response. Oppong and Assuah (2015) compared the Poisson regression with the NB model with the aim of identifying the best fitting model for road crash fatality counts. The NB was identified to fit the data best. Likewise, Ackaah and Salifu (2011) used the NB model in forecasting injury prone road crashes occurring along rural areas in Ghana. Salifu (2004) did analyse road crashes occurring at T-junctions and X-junctions using the Negative Binomial (NB) model in the presence of covariates falling in the category of traffic flow variables and road features. However, these traditional models fail when most of the observed counts are zeros. The zero-inflated models have been devised to address such situations by modelling zero counts separately. Shankar *et al.* (1997) utilized the zero-altered count models on road crash frequency distinguishing between safe road sections with a likelihood of nearly zero and those that are not safe but recording no road crash. Qin *et al.* (2004) investigated the association which exists between crash frequency and selecting exposure measures using the ZIP model to predict crash counts as a function of annual average daily traffic segment length, speed limit and roadway width.

This study seeks to investigate the most appropriate model to examine the factors that significantly influence the mortality of road crash accidents in Ghana.

2 Data Description and Methodology

This research relies on secondary data obtained from the BRRI on crash reports from the Motor Traffic and Transport Unit (MTTU) of the Ghana Police Service. A total of 6239 crashes were observed on 14 national roads in Ghana for the year 2013. Out of the cases reported, a total of 1521 people were killed and 7402 were injured. In addition, 22 driver-related, traffic, weather and road related variables were recorded. Summary statistics on the number of people killed showed that the mean (0.2438) and variance (0.7855) differ. This gives an indication of possible overdispersion in the counts. Also, the histogram in Figure 1 shows a skew distribution with over 82% of zero observation.

The two phenomena, overdispersion and zero-inflation, exhibited by the data need to be accounted for by our model. Furthermore, an appropriate model building process is required to select the most significant factors of the 22 observed.

This study seeks to model the mortality as a result of road accident crashes in Ghana. Since the outcomes are counts, the conventional univariate Poisson model is typically what comes to mind. However, due to the presence of possible overdispersion and excessive amount of zero counts, the Negative Binomial and Zero-Inflated models are also considered. Since both can occur simultane-

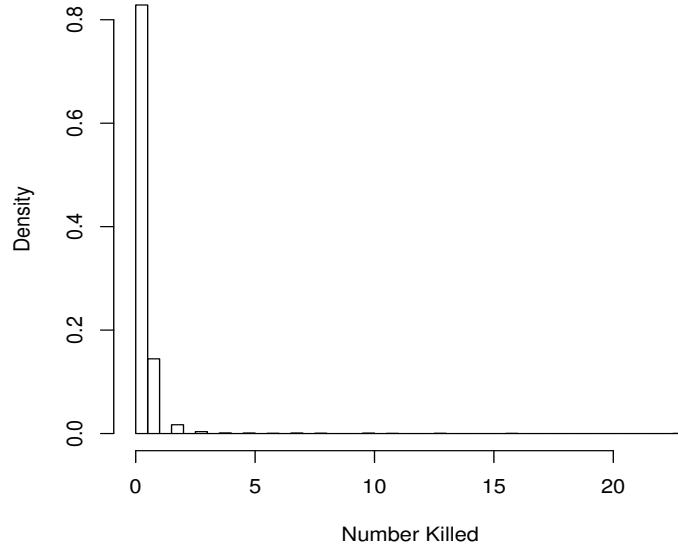


Figure 1: Histogram of Number of Death Cause by Road Crash.

ously, extensions such as Zero-inflated Poisson (ZIP) and Zero-inflated Negative binomial (ZINB) are obvious candidate models as well.

To address the issue of overdispersion, approaches such as the quasi-likelihood Poisson GLM and Negative Binomial (NB) models have been proposed. The NB model is an extension of the Poisson regression and is commonly used to address overdispersion and is reviewed here.

Let $\mathbf{X} = (X_1, X_2, \dots, X_p)$ be the vector of p regressors and let $Y = (Y_1, Y_2, \dots, Y_n)$ be the response vector with n as the number of observations. Let $Y_i \sim \text{Pois}(\theta_i)$, with random mean, $\theta_i \sim \Gamma(\frac{\mu_i}{\lambda}, \lambda)$. Then the marginal distribution of $Y_i (i = 1, 2, \dots, n)$ is the negative binomial with probability function given by

$$p(y_i, k, \mu) = \frac{\Gamma(y_i + k)}{\Gamma(k)\Gamma(y_i + 1)} \left(\frac{1}{\mu_i k^{-1} + 1} \right)^k \left(\frac{\mu_i k^{-1}}{\mu_i k^{-1} + 1} \right)^{y_i} \quad (y_i = 0, 1, 2, \dots) \quad (1)$$

The first two marginal moments of Y are

$$\begin{aligned} E(Y_i) &= \mu_i \\ \text{Var}(Y_i) = \sigma_i^2 &= \mu_i(1 + \mu_i k^{-1}) \end{aligned} \quad (2)$$

where k^{-1} is known as the dispersion parameter and will reflect the extent of overdispersion present in empirical data. The NB distribution converges to the Poisson as $1/k$ approaches zero.

Similar to the conventional Poisson regression, the NB regression model, with k^{-1} held fixed, relates

the expectation of Y to sets of predictors through a log-link. That is,

$$\log(\mu_i) = \beta' \mathbf{X}. \quad (3)$$

To handle the occurrence of excess zero counts, we use the Zero-Inflated Negative Binomial (ZINB) regression model, a natural extension of the Zero-Inflated Poisson (ZIP) regression model. These models assume that two processes generate the counts. The first process generates only zeroes with probability π_i and is modelled by a logistic regression. The second process generates both zeroes and positive counts with probability $1 - \pi_i$. Thus,

$$P(Y = y_i) = \begin{cases} \pi_i + (1 - \pi_i) \left(\frac{1}{\mu_i k^{-1} + 1} \right)^k & \text{if } y_i = 0 \\ (1 - \pi_i) \frac{\Gamma(y_i + k)}{\Gamma(k) \Gamma(y_i + 1)} \left(\frac{1}{\mu_i k^{-1} + 1} \right)^k \left(\frac{\mu_i k^{-1}}{\mu_i k^{-1} + 1} \right)^{y_i} & \text{if } y_i = 1, 2, \dots \end{cases} \quad (4)$$

The marginal mean and variance of Y are respectively

$$\begin{aligned} E(Y_i) &= (1 - \pi_i) \mu_i \\ \text{Var}(Y_i) &= (1 - \pi_i) \mu_i \left(1 + \mu(\pi_i + k^{-1}) \right) \end{aligned} \quad (5)$$

It can be observed that $\text{Var}(Y) > E(Y)$ which implies that this model can also cater for overdispersion.

The ZINB model is fitted by assuming the following two models:

$$\begin{aligned} \text{logit}(\pi_i) &= \alpha' \mathbf{Z} \\ \log(\mu_i) &= \beta' \mathbf{X} \end{aligned} \quad (6)$$

In assessing the effect of the covariates, the zero-inflated model displays an aversion for a particular covariate when the covariate shows an increasing co-efficient in the zero process and a decreasing effect in the parent count process. An attraction is also said to occur when the zero part for the covariate shows a negative coefficient implying a decreasing effect and a positive sign indicating an increasing effect in the parent count process.

To statistically validate zero-inflated probability models, Vuong (1989) proposed a test based on the t -statistic to distinguish between the parent count model and the zero-inflated count model. The test statistics for the Vuong-test, is defined as

$$V^* = \frac{\bar{m} \sqrt{n}}{S_m} \quad (7)$$

with $m = \ln \left(\frac{f_1(\cdot)}{f_2(\cdot)} \right)$ where $f_1(\cdot)$ is the density function of the zero-inflated specification, $f_2(\cdot)$ is also the density of the parent distribution, \bar{m} is the mean of m , S_m is the standard deviation and n is the sample size. The idea is that, if $f_1(\cdot)$ and $f_2(\cdot)$ are not statistically different, we expect

that the mean ratio of the densities should be approximately unity. The test thus compares the null hypothesis that the parent model and the zero-inflated model are equally close to the true model against the alternative hypothesis that the zero-inflated model is significantly closer to the true model as compared with the parent model. Hence the decision goes in favor of the parent model if the p -value of the test is greater than the 0.05 α -level of significance.

The Akaike Information Criterion (AIC, Akaike 1994) and the Bayesian Information criterion (BIC) are used for selecting the best model from several fitted models. The AIC gives a fair idea of the information lost when a model is fitted to a set of data generated from the unknown true model by considering the complexity of the model and its goodness of fit. The AIC for a model with p parameters and likelihood, L at convergence is defined as

$$\text{AIC} = -2 \ln(L) + 2p. \quad (8)$$

As with AIC, the BIC model selection criteria penalizes model with additional parameters. We define,

$$\text{BIC} = -2 \ln(L) + p \times \ln(n) \quad (9)$$

with n representing the sample size, and p , the number of free parameters. Out of several models fitted for a given dataset, the preferred model is one with smallest AIC and BIC values.

2.1 Estimation of Model Parameters

The goal of this study is to model the frequency of vehicular crash mortalities in Ghana in the presence of some covariates using the NB and ZINB model specifications. Maximum likelihood estimation technique is commonly used for obtaining estimates of parameters in these models. The basic principle behind this method is to estimate parameters that maximize the likelihood of the observed data. Often, it is much easier to maximize the log-likelihood.

For independent sample observations (y_i, x_i) , the log-likelihood function of the NB model is given by

$$l(\boldsymbol{\beta}, \phi) = \sum_{i=1}^n \left[\ln \left(\sum_{j=0}^{y_i-1} \left(j + \frac{1}{\phi} \right) \right) - \ln(y_i!) - \left(y_i + \frac{1}{\phi} \right) \ln \left(\phi \exp(\boldsymbol{\beta}' \mathbf{X}) + 1 \right) + y_i \ln(\phi) + y_i \boldsymbol{\beta}' \mathbf{X} \right] \quad (10)$$

where $\phi = k^{-1}$. By solving the score equation, first derivatives of the log-likelihood with respect to the parameters, we obtain estimates of the parameters. Standard errors of the estimated parameters are obtained from the inverse of the observed Fisher's information matrix.

For the ZINB, the log-likelihood is of the form:

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi) = \sum_{y_i=0} \ln \left[e^{\boldsymbol{\alpha}' \mathbf{Z}} + \left(1 + \phi e^{\boldsymbol{\beta}' \mathbf{X}} \right)^{-1/\phi} \right] + \sum_{y_i>0} \ln \left[\frac{\Gamma(\frac{1}{\phi} + y_i)}{\Gamma(\frac{1}{\phi})} \right] - \ln(y!)$$

Table 1: A Comparison of Log-likelihood, AIC and BIC values

Criterion	Poisson	NB	ZIP	ZINB
Log-likelihood	-3544.6	-3215.5	-3137.4	-3049.6
Df	54	55	91	79
AIC	7197.1	6541.0	6456.7	6257.2
BIC	7559.4	6910.0	7067.2	6787.2
Pearson	12356.0	8999.9	7858.0	8042.7
Df	6000	6000	5963	5976
Pearson/Df	2.0594	1.5000	1.3178	1.3458

$$-\left(y_i + \frac{1}{\phi}\right) \ln \left[1 + \phi e^{\beta' \mathbf{X}}\right] + y_i \beta' \mathbf{X} - \sum_{i=1}^n \ln \left[1 + e^{\alpha' \mathbf{Z}}\right] \quad (11)$$

Parameter estimates and standard errors are obtained in a similar way as discussed for the NB regression model above. The estimation of these models are carried out in SAS.

3 Analysis and Discussion of Results

The Poisson, Negative Binomial and their extended versions, Zero-Inflated Poisson and Zero-Inflated Negative Binomial models were fitted to the data. In all, a total of 22 explanatory variables namely the crash day, crash time, weather, light conditions, road description, road surface type, shoulder type, road separation, traffic control, location type, collision type, hit and run, road environmental landmark, road width, road works, driver's sex, driver's age, driving under influence, driver error, vehicle type, vehicle maneuver, and vehicle defects were used. The large number of covariates posed additional problems to the data analysis. To carefully select which of these regressors were significant in explaining the number of people killed, we applied both backward and forward model selection. The log-likelihood, AIC and BIC of the final models are displayed in Table 1. Data manipulations were done in R and statistical analyses were conducted using SAS. The codes used to generate our results can be provided by the authors upon request.

The Poisson regression model performed badly among all the models. From the data, it was observed that the mean number of road accident deaths was 0.244 and variance of 0.7855. Since the mean and variance are not equal, over-dispersion is suspected. The Pearson statistic tells how well the model performs in predicting the observed count response when covariates are considered. It also helps to determine whether there is evidence of overdispersion or underdispersion in the model predicted. Ideally specified model would exhibit equidispersion, where an expected value of 1 is obtained when we divide the Pearson statistic by its degree of freedom. Overdispersion is characterized with a

bigger Pearson value over the degrees of freedom, underestimated standard errors and increase in number of significant variables which in turn results in misleading inferences. Indeed, the large ratio (Pearson/DF) away from 1 gives an indication of the presence of overdispersion in the response that needs to be accounted for. Next, a NB model was fitted to the data. Some consistency was observed in the variable selection since the variables, which were significant after variable selection for the NB model, were also significant for the Poisson model. Furthermore, compared to the Poisson case, we realized a better Pearson Chi-Square on degrees of freedom for the NB was considerably lower than that of the Poisson Model. Also, the AIC and BIC values for the NB were certainly smaller than that of the Poisson model.

Over three fourth (82.85%) of the observed death counts were zeroes. Some of the zeroes could truly be cases where the crashes were non-fatal crashes with no death. Other zero counts could be because of errors in reporting the fatalities. Regardless of the fact that the NB was preferable to the Poisson model, we are still faced with the problem of excess zeros in data. Thus, ZIP model was fitted to handle overdispersion and also account for some aspects of overdispersion. The AIC for the ZIP was smaller than the case of the conventional Poisson. In addition, we performed the Vuong test ($V^* = 5,7624$), which was significant and emphasized the superiority of the ZIP model over the conventional Poisson model. Next, compared with the NB model, we observed that the NB had a larger AIC and BIC value with the ZIP performing better. Thus, it appears the ZIP model corrected for the excess zeroes and did adequately capture the dispersion in the data as compared to the Poisson and NB case. Finally, the ZINB model was considered. The AIC for the ZINB and that of the conventional NB model were significantly different. The Vuong test ($V^* = 4.8197$) was significant at 0.05 significance level. This is an indication that the ZINB model performs quite better than the NB. Therefore, amongst all the models considered, we selected the ZINB model as the best fitting model for the given data set.

The parameter estimates and respective standard errors of the ZINB model are presented in Table 2. For the variable Day with Wednesday as the reference variable, there was significant difference in fatalities occurring between Thursday and Wednesday with coefficient estimate of -0.3494 and a p -value of 0.03494. This implying that Thursday has a decreasing effect on the log of expected count of persons killed while holding other variables in the model constant. We therefore deduce that the incident density ratio of Thursday relative to Wednesday is approximately 0.7051 with an effect of 29.49% decrement on the expected number of persons killed. Amongst the days, Friday and Sunday also exhibits a decreasing effect whilst Monday, Tuesdays and Saturdays also displays an increasing effect compared to Wednesday, although these were not statistically significant. We infer that, it is less dangerous to drive on Thursdays as compared with the other days of the week.

For the Time of the Day variable with the reference category as Night, we observe that Morning, Afternoon, and Evening tend to decrease the log of expected count of persons killed. Out of these three variable levels, Morning and Afternoon were found to be significant with estimates -0.3069 and

-0.3421 respectively. The incident density ratio of Morning relative to Night is 0.7357 and that of Afternoon relative to Night is 0.7103. Hence, Morning and Afternoon are associated with percentage decrements of 26.43% and 28.97% respectively. We can conclude from the results that, more deaths are recorded at Night hours followed by Evening hours thus making it unsafe to drive or travel at Evening and Night hours as compared with Morning and Afternoon hours.

Next, for the Collision Type with Slide collision as the reference variable, we observe that Hit Animal, Hit Object On Road, Hit Parked Vehicle, Hit Pedestrian, and Ran Off collision exhibits a decreasing effect on the log count of persons killed whilst Head On, Hit Object Off Road, and Rear End collisions all increases the log of expected count of number of deaths. We also notice that all the levels of the variable were not significant at the 0.05 level of significance except for Hit Pedestrian and Ran Off collisions giving p -values of 0.0011 and 0.0005, respectively. The estimated coefficients for these variables are recorded as -0.5329 and -0.7475, respectively. Thus we infer that, with all other variables held fixed, Hit Pedestrian displays a decrement of 41.317% in death. This could probably be due to small number of people involved in a hit and run situations. Ran Off collision also indicates a 56.65% decrease in accident mortality in comparison to Slide collision.

Turning to the Road Environmental Landmark covariate, with the reference variable as village, we observed that Rural and Urban Landmark significantly decreases the log mean count of persons killed. We observe estimates of -0.5591 and -0.6503 respectively with both p -values less than 0.0001 for the two levels. Furthermore, we obtain the incident density ratio for Rural relative to Village Landmark as 0.5717 with a decrease of 42.83%.

Table 2: Parameter estimates and standard errors (s.e) for the ZINB model (continue to next pages).

Variables	Levels	Negative-Binomial Regression Part		Logistic Regression Part	
		Estimate(s.e)	$Pr(> z)$	Estimate(s.e)	$Pr(> z)$
Intercept	-	-0.5176(0.7339)	0.4807	0.4256(1.5464)	0.7831
Day	Friday	-0.1109(0.1524)	0.4668	-0.3916(0.3144)	0.2129
	Monday	0.2635(0.1481)	0.0752	-0.0471(0.2995)	0.8750
	Saturday	0.1062(0.1436)	0.4597	0.2640(0.2889)	0.3608
	Sunday	-0.0563(0.1521)	0.7112	-0.4053(0.3047)	0.1834
	Thursday	-0.3494(0.1682)	0.0377	-1.0457(0.3803)	0.0060
	Tuesday	0.0703(0.1552)	0.6506	-0.0918(0.3248)	0.7776
	Time of the Day	Afternoon	-0.3421(0.1357)	0.0117	-0.5635(0.4630)
	Evening	-0.1290(0.1336)	0.3342	0.6103(0.2824)	0.0307
	Morning	-0.3069(0.1428)	0.0316	-0.4862(0.4724)	0.3033
Collision type	Hit Animal	-0.1492(0.7538)	0.8431	-	-
	Head on	0.1569(0.1634)	0.3371	-	-
	Hit object off road	0.5111(0.4338)	0.2388	-	-
	Hit object on road	-1.2241(0.8882)	0.1681	-	-
	Hit parked vehicle	-0.2082(0.2848)	0.4648	-	-
	Hit Pedestrian	-0.5329(0.1632)	0.0011	-	-
	Other	-0.2650(0.3385)	0.4337	-	-
	Ran off road	-0.7475(0.2153)	0.0005	-	-
	Rear end	0.0459(0.1773)	0.7957	-	-
	Right angle	-0.3199(0.2476)	0.1963	-	-
Road environmental landmark	Rural	-0.5991 (0.1242)	< .0001	-0.5693(0.3144)	0.0702
	Urban	-0.6503(0.0929)	< .0001	0.5709(0.1943)	0.0033
Shoulder type	No shoulder type	0.2237(0.1780)	0.2089	-	-
	Paved/Tarred	0.5158 (0.1522)	0.0007	-	-
Driver's sex	Female	-0.7419(0.3572)	0.0378	-	-
Vehicle type	Bus	0.8831(0.7037)	0.2095	1.4082(1.3385)	0.2928
	Car	0.2753(0.6976)	0.6932	1.3494(1.3223)	0.3075
	Cycle	-0.2120(0.7242)	0.7697	-13.0049(185.4419)	0.9441
	HGV	0.4164(0.6971)	0.5503	0.4414(1.3216)	0.7384
	Motor cycle	-0.1440(0.6964)	0.8362	-2.2473(1.3961)	0.1075
	Minibus	0.5990(0.7004)	0.3924	0.6544(1.3279)	0.6221
	Other	0.2977(0.7798)	0.7027	0.0467(1.4795)	0.9748
	Pick up	0.4337(0.7215)	0.5478	1.4246(1.3572)	0.2939
	Tractor	-0.2358(0.8674)	0.7858	1.0752(1.5856)	0.4977
Light condition	Day	-	-	1.1492(0.4218)	0.0064
	Night light off	-	-	0.8504(1.3152)	0.5179

	Night light on	-	-	-0.3611(0.4468)	0.4189
Road Separation	Median	-	-	0.5448(0.1876)	0.0037
Collision Type	Hit Animal	-	-	-0.1147(1.0057)	0.9092
	Head on	-	-	-1.8899(0.3084)	< .0001
	Hit object off road	-	-	0.4075(0.6036)	0.4996
	Hit object on road	-	-	-1.0082(1.3777)	0.4643
	Hit parked vehicle	-	-	-0.9509(0.4456)	0.0328
	Hit Pedestrian	-	-	-16.0910(5.3426)	0.8786
	Other	-	-	-3.4056(1.6125)	0.0347
	Ran off road	-	-	-1.5400(0.4270)	0.0003
	Rear end	-	-	0.1159(0.2546)	0.6489
	Right angle	-	-	-0.6429(0.3970)	0.1054
Driver's error	Care	-	-	-0.3834(0.7823)	0.6240
	Too close	-	-	1.5733(1.1826)	0.1834
	Too fast	-	-	-1.3590(0.8109)	0.0938
	None	-	-	-0.2879(0.7780)	0.7113
	Inexperience	-	-	1.4241(1.1422)	0.2125
	Improper overtaking	-	-	-0.4962 (0.8595)	0.5637
	Other	-	-	-0.5453(0.7965)	0.4936
	No signal	-	-	0.5119(1.3490)	0.7043
Fatigue	-	-	-1.3131(1.7179)	0.4447	
Dispersion, <i>k</i>	-	0.5462(0.0727)	-	-	-

Urban relative to Village Landmark also has an incident density ratio of 0.5219 representing 47.81% decrease in the expected count of persons killed. From the analysis, we can confidently state that more deaths are associated with crashes occurring at Village Landmarks as compared with Rural and Urban Landmarks.

For the Shoulder Type variable, with Unpaved Shoulder Type as the reference level, we first notice an increasing effect for the two variable levels Paved and No Shoulder Types. We observed that Paved Shoulder Type with an estimated coefficient of 0.5158 was highly significant with a *p*-value of 0.0007. Hence we deduce that the increasing effect of Paved Shoulder Type on the expected number of persons killed is about 67.50%. Moreover, we infer that the risk of driving on Paved Shoulder Type roads is probably higher than Unpaved Shoulder Type roads followed by No Shoulder Type roads.

The Driver's Sex with variable levels, Male and Female, had significant effect on the log-count of persons killed. With the reference level as Male, we observe a negative coefficient of -0.7419 and a *p*-value of 0.0378 for Female and this seems to significantly decrease the log expected count. The decrement is represented by a percentage of 52.38% so the results suggest that Female drivers

reduces the probability of death counts as compare with Male drivers.

For Vehicle Type, none of the variable levels were found to be significant even though Bus, Car, HGV, Minibus, Pick Up and Other Vehicle Type appear to have an increasing effect whilst Cycle, Motor Cycle, Tractor displays a decreasing effect on death counts.

4 Conclusion

In this research, we have analyzed the effect of risk factors that could explain the number of fatalities in car crashes in Ghana. A number of possible candidate models were fitted and through a careful modeling selection process, the ZINB model was identified as the “best” fitting model to the data.

The analyses evidently endorses the public’s perception concerning some of the crash variables such as the Day, Time, Collision Type, Road Description, Driver’s Age, Driver’s Sex, and Shoulder Type and their effects on road crash fatality count. The findings suggest that it is more dangerous to drive on Saturday and less risky to use the highways on Thursday. The effect of Saturdays on the number of deaths could be because usually several programs such as funerals and wedding are organized on Weekends and this could influence a lot of people to use the roads on weekends hence fatality counts are likely to increase on these days. We also realized that more deaths are recorded at Night and Evening hours. Lack of streetlights may dim the vision of drivers during evening and night hours. Moreover, drivers may lose concentration due to fatigue or sleeping and this may lead to fatal crashes during these hours. It is also noticed that crashes occurring along Village landmarks are more likely to result in death as compared with crashes occurring along Urban and Rural landmarks. This may be attributed to the fact that along these routes, there are but few traffic regulating mechanisms to check drivers for overspeeding hence such collisions are highly likely to result in more victims killed. In addition, many drivers ignore the speed limit and tend to overspeed on paved shoulder type highways increasing the tendency of fatal crashes.

The youth and teenagers are known for their exuberance and often drive recklessly while adults with lots of experience are mostly careful therefore, the higher the driver’s age, the less likely it is observe an increase in number of persons killed. As well, female drivers reduce the risk of fatality as compare with male drivers. This could be attributed to their patience and cautiousness when driving so as to obey road regulations such as avoiding overspeeding.

We recommend that authorities responsible managing road should focus on installing reflective markings on the shoulders of roads and increase education of drivers in adhering to road regulations while also paying keen attention to road environmental landmarks. The Law enforcement agencies per reports have identified crash black spot in Ghana over several years of observation. Knowing the day, time and other factors which has influence on crash fatality implies intensifying their patrol around such times in order to check and regulate vehicle movements at such spots. Reflective road markings

would aid driver's vision at night since more fatalities are experienced during dark hours. Authorities should intensify educating road users especially drivers when driving at night with the intension of reducing road crashes.

To comment on the computational work to get the MLE's for the parameters, it is obvious that the computational cost for the more complex models would be higher. However, with the computational efficiency and speed of R and SAS softwares, the differences in computational time between these models are almost negligible. In terms of estimation of model parameters, other estimation strategies such as Bayesian MCMC and Psuedo-likelihood estimation techniques could be explored in the future. Furthermore, since the focus of this empirical research is making inference and not predictions, assessment of the predictive power of the models is beyond the scope of this manuscript. However, this can be considered in a future study.

Author's Contributions

ES-W, SI and ENNN conceived the idea and planned the study. ES-W and AAK collected and processed the data. ES-W and SI performed the analysis. SI and ENNN suggested the statistical methodology and supervised the entire study. ES-W, AAK and SI wrote the manuscript. All authors agree to be accountable for all aspects of the work and jointly own the work. All authors read and approved the final paper.

Acknowledgements

The authors are grateful to Mr Kwadwo Poku Agyemang, Mr Yaw Preprah and Mr Simon Ntramah of the Building and Road Research Institute (BRRI) who helped to acquire the data.

Competing interests

The authors declare that they have no competing interests.

References

- Ackaah, W., and Salifu, M. (2011). Crash prediction model for two-lane rural highways in the Ashanti region of Ghana. *International Association of Traffic and Safety Sciences Research*, **35**, 34–40.
- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons: New York.

- Akaike, H. (1994). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.
- Building and Road Research Institute (2006). *Estimation of the cost of Road Traffic accidents in Ghana*. Council for Scientific and Industrial Research, Ghana.
- Ghana News Agency (2013). *Road crashes cost the nation GHC948, 224 million annually - NRSC*. [Online]. Available at:
<http://ghananewsagency.org/social/road-crashes-cost-the-nation-gh-948-224-million-annu>
- Motor Traffic and Transport Unit (2013). *Motor Accident Returns Full Year 2013, Nationwide (Annual Report)*. Ghana Police Service, Ghana.
- Motor Traffic and Transport Unit (2013). *Critical Accident Black Spots, Nationwide*. Ghana Police Service, Ghana.
- National Road Safety Commission (2012). *Annual Report*. Ministry of Transportation, Ghana.
- Obeng, D.A., and Salifu, M. (2013). Modeling Risk Factors of Pedestrian Accidents on Trunk Roads in Ghana. *International Refereed Journal of Engineering and Science (IRJES)*, **2(5)**, 55–64.
- Opong, R.A. (2012). *Statistical Analysis of Road Accidents fatality in Ghana using Poisson Regression (Final Report)*. Department of Mathematics, Kwame Nkrumah University of science and technology, Ghana.
- Opong, R. A., and Assuah, C. K., (2015). Comparative Assessment of Poisson and Negative Binomial Regressions As Best Models For Road Count Data. *International Journal of Scientific Research and Engineering Studies*, **11 (2)**, 28–32.
- Opong, R.A., and Asiedu-Addo S.K. (2014). Analysis of Vehicular Type as a Risk Factor of Road Accidents' Fatality in Ghana. *International Journal of Modern Sciences and Engineering Technology (IJMSET)*, **1(5)**, 106–114.
- Salifu, M. (2004). Accident Prediction Models for Unsignalised Urban Junctions in Ghana. *International Association of Traffic and Safety Sciences (IATSS) Research*, **28(1)**, 68–81.
- Shankar, V., Milton, J. and Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis and Prevention*, **29(6)**, 829–837.
- Qin, X., Ivan, J.N., and Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention*, **36(2)**, 183–191.
- Vuong, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypothesis. *Econometrica*, **57**, 307–334.

Wang, K., Lee, A.H., Yau, K.K.W., and Carrivick, P.J.W. (2003). A Bivariate Zero-Inflated Poisson Regression Model to Analyze Occupational Injuries. *Accident Analysis and Prevention*, **35(4)**, 622–629.

World Health Organization (WHO) (2004). *World Report on road traffic injury prevention* [Online]. Available at:
http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/.

World Health Organization (WHO) (2013). *Road traffic injury* [Online]. Available at:
<http://www.who.int/mediacentre/factsheets/fs358/en/>.

World Health Organization (2015). *Global Status Report on Road Safety 2015* [Online]. Available at: http://apps.who.int/iris/bitstream/10665/189242/1/9789241565066_eng.pdf?ua=1

Zhang, Y. (2011). Exploring Single Vehicle Crash Severity on Rural, Two-Lane Highways with Crash-Level and Occupant-Level Multinomial Logit Models. *Masters thesis* Available at:
<https://pdfs.semanticscholar.org/b5d3/dcfadeeb268ef47327aba91e77ba4705db4d.pdf>