Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 53 (2012) 1218 – 1227

Procedia
Social and Behavioral Sciences

SIIV - 5th International Congress - Sustainability of Road Infrastructures

Modeling vehicle-pedestrian crashes with excess zero along Malaysia federal roads

Mehdi Hossein Pour[a] *, Joewono Prasetijo[a], Ahmad Shukri Yahaya[a], Seyed Mohammad Reza Ghadiri[a]

School of Civil Engineering, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia

Abstract

This study focused on pedestrian crashes occurred in a 543-km sample of Malaysia federal roads in which roadways mostly pass through rural areas. To do this, four count models including Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB) models were developed to establish the relationship between vehicle-pedestrian crashes and the relevant road infrastructure and environmental features. Data on pedestrian crashes were collected for a four-year period from 2007 through 2010. The result indicates that ZIP model performed best, in terms of the comparative criteria.

Keywords: Vehicle-pedestrian crashes, crash prediction models, two-part models, model selection criteria.

1. Introduction

Globally over 1.2 million people are killed and more than 20 million injured in crashes every year. The global economic losses due to road traffic crashes exceed US$ 500 billion [1]. In Malaysia, traffic accident has become a major social-economic problem and considered as a main cause of death. Road accidents result in annual economic cost of 9 billion ringgit to the economy. Every year, over 6,000 people are killed in road accidents nationwide. This amount has been steadily increasing since 2003 and reached 6,872 for the year 2010 while around 21,397 suffer injuries. Meanwhile, the number of pedestrian-involved fatalities was recorded about 626

* Corresponding author. Tel.: +60-174-577-157.
E-mail address: h_mehdi61@yahoo.com

 doi:10.1016/j.sbspro.2012.09.970

for 2010 by representing about 9% of total road user fatalities [2]. To address this problem, the Malaysian transportation agencies and road safety authorities have made considerable efforts to improve the safety of road network and make a safer environment for road users. For example, the government launched the Road Safety Plan 2006-2010 in 2006 to reduce fatality rates by the end of year 2010. Furthermore, the government has approved and implemented 'Road Safety Plan 2011 - 2020' to reduce accident fatalities by 50% in the year 2020 as compared to 2010. To this end, understanding the effects of contributing factors on crashes can assist to develop more effective countermeasures. The suggestion and implementation of cost-effective program requires modeling tools to properly identify the safety performance of road elements and their effect on crash occurrence. In recent years, considerable studies have been conducted to develop crash prediction models to be served as safety performance functions of traffic and geometrics elements [3-8]. Regarding pedestrian-involved crashes, we can refer to some studies conducted by Brüde and Larsson [9], Elvik [10]. As an example, Graham and Glaister [11] examined the effect of urban scale, density and land-use mix on the occurrence of vehicle-pedestrian casualties by developing a spatial model at a disaggregate level. Their findings indicate that the pedestrian casualties is more likely to occur in residential than in economic zones and a quadratic relationship is found between urban density and pedestrian casualties. Lee and Abdel-Aty [12] analyzed vehicle–pedestrian crashes by using log-linear models at intersections in Florida over 4 years, 1999–2002. The study found that pedestrian and driver demographic factors, traffic and environmental characteristics are associated with the frequency and injury level of pedestrian crashes. An ordered probit model was also used to estimate the likelihood of injury severity of pedestrians involving in crashes. Wier, et al. [13] developed simple bivariate models to predict vehicle-pedestrian injury collisions based on changes in traffic volume based on environmental and population data for 176 census tracks in San Francisco, California census tracts. The results show that factors like traffic volume, proportion of neighborhood commercial use, residential–neighborhood commercial use, arterial streets without public transit, employment, resident population, and population below poverty level have a positive correlation while the fraction of people aged 65 and older are negatively associated with pedestrian crashes. Torbic, et al. [14] developed a methodology for identifying the effects of existing road features and proposed improvements on pedestrian safety in urban and suburban arterials. The results showed that daily pedestrian crossing volume tends to have significant relationship to pedestrian crashes at signalized intersections. In addition, the greater the ratio of minor road traffic volume to major road traffic volume, the higher vehicle–pedestrian crash frequency was observed. Ukkusuri, et al. [15] developed models for two area-type levels named census tract level and zip code level for identifying (i) the relationship between land use and road design with pedestrian safety and (ii) the effect of the level of spatial aggregation on the frequency of pedestrian accidents for New York City over a five-year period. The authors concluded that the census tract analysis provides more insightful and better results than the analysis at the zip code level. Their findings also show that tracts with higher proportion of industrial, commercial, and open land use have greater likelihood for crashes while tracts with a greater proportion of residential land use have significantly lower likelihood of pedestrian crashes. This paper focuses on vehicle-pedestrian crashes on Malaysia federal roads where roadways mostly pass through rural areas with low pedestrian activities resulting in a large number of sections with no pedestrian crash which implies a mass of zero crash counts. The objective of this paper is twofold. (i) To develop and investigate a set of crash prediction models for estimating the safety potential of pedestrian activities on the road links. (ii) To examine the effects of various roadway characteristics on the incidence of pedestrian crashes. To accomplish the objectives of this study, Poisson, Negative Binomial (NB), zero-inflated, and zero-inflated negative binomial models were developed and compared using data collected on considered roadways.

2. Methodology

2.1. Count models for pedestrian crashes

Due to the random, discrete, and positive nature of crash data, count-data modeling techniques such as Poisson and negative binomial regression models are used in such cases. The Poisson regression model has been traditionally considered as the starting point in modeling crash data, with the assumption of the mean of crash counts being equal to its variance (that is, equal-dispersion) [16]. In the Poisson regression model, the probability of n_i pedestrian crashes occurring at a given road site *i*, $Pr(n_i|\mu_i)$, can be estimated by the equation:

$$\Pr\,(n_i\|\mu_i) = \frac{\exp\,(-\mu_i)\mu_i^{n_i}}{n_i!} \tag{1}$$

$$\mu_i = \exp\,(\beta_0 X_{i0} + \beta_0 X_{i0} + \beta_0 X_{i0} + \cdots + \beta_0 X_{i0} = exp\left(\sum_{j=1}^{M} \beta_j X_{ij}\right) \tag{2}$$

where μi is the expected number of pedestrian crashes at the given site, X_j is a vector of covariates (road geometric, environmental and traffic information) and β_j is a vector of estimable regression coefficients. However, in much of the crash data, the variance is greater than the mean, well known as over-dispersion. The overdispersion is a result of extra variation in crash means across sites which can be caused by various factors such as model misspecification, omission of important covariates, and excess zero counts. In such a case, applying a Poisson regression model for pedestrian crash data would result in underestimation of the standard error of the regression parameters and narrower CIs, which can lead to a biased selection of covariates [17]. Hence, to this end, a gamma distributed error term was included in the Poisson model to serve as a mixture Poisson-gamma or negative binomial model such that:

$$\mu_i = \exp(\beta X_i + \varepsilon_i) \tag{3}$$

where the $\exp(\varepsilon_i)$ is gamma distributed with mean 1 and variance α (the overdispersion parameter). The NB model takes the unobserved heterogeneity of the Poisson mean into account by allowing the variance to differ from the mean as below:

$$Var[y_i] = E[y_i][1 + \alpha E[y_i] = E[y_i] + \alpha E[y_i]^2 \tag{4}$$

If overdispersion, α, equals 0, the negative binomial reduces to the Poisson model. The NB model assumes that unobserved heterogeneity across road sections follows gamma distribution, while crashes within sites are Poisson distributed. The maximum likelihood function of the binomial regression model is shown by:

$$L(\mu_i) = \prod_i^N \frac{\Gamma(\theta+n_i)}{\Gamma(\theta).n_i!}\left[\frac{\theta}{\theta+\mu_i}\right]^{\theta}\left[\frac{\mu_i}{\theta+\mu_i}\right]^{n_i} \tag{5}$$

where θ is inverse dispersion parameter ($1/\alpha$), $\Gamma(.)$ is a value of gamma distribution, and N is the total number of roadway sections (N=543).

The choice between this negative binomial model and the Poisson model can largely be determined by the statistical significance of the estimated coefficient α. If α is not significantly different from zero, the negative binomial model simply reduces to a Poisson model with $var[n_i] = E[n_i]$. If α is significantly different from zero, the negative binomial model is the correct choice [18].

In some cases, excess zeros in crash data exist and considered as a result of overdispersion. In such a case, the NB model cannot be used to handle the overdispersion which is due to high amount of zeros. To do this, zero-inflation (ZI) models including Zero Inflated Poisson (ZIP) and Zero Inflated Negative Binomial (ZINB) models can be alternatively used. The ZIP model, introduced by Lambert [19], is served as a dual-state method for modeling data characterized by a significant amount of zeros or more zeros than the one would expect in a traditional Poisson or negative binomial model, while the ZINB model, introduced by Greene [20], is a more flexible model that can be used to handle overdispersion caused by both unobserved heterogeneity and excess zeroes which commonly arises in crash data. Both the ZIP and ZINB models assume that all zero counts come from two different processes: (i) the process generating excess zero count (zero-crash-state) derived from a binary model, and (ii) the process generating non-negative counts for pedestrian crashes including zero values, which estimated from the Poisson/NB distribution [21]. As the description of ZIP model, let P_i be the probability of being excess zero for the section *i*, and (1 - P_i) be the probability of crash counts derived from the Poisson distribution. In general, the probability density function for the ZIP model is given below:

$$P(Y = y_i) = \begin{cases} P_i + (1 - P_i)e^{u_i} & y_i = 0 \\ (1 - P_i)\dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} & y_i > 0 \end{cases} \tag{6}$$

where y is the number of pedestrian crashes for section *i* and u_i is the expected crash frequency in section *i* as a function of road section covariates, $u_i = \exp(X_i'\beta)$. The probability of being in the zero-crash-state, P_i, is often fitted using logistic regression model, as follows:

$$logit(P_i) = \ln\left(\frac{P_i}{1 - P_i}\right) = \gamma_0 + \gamma_1 Z_1 + \cdots + \gamma_N Z_N \tag{7}$$

where $Z = (Z_1, Z_2, \ldots, Z_N)$ is a function of some explanatory variables and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_N)$ is the estimable coefficients.

Similar to ZIP model, the probability density function for the ZINB given mean, μ_i, and dispersion parameter, α, is given by:

$$P(Y = y_i) = \begin{cases} P_i + (1 - P_i)\dfrac{1}{(1 + \alpha\mu_i)^{1/\alpha}} & y_i = 0 \\ (1 - p)\dfrac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{\Gamma(y_i + 1)\,\Gamma\left(\frac{1}{\alpha}\right)}\dfrac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{y_i + \frac{1}{\alpha}}} & y_i > 0 \end{cases} \tag{8}$$

The maximum likelihood method is used to estimate the parameters of the ZI models. In general speaking, the probability of observing zero counts in ZI models is the sum of the probability of observing an excess zero in the first process (estimated from logistic regression) and the probability of observing a zero in the second process (estimated from a count model) [22].

2.2. Model development and selection

We apply four count approaches to model the number of pedestrian crashes occurring on the considered roadways. These models are the standard Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) models. The modeling approach is based on a stepwise backward procedure by starting with all variables and removing non-significant variables with *p*-value more than 0.05 at each step.

For comparing the Poisson versus NB and ZIP versus ZINB, since these models are nested together, we test if there is overdispersion due to heterogeneity by testing the significance of dispersion parameter. A significant value for dispersion parameter indicates that the overdispersion in the crash data is due to unobserved heterogeneity, which implies the NB and ZINB models are plausible models to the Poisson and ZIP models, respectively. However, an extra test is needed to check whether excess zero counts may be the cause of overdispersion or not. To do this, since the Poisson and NB are not nested within the ZIP and ZINB models, respectively, we use the Vuong test to compare the non-nested models. Given that $P_1(y_i|x_i)$ and $P_2(y_i|x_i)$ are the predicted probability of the Poisson/NB and PH/NBH, respectively. The Vuong test can be expressed as [23]:

$$V = \frac{\bar{m}\sqrt{n}}{SD(m)} \tag{9}$$

$$m_i = ln\left(\frac{\sum_i P_1(y_i|x_i)}{\sum_i P_2(y_i|x_i)}\right) \tag{10}$$

where $\bar{m}$ is the mean of m_i, $SD(m)$ is the standard deviation of m_i. The Vuong test (V) follows a standard normal distribution. If V is greater than 1.96 then it favors the ZIP/ZINB and V is lower than −1.96 favors the Poisson/NB model. If -1.96 < V < 1.96, it indicate that neither model is preferred over the other.

In addition, we use two information-criteria measures named the Akaike (AIC) and Bayesian (BIC) information criteria to compare both nested and non-nested models. The AIC and BIC are defined as follows:

$$AIC = -2 * LL + 2P \tag{11}$$

$$BIC = -2 * LL + (\ln(n))P \tag{12}$$

where *LL* is the logarithm of maximum likelihood estimation for each model, *P* is number of the model parameters, and n is the number of observation (n=543). A model with the lower AIC and BIC values is preferred.

2.3. Study area and data collection

Malaysia federal road network was selected as the study area of this study. The network accounts for approximately 40% of the fatal accidents reported in the whole country [24]. The selected roadways include five-hundred forty-three sections, each with a fixed length of 1 km, including F2, F3, F4, F67, and F76 located in the states of Perak, Kedah, Kelantan, Pahang, and Terengganu. Most parts of the selected roadways pass through rural areas, and also experiences high accident frequencies. The data used in this study came from three sources: the Malaysia Institute of Road Safety Research (MIROS), the Highway Planning Unit, and the Malaysia Royal Police, as described below:

The first database includes a list of road geometric and roadside characteristics such as curvature, land use, shoulder width, number of lanes. The second database contains traffic data including two variables: average daily traffic (ADT) and average daily heavy vehicle traffic. The third database consists of pedestrian crash data including location and time of pedestrian crashes that occurred on the selected roadways during a 4-year period (2007-2010). The data were obtained from the MIROS provided by the Royal Malaysia Police database (POL. 27). Combining crash data with site-specific roadway geometry and traffic flow data provides us a database that contains four years of pedestrian crash data (2007-2010) for the selected roads. The summary statistics of the roadway sections are shown in Table 1.

Table 1 A summary of some predictor variables

Variable	Type	Description
Traffic flow	Continuous	average daily traffic including all motorized forms of traffic (i.e. motorcyclists, cars, bus, trucks)
Heavy vehicle flow	Continuous	average daily heavy vehicle including bus and truck
Speed limit	Continuous	Actual posted numerical speed limit (ranging from 50 to 90)
No. of lanes	Count	Total number of lanes in the direction of travel
Shoulder width	Continuous	Road shoulder width (ranging from 0 to 2.4 m)
Curvature	Continuous	Horizontal curvature (1/km)
Vertical alignment	Qualitative	Indicator of vertical slope along road's length (1 for flat, 2 for rolling/undulating)
Access point	Count	Number of intersection and minor access points per km along roadway
Land use	Qualitative	Level of activity along roadway (1 for no activity level, 2 for low activity level (e.g., school, factory), 3 for high activity level (e.g., residential or commercial)

3. Results and Discussion

We firstly draw a histogram plot for the pedestrian crash distribution as shown in Fig 1. The figure shows that the number of zeros is 455 (approximately 84%), while there are 88 observations for positive crash counts (approximately 16%). It confirms the presence of excess zeros in pedestrian crashes. Moreover, the ratio of the variance to mean is 1.6 indicating the crash data is over-dispersed implying the Poisson model doesn't adequately fit the overdispersed crash data; then in this case, alternative models such as NB model and ZI models (ZIP and ZINB) can be alternatively used.

In order to assess the source of overdispersion, we developed three models: NB, ZIP and ZINB. As mentioned before, NB model is a superior model for dealing with overdispersion due to unobserved heterogeneity; ZIP model is a mixture Poisson model for handling overdispersion arising from excess zero in the crash data; and ZINB is a model for handling overdispersion resulting from both excess zero and unobserved heterogeneity in the crash data. We summarized some results of considered models in Table 2. The results show that, in the NB and ZINB model, the dispersion parameter, α, is not statistically significant (*p*-value is 0.39 and 0.27 for the NB and ZINB models, respectively) which implies overdispersion is not due to unobserved heterogeneity in the crash data, but it may be caused by excess zero. To confirm this, we compare the Poisson and NB versus ZIP model by Vuong test to determine if the over-dispersion parameter is significantly due to excess zero. The ZINB model was excluded because the dispersion parameter estimate was not statistically significant. The Vuong statistic for the Poisson versus ZIP (1.78, *p*-value = 0.037) and NB versus ZIP (1.61, *p*-value =0.05) favors the ZIP model. The test results indicate that over-dispersion in crash data is due to excess zeroes rather than unobserved heterogeneity.

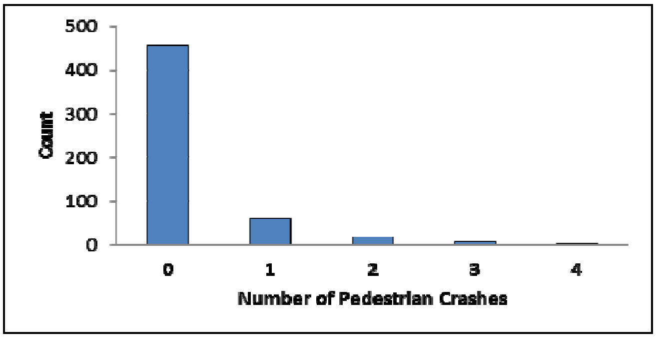

Fig. 1 Histogram of observed pedestrian crashes for the 543 sections during 2007-2010

Table 2 also shows the actual and predicted count categories of pedestrian crashes for all models. The comparison for the Poisson model reveals the Poisson model is not an appropriate model for predicting crashes with excess zero counts. On the other hand, for the remaining models, the predicted frequencies are nearly close to the observed ones, especially for the ZIP and ZINB models which outperform the NB model for all count categories. Next we compare all developed models using AIC and BIC. These information-based criteria favor the ZIP model over all other models. As a result, we select the ZIP model as the best among the four models and tabulate its results in Table 3. Results for the Poisson, NB, and ZINB are not presented, since the ZIP model was found to be preferred over the Poisson, NB, and ZINB models. In the count state, logarithm of ADT (Ln(ADT)), logarithm of heavy vehicle traffic (Ln(heavy vehicle)), speed limit, and land use were found to be significant. Out of these variables, "Ln(ADT)" and "Land use" are positively associated with pedestrian crashes, as expected. On the other hand, "Speed limit" and "Ln(heavy vehicle)" have a negative sign.

The findings show that, PCs increase as traffic flow increase, as expected. Logarithm of heavy vehicle traffic has a negative impact on the PCs; it means that sections with higher heavy traffic flow are less likely to have PCs. One reason for this finding is that heavy vehicle with low speed lead the following vehicles to reduce their speed and subsequently lessen the probability of other vehicles to hit pedestrians. The results show that road sections with higher speed limits tend to have lower probability of having pedestrian crashes. It means that speed limit is negatively related to pedestrian crashes. This is expected, because the roads with higher speed limits have safer designs and environmental conditions, while lower speed limits are assigned to those road sections with unsafe design or condition. This result is consistent with the previous studies [25, 26]. However, higher amount of speed limit should not be interpreted as a measure to reduce PCs. The factor land use has a positive effect on the PCs. Residential/commercial areas, as a high level of activity, are significantly associated with PCs, while areas with lower activity appear not to be significantly associated with pedestrian crashes (with *p*-value 0.38). As a result, land use can served as a surrogate variable to pedestrian exposure, whereas data regarding pedestrian activities was not available in the dataset. In the zero state (logistic model), the probability of a section having no-crash is dependent on only one variable, namely Ln(ADT). In other words, the probability of having at least one crash increases as traffic flow increase. A advantage of ZIP as a two-state model is that it allows for different sets of variables in both model states, the zero state and the count state, and thus for distinguishing factors affecting those two parts.

Table 2 The actual and predicted frequencies of pedestrian crashes by developed models and their goodness-of-fit results

Count	Actual	Poisson	NB	ZIP	ZINB
0	455	444	452	452	453
1	61	78	69	65	64
2	17	16	15	18	18
3	8	4	5	6	6
4	2	1	2	2	2
Parameters		6	7	8	9
Log-likelihood		-279.84	-276.77	-271.8	-271.5
α *p*-value		-	0.39	-	0.27
AIC		571.69	567.54	559.6	561
BIC		597.47	597.62	594	599.7

Note: NB = negative binomial, ZIP = zero-inflated Poisson model, ZINB = zero-inflated negative binomial model, "χ^2 *p*-value" = corresponding p-value for Pearson's goodness of fit measure, and "α *p*-value" = corresponding p-value for dispersion parameter for the NB and ZINB models.

Table 3 Estimation of zero-inflated Poisson regression model of pedestrian crashes

Variable		Coefficient	*p*-value
Count state as Poisson model			
Constant		4.023	0.105
Ln(ADT)		0.442	0.156
Ln(Heavy Vehicle)		-0.885	0.005
Speed limit		-0.037	1.24e-05
Land use	1: No activity	R	-
	2: Low activity	0.240	0.377
	3: High activity	0.793	0.003
Zero state as logistic model			
Constant		26.524	0.001
Ln(ADT)		-2.932	0.001
Number of observation		543	
Log-likelihood at convergence		-271.8	
Vuong test (ZIP vs Poisson)		1.78	
(ZIP vs NB)		1.61	

4. Conclusion

This paper aimed to focus pedestrian crashes on 543-km sections of Malaysia federal roads. The selected roadways pass mostly through rural areas in which the level of pedestrian activities is very low and rarely occurs. It implies an expectation of considerable number of sections with no pedestrian crashes during the study period. For modeling count data in the presence of excess zeros, as an important cause of overdispersion, standard count models like Poisson and NB models are found to be misspecified. As an alternative, ZIP model can be used to handle over-dispersion arising only from excess zeroes. First, we developed a NB model to test whether the dispersion parameter in the crash data is due to unobserved heterogeneity or not. The result showed that dispersion parameter was statistically rejected at 0.05 significance level implying overdispersion in crash data may be resulted from excess zeros. To do this, the Vuong test was conducted for two pairs: Poisson versus ZIP and NB versus ZIP. The test gave an advantage to the ZIP model. Moreover, the ZIP model was found to be the most suitable model, since AIC, BIC values were smallest in comparison to the considered models (Table 2). This paper also examined factors influencing on PCs. The factors total traffic volume, heavy traffic flow, speed limit, and land use appear to increase probability of the count process. For the dichotomous process, only traffic flow was found to be positively associated to the probability of at least one PC. To conclude, high amount of zeros in pedestrian crashes should not be ignored for rural roads since they result from sites with no pedestrian activities; and to handle the issue, a zero-inflated Poisson model can be proposed.

Acknowledgements

The authors would like to thank the Malaysia Institute of Safety Research and the Highway Planning Unit for the data used in this research. This study is a part of a research project sponsored by the Universiti Sains Malaysia. The authors acknowledge the support and assistance.

References

[1] WHO (1999). The world health report 1999: Making a difference. World Health Organization.

[2] RMP (2010). Statistical Report on Road Accidents in Malaysia.

[3] Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering, 122, 105-113.

[4] Karlaftis, M. G., & Golias, I. (2002). Effects of road geometry and traffic volumes on rural roadway accident rates. Accident Analysis & Prevention, 34, 357-365.

[5] Chin, H. C., & Quddus, M. A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis & Prevention, 35, 253-259.

[6] Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. Accident Analysis & Prevention, 39, 657-670.

[7] Park, B. J., & Lord, D. (2009). Application of finite mixture models for vehicle crash data analysis. Accident Analysis & Prevention, 41, 683-691.

[8] Dinu, R. R., & Veeraragavan, A. (2011). Random parameter models for accident prediction on two-lane undivided highways in India. Journal of Safety Research, 42, 39-42.

[9] Brüde, U., & Larsson, J. (1993). Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit? Accident Analysis and Prevention, 25, 499-509.

[10] Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. Accident Analysis and Prevention, 41, 849-855.

[11] Graham, D. J., & Glaister, S. (2003). Spatial variation in road pedestrian casualties: The role of urban scale, density and land-use mix. Urban Studies, 40, 1591-1607.

[12] Lee, C., & Abdel-Aty, M. (2005). Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. Accident Analysis and Prevention, 37, 775-786.

[13] Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis and Prevention, 41, 137-145.

[14] Torbic, D. J., Harwood, D. W., Bokenkroger, C. D., Srinivasan, R., Carter, D., Zegeer, C. V., & Lyon, C. (2010). Pedestrian safety prediction methodology for urban signalized intersections. 65-74.

[15] Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. Safety Science, 50, 1141-1151.

[16] Miaou, S. P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis and Prevention, 26, 471-482.

[17] Miranda-Moreno, L. F., & Fu, L. (2006). A comparative study of alternative model structures and criteria for ranking locations for safety improvements. Networks and Spatial Economics, 6, 97-110.

[18] Kim, D. G., Washington, S., & Oh, J. (2006). Modeling crash types: New insights into the effects of covariates on crashes at rural intersections. Journal of Transportation Engineering, 132, 282-292.

[19] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 1-14.

[20] Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. NYU Working Paper No. EC-94-10,

[21] Welsh, A., Cunningham, R. B., Donnelly, C., & Lindenmayer, D. B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling, 88, 297-308.

[22] Rose, C. E., Martin, S. W., Wannemuehler, K. A., & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of Biopharmaceutical Statistics, 16, 463-481.

[23] Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society, 307-333.

[24] IRTAD (2010). International Road Traffic and Accident Database (IRTAD) Annual Report 2010.

[25] Qin, X., Ivan, J. N., Ravishanker, N., & Liu, J. (2005). Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov Chain Monte Carlo modeling. Journal of Transportation Engineering, 131, 345.

[26] 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.