Young African Leaders Journal of Development

Volume 3 Article 17

1-1-2021

Hurdle Negative Binomial Model for Motor Vehicle Crash Injuries in Namibia

Ndilimeke Shiyuka

HURDLE NEGATIVE BINOMIAL MODEL FOR MOTOR VEHICLE CRASH INJURIES IN NAMIBIA

Ndilimeke Shiyuka,
Statistician, Namibia Statistics Agency

ABSTRACT

The study was based on a quantitative research for all road crash injuries recorded from 2011-2016 secondary data, with number of injured persons per crash as the dependent variable. The Hurdle Negative Binomial two-way model looked at crashes with injuries=0 (Ordinary Ratio (OR)), and those with injuries >0 (Relative Ratio (RR)): crash type with vehicle to vehicle (OR=0.5), Cause of crash with driver behaviour (OR=0.1), and Time of crash with peak time (OR=0.3); Months of crashes with holiday month (RR=0.2) Day of crash with Weekend (RR =0.1) Region with Northern regions (RR= 0.3), Type of crash with vehicle by vehicle (RR =1.6), Crash cause with driver behaviour (RR= 0.2), with P< 0.001 for all variables had higher probability. Emphasis should be placed on driver behaviour and campaigns should focus more on school holiday months and weekends, due to the fact that a large number of crashes occur more on weekends and school holiday months.

Keywords: *Statistical model, Motor vehicle crash, Quantitative research, Driver behaviour*

INTRODUCTION

It is estimated that injuries as a result of road traffic crashes results in approximately 1.25 million deaths annually and another 20 to 50 million people sustain injuries (WHO, , 2015). The World Health Organisation (WHO) further states that Road traffic injuries are a leading cause of death, and the main cause of death among those aged 15–29 years (WHO, 2015), making this age group an accident high risk. Additionally, the Motor Vehicle Accident fund has recorded that 31% of road crash fatalities were people aged 16- 30.

Current trends suggest that by 2030 road traffic deaths will become the fifth leading causes of deaths unless urgent action is taken (World Health Organization, 2013). In fact for Namibia we have already reached this milestone according to a study done by (MVA Fund, 2015), Motor vehicle crashes were the 5th cause of death in Namibia. A study on fatal injuries of United States of America citizens abroad indicated that: The highest age adjusted proportional mortality ratios where highest in Africa. Also, the death rate in Africa due to road crashes is at 24.1 per 100 000 inhabitants (Guse, et al 2007), in 2013 Namibia stood at 29.9 deaths per 100 000 population (Motor Vehicle Accident Fund, 2015).

Over a five year period between 2011 and 2015, the number of injured persons has increased by 30%, it was reported that there was 3% increase between 2012 and 2013, 18% increase between 2013 and 2014, and a further 6 % increase between 2014 and 2015 (MVA, 2015). Furthermore, road crash injuries keeps on fluctuating with time, the number of injured person was 5652 in 2012, 5845 in 2013, 6314 in 2014 and 7333 in 2015, indicating 347 injuries per 100, 000 populations in 2015. On average 20 people are injured daily in Namibia due to road crashes, the injuries vary from slight to severe. The research aims to analyse the different variables that are associated with the increase of crashes to close this gap and alert organisations, public and private sectors on the seriousness of road fatalities to vender resources in the subject. Especially resources in research to further scrutinize and understand crash causes and causes of injuries. It is essential that analysis is done with the available data, to study the relations among variables to determine causes of crashes.

METHODS

Data

The study was a quantitative study for all road crash injuries recorded between 2011 and 2016. Secondary data from the MVA Fund crash and claims web based system was retrieved and exported to Micro soft excel, the analysis of this study has been done in SPSS and R statistical packages. The study analysed injuries from crashes where one or more injuries or a deaths has occurred. No sampling was necessary since all injuries that occurred between 2011 and 2016 and are recorded on the MVA Fund database and will be used for analysis. Poisson, Negative Binomial, Zero-Inflated Poisson, Zero-inflated negative binomial, Poisson Hurdle, and Hurdle negative binomial models were each fitted with "MASS" and "pscl" in R 3.3.2 packages using the glm, nb, zeroinfl, hurdle functions to fit all the models in order to choose the best.

Hurdle Negative Binomial Generalised linear model

Hurdle Negative binomial models is mixed by a binary outcome of the count being below or above the hurdle (the selection variable), with a truncated model for outcomes above the hurdle (Saffari et al., 2012). The hurdle negative binomial can handle excess zeros and the analysis of under- dispersion and over- dispersion.

Tests For Association

Tests for associations between the dependent variable and the independent variables was essential, since the research paper has both numeric and categorical variables different tests were performed. The variable "Persons injured" was recoded into categorical variable in order to carry out a chi- Squared test, between persons injured and the categorical variables as illustrated in Table 1. At 5% level of significance the P value for all the variables is less than 0.05 meaning that there is enough evidence to prove that there is association between the Persons Injured and month, Day of crash, Region, Crash type, Crash cause and time the crash has occurred. There is a high correlation of 0.874 between the Crash cause and the dependent variable, this is the same for crash type as well. Although there is correlation between month and persons injured, the relationship is close to zero.

Table1: Test for association Persons Injured~ categorical variables

	Chisquare test	P Value	Significance	R^2
Month	156.820	<0.001	***	0.085
Day of crash	130.620	<0.001	***	0.077
Region	1013.120	<0.001	***	0.215
Crash Type	4919.140	<0.001	***	0.474
Crash Cause	2745.520	<0.001	***	0.874
Grouped time	664.710	<0.001	***	0.354

Variable is significant

Model Selection

Table 2. show the different generalised linear models used by the author in this study, the Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) are goodness-of-fit criteria used for model selection the fact that the goodness of fit statistics is greater than 1 shows that there was over

dispersion in the data (Yesilova et al., 2010). The Poisson Regression (PR), Negative Binomial (NB), Zero-Inflated Poisson (ZIP), Zero- Inflated Negative Binomial (ZINB), Hurdle Poisson (HP), and Hurdle Negative Binomial (HNB) are given in the table that produced widely different results. The model with the lowest AIC is the HNB which is bolded out in Table 2. Therefore, the Hurdle Negative Binomial was chosen as the best model.

Table 2: AIC comparison model

	Model Name	Log-likelihood	AIC
Model 1	Poisson		79890.16
Model 2	Negative binomial	-73812.24	73832.24
Model 3	Zero-inflated Poisson model	-3.89e+04	77838.93
Model 4	Zero-inflated negative binomial	-3.639e+04	72815.07
Model 5	Hurdle Poisson	-3.733e+04	74695.09
Model 6	**Hurdle negative binomial**	**-3.203e+04**	**64089.84**

RESULTS AND DISCUSSIONS

Results

Table 3 shows the results of the HNB Regression Model. The Ordinary ratio (OR) shows model estimates for crashes with zero injuries or the zero outcomes while the Relative Ratio (RR) shows model estimates for crashes that had one or more injuries. When the p-value of any model shows to be more than the assigned level of significance, we reject the null hypothesis with the implication that there is no significance different between the two variables of test. The data is analysed based on the two separate models produced by the HNB; the Model 1 Ordinary ratio which focus on crashes where zero persons are injured, and model 2 the Relative Ratio where one or more people have been injured.

Model 1 - Ordinary Ratio

There was no significant difference between holiday months and non-holiday months for crashes that did not have any injuries (p=0.66143). Likewise, for weekends OR results indicate that whether it is weekend or not there will still be crashes where no injuries were observed with (p= 0.0786). Also, there is no significance difference (p=0.6788) for northern regions in comparison with other regions in the rest of the country. On the other hand, for types of crashes, types that involved vehicles by vehicle where significantly high, thus if you increase vehicle by vehicle crashes by one point the odds that there would be zero injuries would increase by 0.4 if other variables are held constant (OR = 0.487, 97.5% CI: 0.3948, 0.5789).The causes of crashes that occurred as a result of driver behaviour are 0.147 more than those that have nothing to do with the driver (OR= 0.147, 97.2% CI: 0.0534.0.2409). For crashes where fatalities were recorded, the crashes that had fatalities were 1.835 times way less compared to those that had none, in other words for crashes with more fatalities decreases the number of injuries (OR= -1.835, 97.5% CI: -1.9269, -1.7433). With regards to the number of vehicles involved there was no significance difference between the number of vehicles involved, whether one or more vehicles are involved the probability of have zero injuries is there (p= 0.7885). On the other hand when we look at time, the non-injuries where 0.3 times higher during peak time as compared to off peak time (OR = 0.326, 97.5% CI: 0.2371, 0.4152).

Model 2 - Relative Ratio

The intensity of injured person's due to crashes is significantly associated with the month, (crashes that had one or more persons injured). During holiday months it is 0.2 times more likely to have a crash with injured people as compared to months with no school holidays, if you are to increase the number of holiday months by one unit, the number of injuries will increase with 0.2 (RR = 0.209, 97.5%, CI: 0.1460, 0.2715). This is similar to days of the week, weekends are 0.1 times more likely to have injuries intensity as compared to other days of the week (RR = 0.106, 97.5% CI: 0.0434, 0.1682). When it comes to the five Northern regions, they are significantly high in the number of injuries when contrasted with other regions in the country (RR = 0.295, 97.5% CI: 0.2305, 0.3595). The model indicates that there is significance difference in the types of crashes, if vehicle to vehicle crashes increase by one unit, the number of injuries will increase by 1.6 while holding other variables constant, accordingly, the higher the number of vehicle by vehicle crashes, the higher the number of injuries (RR = 1.582, 97.2% CI: 1.5094, 1.6546). With the causes of crashes that had something to do with driver behaviour, they were significantly higher than those with no environmental or vehicle factors with (RR = 0.224, 97.5% CI: 0.1562, 0.2915). With regard to fatalities, high injury crashes tent to involve 1.0 times more deaths per crash as compared to those without fatalities (RR = 1.01, 97.5% CI: 0.8974. 1.1227). With the number of vehicles involved, single vehicle crashes involve less injuries compared to crashes with more than one vehicle (RR = -0.216, 97.5% CI: -2.2860, -0.1454). Moreover, the time the crash occurs is insignificant to the number of injuries that occur, at 97.5% confidence interval peak time is not significantly different from off peak time in terms of the number of injuries (p= 0.4054) unlike with zero injuries were the peak time yielded less injuries.

Table 3: *Regression Estimates from the Hurdle Negative Binomial Regression Model*

	Hurdle Negative Binomial regression model			
Variable	Injured persons probability (Zero people injured in a crash)		Injured Persons Intensity(# of Injures Persons >= 1)	
	OR	97.5% CI	RR	97.5% CI
Number of persons injured percrash	1.596***	(1.4718, 1.72049)	-4.427***	(-6.4427, -2.4113)
Month				
None Holiday month	1		1	
Holiday Month	-0.017	(-0.0955, 0.0606)	0.209***	(0.1460, 0.2715)
Day of Crash				
Non-Weekend	1		1	
Weekend	0.069	(-0.0079, 0.1469)	0.106***	(0.0434, 0.1682)
Region				
Non-Northern	1		1	
Northern	0.017	(-0.0635, 0.0976)	0.295***	(0.2305, 0.3595)
Crash Type				
Vehicle and other factors	1		1	
Vehicle by Vehicle	0.487***	(0.3948, 0.5789)	1.582***	(1.5094, 1.6546)
Crash Cause				
Non Driver	1		1	
Driver behaviour	0.147**	(0.0534, 0.2409)	0.224***	(0.1562, 0.2915)
Fatalities				
NoDeaths	1		1	
Deaths	-1.835***	(-1.9269, -1.7433)	1.01***	(0.8974, 1.1227)
Vehicles Involved				
More than one	1		1	
Single Vehicle	0.015	(-0.0875, 0.1153)	-0.216***	(-0.2860, -0.1454)
Time				
Off Peak time	1		1	
Peak time	0.326***	(0.2371, 0.4152)	0.029	(-0.0387 0.0959)

	Hurdle Negative Binomial regression model			
Variable	Injured persons probability (Zero people injured in a crash)		Injured Persons Intensity(# of Injures Persons >= 1)	
	OR	97.5% CI	RR	97.5% CI
Number of persons injured percrash	1.596 ***	(1.4718, 1.72049)	-4.427 ***	(-6.4427, -2.4113)
Month				
None Holiday month	1		1	
Holiday Month	-0.017	(-0.0955, 0.0606)	0.209 ***	(0.1460, 0.2715)
Day of Crash				
Non-Weekend	1		1	
Weekend	0.069	(-0.0079, 0.1469)	0.106 ***	(0.0434, 0.1682)
Region				
Non-Northern	1		1	
Northern	0.017	(-0.0635, 0.0976)	0.295 ***	(0.2305, 0.3595)
Crash Type				
Vehicle and other factors	1		1	
Vehicle by Vehicle	0.487 ***	(0.3948, 0.5789)	1.582 ***	(1.5094, 1.6546)
Crash Cause				
Non Driver	1		1	
Driver behaviour	0.147 **	(0.0534, 0.2409)	0.224 ***	(0.1562, 0.2915)
Fatalities				
NoDeaths	1		1	
Deaths	-1.835 ***	(-1.9269, -1.7433)	1.01***	(0.8974, 1.1227)
Vehicles Involved				
More than one	1		1	
Single Vehicle	0.015	(-0.0875, 0.1153)	-0.216 ***	(-0.2860, -0.1454)
Time				
Off Peak time	1		1	
Peak time	0.326 ***	(0.2371, 0.4152)	0.029	(-0.0387 0.0959)

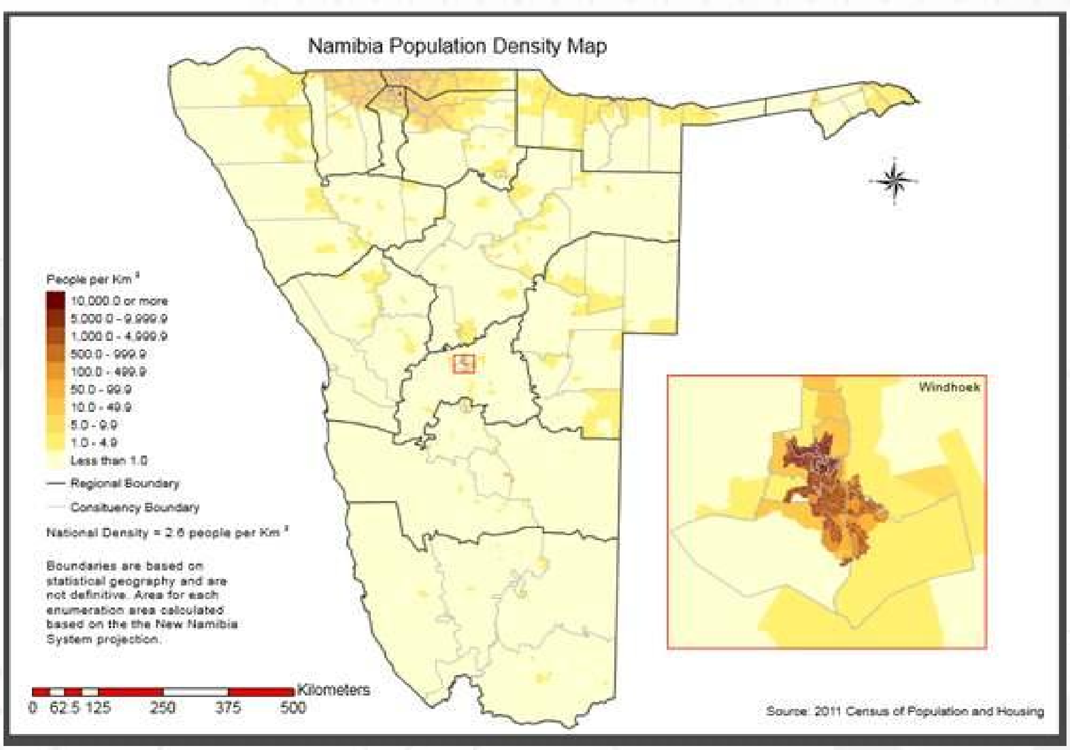
DISCUSSION

The chi-square and the Pearson test indicated that there was association between the predicted and the predictor variables, of which all were found significant at 5% level of significance. All the variables had R2 less than 0.5 except for the causes of crashes that had R2 = 0.874 making it the best predictor for road crash Injuries. Therefore, it is reasonable to conclude that the cause of crash can predict that the crash will have injuries.

The two-way model looked at crashes that had zero injuries (OR) and those with one or more injuries (RR), on the risk factors of the two there were some similarities found. For example, for CZI (Crashes with zero injuries) there was no significant difference between school holiday months and other months of the year, whereas crashes that had Injuries Greater than or Equal to one (IGE1) school holiday months (which also coincides with moths that have public holidays) had more crashes, concluding that, January, April, May, August and December which consist of school holidays and public holidays increase the number of injuries in a crash. Similarly, Sukhai et al., (2011) declares that South Africa has a significant peak in December when it comes to road traffic crash fatalities. The peak is largely explained by traffic flow factors, increased alcohol consumption during the holiday may also be a risk factor. Also, weekends proved to have had more IGE1s comparing to weekdays. The work of Elliot (2009) suggest that 52% of injuries that occurs during weekends with most crashes happening on Saturdays and Sundays.

One very interesting fact of the data is that peak hours indicate many CZI, meaning that although there is many crashes during the peak hours they are so minor that people do not get injured or die. Kingham et al., (2011) report their findings that crash rates are not occurring at a uniform rate throughout the day, with comparative increases in crash rates occurring during morning rush hour, and during the 'school runs'. However this study confirms that, crashes that occur during peak time caused very few injuries, in layman terms, if a crash occurs at pick time, the probability that no one gets injured is high. It is worth noting that at this point we are only looking at injuries as our predicted variable, the results may yield differently if deaths were the variable of interest.

Figure 1: Namibia's population density



The Northern regions are prompted to IGE1s more than the CZI, the vast country makes it difficult to tell the exact regions the crash has occurred, some of the very severe crashes occurred between roads that are connecting towns (MVA Fund, 2016).

Another fascinating fact is the relationship between the number of none injured persons and fatalities by crash; crashes with fatalities decrease the number of none injured persons denoting that if the number of people that died increase, the number of none injured persons will decrease. This is due to the fact that, if there are fatalities in a crash it indicates the severity of the crash, and the more severe the crash the higher the chances of injuries and the chances of deaths in that crash. As the number of deaths increases so does the number of injuries, so the number of fatalities work in conjunction with the number of injuries.

Crashes that occurred as a result of driver behaviour such as, Reckless and Negligence driving, speed and being intoxicated compared to issues like tyre burst, mechanical failure, poor road design, pedestrian, animal, poor weather condition indicated that the number of injured persons is predicted by the cause of crash and that injuries will even be more for crashes that occurred as a result of driver behaviour or circumstances. Ultimately, the behaviour and attitudes of the driver will determine how severe the crash will be, should a crash occur. The other observation was among the types of crashes, crashes resulted in vehicles crash only such as collisions with other vehicles and roll over exhibited crashes with more injuries with a gigantic difference for both the CZI crashes and with crashes that have had 1 and more injuries (IGE1), that is to say the type of crash does predict injuries, and if a crash involves only vehicles for instance collisions with other vehicles and roll overs increase the number of injuries more than the other types of crashes (collision with animals, collision objects, pedestrians, falling from moving vehicle, cyclist and motorcyclists). In contrast, Single vehicle crashes were significantly lower among IGE1 and insignificantly more with CZI, what this means is that crashes that involved one vehicle have the potential to reduce injuries, the fact is that with a single vehicle it is usually that the number of casualties are few for example: in a pedestrian crash, only one person gets either injured or killed, and most vehicles have a maximum of 5 occupants.

CONCLUSION

The risk factors associated with road traffic injuries were clearly stipulated in this study: Month, Day of crash, Region, Crash type, crash cause, number of vehicles involved, and the time the crash occurred.

The Risk factors that influence crashes have been determined and found as follows: The types of crashes for vehicle to vehicle collisions and roll overs posted a greater probability to injuries as compared to pedestrians, collision with fixed objects, fall from moving vehicles and collisions with animals. Causes of crashes were another estimator of injuries; the study found that driver behaviour is a much larger contributor to injuries than other factors such as reckless and negligent driving, speed, poor visibility and being intoxicated. Other causes like the tyre bursts, mechanical failures, weather conditions and road designs are not the huge contributors to injuries. The month in which a crash has occurred is significant when specifically looking at school holiday month which are also months in which most public holidays fall (January, April, May, August and December), these holidays showed a great concern in the number of injuries in a crash. The weekend also proved to increase the number of injuries per crash. Finally, the geography of the crash plays a role in the number of injured persons per crash. The Oshana, Omusati, Ohangwena, Oshikoto, Kavango and Otjozondjupa regions contribute more to the number of injuries compared to other regions. Crashes with fatalities are also good predictors of road crash injuries.

RECOMMENDATION

l Policy makers should concentrate on crashes reduction mechanisms that focuses on driver behaviour, issues of speed and being intoxicated post a greater risk to injuries. Since speed is the second cause of crash after reckless and negligent driving, laws can be passed that regulates speed; for instance, introduce the arrest on the scene mechanism, where if a driver is found driving beyond the speed limit, should be arrested and should appear in court before continuing with their destination.

l The introduction of the penalty points system in Namibia may just be the answer, seeing that driver behaviour poses a huge risk to the increase of crashes. Novoa (2010), did a study to assess the effectiveness of the penalty points system introduced in some European countries showed that in Italy, Ireland and Spain indicated a reduction in the number of people injured by 19%, 36% and 12% respectively.

- School holiday months coincided with public holidays, weekends revealed high intensity of injuries. Campaigns should be concentrated during this time periods to check for speed and whether the driver is intoxicated while operating a vehicle. This interactions should be set up any time of the day, the study showed that peak time are not the dangerous times to drive early morning hours and very late at night is a time to explorer.
- As suggested by Reynolds et al., (2009) a construction of a bicycle and pedestrian only lane plus sufficient pedestrian and cyclist crossing in towns may be the key to reducing pedestrian and cyclist types of crashes that contribute 28% to crashes. Furthermore, almost half of all deaths on the world's roads are road users with the least protection – motorcyclists, cyclists and pedestrians (WHO, 2015).

REFERENCES

Guse, C. E., Cortés, L. M., Hargarten, S. W., & Hennes, H. M. (2007). Fatal injuries of US citizens abroad. Journal of travel medicine, 14(5), 279-287.

Kingham, S., Sabel, C. E., & Bartie, P. (2011). The impact of the 'school run'on road traffic accidents: A spatio-temporal analysis. Journal of transport geography, 19(4), 705-711.

Motor Vehicle Accident Fund (2016) 2015 crash and claim report Windhoek. MVAFund

Nantulya, V. M., & Reich, M. R. (2003). The neglected epidemic: road traffic injuries in Developing countries. Bmj, 324(7346), 1139-1141.

Novoa, M. A., Perez, K. Santamarina-Rubio, E., Mari-Dell'Olmo, M., Ferrando, J., Peiro, R., Tobias, A., Zori, P., & Borrell, C. (2010). The Impact of the Penalty Points System on road traffic injuries inSpain: A time series study. American Journal of Public health. 100(11): 2220-2227 DOI: 10.2105/ajph.2010.192104

Reynolds, C. C., Harris, M. A., Teschke, K., Cripton, P. A., & Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environmental health, 8(1), 47.

Saffari, S. E., Adnan, R., & Greene, W. (2012). Hurdle negative binomial regression model with right censored count data. SORT, 36(2), 181-194.

Sukhai, A., Jones, A. P., Love, B. S., & Haynes, R. (2011). Temporal variations in road traffic fatalities in South Africa. Accident Analysis & Prevention, 43(1), 421-428.

World Health Organization. (2013). Global Status Report on Road safety. WHO.

World Health Organization. (2015). Global Status Report on Road safety. WHO.

Yesilova, A., Kaydan, M. B., & Kaya, Y. (2010). Modeling insect-egg data with excess zeros using zero-inflated regression models. Hacettepe Journal of Mathematics and Statistics, 39(2), 273-282.