

Vehicle Road Accident Prediction Model Along Federal Road FT050 Kluang-A/Hitam-B/Pahat Route Using Excess Zero Data

J. Prasetijo^{1*}, W. Z. Musa², Z. Mohd Jawi³, Z. F. Zainal², K. Anuar², M. A. Ridho¹ and M. A. M. Aron¹

¹Department of Transportation Engineering Technology, Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia, Johor D.T., Malaysia

²Department of Infrastructure Engineering and Geomatics, Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, Johor D.T., Malaysia

³Vehicle Safety & Biomechanics Research Centre, Malaysian Institute of Road Safety Research (MIROS), Kajang, Selangor D.E., Malaysia

*Corresponding author: joewono@uthm.edu.my

ORIGINAL ARTICLE

Open Access

Article History:

Received
3 Jul 2018

Received in
revised form
29 Sep 2018

Accepted
10 Oct 2018

Available online
1 Jan 2019

Abstract – Traffic accidents have become a major socio-economic problem in Malaysia and are considered the primary cause of mortality. Over 60 percent of these fatal accidents occurred on rural roads. Nearly half of all fatalities took place on federal roads and over a quarter happened on state roads. It is also estimated that about 2 percent of the country's Gross Domestic Product (GDP), or approximately RM 9 billion, is lost through road accidents. Several models therefore have been developed for modelling the occurrence of accidents, but most of these models have plenty of deficiencies. The following study focuses on stochastic regression models, such as Poisson, Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial with excess zero outcomes on the response variables. Furthermore, R-statistical programming is used to specify the regression relationship with a sophisticated result. The method used consider as updating approach in predicting potential road accidents, which can also produce an accuracy probability of hazardous locations. Based on road accident data collected over a five-year period from 2010 to 2014 at Federal Road F0050: Kluang-A/Hitam-B/Pahat in Johor, Malaysia, results of this study show that Zero Inflated model performed better, in terms of the comparative criteria based on the AIC value.

Keywords: Speed profile, accidents, Poisson method, Federal Road Johor

Copyright © 2019 Society of Automotive Engineers Malaysia - All rights reserved.
Journal homepage: www.journal.saemalaysia.org.my

1.0 INTRODUCTION

Malaysia is ranked among the top 25 most dangerous countries for road users, with 30 fatalities per 100,000 individuals according to research by the University of Michigan. A further study conducted by the Transportation Research Institute with data from the World Health Organization (WHO) on 193 countries indicates that Malaysia is the 17th most dangerous country for drivers as of February 2014 (WHO, 2008). There were 1,646 fatal accidents, 2,173 serious accidents, 3,539 minor accidents, and 65,639 wreckage accidents in Johor state from 2010 until 2013 (RMP, 2014). Seven (7) out of the total 17 locations or KM prone to accidents in Johor are located along F0050 Route, namely Kluang- A/Hitam- B/Pahat (RMP, 2014). This study was carried out to investigate the factors contributing to federal road accidents by using the stochastic (P, NB, ZIP, ZINB) regression models. This study also uses R-statistical programming to specify the regression relationship. Traditionally, simple risk measures such as observed accident counts, rates, costs, and other similar methods have been used to identify locations with the number of accidents exceeding a chosen threshold as “hot spots” (Geurts & Wets, 2003). They are very sensitive to random variation in accident counts and to the regression-to-the-mean problem (Mustakim & Fujita, 2011). Specifically, a targeting method relying on these measurements may produce a large number of misclassifications (e.g., selecting relatively safe locations as hotspots or vice versa) due to the random variation of traffic accidents from year to year (Cheng & Washington, 2005).

Alternatives to statistical modelling approach relating to the traffic safety are used for identification of hazardous locations. This approach is extensively used in traffic safety studies for identifying major contributing factors to crashes and injuries, establishing proper relationships between crashes and explanatory variables, and predicting crash frequency and injury severity. These properties of accident data, along with the fact that counts are nonnegative and discrete, make conventional linear regression with a normally distributed error structure inappropriate for the modelling of accident data. Instead, many research suggested the use of a Poisson or Negative Binomial error structures for modelling the occurrence of traffic (Abdel-Aty & Essam Radwan, 2000; Martin, 2002; Hauer, 2004; Zhang & Ivan, 2005; Xie & Zhang, 2008; Ramírez et al., 2009). These count models are suitable for handling data with a large number of zero counts (zero-inflated models). The Poisson, Negative Binomial, Zero-Inflated models and other related models are collectively called the generalized linear models (GLMs). According to the literature, it is increasingly recognized that the technique of generalized linear models offers the most appropriate method in order to estimate road traffic accidents over the last decade by researchers.

The Negative Binomial or Poisson-Gamma model is an extension of the Poisson model which arises mathematically by assuming that the unobserved crash heterogeneity across sites is gamma distributed while crashes within sites are Poisson distributed (Kim et al., 2006). Due to its effectiveness, researchers have extensively used the Negative Binomial model for predicting accidents at different traffic facilities such as roads, highways, highway ramps, signalized intersections, and unsignalised intersections. Significant studies have been conducted to predict motorcycle crashes and investigate motorized vehicular accidents in Malaysia such as applying Fixed-effects Poisson and Negative Binomial model to the fourteen states in Malaysia (Yaacob et al., 2011), using Generalized Linear Modelling approach to model motorcycle crashes at non-signalized intersections on urban roads in Malaysia (Harnen et al., 2003) and Regression analysis of road accidents at Federal Route 50 (Fajaruddin et al., 2007). However, in some cases, excess zeros in crash data exist and are considered as a result of overdispersion. Therefore, the Poisson model and Negative Binomial (NB) model cannot be

used to handle the overdispersion, which was due to the high amount of zeros. Therefore, Zero-Inflation (ZI) model can be alternatively used. The model was introduced by Lambert (1992), and served as a dual-state method for modelling. In addition, transportation safety analysts have typically justified the use of ZI models because of the improved statistical fit compared to traditional Poisson and Negative binomial models (Shankar et al., 1997; Lee & Mannering, 2002; Shankar et al., 2003). Moreover, the main justification for the use of ZI models has rested on improving the statistical fit compared with traditional Poisson. Despite the fact that the application of ZI model forms did not seem logical, one might argue that this type of model provides improved fit for modelling crash data characterized by a preponderance of zeros especially when no other alternatives readily exists (Lord & Mannering, 2010). The goal of statistical modelling, in general, is to achieve model parsimony, namely to maximize fit while simultaneously minimizing complexity (Lord & Mannering, 2010).

2.0 METHODOLOGY

2.1 FT050 Route: Kluang-A/Hitam-B/Pahat

According to the Royal Malaysia Police headquarters in Bukit Aman, Kuala Lumpur, F0050 was identified as among the most dangerous route in Malaysia based on accident statistics from 2010 to 2014. F0050 is a four lane two-way divided road that runs from Batu Pahat through Ayer Hitam to Kluang. The road has high access density and Annual Average Daily Traffic (AADT) in most kilometres. Route F0050 has registered 9,979 road accidents between the years 2010 to 2014.

2.1.1 Accident Data

Accident data collected from the Royal Malaysia Police were analysed to determine the accident pattern between fatal, serious, minor and wreckage in the research area. With further analysis, the accident data provides more details in order to rank the hazardous locations such as ranking safety performance using accident point weightage. Accident data analysis also provides verification of the overdispersion distribution pattern using overdispersion test.

2.1.2 Site Visit Investigation

The site visit investigation includes site, route and kilometre inspection. This field investigation also involves observation studies, speed studies using GPS and surveying. Pre-analysis observation was conducted to find the possible factors of the accident. The site visit investigation also includes both a drive-over and walk-over inspection.

2.2 Data Collection

2.2.1 Accident Data

The accident data of Route F0050 Kluang-Air Hitam-Batu Pahat were used for modelling accidents on federal roads (Figure 1). Information was collected by the Traffic Police Branch of the Royal Malaysia Police. There were four types of accident provided in the police reports which included fatal, serious, minor and wreckage. In this study, the model applies to a number of accidents with fatality data from kilometre 0 to kilometre 58 (KM 0 – KM 58).

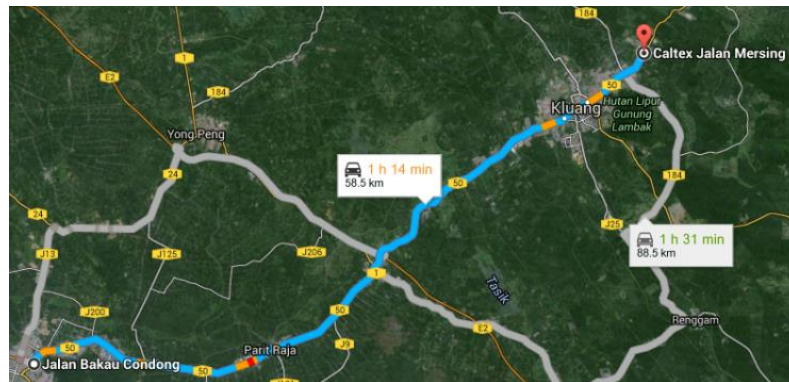


Figure 1: Study location of Federal Route F0050 from km 0 – km 58

2.2.2 Road Features

Road features of route F0050 Kluang-Air Hitam-Batu Pahat were taken as independent variables of the model in this study. The features include shoulder width, lane width, median width and access density.

2.2.3 Traffic Information

Traffic information used in this research included Average Annual Daily Traffic (AADT) and speed of the vehicles. Average Annual Daily Traffic (AADT) was provided by the Ministry of Works Malaysia-Highway Planning Unit-Road Traffic Volume Malaysia for 2010, 2011, 2012 and 2013. The data separately were not readily available by kilometre in the Highway Planning Unit. The organization conducts a bi-monthly study each year at a specific station in April for the first-half data and in October for the other half-year data. Data were collected hourly from 0600 until 2200 hours for every type of vehicles. Another traffic information data was speed data. Speed data were obtained using Global Positioning System (GPS) devices.

Figure 2 shows the pattern of accidents in federal route 50 from 2010 to 2014. There are different trends between fatal accidents, serious, minor and wreckage from 2010 to 2014. The figure for fatal crashes showed an upward and downward trend each year, while serious type indicated a downtrend. Minor accident was at its peak in 2012 while wreckage type accident only showed a decrease in 2011. Although the data presented different patterns, the numbers were mostly on a decline. It can therefore be concluded that the upgrading works on the F0050 route from two-lane road to a four-lane road and a constructed road median have had a positive impact.

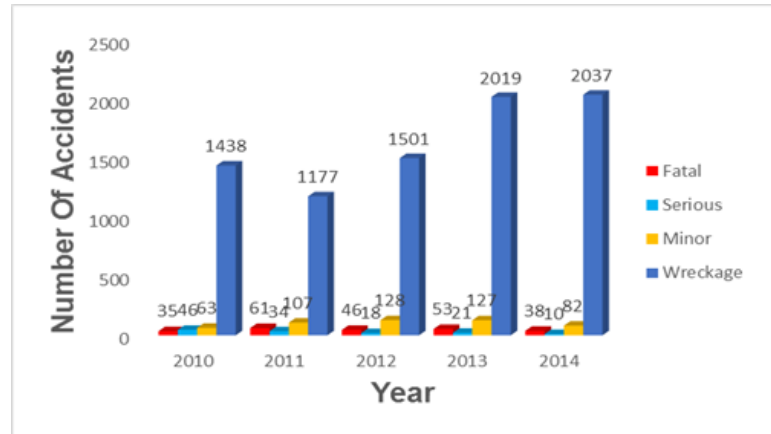


Figure 2: Accident data at F0050 at KM 0 – KM 58 (2010-2014)

2.3 R-studio Programming

Zero-Inflated count data models fit well with the zero-inflation function from the pscl package in R (Zeiles et al., 2008). Hurdle and Zero-Inflated models are able to incorporate over excess zero (Zeiles et al., 2008). Thus, R, accident crashes and statistical modelling are compatible with each other. Table 1 summarizes the model variables used in this study.

Table 1: Summary model variables

	Variable	Type	Description
Dependent	Accident Crashes	Count	Type of accident crashes (Fatal, Serious, Minor, Wreckage)
Independent	Ln AADT	Continuous	Average annual daily traffic per km
Independent	Lane width	Continuous	Lane width (ranging from 3.5m to 3.7m)
Independent	Shoulder width	Continuous	Road shoulder width (ranging from 0 to 2.4m)
Independent	Median width	Continuous	Median width (ranging from 3.5 to 9.0m)
Independent	Access density	Count	Number of intersection and minor access points per km along roadway
Independent	Speed	Continuous	Average speed value from GPS

2.4 Statistical Model

2.4.1 Poisson Regression Model

Poisson regression model is a basic model and is easy to be estimated. Given that the dependent variable is a non-negative integer, most of the recent thinking in the field has used the Poisson regression model as a starting point (Lord & Mannering, 2010). Researchers have found that accident data exhibit characteristics that make the application of the simple Poisson regression problematic even though the Poisson model has served as a starting point for accident frequency analysis for several decades (Lord & Mannering, 2010). Most crash data were over-dispersion due the variance greater than mean and the Poisson cannot handle over-dispersion, under-dispersion and will be negatively influenced by the low sample mean. The data related

to lane width, shoulder width, median width and access density per kilometre are collected from the previous study.

Data were obtained via site observation and with the help of the Google Earth application. The summary of the dependent and independent variables of the models from 2010 to 2014 within 58 sections of federal route F0050 Kluang-A/Hitam-B/Pahat with 1-kilometre length in total is presented in Table 2. The total number of fatal, serious, minor and wreckage accidents taking place in all sections of F0050 from 2010 to 2014 was 9,979.

Table 2: Statistics summary

Variable	Min	Median	Mean	3RD Qu.	Max	Std
Fatal	0.00	3.00	3.25	5.00	10.00	2.41
Ln AADT	7.91	9.43	9.46	9.81	9.81	0.46
Shoulder Width	0.50	1.30	1.47	1.80	2.80	0.54
Lane Width	2.50	3.75	3.52	3.75	3.75	0.42
Median Width	0.50	4.44	2.50	4.44	4.44	1.98
Speed	36.12	85.24	80.15	100.70	140.87	24.59
Access Density	3.00	9.00	9.01	11.00	14.00	2.76

2.4.2 Negative Binomial Regression Model

Negative Binomial was first defined and analysed by Jain and Consul (1971) but was later amended as attribution did not sum to one when the heterogeneity parameter was less than zero. Noticeably, the Negative Binomial fit is preferred over Poisson. Table 3 shows a summary of the correlation between the dependent variable and the independent variable.

2.4.3 Zero-Inflated Poisson Regression Model

The Zero-Inflated Poisson regression model is a mixture of the Poisson model for handling overdispersion arising from excess zero in the accident data. In addition, the advantage of zero-inflated Poisson regression model is that it allows for different sets of variables in model state, the zero state and the count state (Pour et al., 2012). Table 4 shows that the ratio of the variance is 5.81 which is greater to mean as 3.25, indicating that the crash data is over-dispersed and implying the Poisson model does not adequately fit the over-dispersed crash data.

2.4.4 Zero-Inflated Poisson Regression Model

The Zero-Inflated Poisson regression model is a mixture of the Poisson model for handling overdispersion arising from excess zero in the accident data. In addition, the advantage of zero-inflated Poisson regression model is that it allows for different sets of variables in model state, the zero state and the count state (Pour et al., 2012). Further studies of the models used for Federal Roads at state of Johor (Prasetijo & Zainal, 2016a; Prasetijo & Musa, 2016b; Prasetijo et al., 2018). Some findings are also shown in Musa et al. (2017). Table 4 shows that the ratio of the variance is 5.81 which is greater to mean as 3.25, indicating that the crash data is over-dispersed and implying the Poisson model does not adequately fit the over-dispersed crash data.

Table 3: Correlation summary

Variable	Fatal	Ln AADT	SW	LW	MW	SP	AD
Fatal	1						
Ln AADT	-0.003	1					
Shoulder Width (SW)	-0.113	-0.085	1				
Lane Width (LW)	-0.022	-0.088	-0.103	1			
Median Width (MW)	0.047	-0.395	-0.065	-0.437	1		
Speed (SP)	-0.148	0.176	-0.041	0.429	-0.227	1	
Access Density (AD)	0.105	0.318	-0.211	-0.133	-0.055	0.151	1

2.4.5 Zero-Inflated Negative Binomial Regression Model

The zero-inflated negative binomial regression model is a model for handling overdispersion resulting from both excess zero and unobserved heterogeneity in the accident data. The underlying assumption of zero-inflated models is that entities exist in two states (Lord et al., 2007), whereby:

A. True-zero or inherently safe state. In recent years, some have defined it as “virtually safe state” to avoid having to defend the notion that sites can be perfectly safe.

B. Non-zero state, which may happen to record zero accidents in an observation period that follows the Poisson (Zero-Inflated Poisson) or NB (Zero-Inflated Negative Binomial) distribution.

Table 4: Result of over-dispersion existence

Variable	Standard Deviation (std)	Variance	Mean	Over dispersion = [var > mean]
Fatal	2.41	5.813	3.25	Over dispersion

Note: var = variance

Figure 3 shows the result of over dispersion test using coding in R programming. The p-value was really small which means that the data is over-dispersed and we should use negative binomial model.

```

Over-dispersion test

data: poisson
z = 2.2576, p-value = 0.01199
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
0.2000276

```

Figure 3: Results of Over-dispersion Test

3.0 RESULTS AND DISCUSSION

Table 5 shows the results of model evaluation and comparison between the models, namely Poisson (P), Negative Binomial (NB), Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative binomial (ZINB). K is defined as the number of parameters used in the model. The following study is interested in using six (6) parameters for all models.

Table 5: Results of model evaluation and comparison

Models	P	NB	ZIP	ZINB
Log likelihood	-132.82	-254.448	-129.8	-127.2
K (Parameters)	6	6	6	6
N (No. of observation)	58	58	58	58
AIC	277.63	270.45	271.6	266.4
Theta	-	4.634	-	4.635

Note: P = Poisson, NB = negative binomial, ZIP = zero-inflated Poisson, ZINB = zero-inflated negative binomial

The following parameters used are Ln AADT, shoulder width (SW), lane width (LW), median width (MW), speed (SP) and access density (AD). The estimated parameters are the best when the maximum likelihood was obtained according to the definition of the maximum likelihood estimation method. According to the result in Table 5, ZINB is the maximum likelihood and is followed by the ZIP, P and NB. Further evaluation parameter used is the Akaike Information Criteria (AIC). The result shows that AIC ZINB = 266.4, ZIP = 271.6, Negative Binomial = 270.45 and Poisson = 277.63. The smaller the AIC value, the better the model. Based on the result, AIC ZINB is the smallest with 266.4, compared to the other models.

Table 6 shows the model comparison using the Vuong test statistic. The Vuong test statistic compares between each of the models. The tests compare the fitness of ZIP or ZINB model versus P or NB models or between each model. The test has its own standard distribution with three possible outcomes; either an absolute value which is less than 1.96, a large positive value where model 1 is preferred or large negative value where model 2 is preferred. Table 7 shows comparison of the least preferable Poisson models with the other models. It can be seen that Negative Binomial (NB) is the most preferred compared to the other models. Another comparison is made between the ZIP and ZINB models. The Vuong statistics shows that ZINB and NB models were acceptable and better than the ZIP and Poisson models.

Table 6: Comparison Vuong Test of all models

Models	P			NB			ZIP			ZINB		
Comparison with	NB	ZIP	ZINB	P	ZIP	ZINB	P	NB	ZINB	P	NB	ZINB
Raw	-1.51	-0.71	-1.51	1.51	0.94	0.49	0.71	-0.94	-0.94	1.51	-0.49	0.94
AIC corrected	-1.51	-0.36	-1.18	1.51	1.31	4.93	0.36	-1.31	-0.94	1.18	-4.93	0.94
BIC corrected	-1.51	0.02	-0.84	1.51	1.69	1.01	-0.02	-1.69	-0.94	0.83	-1.01	0.94

Note: 1. V is large +ve value, then model 1 preferred. 2. V is large -ve value, then model 2 is preferred

Table 7: Estimation results of Poisson, Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial Regression Models

Models/ Variable	P	NB	ZIP	ZINB
Ln AADT	0.013473	-0.028873	0.173530	-0.028856
Shoulder Width	-0.123252	-0.141208	-0.063156	-0.141201
Lane Width	0.152713	0.150367	0.414435	0.150410
Median Width	0.015727	0.009966	0.096246	0.009976
Speed	-0.005967	-0.006037	-0.005150	-0.006037
Access Density	0.0033782	0.037495	-0.006335	0.037491

Table 8: Results of accident models

Models/ Variable	P	NB	ZIP	ZINB
Ln AADT	0.9454	0.910	0.499	0.9146
Shoulder Width	0.4104	0.463	0.705	0.4691
Lane Width	0.5034	0.619	0.107	0.6344
Median Width	0.7397	0.873	0.114	0.8753
Speed	0.0755	0.175	0.129	0.1646
Access Density	0.2371	0.322	0.852	0.3399

Table 7 shows the estimation result for each model. Instead of the following variables, LW, MW, AD which is positively associated with fatal crash data as it was expected. As in the value in Table 7 and probability results in Table 8, Poisson and Zero-Inflated Poisson model namely AADT, LW, MW and AD are the main parameters that contribute to the fatal crash. However, Negative Binomial and Zero-Inflated Negative Binomial model where the parameters include LW, MW and AD contribute higher impact to the fatal crash data positively.

On the other hand, speed indicated negative signs for all models. The findings show that as NB and ZINB models were selected as the best model where fatal crash increases as AD increases as expected. Furthermore, speed, AADT and SW have negative response to the fatal crashes. This means that most sections with high speed, high AADT and high SW are less likely to result in fatal accidents. One of the reasons for this finding is that from kilometre 0 until kilometre 58 along federal route F0050, each lane has high access density originating from the rural or business areas, especially kilometre 0 to kilometre 7 (KM 0 – KM 7) and kilometre 21 to kilometre 24 (KM 21 – KM 24).

Therefore, high access in each section interferes with the route flow and could increase the probability of crash occurrence due to the increasing number of conflicts. It can be concluded that the high access density results in the low speed of vehicle as many drivers are aware of the many intersections, with vehicles moving in and out and thus reduce their speed. Hence, there is less probability of fatal crashes which means that speed is negatively related to fatal road accidents.

4.0 CONCLUSION

This study has focused on the fatal crashes along the Federal Road in Johor, namely F0050 (federal route 50) from kilometre 0 (KM0) to kilometre 58 (KM58). As federal route 50 is one of the most dangerous with high occurrence of crashes in Johor, the following model is considered as an important approach to develop a model that can properly handle the highly stochastic nature of the crash events and determine the most relevant contributing factors along the road sections. Based on the above results and discussion, NB and ZINB are chosen over Poisson and ZIP regression models. Both NB and ZINB regression models provided better results while AIC value for ZINB and NB was lower, although its likelihood was higher among other regression models.

ACKNOWLEDGEMENTS

The authors would like to thank the Malaysian Institute of Road Safety Research (MIROS) – ASEAN NCAP (New Car Assessment Program) for supporting the study through the ASEAN NCAP Collaborative Holistic Research (ANCHOR) grant, Universiti Tun Hussein Onn Malaysia Contract grant U395 and all individuals, organization that has made this study possible. Thank you are also extended to the Focus Group of Sustainable Transport and Safety Studies (STSS), Department of Transportation Engineering Technology (STSS), Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia.

REFERENCES

- Abdel-Aty, M.A., & Essam Radwan, A., (2000). Modelling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5), 633-642.
- Cheng, W., & Washington, S. (2005). Experimental evaluation of hotspot identification methods. *Accident Analysis & Prevention*, 37(5), 870-881.
- Fajaruddin, M., Ismail, I., Basil, D.D., & Ambak, K. 2007. Accident investigation, blackspot treatment and accident prediction model at federal route F050 Batu Pahat-Ayer Hitam. *Jurnal Sains dan Teknologi*, 4(1), 47-66.
- Geurts, K., & Wets, G. (2003). *Black spot analysis method: Literature review*. RA-2003-07.
- Harnen, S., Radin Sohadi, R.U., Wong S.V., & Wan Ibrahim, W.H. (2003). Development of prediction models for motorcycle crashes at signalized intersections on urban roads in Malaysia. *Research and Innovative Technology Administration (RITA)*, 1-13.
- Hauer, E. (2004). Statistical road safety modelling. *Transportation Research Record: Journal of the Transportation Research Board*, 1897(1), 81-87. <https://doi.org/10.3141%2F1897-11>
- Jain, G.C., & Consul, P.C. (1971). A generalized negative binomial distribution. *SIAM Journal of Applied Mathematics*, 21(4), 501-513.
- Kim, D.G., Washington S., & Oh, J., (2006). Modelling crash type: New Insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering*, 132(4), 282-292.

- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.
- Lee, J., & Mannering, F. (2002) Impact of Roadside Features on the Frequency and Severity of Runoff Roadway Accidents: An Empirical Analysis. *Accident Analysis and Prevention*, 41, 798-808.
- Lord, D., Washington, S.P., & Ivan, J.N. (2007) Further notes on the application of zero- inflated models in highway safety. *Accident Analysis & Prevention*, 39, 53-57.
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291-305. <https://doi.org/10.1016/j.tra.2010.02.001>
- Martin, J.-L. (2002). Relationship crash rate and hourly traffic flow on inter urban motorways. *Accident Analysis & Prevention*, 34(5), 619-629.
- Musa, W.Z., Prasertijo, J., & Zainal, Z.F. (2017). Road fatality model based on over-dispersion data along federal road route F0050. *MATEC Web of Conferences*, 103. <https://doi.org/10.1051/mateconf/201710308012>
- Mustakim, F., & Fujita, M. (2011). Development of accident predictive model for rural roadway. *International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering*, 5(10), 408-413.
- Pour, M.H, Prasertijo, J., Yahaya, A.S., & Ghadiri, S.M.R. (2012). Modelling vehicle-pedestrian crashes with excess zero along Malaysia federal roads. *Procedia-Social and Behavioral Sciences*, 53, 1218-1227.
- Prasertijo, J., & Zainal, Z.F., (2016a). Development of continuous speed profile using GPS at Johor federal roads F0050, *MATEC Web of Conferences*, 47. <https://doi.org/10.1051/mateconf/20164703024>.
- Prasertijo, J., & Musa, W.Z., (2016b). Modeling zero-inflated regression of road accidents at Johor federal road F001, *MATEC Web of Conferences*, 47. <https://doi.org/10.1051/mateconf/20164703001>
- Prasertijo, J., Guntor, N.A.A., Siang, A.J.L.M., Daniel, B.D., & Sanik, M.E. (2018). Change of road integrated design consistency due to antiskid transverse rumble strips on high-speed federal road FT050. *Journal of Advances in Civil Engineering Materials*, 5(1). doi:10.1520/ACEM20170107
- Ramírez, B.A., Izquierdo, F.A., Fernández, C.G., & Méndez, A.G., (2009). The influence of heavy goods vehicle traffic on accidents on different types of Spanish inter-urban roads. *Accident Analysis & Prevention*, 41(1), 15-24. <https://doi.org/10.1016/j.aap.2008.07.016>
- RMP (2014). *Statistical report on annual road accident in Malaysia*. Kuala Lumpur: Royal Malaysia Police (RMP).
- Shankar, V., Milton, J., & Mannering, F.L. (1997). Modelling accident frequency as zero altered probability processes: An empirical inquiry. *Accident Analysis & Prevention*, 29(6), 829-837.
- Shankar, V.N., Ulfarsson, G.F., R.M. Pendyala, R.M., & Nebergall, M.B. (2003) Modelling crashes involving pedestrians and motorized traffic. *Safety Science*, 41(7), 627-640.
- WHO (2008). *World health statistics 2008*. Geneva: World Health Organization (WHO).

- Xie, Y., & Zhang, Y. (2008). Crash frequency analysis with generalized additive models. *Transportation Research Record: Journal of the Transportation Research Board*, 2061, 39-45. <http://dx.doi.org/10.3141/2061-05>
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression model for count data in R. *Journal of Statistical Software*, 27(8), 1-25.
- Zhang, C., & Ivan, J.N. (2005). Effects of geometric characteristics on head-on crash incidence on two lane roads in Connecticut. *Transportation Research Record: Journal of the Transportation Research Board*, 1908, 159-164. <https://doi.org/10.1177/0361198105190800119>