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

Prasetijo, J.*, Musa, W. Z. 2, Mohd Jawi, Z. 3, Zainal, Z. F. 4, N. B. Hamid5, A. Subramaniyan6, Siang, A.J.L.M. 2, Anting, N. 2, Mohd Hafzi Md, I. 3

1,2,5,6,7,8,9 Department of Transportation Engineering Technology, Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia
2,4 Department of Infrastructure Engineering and Geomatics, Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia
3 Vehicle Safety & Biomechanics Research Centre, Malaysian Institute of Road Safety Research (MIROS), Kajang, Selangor, Malaysia

* joewono@uthm.edu.my

Abstract. Traffic accidents have become a major socio-economic problem in Malaysia as it is 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. Previous studies managed to develop several models 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, in order to specify the regression relationship with a sophisticated result, R-statistical programming is used. The method used is also the 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.

1. Introduction
A research by the University of Michigan ranks Malaysia in the top 25 most dangerous countries for road users, with 30 fatalities per 100,000 individuals. A further study conducted by the Transportation Research Institute based on the data of 193 countries by World Health Organization (WHO), indicates that Malaysia is the 17th most dangerous country for drivers as of February 2014 (WHO, 2008).
There were 1646 fatal accidents, 2173 serious accidents, 3539 minor accidents, and 65639 wreckage accidents in Johor state from 2010 until 2013 (RMP). Seven (7) out of the total 17 locations or KM prone to accidents in Johor are located along F0050 Route, namely Kulang- A/Hitam- B/Pahat (RMP). 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, the practitioners used the simple risk measures such as observed accident counts, rates, costs, and other similar methods to identify locations with the number of accidents exceeding a chosen threshold as “hot spots” (Geurts et al., 2003). They are very sensitive to random variation in accident counts and to the regression-to-the-mean problem (Mustakim et al., 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 et al., 2005).

To identify the hazardous locations, the practitioners use the alternatives to the statistical modelling approach relating to the traffic safety. This approach is used extensively 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 et al., 2000; Martin, 2002; Hauer, 2004; Zhang et al., 2005; Xie et al., 2008; Ramirez et al., 2009). These count models are suitable for handling data with a large number of zero counts (zero-inflated models). Theoretically, the generalized linear models (GLMs) category covers models such as the Poisson, Negative Binomial, Zero-Inflated models and other related models. According to the literature, over the last decades, the previous researchers recognized that the technique of generalized linear models offers the most appropriate method in order to estimate road traffic accidents.

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 non-signalized intersections. Previous researchers also conducted a significant amount of studies 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). These are done by using Generalized Linear Modelling approach to model motorcycle crashes at non-signalized intersections on urban roads in Malaysia (Harmen et al., 2003) and Regression analysis of road accidents at Federal Route 50 (Fajaruddin, 2007). However, in some cases, excess zeros in crash data exist and they are the results of over dispersion. Therefore, the Poisson model and Negative Binomial (NB) model are not suitable for handling the over dispersion, which was due to the high amount of zeros. Therefore, Zero-Inflation (ZI) model is an alternative to consider. Lambert (1992) introduced the model and it can serve 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 et al., 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 et al., 2010). The goal of statistical modelling, in general, is to achieve model parsimony, namely to maximize fit while simultaneously minimizing complexity (Lord et al., 2010).

2. Methodology

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

According to the Royal Malaysia Police headquarters in Bukit Aman, Kuala Lumpur, F0050 is identified as among the most dangerous route in Malaysia based on accident statistics from 2010 – 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 9979 road accidents between the years 2010 to 2014.

2.1.1. Accident Data

The author analyzed the data collected from the Royal Malaysia Police 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 kilometer inspection. This field of investigation also involves the observation studies, speed studies using GPS and surveying. The author conducted the pre-analysis observation 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

To model the accidents on federal roads, the author used the accident data of Route F0050 Kluang-Air Hitam-Batu Pahat. The Traffic Police Branch of Royal Malaysia Police is the agency that manage to collect the information. 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 kilometer 0 to kilometer 58 (KM 0 – KM 58).

![Figure 1](image.jpg)

**Figure 1.** Study location of Federal Route F0050 from km 0 – km 58.
2.2.2. Road Features
The author used the road features of route F0050 Kluang-Air Hitam-Batu Pahat as the 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. The Ministry of Works Malaysia-Highway Planning Unit-Road Traffic Volume Malaysia for 2010, 2011, 2012 and 2013 provided the Average Annual Daily Traffic (AADT). The data were not readily available separately by the kilometer 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. The author managed to collect the data for every type of vehicles, in the hourly manner from 0600 until 2200 hours. Another traffic information data was speed data and the author used the Global Positioning System (GPS) devices to obtain these data.

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. As a conclusion, 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. In this study, the independent variables of the model are the Road features of route F0050 Kluang-Air Hitam-Batu Pahat. The features include shoulder width, lane width, median width and access density.

![Figure 2](image)

**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. Summary Model Variables

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

2.4. Statistical Model

2.4.1. Poisson Regression Model

Poisson regression model is a basic model that is easy for estimation. Given that the dependent variable is a non-negative integer and most of the recent thinking in the field used the Poisson regression model as a starting point (Lord et al., 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 et al., 2010). Most crash data fall into the over-dispersion category because of the variance, which are greater than mean. The Poisson is not able to handle over-dispersion, under-dispersion and will influence in a negative way by the low sample mean. The researchers from previous study collected data related to lane width, shoulder width, median width and access density per kilometer.

The site observation, with the help of the Google Earth application managed to obtain the data sufficiently. Table 2 presents 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. The total number of fatal, serious, minor and wreckage accidents taking place in all sections of F0050 from 2010 to 2014 was 9979.

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 analyzed by Jain et al. (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 (Hossein 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 (Hossein et al., 2012). Prasetijo et al., (2016a, 2016b and 2018) presented the further studies of the models used for Federal Roads at state of Johor and Musa et al. (2017) also shown several findings. Several studies relate to probabilistic of pedestrian and vehicles movement behavior are also studied with non-zero inflated (Daniel et al., 2016, Ambak et al., 2016. 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.

| 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 significantly small which means that the data is over-dispersed and the author should use the negative binomial model.

3. 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 represents 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 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 definition of the maximum likelihood estimation method is used to obtain the maximum likelihood. According to the result in Table 5, ZINB is the maximum likelihood and the next in
line are 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 is clear that Negative Binomial (NB) is the most preferred compared to the other models. The author managed to make a comparison 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 with | P | NB | ZIP | ZINB |
|-------------|---|----|-----|------|
| Raw | -1.51 | -0.71 | -1.51 | 1.51 | -0.94 | -0.94 | 1.51 | -0.49 |
| 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. 3. V is large 0 value, then models are equal in preference.

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 | -0.006335 | -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 that 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 whenever the fatal crash increases due to the increasing of AD, the best models are the NB and ZINB models. 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 could increase the probability of crash occurrence due to the increasing number of conflicts. It is clear that the high access density results in the low speed of vehicles because 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 there is no relation between speed and the fatal road accidents.

4. Conclusion
This study 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 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. The results from this study tend to strengthen the selection of the NB and ZINB models over the 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.

Acknowledgments
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 I) grant, Universiti Tun Hussein Onn Malaysia Contract grant U395 and all individuals, organization that has made this study possible. The author would also like to thank 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
[1] World Health Organization, World Health Statistics 2008. World Health Organization (WHO), Geneva.
[2] Royal Malaysia Police (RMP), (2014). Statistical Report on Annual Road Accident in Malaysia.
[3] K. Geurts and G. Wets (2003). Black Spot Analysis Method: Literature Review, RA-2003-07, p.7-13.
[4] F. Mustakim and M. Fujita (2011). “Development of accident predictive model for rural roadway.” International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering. Vol: 5, No: 10.
[5] Cheng, W. and Washington, S., (2005). Experimental Evaluation of Hotspot Identification Methods, Accident Analysis and Prevention, 37, p.870-881.
[6] Abdel-Aty, M.A., E. Radwan, E.A., (2000). Modelling Traffic Accident Occurrence and Involvement, Accident Analysis and Prevention, 32, p.633-642.
[7] Martin, J.-L., (2002). Relationship Crash Rate and Hourly Traffic Flow on Inter Urban Motorways. Accident Analysis and Prevention, 34, p. 619-629.
[8] Hauer, E., (2004). Statistical Road Safety Modelling. In: Proceedings of the 83rd TRB Annual Meeting, Washington, DC, USA, January 11-15.
[9] Zhang, C. and 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, No. 1908, Transportation Research Board of the National Academies, Washington, D.C., p.159-164.
[10] Xie, Y. and Zhang, Y. (2008), Crash Frequency Analysis with Generalized Additive Models. Transportation Research Record: Journal of the Transportation Research Board, No.2061, Transportation Research Board of the National Academies, Washington, D.C., p.39-45
[11] 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 and Prevention 41, p. 15-24.
[12] Kim, D.G., Washington S., and Oh, J., (2006). Modelling Crash Type: New Insights into the Effects of Covariates on Crashes at Rural Intersections. Journal of Transportation Engineering 132, p. 282-292.
[13] Wan Fairos Wan Yaacob, Mohamad Alias Lazim and Yap Bee Wah, 2011. Applying Fixed Effects Panel Count Model to Examine Road Accident Occurrence. Journal of Applied Sciences, 11: 1185-1191.
[14] Harnen, R. S. Radin Umar, S.V. Wong and W.I. Wan Hashim, 2003. Development of Prediction Models for Motorcycle Crashes at Signalized Intersections on Urban Roads in Malaysia. Research and Innovative Technology Administration (RITA).
[15] Fajaruddin, M., Imran Ismail, 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.
[16] Lambert, D., 1992.Zero-inflated Poisson regression, with an application to defects in Manufacturing. Technometrics, 1-14.
[17] Shankar, V., J. Milton, and F.L. Manering (1997) Modelling Accident Frequency as Zero Altered Probability Processes: An Empirical Inquiry. Accident Analysis & Prevention, Vol. 29, No. 6, pp. 829-837.
[18] Lee, J. and Manering, 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.
[19] Shankar, V.N., G.F. Ulfarsson, R.M. Pendyala, and M.B. Nebergall (2003) Modelling Crashes Involving Pedestrians and Motorized Traffic. Safety Science, Vol. 41, No. 7, pp.627-640.
[20] D. Lord and F. Manering, (2010). The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. Forthcoming in Transportation Research, Part A. DOI 10.1016/j.tra.2010.02.001.
[21] A, Zeileis., C, Kleiber. and S, Jackman. (2008). Regression Model for Count Data in R. Journal Of Statistical Software, 27(8).
[22] Jain, G. C., & Consul, P. C. (1971). A Generalized Negative Binomial Distribution. SIAM Journal of Applied Mathematics, 21(4), 501-513.
[23] Lord, D., Washington, S.P. and Ivan, J.N. (2007) Further Notes on the Application of Zero•Inflated Models in Highway Safety. Accident Analysis and Prevention. 39, 53-57.
[24] M. Hossein, J. Prasetijo, A.S. Yahaya and S.M.R. Ghadiri, (2012). Modelling Vehicle Pedestrian Crashes with Excess Zero along Malaysia Federal Roads, Procedia-Social and Behavioral Sciences 53(2012) 1218-1227.
[25] Prasetijo, J., and Zainal, Z. F., (2016a). Development of Continuous Speed Profile Using GPS at Johor Federal Roads F0050, MATEC Web of Conferences, vol. 47,. EDP Sciences, doi: 10.1051/matecconf/20164703024.

[26] Prasetijo, J., and Musa, W. Z., (2016b). Modeling Zero-Inflated Regression of Road Accidents at Johor Federal Road F001, MATEC Web of Conferences, vol. 47,. EDP Sciences, doi: 10.1051/matecconf/20164703001.

[27] Musa, W. Z., Prasetijo, J., and Zainal, Z. F. (2017). Road Fatality Model Based on Over-Dispersion Data Along Federal Road Route F0050, MATEC Web of Conferences, vol. 103,. EDP Sciences, doi: 10.1051/matecconf/201710308012.

[28] Daniel, B.D., Nor, S.N.M., Md Rohani, M., Prasetijo, J., Aman, M.Y., Ambak, K. (2016). Pedestrian Footpath Level of Service (FOOT-LOS) Model for Johor Bahru, MATEC Web of Conferences, vol. 47, EDP Sciences, doi: 10.1051/matecconf/20164703006.

[29] Ambak, K., Kasvar, K.K., Daniel, B.D., Prasetijo, J., Ghani, A.R.A. (2016). Behavioral Intention to Use Public Transport Based on Theory of Planned Behavior, MATEC Web of Conferences, vol. 47, EDP Sciences, doi: 10.1051/matecconf/20164703008.

[30] Prasetijo, 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, ASTM International , 5, 1, ISSN:23791357.