Road Fatality Model Based on Over-Dispersion Data Along Federal Route F0050

Wan Zahidah Musa^{1,*}, Joewono Prasetijo¹, and Zaffan Farhana Zainal¹

¹ Smart Driving Research Center (SDRC), Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia

Abstract. According to The World Health Ranking 2011 has ranked Malaysia as 20th in its list of countries with the most deaths caused by road accidents. Road accidents also have been identified as the prime cause of death in Malaysia after the heart disease, stroke, influenza and pneumonia. To date, previous researches from Malaysian Institute of Road Safety (MIROS) have reported that averages of 18 people were killed on Malaysian road daily. There are many kinds of models that have been developed in modelling the circumstance of accidents. The most widely applied was Poisson and Negative Binomial regression models while Zero-inflated Poisson and Zero-inflated Negative Binomial are the modification of Poisson and Negative Binomials regression models. This study interested to focus on road F0050 as statistic data from Royal Malaysian Police 2014 list F0050 as one of the high accident road in Malaysia from kilometre 0 until kilometre 58. R programming was chose to analyse the relationship between road fatality and its factor (annual average daily traffic (AADT), speed, shoulder width, lane width). Negative binomial and Zero-inflated negative binomial (ZINB) were shown to be preferred modelling methods for this study. Significant positive relationships were also identified between road fatality and annual average daily traffic (AADT) and lane width. This relationship can be a helpful support to the decision making of accident management for road F0050.

1 Introduction

Traffic accidents have been recognized as one of the major causes for human and economic losses both in developed and developing countries. In 2014, Malaysia launched a new road safety plan, aligned with the Global Decade of Action for Road Safety which aims at having the predicted number of fatalities by 2020 [1]. Most recent safety data from Royal Malaysian Police found that there were 6674 road deaths in 2014 which were 3.5% less compared to 2013. It has been conclusively shown that this is the largest reduction in the number of road fatalities in the past 10 years [2]. Traffic accidents represent a significant cost for the Malaysian society which based on a willingness-to-pay estimation, traffic crashes cost each year around 1.6% of the national gross domestic product. In a following-

* Corresponding author: wanzahidahmusa@yahoo.com

up study, [3, 4] found that the statistical value of life in Malaysia is RM 1.3 million, around EUR 330 000.

Fig. 1. Road fatalities figures by road user group (total fatalities) 2014 [ITF, 2015].

Fig. 2. Road fatalities figures by road type [ITF, 2015]

Zero-Inflated models has a pivotal role in predicting probability of crashes and identify the most relevant contributory factors at the hazardous sections as this models can handle an over-dispersion data and high excess zero properly [5-15]. Malaysia has it owns plan and target that set by the Global Decade of Action for Road Safety 2011-2020 to reduce the predicted number of road deaths in 2020 by 50%. MIROS present a comprehensive review on the absence of a comprehensive road safety programme, there would be 8760 road fatalities in 2015 and 10716 in the year 2020 based on their 2015 annual report. This shows an effective reduction of 22% in the number of road fatalities when compared to 2010 level.

Table 1. Fatalities target [ITF.2015]

	2015	2020
Predicted number of fatalities (business as usual)	8760	10716
Fatalities targets with the implementation of the plan	6570	5358

Over dispersion in Poisson models occurs when the response variance is greater than the mean [16]. The main issue in need of future effort is the reliability test of some statistical models (Poisson, Negative Binomial and Zero Inflated) in yielding better results in traffic safety. Although these types of model have been well coordinated in many other fields such as economics and sociology, it is only less than 10 years when traffic safety analysts first reported relevant results in well-known publications. Especially, while most applications are limited to types of vehicles, types of crashes, varying-intercept model, and varying-slope model deserves more effort because of the complexity in model estimation and result interpretation. Therefore, this study is the new approach of quantifying road safety performance at the heterogeneous traffic conditions at route F0050.

2 Method

2.1 Accidents

Fig. 3. Road accidents in front of traffic light UTHM at kilometre 19

Fig. 4. Road accidents in front of shell gas station: Parit Raja at kilometre 12

Fig. 5. Road accidents in front of Batu Pahat Mall at kilometre 4

Up to now, there are three types of road accident that applied in Malaysia which was fatal, serious injured and slight/minor injuries. Road fatality was defined as death resulting from a road crash within 30 days after the crash. While serious injured is when a person

injured as a result of a road crash. Lastly, a slight injury is any injury that does not fit under death or serious injury. This paper attempts to show an interest only to road fatality and investigates the relationship with the factors that determine the fatality.

2.2 Study Location: Federal Route F0050

2.2.1 Route F0050: Batu Pahat-Ayer Hitam-Kluang

Fig. 6. Hotspot accident locations at kilometre 23

According to [2], F0050 was listed among the highest accident road in Malaysia. As shown in Figure 6 and Figure 7, there are the hotspots locations of accident happen. F0050 is a four lane two ways divided road that runs from Batu Pahat through Ayer Hitam to Kluang. Most of the hazardous locations have high access points and annual average daily traffic(AADT).

Fig. 7. Hotspots location of accident at kilometre 22

2.3 Stochastic Regression Models

2.3.1 Poisson, Negative Binomial, Zero-inflated Poisson, and Zero-inflated Negative Binomial

The data of crashes was under characteristic of discrete, sporadic and random. Stochastic regression models were applied to describe the occurrence of accidents instead of deterministic models [11]. The most widely applied was Poisson and Negative Binomial regression models while Zero-inflated Poisson and Zero-inflated Negative Binomial are the modification of Poisson and Negative Binomials regression models. They were proposed to handle over dispersion and excess zeros in the data set.

3 Results and discussion

Table 2. Variables in the final model

Independent Variable	Location	Dependent Variable
LN AADT, SP, SW, LW	F0050	Fatal/Yr/Km

Table 3. Results of Pr(>|z|) modeling accidents for model

	P2	NB2	ZIP2	ZINB2
LN.AADT	0.842	0.963	0.686	0.961
SP	0.116	0.240	0.097	0.223
SW	0.244	0.319	0.257	0.320
LW	0.704	0.765	0.461	0.768

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4. Summary statistics of model

Variable	Variance	Mean	S.D	Min	Max
Fatal /Yr/Km	5.8135	3.254	2.4111	0	10
LN AADT	-	9.469	0.4611	7.91	9.81
SP	-	80.15	24.5998	36.12	140.87
SW	-	1.475	0.5409	0.50	2.80
LW	-	3.521	0.4211	2.50	3.75

Table 5. Correlation summary F0050 for model

Variable	F	LN.AADT	SP	SW	LW
F	1				
LN.AADT	-0.003	1			
SP	-0.148	0.176	1		
SW	-0.113	-0.085	-0.041	1	
LW	-0.022	-0.088	0.429	-0.103	1

Table 6. Over-dispersion test for model

Location	Dispersion Test (Poisson)
F0050	Over dispersion test data: poisson z = 2.5283, p-value = 0.005731 alternative hypothesis: true dispersion is greater than 1 sample estimates: dispersion 1.72433

Table 7. Estimation results for model

	P2	NB2	ZIP2	ZINB2
LN.AADT	0.032	0.010	-0.073	0.010
SP	-0.005	-0.005	-0.005	-0.005
SW	-0.167	-0.185	-0.166	-0.185
LW	0.074	0.077	0.147	0.078

Table 8. Results of model evaluation and comparison model

	P	NB	ZIP	ZINB
LL	-133.57	-127.69	-131	-127.7
K	4	4	4	4
N	58	58	58	58
AIC	275.14	267.39	270	263.4
BIC	274.19	262.43	269.05	262.45
Theta	-	4.46	-	4.464

Table 9. Comparison for best performance for model

	P	NB	ZIP	ZINB
AIC	-	**	*	***
BIC	-	***	*	**
Vuong	-	***	*	**
LL	-	***	*	**

Criteria: '***' (Best & Fit), '**' (Fit), '*'(Average), '-'(Low)

Stochastic regression model was used to analysis all of the collected data which was Poisson model, Negative Binomial model, Zero-inflated Poisson and Zero-inflated Negative Binomial. Table 7 above illustrates the estimation results of each model that was analyzed using R programming. In order to see the performance of the model the estimation value of variables has been examined. Models with logical algebraic signs of the variables were selected. As shown in Table 7 provides the summary of how the dependent variable associates with independent variables. Interestingly as can be seen, mostly from all model that the variables speed (SP) and shoulder width (SW) have negative association with fatality crashes. While for the variables AADT (LN.AADT) and lane width (LW) have majorly positive association with the accident. Taken together, these results suggest that there is preliminary evidence that in route F0050 with too many access points with increase of average annual daily traffic (AADT) have more contribution to fatality crashes. This proved by the most hazardous sections along this route was at main point by local people mostly around business and residential area such at kilometre 2,4,7,19,22,23,48,57 and 58. High AADT leads to lower speeds and trigger driver to take over the front car and this increase the accidents in federal route F0050. It is interesting to note that in most study found that lane width (LW) is negative association with fatality but incompatible with result found in Table 7. Lane width (LW) increases the likelihood of fatality in route F0050, although when lane width (LW) of lanes increases the traffic flows gets more convenient but the traffic conflicts and driving maneuvers such as lane changing increase and hence increases the probability of accident happened. As comparison of the shoulder width, a negative sign of shoulder width suggests that more is the shoulder width less will be the accident and it is logically acceptable.

Table 8 presents the results of model evaluation and comparison between the models. The best log-likelihood has the maximum value. In Table 8, negative binomial (NB) has the higher value which was -127.69 followed by zero-inflated negative binomial (ZINB) (-127.7), zero-inflated Poisson (ZIP) (-131) and Poisson (P) (-133.57). While for the akaike information criteria (AIC), the lesser the fittest. Zero-inflated negative binomial (ZINB) has the least akaike information criteria (AIC) value (263.4); followed by negative binomial (NB) (267.39), zero-inflated Poisson (ZIP) (270) and Poisson (P) (275.14). Comparisons of best performance of model are set out in Table 9. Data from this table can be compared with the results from Table 8 which shows the comparison to choose the best and fittest performance of model. What is interesting in this data that negative binomial (NB) has the

most points of criteria for the best and fittest model followed by zero-inflated negative binomial (ZINB), zero-inflated Poisson (ZIP) and lastly Poisson (P). Although negative binomial (NB), has the most points, the different value of NB performance and ZINB performance in this model just 0.01. Interestingly, these summarised that for this data was being under over-dispersion but not have high excess zero. In the final part for these studies both NB and ZINB model was choose as the best model.

4 Conclusions

This study set out to determine a model that can properly handle the highly stochastic nature of crash event and finding the most relevant contributory factors at the hazardous road sections at Johor federal route. In general, Zero-inflated regression and negative binomial was found to determine an empirical association between fatality road accidents, route geometric and traffic parameters. The results suggest that valuable insight into the underlying association between probability factors and vehicle crash.

The author would like to thanks the Ministry of Education Malaysia in funding this study under the Fundamental Research Grant Scheme (FRGS) Vot. 1452. Further thanks are also for Smart Driving Research Center (SDRC), Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) for support and providing facilities to accomplish the study.

References

- [1] OECD/ITF, *Road Safety Annual Report 2015*, OECD Publishing, Paris, (2015)
- [2] RMP, *Statistical Report on Road Accidents in Malaysia*, (2014)
- [3] G. Nor, M. Fauzi, Value of life of Malaysian motorists : Estimates from a nationwide survey, *Journal of the Eastern Asia Society For Transportation Studies*, **5**, (2003)
- [4] M. Yusoff, Malaysian value of fatal and non-fatal injury due to road accident : The willingness to pay using conjoint analysis study, *Proceedings of the Eastern Asia Society for Transportation Studies*, **8**, (2011)
- [5] C. Pudprommarat, M. Khamkong, P. Bookkamana, Zero-inflated poisson regression in road accidents on major road in the North of Thailand, *IRCNSA Proceedings*, 323-330 (2005)
- [6] A.H. Lee, M.R. Stevenson, K. Wang, K.W. Yau, Modeling young driver motor vehicles crashes: Data with extra zeros, *Accident Analysis and Prevention*, **34**, 515-521 (2002)
- [7] S.P. Miaou, The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, *Accident Analysis and Prevention* **26**, 471-482 (1994)
- [8] E. Ayati, E. Abbasi, Modelling accidents on mashad urban highways, *Open Journal of Safety Science and Technology*, **4**, 22-35 (2014)
- [9] M. Asrul, N.N. Naing, Mortality in hospitalized human immunodeficiency virus (HIV) and tuberculosis (TB) infection patients: Zero-inflated negative binomial death rate (ZINBDR) models, *Applied Mathematical Sciences*, **7**, 37-47 (2013)
- [10] A.K. Sharma, V.S. Landge, Zero inflated negative binomial for modelling heavy vehicle crash rate on Indian rural highway, *International Journal of Advance in Engineering & Technology*, **5**(2), 292-301 (2013)
- [11] N.J. Garber, L. Wu, *Stochastic Models Relating Crash Probabilities with Geometric and Corresponding Traffic Characteristics Data*, Research Report No. UVACTS-5-15-74, (2001)

- [12] M.H. Pour, J. Prasetyo, A.S.Yahaya, S.M.R. Ghadiri, Modeling vehicle-pedestrian crashes with excess zero along Malaysia Federal Roads, *Procedia-Social and Behavioral Sciences* **53**, 1218-1227 (2012)
- [13] D. Lord, S.P. Washington, J.N. Ivan, Further notes on the application of zero-inflated models in highway safety, *Accident Analysis and Prevention* **39**, 53-57 (2007)
- [14] J. Prasetyo, W.Z.Musa, Modeling zero-inflated regression of road accidents at Johor Federal Road F001, *MATEC Web of Conferences*, **47**, 03001, (2016)
- [15] J. Prasetyo, W.Z.Musa, Z.F. Zainal, Identification of hazardous road sections using over dispersion-excess zero data of vehicle accident at Johor Federal Roads, *Proceedings International Conference on Engineering of Traumatology*, **1**(1), (2015)
- [16] J.M. Hilbe, *Negative Binomial Regression*, 2nd ed., *California Institute of Technology and Arizona State University*, (2011)