Tanzania Journal of Science 47(1): 308-314, 2021
ISSN 0856-1761, e-ISSN 2507-7961

Modelling Road Traffic Accidents Counts in Tanzania: A Poisson Regression Approach

Thadei D. Sagamiko[1*] and Nyimvua Shaban Mbare[2]
[1]*Department of Physics, Mathematics and Informatics, Dar es Salaam University College of Education,University of Dar es Salaam, P. O. Box 2329, Tanzania*
[2]*Department of Mathematics, University of Dar es Salaam, P. O. Box 35062, Tanzania*
E-mail addresses: tsagamiko@gmail.com; shabanmbare@gmail.com
**Corresponding author*
Received 29 Aug 2020, Revised 1 Feb 2021, Accepted 2 Feb 2021, Published Feb 2021
https://dx.doi.org/10.4314/tjs.v47i1.26

Abstract

Road traffic accidents have become serious threats to Tanzanians in recent years. The outcry emanates from the increasing prevalence of negative effects of accidents on human lives, properties, environments and the economy. Poisson regression model was used to study the relationship between road accidents and the factors facilitating them in Tanzania. Count data on yearly road traffic accidents for Tanzania covering the period 1993 to 2019 were used. Due to over-dispersion of Poisson regression model, quasi-Poisson regression model was found the most appropriate approach for the analysis of these data. Results indicated that all predictors are significant under Poisson regression model with p-value less than 0.05 but high speed was found insignificant using quasi-Poisson regression model. All factors causing road accidents predicted minor increase of accidents, showing that current control measures on road accidents are likely to be effective.

Keywords: Road accidents; Poisson regression, Over-dispersion; Deviance, Variance inflation factor.

Introduction

Road traffic accidents are major public health threats, and without proper preventive measures, they are projected to increase significantly worldwide over the next twenty years (Zhang 2019). Accidents due to road traffic are recognized as a major public health problem in developing countries and continue to cause morbidity, mortality and disability in Sub Saharan Africa (Tarimo 2012). Following the current trends of road traffic accidents, it was estimated that by the year 2020, deaths due to road accidents would increase by 83% in low-income and middle-income countries (if no appropriate actions are taken), and to decrease by 27% in high-income countries (Zhang 2019).

The problem of deaths and injuries from road traffic accidents is acknowledged to be a global phenomenon and traffic safety regulations have been major concerns since the begining of automobile age, almost one hundred years ago (Komba 2006). It is approximated that 1.24 million people die worldwide annually on the roads, almost equal to the number of deaths caused by Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS), tuberculosis and malaria combined (WHO 2009). In addition, road traffic crashes are estimated to cause 20 to 50 million non-fatal injuries to people every year. Deaths and disabilities due to road traffic injuries affect all age groups, but the most affected are people in the young and productive years of their lives. It was estimated that road traffic injuries will move up in the ranking of leading causes of death from tenth in 2004 to fifth in 2030 (WHO 2013). Economically disadvantaged families are hardest hit by both direct medical costs and indirect costs such as lost wages that result from these injuries. At the national level, road traffic injuries lead to considerable financial costs, particularly in developing countries (WHO 2009).

Several studies have been done on the road traffic accidents globally. For instance, Tortum et al. (2012) studied and modelled road traffic accidents using a logarithmic-linear regression in Turkey to determine the main factors contributing to the road traffic accidents. They found that, the main factors contributing to road traffic accidents in Turkey were road defects, vehicle defects, passenger errors, driver defects and pedestrian errors, among general factors, the driver defects had highest effects, while road defects had the lowest effects on road traffic accidents. Kumar and Umadevi (2011) studied various causes of road accidents in Chennai City (India) and developed a dynamics model for reducing severity of road accidents by considering human, road, vehicle and environmental factors. It was revealed that human factors were significant and accounted for 95% of all road traffic accidents.

The study by Nyakyi et al. (2014) formulated a multiple regression model that analyzed the factors that lead to motorcycle accidents in Kilimanjaro and Arusha regions. The study analyzed more than six factors which were mechanical defects, legal status of not owning license, legal status of owning license, rough road, tarmac road, driving experience, wrong overtaking, high speed, and personal status factors (alcohol or drug intake). The results from the study showed that the number of motorcycle accidents in both Arusha and Kilimanjaro regions had strong relationship with experience of drivers, tarmac road, and personal status. The study also observed that there was a strong relationship between number of motorcycle accidents, high speed, and wrong overtaking in Arusha region.

Road traffic accidents in Tanzania have become one of the growing concerns to most Tanzanians in recent times, and the Government of Tanzania has been taking several measures to reduce the road carnages such as introduction of speed limits (Komba 2006). Yet, every year the Tanzania National Road Safety Council (TNRSC), Land Transport Regulation Authority (LATRA) and other organizations have been reporting the occurrence of road traffic accident casualties. This study considers a Poisson regression model of factors affecting road accidents for the entire country (Tanzania mainland and Zanzibar) from 1993 to 2019 the case which is not treated in the existing literature for road accidents count data in Tanzania. The Poisson regression model is used as a standard model for count data which is derived from the Poisson distribution by allowing the intensity parameter μ (mean) to depend on regressors (Cameron and Trivedi 1986).

Materials and Methods

Data

The current study used secondary data from National Bureau of Statistics (NBS) and twenty seven (27) observations were collected for each factor (explanatory variables) for the years from 1993 to 2019. Table 1 gives summarized description for data and variables used.

Table 1: Road accidents data summary in Tanzania from 1993 to 2019

Variables	Unit	Descriptions
Y	Number	Total road accidents in Tanzania
X_1	Number	Reckless driving
X_2	Number	Careless pedestrians
X_3	Number	High speed
X_4	Number	Defective motor vehicle
X_5	Number	Motor cyclists
X_6	Number	Other factors

Source: National Bureau of Statistics .

Data analysis

The data were analyzed using Excel and R softwares by examining the behavior of the distribution from histogram as indicated in Figure 1 and testing for multicollinearity using Variance Inflation Factor (VIF) as shown in Table 2. Counts follow a Poisson distribution rather than a normal distribution which is applied in linear regression models (Cameron and Trivedi 1986). The variance inflation factor used for testing presence of multicollinearity is given by the formula

$$\mathrm{VIF} = \frac{1}{1 - R_j^{2}} \qquad (1)$$

where; R_j^2 is the coefficient of determination of a regression of an explanatory variable j on all other predictors (Nwankwo and Nwaigwe 2016). A VIF value greater than 5 indicates presence of correlation between a given explanatory variable and other variables in the model.

Poisson regression model

The technique of generalized linear models (GLMs) was applied to model the data in particular Poisson regression model which is a standard model for count data derived from Poisson distribution with a parameter μ, which is the expected accident frequency (or number of accidents) for the i^{th} road section during a period of time.

$$p(y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \qquad i = 0,1,2,\ldots \quad (2)$$

Where: $p(y_i)$ is the probability of y accidents occurring at i^{th} road section during a period of time. In Poisson regression model, the expected accident frequency is assumed to be a function of predictors such that

$$\mu = \exp(\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \theta_5 X_5 + \ldots \theta_n X_n) \qquad (3)$$

where; $X_1, X_2, X_3 \ldots X_n$ are predictors and $\theta_0, \theta_1, \theta_2, \ldots, \theta_n$ are model coeficients which are estimated by maximum likelihood methods as their results indicated in Table 3 (Lawless 1987). Thus, the number of road traffic accidents Y is modelled as a loglinear function of the predictors as shown in Equation (4).

$$Ln(Y) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \theta_5 X_5 + \theta_6 X_6 + \varepsilon, \qquad (4)$$

Where: ε is the error term.

Goodness of fit test statistics

The model was tested to check whether it fits well the data by using residual deviance which is the change in the deviance between the given model and the saturated model (the model that fits the data perfectly). Residual deviance has a degree of freedom $n-p$, where n is the number of observations and p is the number of parameters in the model. If the residual deviance is greater than its corresponding degrees of freedom then the model does not fit the data well (Table 4 shows this). For a fitted Poisson regression model, the deviance D is given by

$$D = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\mu_i}\right) - \left(y_i - \mu_i\right)\right] \quad (5)$$

Over-dispersion in data

One feature of the Poisson distribution is that the mean equals the variance. However, over-or under-dispersion happens in Poisson models where the variance is larger or smaller than the mean (Dean and Lawless 1989). In this study, the model was tested for over-dispersion by comparing the mean and variance of a dependent variable as shown in Table 5. The variance is greater than the mean implying that there is over-dispersion which has an effect on standard errors (Tortum et al. 2012).

The quasi-Poisson family was introduced to correct standard errors in the Poisson regression model by replacing the Poisson family in the generalized linear model function in R (See results in Table 6) and the exponential of resulting coefficients were used for predicting road accidents as indicated in Table 7.

Results and Discussion

The Excel output for Figure 1 and R output for Tables 2 to 7 reveal the following:

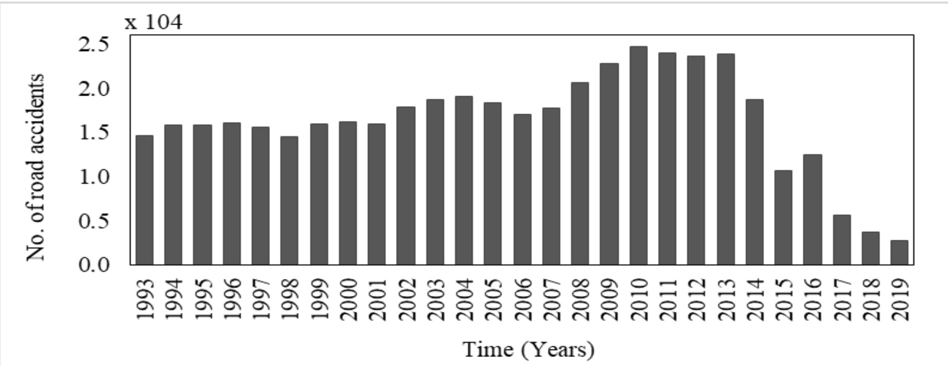

Figure 1: Histogram for road accidents data.

Figure 1 describes the Poisson distribution of the data which gives the reason of not using linear regression model that assumes normality. Table 2 shows that all the variables have VIF values less than 5 implying that there is no mutlicollinearity among the explanatory variables. Thus, all variables can be included in the subsequent analysis and modelling with the Poisson regression (Nwankwo and Nwaigwe 2016).

Table 2: Collinearity statistics

Model	Tolerance	VIF
Reckless driving	0.306	3.263
Careless pedestrians	0.447	2.239
High speed	0.282	3.544
Defective vehicle	0.201	4.982
Motor cyclists	0.225	4.435
Other factors	0.364	2.750

Results in Table 3 indicate that all predictors are significant since their corresponding p-values are less than 0.05. Also, the positive coefficient of predictors show that as a particular predictor increases so does the log number of accidents. That is to say, increasing the factors reckless driving, careless pedestrians, high speed, defective vehicles and other factors causes the increase of traffic accidents.

Table 3: Poisson regression model parameter estimation

Parameters	Estimate	Standard error	Z value	Pr(>|z|)
Intercept	8.228	0.0007741	1062.868	<2e-16
Reckless driving	0.00007653	0.0000009951	76.902	<2e-16
Careless pedestrians	0.0001111	0.000004305	25.818	<2e-16
High speed	0.00001216	0.000003690	3.296	0.000981
Defective vehicle	0.0001727	0.000003585	48.176	<2e-16
Motor cyclists	0.00006248	0.000002037	30.669	<2e-16
Other factors	0.0001007	0.000001477	68.155	<2e-16

Table 4 shows that the value of residual deviance is greater than its corresponding degree of freedom, and therefore, the model does not fit the data. Thus, the study checks whether there is over-dispersion or not by examining the relationship between mean and variance of the response variable as shown in Table 5. Results from Table 5 reveal that the variance is greater than the mean indicating the presence of over-dispersion which has the effect on standard errors of model parameters as they become smaller than they should be. One of the ways to address this problem is the introduction of quasi-Poisson to replace Poisson family in the genaralized linear model (glm) function if R Software was used and the results are as shown in Table 6.

Table 4: Goodness of fit test statistics

Residual deviance	df
2945.6	20

Table 5: Test for over-dispersion

Mean	Variance
16,385	32,191,076.87

Table 6: Quasi-Poisson regression model parameter estimation

Parameters	Estimate	Standard error	t value	Pr(>|t|)
Intercept	8.228	0.09162	89.810	<2e-16
Reckless driving	0.00007653	0.00001178	6.498	0.00000247
Careless pedestrians	0.0001111	0.00005095	2.182	0.041245
High speed	0.00001216	0.00004367	0.278	0.783491
Defective vehicle	0.0001727	0.00004242	4.071	0.000596
Motor cyclists	0.00006248	0.00002411	2.591	0.017445
Other factors	0.0001007	0.00001748	5.759	0.0000123

Table 6 shows the effect of introducing quasi-Poisson family to replace Poisson distribution in the glm fuction for the purpose of correcting standard errors which were smaller than they were supposed to be due to over-dispersion. By comparing the results from Tables 3 and 6, parameters in both cases have the same estimates/coefficients but with difference in their standard errors. The standard errors in Table 6 are greater than those in Table 3 indicating the impact quasi-Poisson.

Table 7 shows how road accidents can be predicted using explanatory varibles by taking exponentials of their estimates. Road accidents with respect to the expalanatory variables are predicted as follows:
Reckless driving with exponential estimate of 1.00007653 implies that the number of road accidents next year caused by reckless driving will increase by 0.007653%. Careless pedestrians with exponential estimate of 1.0001111 implies that the number of road accidents next year caused by careless pedestrians will increase by 0.01111%. High speed with exponential estimate of 1.00001216 implies that the number of road accidents next year caused by high speed will increase by 0.001216%. Defective vehicles with exponential estimate of 1.0001727 implies that the number of road accidents next year caused by defective vehicles will increase by 0.01727%. Motor cyclists with exponential estimate of 1.00006248 implies that the number of road accidents next year caused by motor cyclists will increase by 0.006248% and other factors with exponential estimate of 1.0001007 implies that the number of road accidents next year caused by other factors will increase by 0.01007%.

Generally, this shows that careless pedestrians will contribute most in the occurrence of road accidents, while high speed will take the least part. These results differ with that of Barengo et al. (2006) who carried out the study on the changes of traffic accidents in Dar es Salaam region from 1999 to 2001 and explored their underlying causes. Their study revealed that speeding, careless driving and mechanical defects of the cars (defective vehicles) were most contributing factors. The study by Boniface et al. (2016) have different findings from this study on which motorcycles were found responsible for the majority of road traffic crashes accounting for 53.4% of all cases. This is because their study considered patients involved in motor traffic crashes and attended in six public hospitals of Tanzania mainland for only one year in 2014, the reason which is also stated in their discussion section. However, the observed decreasing cases of road accidents from 2013 to 2019 may be caused by enforcement of traffic regulations and rules and recruitment of more police officers.

Table 7: Road accidents predictions

Parameters	Estimate	Exp (Estimate)
Reckless driving	0.00007653	1.00007653
Careless pedestrians	0.0001111	1.0001111
High speed	0.00001216	1.00001216
Defective vehicle	0.0001727	1.0001727
Motor cyclists	0.00006248	1.00006248
Other factors	0.0001007	1.0001007

Conclusion

The objective of this study was to develop the statistical model of road accidents in Tanzania. The Poisson regression model was proposed for establishing the relationship between road accidents and their contributing factors which are reckless driving, careless pedestrians, high speed, defective motor vehicles, motor cyclists and other factors including slippery road and poor visibility. The goodness of fit test was carried out and it was found that there was over-dispersion which was improved by introducing quasi-Poisson family in the generalized linear model function. Finally, all the factors causing road accidents predicted minor increase of accidents, and due to that the study recommends authorities to continue on monitoring traffic control regulations which have been put in place.

Acknowledgements

Authors are grateful to National Bureau of Statistics for data and reports on road traffic incidents in Tanzania.

References

Barengo NC, Mkamba N, Mshana SM and Miettola J 2006 Road traffic accidents in Dar es Salaam, Tanzania during 1999 and

2001. *Int. J. Inj. Control Safe Promot.* 13(1): 52-54.

Boniface R, Museru L, Kiloloma O and Muthari V 2016 Factors associated with road traffic injuries in Tanzania. *Pan Afr. Med. J.* 23(1).

Cameron AC and Trivedi PK 1986 Econometric models based on count data: comparisons and applications of some estimators and tests. *J. Appl. Econometr.* 1: 29-53.

Dean C and Lawless JF 1989 Test for detecting over-dispersion in Poisson regression models. *J. Am. Stat.. Assoc.* 84(406): 467-472.

Komba DD 2006 Risk factors and road traffic accidents in Tanzania Master Thesis, Norwegian University of Science and Technology (NTNU), 58-64.

Kumar ME and Umadevi G 2011 Application of system dynamic simulation modelling in road safety. In 3rd International Conference on Road Safety and Simulation Purdue University Transportation Research Board 2011.

Lawless JF 1987 Negative Binomial and mixed Poisson regression. *Can. J. Stat.* 15(3): 209-225.

Nwankwo CH and Nwaigwe GI 2016 Statistical model of road traffic crashes data in Anambra State Nigeria: a Poisson regression approach. *Eur. J. Stat. Probab.* 4(1): 41-60.

Nyakyi VP, Kuznetsov D and Nkansah-Gyekye Y 2014 Validating regression models to assess factors for motorcycle accidents in Tanzania. *Sci. J. Appl. Math. Stat.* 2(5): 97-104.

Tarimo HT 2012 Motor traffic injuries in Sub-Saharan Africa. *Dar es Salaam Med. Students' J.* 19(1): 13-16.

Tortum A, Codur MY and Kilinc B 2012 Modelling traffic accidents in Turkey using regression analysis. *Iğdır Univ. J. Inst. Sci. Technol.* 2: 69-78.

WHO (World Health Organization) 2009 Global status report on road safety. Geneva World Health Report: Making a difference. Geneva, World Health Organization. www.who.int/violence_injury_prevention/road_safety_status/en Retrieved on 25/06/2020.

WHO 2013 Global status report on road safety 2013*:* supporting a decade of action: summary (No. WHO. NMH. VIP 13.01). Geneva, World Health Organization. *www.who.int/violence_injury_prevention/road_safety_status/2013/en/* Retrieved on 31/01/2021.

Zhang H 2019 Traffic organization optimization of urban road combined intersections. *J. Transport. Technol.* 9: 325-330.