Accident Analysis and Prevention 62 (2014) 209–222

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads

Mehdi Hosseinpour, Ahmad Shukri Yahaya*, Ahmad Farhan Sadullah

School of Civil Engineering, Universiti Sains Malaysia, USM Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia

ARTICLE INFO

Article history:
Received 13 March 2013
Received in revised form 1 October 2013
Accepted 1 October 2013

Keywords:
Head-on crashes
Count-data models
Injury severity
Ordered probit model
Random effects

ABSTRACT

Head-on crashes are among the most severe collision types and of great concern to road safety authorities. Therefore, it justifies more efforts to reduce both the frequency and severity of this collision type. To this end, it is necessary to first identify factors associating with the crash occurrence. This can be done by developing crash prediction models that relate crash outcomes to a set of contributing factors. This study intends to identify the factors affecting both the frequency and severity of head-on crashes that occurred on 448 segments of five federal roads in Malaysia. Data on road characteristics and crash history were collected on the study segments during a 4-year period between 2007 and 2010. The frequency of head-on crashes were fitted by developing and comparing seven count-data models including Poisson, standard negative binomial (NB), random-effect negative binomial, hurdle Poisson, hurdle negative binomial, zero-inflated Poisson, and zero-inflated negative binomial models. To model crash severity, a random-effect generalized ordered probit model (REGOPM) was used given a head-on crash had occurred. With respect to the crash frequency, the random-effect negative binomial (RENB) model was found to outperform the other models according to goodness of fit measures. Based on the results of the model, the variables horizontal curvature, terrain type, heavy-vehicle traffic, and access points were found to be positively related to the frequency of head-on crashes, while posted speed limit and shoulder width decreased the crash frequency. With regard to the crash severity, the results of REGOPM showed that horizontal curvature, paved shoulder width, terrain type, and side friction were associated with more severe crashes, whereas land use, access points, and presence of median reduced the probability of severe crashes. Based on the results of this study, some potential countermeasures were proposed to minimize the risk of head-on crashes.

1. Introduction

Every year, more than 400,000 crashes occur in Malaysia, leading to approximately 7000 deaths and as much as nine billion ringgit in losses to the country's economy (RMP 2011). Of these crashes, head-on collisions account for about 12% of all reported crashes, while they are responsible for nearly 17% of fatal crashes (IRTAD 2011). A head-on collision occurs when a vehicle crosses a centreline or median either intentionally or unintentionally and collides with an opposing vehicle with an impact angle of zero (Gårder, 2006). As stated in other studies (Cerrelli, 1997; Abdelwahab and Abdel-Aty, 2004; Zhang and Ivan, 2005; Conroy et al., 2008; Bham et al., 2012), head-on crashes are among the most severe collision types and are of great concern to road safety authorities. For example, Wegman (2004) reported that head-on crashes are responsible for nearly 25% of fatal crashes occurring on rural roads in OECD member countries. According to U.S. statistics on traffic accident fatalities for the year 2005, head-on crashes comprised only 2% of total crashes, but they accounted for 10.1% of fatal crashes (NHTSA 2007). These statistics justify the need for increased efforts to reduce both the frequency and the severity of this collision type by implementing cost-effective countermeasures. To do so, it is necessary to first estimate the safety effects of roadway elements on the occurrence of head-on collision, which can be performed by developing crash prediction models that relate crash outcomes to a set of roadway geometric and environmental characteristics (Kumara and Chin, 2003). The outcomes of these models can assist transportation engineers and researchers in reducing the number of crashes. Among other techniques, statistical models are widely used to determine the critical factors associated with crash occurrence and to identify accident prone-locations. In this domain, there is a large body of research that has focused on road safety

* Corresponding author. Tel.: +60 04 5996270/+60 012 5011718.
E-mail addresses: shukri@eng.usm.my, shukriyahaya@gmail.com (A.S. Yahaya).

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.10.001

modelling. With respect to head-on collisions, a number of studies have attempted to isolate factors that are significantly associated with the crash occurrence. For example, Abdelwahab and Abdel-Aty (2004) evaluated the effect of increased percentage of light truck vehicles (LTVs) on head-on fatal crashes. Time series models were used to forecast the future fatality trends of head-on crashes based on a crash database obtained from the Fatality Analysis Reporting System (FARS) over the period of 1975–2000. The researchers forecasted that the annual deaths in head-on crashes would have increased over a 10-year period since the year 2000, and would have reached 5324 by 2010, representing an 8% increase since the year 2000. Overall, the modelling results showed that head-on fatal crashes were affected by the increased percentage of LTVs in traffic.

Zhang and Ivan (2005) assessed the effects of roadway geometric characteristics on the frequency of head-on crashes on two-lane rural roads in Connecticut. A negative binomial regression model was applied to estimate the crash outcomes as a function of road traits. Based on the modelling results, five variables were found to be significantly associated with head-on crashes: average annual daily traffic (AADT), speed limit, sum of absolute change rate of horizontal curvature, maximum degree of horizontal curve, and sum of absolute change rate of vertical curvature. Of these factors, posted speed limit and AADT were negatively associated with head-on crashes while the others had a positive effect. Gårder (2006) analysed head-on crashes that occurred on two-lane rural highways in Maine during the period 2000–2002. The analysis results showed that most of the head-on crashes were attributed to driver error and violations such as distraction/inattention, excessive speeds, fatigue, and alcohol/drug use. In addition, higher speed limits, more travel lanes, wider shoulder widths, and higher AADT were found to contribute to crash severity. Ye et al. (2009) developed a multivariate Poisson regression approach to model crash frequencies by collision type using crash data for 165 rural intersections in Georgia. The modelling results showed that posted speed limit and traffic volume on both the major and minor roads had a positive effect on the number of head-on crashes. Using crash data for rural, two-lane highways in Minnesota, Geedipally et al. (2010) estimated the proportion of crashes by collision type that occurred during the 5-year period between 2002 and 2006. The results showed that head-on crashes were affected by AADT, truck percentage, and shoulder width. Bham et al. (2012) analysed single and multivehicle collisions using data collected on urban highways in Arkansas from 2005 to 2007. A multinomial logistic regression model was applied to determine the impacts of several factors on crash outcomes for six collision types. Slowing or stopping and driving under the influence of alcohol were found to be significantly associated with head-on collisions. In addition, the authors noted that head-on collisions contribute to a higher risk of severe injuries compared to other crash types.

However, the number of studies focusing on head-on crashes is still limited when compared to those devoted to total or other collision types. This may be due to the lack of detailed data and the rarity of head-on crashes that, together, could be a barrier to accurately identifying the potential causes of crash outcomes. This study attempts to fill that gap and extends the related literature by analyzing head-on crashes. The objective of this study is to evaluate the effects of various roadway geometric design, the environment, and traffic characteristics on the frequency and the severity of head-on collisions. Seven count-data regression models including Poisson, standard negative binomial (NB), random-effect negative binomial (RENB), hurdle Poisson (HP), hurdle negative binomial (HNB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB) regression models were developed and compared to model the crash frequency. A random-effect generalized ordered probit model (REGOPM) was applied to model the severity, given that a head-on crash had occurred. To accomplish the objective of this study, 4 years of crash records (2007–2010) and data on the roadway geometry, and environmental and traffic attributes were collected on 543 km of Malaysia federal roads on which the rate of traffic crashes are considerably higher than on other road types (e.g., expressways, state roads, and municipal roads). This paper is organised as follows: the second section describes the case studies and characteristics of the data used. The third section explains the methodology applied to model the head-on crash frequency and severity. The results of the comparison analysis and an interpretation of the parameter estimation are then presented in the fourth section. The fifth and final section summarizes the findings and presents the conclusion.

2. Study area and data collection

This study is a part of more comprehensive research conducted to investigate the impacts of road site-specific characteristics on traffic crashes by collision type (e.g., head-on, rear-end, pedestrian, etc.). As noted earlier, Malaysian federal road network was selected as the study area. This road system is among the most important transport systems throughout the country. However, it experiences the highest rate of traffic crashes compared to other road types where it accounts for over 40% of all accident fatalities nationwide (IRTAD 2011). Among the candidate federal roads, those finally selected were on conditioned that data on roadway characteristics, traffic flow, and crash information were available and complete. Based on the conditions, the study area finally consisted of 543 km sections from five federal roads including Malaysia federal road 2 (F2), Malaysia federal road 3 (F3), Malaysia federal road 4 (F4), Malaysia federal road 67 (F67), and Malaysia federal road 76 (F76) located in the states of Perak, Kedah, Kelantan, Pahang, and Terengganu in Peninsular Malaysia. The road segments considered are representative sample to Malaysia federal road system in which most of its roadways pass through rural and semi-urban areas.

To investigate the effects of road geometric, environmental, and traffic characteristics on head-on crash frequency and severity, required data were collected from three sources: Malaysia Institute of Road Safety Research (MIROS), Highway Planning Unit (HPU), and Royal Malaysia Police (RMP). The first database, which obtained from the MIROS, originally collected from the Malaysian Pilot study by International Road Assessment Programme (IRAP) carried out in 2007. The database includes a list of road geometric and environmental characteristics such as curvature, land use, shoulder width, number of lanes, etc. The second database, collected from the HPU, contains traffic data for a 4-year period from 2007 to 2010, including average daily light-vehicle traffic (LVT) and average daily heavy-vehicle traffic (HVT). The third database consists of crash data including location and time of crashes occurred on the considered segments between 2007 and 2010. The data were collected from the MIROS based on data originally provided by the Royal Malaysia Police crash database. Merging these three databases provided the study with a data set containing four years (2007–2010) of crashes along with site-specific road and traffic characteristics for the considered road segments. More than one year of crash records was used to reduce the variability of the crash frequency from year to year. With these data at hand, the next step is to divide the study area into homogeneous segments. To do this, the study sections were split into homogeneous segments in terms of traffic flow, land use, and cross-sectional characteristics such as shoulder width, the number of lanes, and median. After the segmentation process, the 543 km sections were segregated into 448 homogeneous segments with the length ranged between 1 km and 7 km, and an average of 1.2 km. For a specific variable on each segment, the characteristic with the largest proportion was determined as the

Table 1
Descriptive statistics of continuous variables.

Variable	Description	Min.	Max.	Mean	Std. Dev.
Head-on crash frequency	The number of head-on crashes occurring on the study segments during the 4-year period (2007–2010)	0	8	1.176	1.483
Segment length	The length of segment (km)	1	7	1.21	0.690
LVT	Average daily traffic including light vehicles (e.g., motorcycles, passenger cars, light vans, SUVs, etc.)	2790	27470	9294	6563
HVT	Average daily heavy vehicle traffic including bus, tractor, lorry, large van, truck	456	4789	1335	659
Speed limit	Actual posted speed limit (ranging from 50 to 90 km/h)	50	90	82	11.9
Paved shoulder width	Paved shoulder width (ranging from 0 to 2.4 m)	0	2.4	1.26	0.59
Unpaved shoulder width	Unpaved shoulder width (ranging from 0 to 2.4 m)	0	2.4	1.15	0.71
Curvature	Horizontal curvature (1/km)	0.091	13.386	3.090	2.682
Access point	Number of intersection and minor access points per km	0	9	1.08	1.51

Table 2
Descriptive statistics of categorical variables.

Variable	Description	Observation (proportion) in Sample		
		1	2	3
Area type	Level of roadside development (1 for rural, 2 for semi-urban, 3 for urban)	388 (87%)	46 (10%)	14 (3%)
Land use	Level of activity along roadway (1 for no activity level, 2 for low activity level (e.g., educational, industrial), 3 for high activity level (e.g., residential or commercial)	279 (62%)	103 (23%)	66 (15%)
No. of lanes	Number of lanes for each travel direction (1 for one-lane, 2 for two and more lanes)	397 (89%)	51 (11%)	– –
Terrain type	Indicator of vertical gradient along roadway (1 for flat terrain, 2 for undulating terrain)	320 (71%)	128 (29%)	– –
Side friction	Level of interaction between roadside activities (e.g., parking, bus stopping, trading) and through traffic (1 for low or non-interaction, 2 for high interaction)	377 (84%)	71 (16%)	– –
Median type	Indicator of two opposing traffic flows are separated or not (1 for unseparated, 2 for separated)	417 (93%)	31 (7%)	– –

representative characteristic of that variable within the segment. Descriptive statistics of the data are presented in Tables 1 and 2 for continuous and categorical variables, respectively.

During the 4-year period, 11,099 crashes occurred on the study segments, resulting in 1421 injuries (slight & severe) and 472 fatalities. Fig. 1 shows proportions of crash frequency and fatality by collision type on the considered segments. As seen in the figure, rear-end crashes are the most frequent collision type ($N_{RE} = 3005$, 27% of the total), yet account for less than 10% of fatal crashes. Angular crashes are responsible for nearly 18% of the total crashes, but contribute to 24% of the fatal crashes. On the other hand, head-on crashes accounted for only 5% of the total crashes, while they comprised approximately one-thirds of all fatal crashes. The reason for over-representation of head-on crashes compared to other collision types is that the relative speed of vehicles approaching from opposite directions is high. In such conditions, drivers do not have sufficient time to avoid a collision with an oncoming vehicle; hence they are more likely to be involved in severe injuries. As a result, head-on crashes are the most fatal type of collision among others. In the case of this study, a total of 527 head-on crashes occurred on the considered segments. Of these, 225 (43%) resulted in no injury, 71 (13%) involved slight injury, 87 (17%) involved serious injury, and 144 (27%) turned out to be fatalities. The distribution of head-on crashes is shown in Fig. 2. The figure illustrates that there are 194 segments for which no head-on crashes occurred during the 4-year period. This indicates potentially the presence of excess zeros in the crash data. In addition, the observed variance and mean are 2.20 and 1.18, respectively. The presence of overdispersion can be assessed by calculating the ratio between the variance and mean of the crash data. The ratio was found to be 1.86, implying that the crash data is overdispersed.

3. Methodological approach

Crash data are typically classified according to injury severity (e.g., slight, serious, fatal) or collision type (e.g., head-on, sideswipe, etc.). There are many studies in the literature that have modelled the crash frequency and severity either independently or simultaneously. A traditional and common approach to model crash frequency by severity or collision type is to develop univariate models, such as Poisson or negative binomial models, to predict the number of crashes for different categories independently (Shankar et al., 1995; Hauer et al., 2004; Noland and Quddus, 2005; Qin et al., 2005; Kim et al., 2006; Jonsson et al., 2007). However, as stated by other researchers (Ma and Kockelman, 2006; Park and Lord, 2007; Ma et al., 2008; El-Basyouny and Sayed, 2009; Ye et al., 2009), a critical drawback to using univariate modelling approaches is that they cannot consider a possible correlation that may exist among crash categories. Neglecting to account for such correlations may reduce precision of parameter estimates and model efficiency (Aguero-Valverde and Jovanis, 2009). To address this problem, some other researchers have recently applied multivariate modelling approaches which model crash counts by collision type or severity simultaneously rather than separately. These models include multivariate Poisson (MVP) models (Ma and Kockelman, 2006; Ye et al., 2009), multivariate negative binomial (MVNB) models (Rodgers and Leland, 2005; Anastasopoulos et al., 2012), and multivariate Poisson log-normal (MVPL) models (Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009).

Nevertheless, despite the effectiveness of multivariate models to capturing correlations among crash categories, they still exhibit some drawbacks. One of their limitations is that they require the

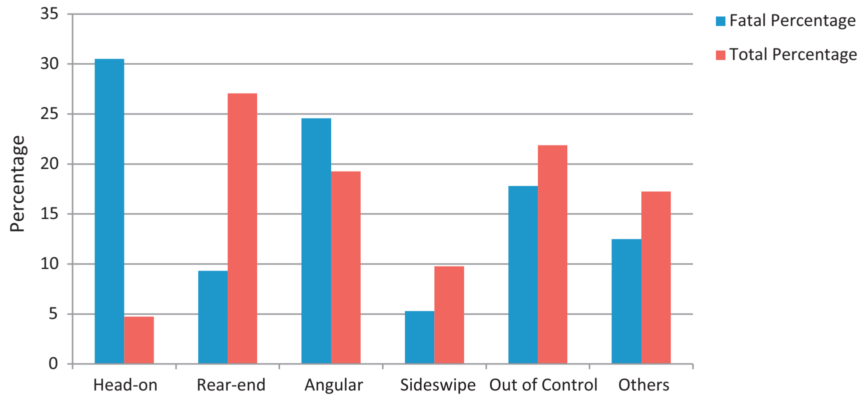

Fig. 1. Percentage of crash fatality and frequency by collision type on the study segments.

same set of covariates for all response categories. However, a specific explanatory variable influencing a certain severity level does not necessarily influence the other ones (Wang et al., 2011). Multivariate models have rarely been applied in the literature mainly because their process for estimating the model parameters is not straightforward; rather, an extensive computational procedure is required to formulate the model. In addition, no built-in functions are available in most of the existing statistical software to fit such models (Park and Lord, 2007).

A major concern for both univariate and multivariate models is that they fail to correctly predict crash frequencies by category in which the crash data are characterized by small sample size and low sample mean (Lord, 2006; Lord and Mannering, 2010; Wang et al., 2011). More specially, by disaggregating the total crashes into different collision types (e.g., head-on, rear-end, etc.), it is more likely that there will be many road segments for which no crashes are observed as the number of crashes is modelled for each collision type by different severity levels either independently or simultaneously. In such cases, discrete choice models (e.g., multinomial or ordered models) can be alternatively applied to model the crash severity where the number of crashes can be initially fitted by univariate count models such as Poisson or NB models (Geedipally et al., 2010; Bham et al., 2012). This approach is known as a two-stage technique. In this context, some researchers have modelled the crash frequency and severity separately (Carson and Mannering, 2001; Lee and Mannering, 2002; Wang et al., 2011; Qi et al., 2013).

As a result of above discussion, this study used a two-stage approach to evaluate the effects of various segment-specific characteristics on the frequency and severity of head-on crashes. First, the crash frequency was modelled by developing and comparing a set of count-data models to predict the number of head-on crashes that occurred on the study segments. Next, the injury severity of the crashes was fitted using a REGOPM. More details on both approaches are presented in the following sections.

3.1. *Modelling the head-on crash frequency*

3.1.1. *Standard count models*

Count-data models are generally used to model traffic crashes that are discrete, random, and non-negative integers. The Poisson regression model is considered as the starting point in modelling count data (Miranda-Moreno and Fu, 2006; Khan et al., 2011). The Poisson distribution requires the variance of the count data is equal to the mean. However, in most of the crash data, the variance is greater than the mean, known as over-dispersion. The overdispersion is a result of extra variation in crash counts across road segments and could result form various factors such as omission of important covariates, model misspecification, and excess zero counts (Mitra and Washington, 2007). In such a case, applying a Poisson regression model would lead to underestimate the standard error of the parameters, causing a biased selection of parameters (Khan et al., 2011). A common approach to deal with extra variation in crash data is to apply a negative binomial regression model. The NB model accommodates the overdispersion by including an error term, $\exp(\varepsilon_i)$, in the Poisson model and hence allows the variance to differ from the mean as such:

$$\mu_i = \exp(\beta X_i + \varepsilon_i) = \exp(\beta X_i) \cdot \exp(\varepsilon_i) \quad (1)$$

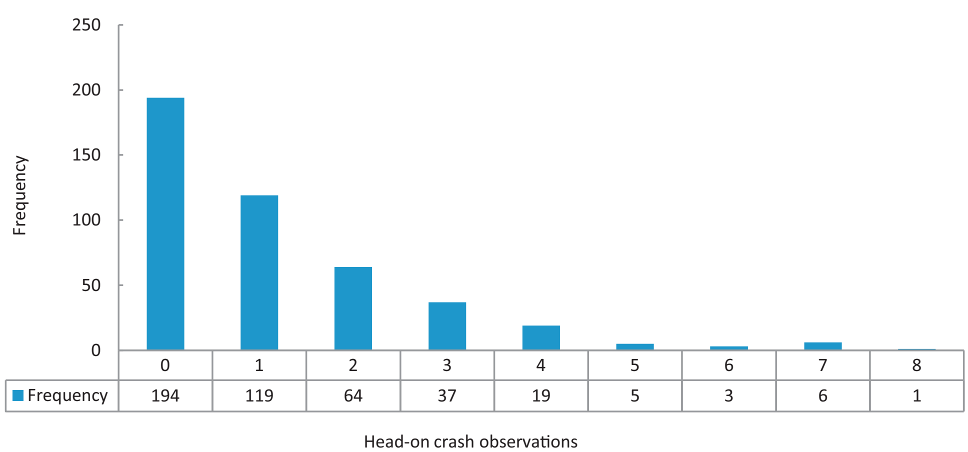

Fig. 2. Distribution of head-on crashes on the study segments during the 4-year period.

$$Var[y_i] = E[y_i] \cdot (1 + \alpha \cdot E[y_i]) = E[y_i] + \alpha \cdot E[y_i]^2 \quad (2)$$

where X_i is a vector of covariates for segment i, and β_j is a vector of estimable regression coefficients, y_i and μ_i are the observed and predicted number of head-on crashes on road segment i, respectively, α is dispersion parameter, $\exp(\varepsilon_i)$ is gamma-distributed error term with mean 1 and variance α.

The probability of y_i, $\Pr(Y_i = y_i)$, can be estimated by Eq. (3):

$$Pr(Y_i = y_i) = \frac{\Gamma(y_i + (1/\alpha))}{\Gamma(1/\alpha) y_i!} \left(\frac{\mu_i}{\mu_i + (1/\alpha)} \right)^{y_i} \left(\frac{(1/\alpha)}{\mu_i + (1/\alpha)} \right)^{1/\alpha} \quad (3)$$

where $\Gamma(\cdot)$ is gamma function.

The parameters of the NB model are estimated by standard maximum likelihood methods through maximizing the logarithm of likelihood function (Eq. (3)) as given by:

$$L(\mu_i) = \prod_i^N \frac{\Gamma(y_i + (1/\alpha))}{\Gamma(1/\alpha) y_i!} \left(\frac{\mu_i}{\mu_i + (1/\alpha)} \right)^{y_i} \left(\frac{(1/\alpha)}{\mu_i + (1/\alpha)} \right)^{1/\alpha} \quad (4)$$

where $L(\mu_i)$ is the maximum likelihood estimator of μ_i and N is total number of observations.

The superiority of NB model over Poisson regression model depends on the value of dispersion parameter α. If α is not statistically different form zero, the NB model reduces to the Poisson model. Otherwise, the NB model is selected as the appropriate choice.

A basic assumption of standard Poisson and NB models, as likelihood-based models, is that observations are independent. However, this assumption is often violated for road crash data in which observations coming from a same group are correlated spatially or/and temporally. Such correlations arise from unobserved factors which may exist in same groups. For example, crash counts may be independent across road segments, but those occurring on a specific road segment over successive time periods are likely to be serially correlated. Ignoring such possible correlations violates fundamental assumption of "independence of errors" in standard count models, and also underestimates standard errors of model parameters.

One way to account for within-site correlation (or heterogeneity) as well as between-site unobserved heterogeneity is to treat road crashes as panel data and apply a random-effect NB model. The RENB model assumes that the within-segment heterogeneity is uncorrelated with the explanatory variables. If this assumption does not hold, a fixed-effect specification should be used. The selection between FE and RE models is based on Hausman test, which determines which model is more appropriate than the other (Hausman et al., 1984). If the test is insignificant, the RE model is preferred over the FE model, implying that there is no correlation between independent variables and random effects (Kweon and Kockelman, 2005). In the context of road safety, there are some studies which have applied random-effect models (Kumara et al., 2003; Huang and Chin, 2010; Quddus, 2013). For example, Chin and Quddus (2003) used a RENB model to investigate variables contributing to traffic accidents at signalized intersections in Singapore. Kweon and Kockelman (2005) applied fixed- and random-effect models along with several other count models to examine safety effects of speed limit changes on high speed highways. A RENB model was selected among other alternative models to fit the data.

Let y_{it} be the number of head-on crashes on segment i in year t, the framework for random-effects negative binomial (RENB) model is given as:

$$\ln(\mu_{it}) = X_{it}\beta + \nu_i \forall i, t \quad (5)$$

where μ_{it} is the expected number of head-on crashes for segment i in the year t, X_{it} is a vector of explanatory variables for segment i, β is a vector of estimable parameters, and ν_i is a segment-specific random effect such that $\exp(\nu_i)$ is gamma distributed with mean 1 and variance α.

In the RENB model, the variation of segment effects over time is allowed to vary randomly by assuming that $1/(1+\alpha_i)$ follows a beta distribution with parameters (a, b). Therefore, the probability density function of RENB model can be presented as:

$$f(y_{i1}, \ldots, y_{iT} | X_{i1}, \ldots, X_{iT}) = \frac{\Gamma(a+b) + \Gamma(a + \sum_T \mu_{it} + \Gamma(b + \sum_T y_{it})}{\Gamma(a)\Gamma(b)\Gamma(a + b + \sum_T \mu_{it} + \sum_T y_{it})} \times \prod_T \frac{\Gamma(\mu_{it} + y_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)} \quad (6)$$

The model parameters a, b, and β are estimated using the standard maximum likelihood method.

3.1.2. Zero-altered models

As stated earlier, there are 194 segments for which no head-on crashes occurred during the study period; hence, the presence of excess zeros is plausible in this case. Excess zeros could be a source of overdispersion in the crash data. Under such scenarios, the standard Poisson and NB models are not appropriate because they can not predict excess zeros exist. Instead, zero-altered models including zero-inflated and hurdle models are typically used to handle excess zero counts. As such, these models were also applied in this study to determine whether zero counts are a potential source of over-dispersion or not. They are explained in the following subsections.

3.1.2.1. Zero-inflated models. Zero inflated regression models are adopted for modelling data characterized by a significant amount of zeros or more zeros than expected in standard Poisson or NB models. These models allow for zeros to be generated by two different processes: (1) the process that generates structural zeros estimated from a binary distribution (logit or probit distribution), and (2) the process that generates sampling zeros, deriving from the Poisson/NB distribution.

While the zero-inflated Poisson (ZIP) model is used to account for excess zeros, the zero-inflated negative binomial model (ZINB) is applied to accommodate overdispersion arising from both unobserved heterogeneity and excess zeros (Miranda-Moreno and Fu, 2006). There are a number of studies in the literature have used zero-inflated models. For example, Daniel and Chien (2004) utilized both ZIP and ZINB models to identify factors affecting truck crashes on urban arterials. Easa and You (2009) developed four statistical models including Poisson, NB, ZIP, and ZINB models to identify the factors affecting crash frequencies on horizontal curves in Washington State. Chen et al. (2012) applied a ZIP model to relate roundabout safety to predicted speed using Italian roundabout approach-level data.

As the description of ZIP model, let P_{it} be the probability of being excess zero for the segment i in the time period t, and $(1 - P_{it})$ be the probability of crash counts derived from the Poisson distribution. The probability density function for the ZIP model is given as:

$$P(Y = y_{it}) = \begin{cases} P_{it} + (1 - P_{it})\, e^{\mu_{it}} & y_{it} = 0 \\ (1 - P_{it}) \dfrac{e^{-\mu_{it}} \mu_{it}^{y_{it}}}{y_{it}!} & y_{it} > 0 \end{cases} \quad (7)$$

where y_{it} is the number of head-on crashes for segment i in the year t, μ_{it} is the expected crashes for segment i as a function of its covariates $\mu_{it} = \exp(\beta X_{it})$, and P_{it} is the probability of being in the zero-crash-state which is fitted using logistic regression model, as follows:

$$P_{it} = \frac{\exp(K_{it}\lambda)}{1+\exp(K_{it}\lambda)} \tag{8}$$

where $K_{it} = (K_{it1},\ldots, K_{itm})$ is a function of explanatory variables and $\lambda_i = (\lambda_{i1},\ldots,\lambda_{im})$ is the estimable coefficients.

Similar to ZIP model, the probability density function for the ZINB is given by Eq. (9):

$$P(Y=y_{it}) = \begin{cases} P_{it} + (1-P_{it})\dfrac{1}{(1+\alpha\mu_{it})^{1/\alpha}} & y_{it}=0 \\ (1-P_{it})\dfrac{\Gamma(y_{it}+(1/\alpha))}{\Gamma(y_{it}+1)\Gamma(1/\alpha)}\dfrac{(\alpha\mu_{it})^{y_{it}}}{(1+\alpha\mu_{it})^{y_{it}+(1/\alpha)}} & y_{it}>0 \end{cases} \tag{9}$$

where α is dispersion parameter and $\Gamma(\cdot)$ is gamma function for the ZINB model.

3.1.2.2. Hurdle models. Similar to zero-inflated (ZI) models, hurdle models can handle data characterized by extra zeros and fit the response variable as a mixture of binary and count distributions. However, contrary to ZI models, the hurdle models assume that all zeros in the crash data are sampling zeros. That is, it indicates that road segments with no head-on crash are safe over the study period only, not for lifetime (not inherently safe). Hurdle models are interpreted as two state models, a zero state with no crash and second state in which at least one crash occurs. The first part of the model can be modelled using a binary regression framework, such as logit or probit model. Given that a crash occurs, the number of crashes is then modelled by a left truncated Poisson or negative binomial distribution. The density function of hurdle Poisson (HP) model is given as follows:

$$P(Y=y_{it}) = \begin{cases} P_{it} & y_{it}=0 \\ (1-P_{it})\dfrac{e^{-\mu_{it}}\mu_{it}^{y_{it}}}{(1-e^{-\mu_{it}})y_{it}!} & y_{it}>0 \end{cases} \tag{10}$$

where p is the probability of being excess zero for the segment i, $(1-P)$ is the probability of non-zero crash counts, and μ_{it} is the predicted crash counts derived from zero-truncated Poisson regression model. The probability of being in the zero-crash-state, P, is fitted using logistic regression model.

To handle over-dispersion arising from unobserved heterogeneity and excess zeros, hurdle negative binomial (HNB) model is used. Similar to HP model, the overall HNB density is given as:

$$P(Y=y_{it}) = \begin{cases} P_{it} & y_{it}=0 \\ (1-P_{it})\left(1-\dfrac{1}{(1+\alpha\mu_{it})^{(1/\alpha)}}\right)\left(\dfrac{\Gamma(y_{it}+(1/\alpha))}{\Gamma(y_{it}+1)\Gamma(1/\alpha)}\right)\left(\dfrac{(\alpha\mu_{it})^{y_{it}}}{(1+\alpha\mu_{it})^{y_{it}+(1/\alpha)}}\right) & y_{it}>0 \end{cases} \tag{11}$$

where α is dispersion parameter, $\Gamma(\cdot)$ is gamma function, and μ_{it} is the predicted crash counts derived from left truncated negative binomial model.

There are several studies in the literature which have applied zero-inflated models to analyse crash counts with excessive zeros. In terms of hurdle models, these models have been applied in a variety of fields to handle excess zeros such as economics, medical science, environment, industry. However, these models have rarely been adopted in road safety literature (Boucher and Santolino, 2010; Son, 2011; Hosseinpour et al., 2013).

3.1.3. Model comparison and selection

The process of model comparison and selection is based on the presence and the source of overdispersion in the crash data. To verify the existence of overdispersion in the head-on crashes, a Wald t-statistical test on the dispersion parameter and a likelihood ratio test (LRT) were performed where the Poisson, HP, and ZIP models were nested within the NB, HNB, and ZINB models, respectively (Isgin et al., 2008). The LRT is based on differences in the log-likelihoods of two nested models, which follows a Chi-squared distribution with one degree of freedom. The test for NB-based models versus their Poisson counterparts is given as:

$$\text{LRT} = 2(LL_{\text{NB-based}} - LL_{\text{PM-based}}) \cong \chi^2_{(\text{d.f.}=1)} \tag{12}$$

A significant value for both the Wald t-statistical and LR tests confirms the presence of overdispersion in the crash data and that the unobserved heterogeneity is a source of overdispersion. In such cases, the NB-based models would be preferred to the Poisson counterparts.

To check the contribution of excess zeros in overdispersion, a Vuong (1989) test is applied to compare zero-altered models including zero-inflated and hurdle models with single count models (Poisson and NB) since zero-altered models are not nested within single models. For the NB-based models, a significant value for the Vuong test implies that both unobserved heterogeneity and excess zeros account for overdispersion, and thus, zero-altered NB models are preferred to the single NB counterparts. Similarly, the Vuong test is also conducted for Poisson-based models. A significant value for the test indicates that only zero counts contribute to overdispersion, and therefore, two-state models (either HP or ZIP) are preferred to the PM. Furthermore, the Vuong test is used for comparing between hurdle and zero-inflated models (ZIP vs. HP or/and ZINB vs. HNB).

Given that $P_1(y_i|x_i)$ and $P_2(y_i|x_i)$ are the predicted probability of the standard models (Poisson or NB models) and the two-state model (zero-inflated and hurdle models), respectively, the Vuong test can be expressed as

$$m_i = \ln\left(\frac{\sum_i P_1(y_i|x_i)}{\sum_i P_2(y_i|x_i)}\right) \tag{13}$$

$$V = \frac{\bar{m}\sqrt{n}}{SD(m)} \tag{14}$$

where $\bar{m}$ is the mean of m_i and $SD(m)$ is the standard deviation of m_i.

The Vuong test (V) follows a standard normal distribution. If V is greater than 1.96, then the test favours HP/ZIP or ZIP/ZINB over Poisson/NB, and if V is lower than -1.96, the parent Poisson or NB model is favoured. A value of $-1.96 < V < 1.96$ indicates neither model is preferred over the other.

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

$$\text{AIC} = -2LL + 2P \tag{15}$$

$$BIC = -2LL + P(\ln(n)) \tag{16}$$

where LL is the logarithm of the maximum likelihood estimation for each model, P is the number of model parameters, and n is the number of observations ($n = 448$).

A model with the lowest AIC and BIC values is preferred. To decide whether there is a statistically significant difference between two models, Hilbe's AIC and Raftery's BIC rule-of-thumb criteria were adopted in this study (Raftery, 1995; Hilbe, 2011). Table 3 shows the significance levels for both criteria. In this case

Table 3
Significance levels for AIC and BIC (Raftery, 1995; Hilbe, 2011).

ΔAIC for models A and B	Result if A < B	ΔBIC for models A and B	Result if A < B
< 0.0 & ≤2.5	No difference	<0.0 &≤2.0	Weak difference
<2.5 & ≤6.0	Prefer A if $n > 256$	<2.0 &≤6.0	Positive difference
<6.0 & ≤9.0	Prefer A if $n > 64$	<6.0 &≤10.0	Strong difference
10+	Prefer A	10+	Very strong difference

study ($n = 448$), if the difference in the AIC value is greater than 2.5, then the model with lower AIC is favoured over another.

The overall goodness-of-fit statistic for all models considered is measured by deviance statistic. The test is twice the differences in the log-likelihoods of the final model and the null model, which follows the Chi-squared distribution with degrees of freedom equal to the number of predictors in the final model, as follows:

$$D = -2\ln\left(\frac{LL_\beta}{LL_0}\right) = -2\ln(LL_\beta - LL_0) \cong \chi^2_{(\mathrm{d.f.}=p)} \tag{17}$$

where LL_0 is the log-likelihood of the null model (intercept-only model), LL_β is the log-likelihood of the final model, and p is number of predictors in the final model. A significant value for the deviance statistic indicates a good statistical fit.

Two statistical software packages R and STATA were utilized to develop and select the considered models (Statacorp, 2009; R Development Core Team, 2011).

3.2. Modelling the head-on crash severity

As noted before, this study aimed to identify the factors that contribute significantly to the severity of head-on crashes. In terms of injury severity, traffic crashes are considered a categorical and ordered outcome variable (i.e., from no-injury to fatal-injury). Ordered models such as the ordered probit model are commonly used to analyse such ordinal outcomes where unordered multinomial models (e.g., multinomial logit/probit approach) cannot be applied while dealing with categorical dependent variables because they fail to account for the ordinal nature of the data. In addition, multinomial models represent some restrictive assumptions such as the lack of a closed-form likelihood and the independence of irrelevant alternatives (IIA) (Kockelman and Kweon, 2002; Quddus et al., 2002; Pai and Saleh, 2008; Haleem and Abdel-Aty, 2010). As a consequence, this study used the ordered model to analyse the injury severity of head-on crashes. In Malaysia, traffic crashes are scaled into four severity levels: 1—no injury (i.e., property damage only); 2—slight injury (a crash in which at least one person is injured but not killed or seriously injured); 3—serious injury (a crash in which at least one person is seriously injured but not killed); and 4—fatal injury (a crash in which at least one person is killed within 30 days from the date of event as a result of crash). The response variable in this study is ordered with these four categories. For a specific head-on crash occurred on segment i in year t, the observed injury level S_{it} is related to an unobserved (latent) variable S^*_{it} as follows:

$$S_{it} = j \Rightarrow \mu_{j-1} \le S^*_{it} \le \mu_j \Leftrightarrow \begin{cases} j=1 & \text{if} \quad -\infty \le S^*_{it} \le \mu_1 \quad (\text{no injury}) \\ j=2 & \text{if} \quad \mu_1 \le S^*_{it} \le \mu_2 \quad (\text{slight injury}) \\ j=3 & \text{if} \quad \mu_2 \le S^*_{it} \le \mu_3 \quad (\text{serious injury}) \\ j=4 & \text{if} \quad \mu_3 \le S^*_{it} \le \infty \quad (\text{fatal injury}) \end{cases} \tag{18}$$

$$S^*_{it} = X_i\beta + \varepsilon_{it} \forall i, t \tag{19}$$

where μs are unknown threshold parameters to be estimated; S^*_{it} is the latent variable of injury severity; X_i is a vector of covariates for segment i; β is a vector of estimable coefficients; ε_{it} is a random error term that follows a normal distribution.

The predicted probabilities of the injury severity level j ($j = 1, 2, 3, 4$) for given X_i can be estimates as:

$$P(S_i = j) = \Phi(\mu_j - X_i\beta) - \Phi(\mu_{j-1} - X_i\beta) \tag{20}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function. The model parameters (i.e., β and S^*_{it}) are estimated by the maximum likelihood method. The same covariates used in the head-on crash frequency analysis were also used in the severity analysis.

The marginal effects of the standard ordered probit model with respect to explanatory variable l can be estimates as:

$$ME_{jl} = \frac{\partial P(S_i = j|X_i)}{\partial X_{il}} = [\Phi(\mu_j - X_i\beta) - \Phi(\mu_{j-1} - X_i\beta)]\beta_l \tag{21}$$

A common constraint of the standard ordered probit model is the parallel slope assumption (or proportional odds assumption), which means that the coefficient vector β is the same for all categories j. However, this assumption is often violated by two or more included variables. In such cases, the use of a standard ordered model may result in biases in parameter estimates and marginal effects (Boes and Winkelmann, 2004; Clark et al., 2005). One way to address this issue is to use a standard multinomial logit or probit model. However, as mentioned earlier, the multinomial models are incapable of exploiting ordinality of crash severity. In addition, they require more parameters to be estimated, and thus reduce the degrees of freedom for estimation (Kockelman and Kweon, 2002; Rifaat et al., 2012). To overcome the issue, Peterson and Harrell Jr (1990) proposed a partial proportional odds model (PPOM) in cases where the parallel slope assumption is violated by one or a few number of independent variables, and hence the PPOM relaxes the restriction only to those variables which violate the assumption. In the context of road safety, Wang and Abdel-Aty (2008) used the PPOM approach to analyse injury severity of left-turn crashes at intersections. The model was further used by Wang et al. (2009) to identify the effects of factors contributing to injury severity at freeway diverge areas.

The generalized ordered probit model (GOPM) is a more flexible approach to overcome the restriction of parallel slope assumption of the standard ordered models by allowing the threshold parameters to vary across road segments depending on explanatory variables, so that:

$$\mu_{ij} = \tilde{\mu}_j + X'_i k_j \tag{22}$$

where k_j represents the influence parameter of the explanatory variables on the thresholds and $\tilde{\mu}_j$ is a constant term. Now, Eq. (20) can be rewritten as:

$$P(S_i = j) = \Phi(\tilde{\mu}_j - X_i\beta_j) - \Phi(\tilde{\mu}_{j-1} - X_i\beta_{j-1}) \tag{23}$$

The GOPM is nested to the standard ordered probit model in which the later restricts the coefficients to be fixed across the thresholds (Rifaat et al., 2012). A likelihood ratio test can be adopted to test the superiority of the GOPM over its standard counterpart. The test follows a Chi-squared distribution with degrees of freedom equal to $P(J-2)$, where P is the number of explanatory variables (Baba, 2009).

The marginal effects of generalized ordered probit model with respect to explanatory variable l are estimated as:

$$ME_{jl} = \frac{\partial P(S_i = j|X_i)}{\partial X_{il}} = \Phi(\mu_j - X_i\beta)\beta_{lj} - \Phi(\mu_{j-1} - X_i\beta)\beta_{lj-1} \tag{24}$$

From Eq. (24), the coefficients have non-constant effects across injury severity categories. As a result, the marginal effects of GOPM are more flexible than those from the standard OP model. Generalized ordered models have been employed in a number of safety-related studies (Clifton et al., 2009; Quddus et al., 2009; Kaplan and Prato, 2012; Rifaat et al., 2012; Rich et al., 2013).

Unlike modelling the crash frequency in which the subject is road segment, the crash severity is typically analysed at individual crash level. Thus, there are additional factors such as driver characteristics, weather conditions, and vehicle attributes that can be considered in modelling injury severity. However, this study did not consider factors other than the roadway characteristics because the data relevant to such factors were not available in the original data set. Therefore, the scope of this study was limited to investigating the effects of roadway characteristics on both the frequency and severity of head-on crashes. Nevertheless, due to such omitted variables, crashes within road segments are likely to be correlated and similar in terms of unobservable characteristics. This could be a potential source of unobserved segment-specific heterogeneity. Neglecting to consider such an important issue may cause biases in parameter estimates and model results.

One way to lower the likelihood of omitted-variable biases would be the use of a fixed-effect model, which assumes that segment-specific heterogeneity is correlated with the independent variables. However, a fixed-effect model requires a large number of parameters (e.g., one constant term for each segment) to be estimated, which is an obstacle to the model estimation. Therefore, the use of fixed-effects generalized ordered probit model is not appropriate to control for unobserved heterogeneity (Obeng, 2008; Baba, 2009; Brown and Roberts, 2011). A more efficient way is to use a random-effect generalized ordered probit model (REGOPM) to control for the segment heterogeneity.

In the REGOPM, the unobservable segment-specific heterogeneity can be accounted for by including the term θ_i as the segment-specific effect, into Eq. (23) so that:

$$P(S_i = j) = \Phi(\tilde{\mu}_j - X_i\beta_j - \theta_i) - \Phi(\tilde{\mu}_{j-1} - X_i\beta_{j-1} - \theta_{i-1}) \quad (25)$$

As documented by Boes and Winkelmann (2006), Baba (2009), Boes and Winkelmann (2010), treating θ_i as fixed parameters without indicating the relationship between X_i and θ_i results in the incidental parameters problem. To cope with this problem, the term θ_i is treated as random variable and allowed for a possible correlation with $X_i$as:

$$\theta_i = \bar{X}_i'\gamma + \delta_i \quad (26)$$

where $\bar{X}_i$ is the vector averages of X_i over time, γ is the parameter vector, and δ_i is segment-specific time invariant random effect that is assumed to be normally distributed and orthogonal to $\bar{X}_i$.

The REGOPM was estimated using a user-written command named regoprob in STATA version 11 (Boes and Winkelmann, 2006). An advantage of using random-effect model over standard ordered model is that it allows the relationship between roadway characteristics and injury severity to vary across road segments due to a possible heterogeneity in the data.

To the authors' knowledge, there are very few studies in the literature that have extended generalized ordered models to analyse injury severity. As one of those studies, Eluru et al. (2008) applied a mixed generalized ordered logit model to examine non-motorist injury severity in traffic crashes using the 2004 General Estimates System (GES) database in the USA. Castro et al. (2013) utilized a spatial generalized ordered response model to investigate the effects of various factors on the injury severity of traffic crashes occurring on highway segments in Austin, Texas. The spatial generalized ordered model was found to account for both unobserved heterogeneity and spatial dependence in the data. Using the 2010 General Estimates System (GES) database, Yasmin and Eluru (2013) compared the performance of different ordered and unordered response frameworks for modelling driver injury severity. For the ordered response, the frameworks included ordered logit, generalized ordered logit, and mixed generalized ordered logit. With respect to the unordered response, the models consisted of multinomial logit, nested logit, ordered generalized extreme value logit, and mixed multinomial logit model. The results revealed the superiority of the mixed generalized ordered logit over the mixed multinomial logit model in modelling injury severity.

4. Empirical results

4.1. Analysis of head-on crash frequency

The NB model was first developed and compared to its Poisson counterpart to examine which model gives the better fit. The *P*-values for both the likelihood ratio test (NB vs. Poisson) and dispersion parameter were found to be highly significant, which implies the presence of overdispersion as well as the appropriateness of NB model over the Poisson model to fit the overdispersed data. Next, to check the existence of segment-specific heterogeneity, random- and fixed-effect NB models were developed. The Hausman test was then undertaken to select between these two models. The test gave a value of 5.08 with 7 degrees of freedom, which is much lower than the critical one (14.065, 7 degrees of freedom), leading the rejection of FENB in favour of RENB model. This means that segment-specific heterogeneity is not correlated to the explanatory variables. In the next step, dual-state models including HP, HNB, ZIP, and ZINB models were developed and compared with the parent Poisson and NB models to check contribution of excess zeros to extra variation in the data.

For the HNB model, the algorithm for model estimation was not converged during the model calibration. Therefore, the HNB model was excluded from further consideration. With respect to the ZINB model, the dispersion parameter estimate was found to be non-significant at the 5% level. As such, the model was also excluded from further analysis. This indicates that overdispersion could be due either to excess zeros or to unobserved heterogeneity rather than their combination. The comparison results for the remaining models are summarized in Table 4. For Poisson model, overall 8 parameters were associated including one intercept and seven explanatory variables; for NB model, 9 parameters including intercept, dispersion parameter, and seven explanatory variables; for RENB model, 10 parameters including intercept, two beta parameters, and seven explanatory variables; for ZIP model, 19 parameters including ten parameters for count state and nine parameters for zero state; for HP model, 17 parameters including, ten parameters for count state and seven parameters for zero state.

The Vuong test was applied to compare the HP and ZIP models with the parent Poisson and NB models. For pairs of PM vs. HP and PM vs. ZIP, the test showed that the Poisson was rejected in favour of the HP and ZIP models. However, the test for NB against HP and ZIP revealed that NB model was favoured. In addition, the two information criteria (AIC and BIC) are also used to determine the best model among others. These measures are also used to compare the RENB with HP and ZIP because no specific test exists to compare random-effect models with zero-altered models (Kweon and Kockelman, 2005). Both AIC and BIC give advantage to the RENB model over the others. In terms of AIC, the RENB model has the lowest value. Based on Hilbe's rule of thumb for this study (n = 448), if ΔAIC is greater than 2.5, then the model with the lowest AIC is preferred. For this study sample, the minimum difference in AIC was found for RENB versus NB by nearly 30, and thus the RENB is strongly preferred over the other models. The superiority of the RENB model was also supported by the BIC. Based on Raftery's rule of thumb, the minimum ΔBIC was found to be very strong for the RENB versus NB by 24, which favours highly the random-effect model. Overall, the RENB model was selected among others based on the above selection

Table 4
Results of the fitted models for homogeneous segments.

Models	Poisson	NB	RENB	HP	ZIP
No. of observations	1792	1792	1792	1792	1792
No. of parameters	8	9	10	17	19
Log-likelihood at converge	−1398.9	−1227.5	−1211.7	−1267.8	−1293.5
Vuong test vs. Poisson (*p*-value) vs. NB (*p*-value)				5.62 (p < 0.0001); −2.41; 0.0081	4.70 (p < 0.0001); −2.99; 0.0014
AIC	2813.71	2473.02	2443.5	2569.67	2624.94
BIC	2857.64	2522.44	2498.4	2663.02	2729.27

Table 5
Parameter estimates of RENB model for head-on crash frequency.

Covariates	Coeff.	*P*-value	IRR
Intercept	−2.553	0.063	–
Access points	0.132	0.003	1.142
Ln(Heavy Vehicle Traffic)	0.502	0.002	1.652
Horizontal curvature	0.068	0.014	1.071
Unpaved shoulder width	−0.255	0.024	0.775
Paved shoulder width	−0.256	0.045	0.774
Terrain type (1: undulating terrain, 0: flat terrain)	0.312	0.046	1.366
Speed limit	−0.011	0.041	0.989
Segment length (exposure)**	1.000	–	–
Summary statistics			
Total number of observation		1792	
Parameter a (std. error)		19.847 (7.973)	
Parameter b (std. error)		1.552 (0.245)	
Wald test statistic		52.08	
Pseudo R^2		0.020	
Likelihood ratio test vs. pooled (*P*-value)		274.11 (<0.00001)	
AIC		2443.462	
BIC		2498.373	

* Reference category, ** Segment length was modeled as an *offset* variable.

criteria. The superiority of RENB model over the standard and zero-altered count models (e.g., Poisson, NB, HP, and ZIP) indicates that both within- and across-segment heterogeneity is associated with overdispersion, whereas excess zeros is not.

The results of RENB model is presented in Table 5. The *P*-value corresponding to the likelihood-ratio test between the panel model (i.e., an extended NB model in which both within- and between-segment heterogeneities are accounted for, and permitted to vary across segments) and its pooled counterpart (i.e., a NB model in which only between-segment heterogeneity is controlled for) was found to be highly significant, which gives advantage to the RENB model, implying the presence of intra-segment heterogeneity in the crash data. Therefore, crash counts should be treated as panel data rather than independent one. As seen in the table, the Wald statistic, which follows a Chi-square distribution, was 52.08 with 10 degrees of freedom (*P*-value <0.0001), indicating a good statistical fit for the model. Overall, seven variables were found to contribute significantly to the head-on (HO) crash frequency; these variables include the natural logarithm of heavy vehicle traffic, horizontal curvature, access points, and terrain type, which had positive effects on the crash frequency, while posted speed limit, paved and unpaved shoulder width were negatively associated with the outcome. The effect of each significant variable is explained in the following. To ease interpretation of the study variables, the incidence rate ratios (IRR), i.e. exp(·) were estimated and are presented in Table 5. For a given variable with IRR less than 1.0, an increase in value of the variable corresponds to a reduction in the crash frequency and vice versa.

The variables "access points" is among the most significant variables (P = 0.003) affecting HO crash frequency. This variable has a positive effect on the crash outcome. This is because as the number of access points increase, there are more conflicts among vehicles, increasing the risk of HO crashes. Especially in Malaysia, which is a left-hand traffic country (i.e., right hand drive), drivers typically stop in the outside lane (i.e. the closest lane to oncoming traffic) when they intend to turn right into a driveway in the opposite direction; hence they are more likely to be struck by an oncoming vehicle in the opposing lane. A similar finding was obtained by (Zhang and Ivan, 2005), who found that the number of access points affects head-on crashes. In addition, a HO crash may occur when a driver is forced to cross the centreline to avoid a collision (e.g., sideswipe or angle crashes) with other vehicles merging from driveways into the main lane and thus increases the risk of hitting a vehicle coming from the opposite direction. The incidence rate ratios (IRR) for "access point" is about 1.14, indicating that one unit increase in the number of access points corresponds to a 0.14%

Table 6
Results of REGOPM for head-on crash severity.

Variable	Slight injury		Serious injury		Fatal injury	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Constant	−0.573**	0.257	−2.49***	0.308	−2.351***	0.39
Horizontal curvature	0.238***	0.053	0.528***	0.058	0.369***	0.051
Paved shoulder width	0.982***	0.15	1.173***	0.167	0.491**	0.205
Access points	−0.103	0.066	−0.426***	0.081	−0.486***	0.099
Land use: (1) No activity	R		R		R	
(2) Low activity	−0.773***	0.218	0.158	0.227	−0.347	0.303
(3) High activity	−1.576***	0.288	−0.837***	0.292	−0.353	0.362
Terrain type (1: undulating terrain, 0: flat terrain)	−1.372***	0.258	−0.165	0.245	0.497**	0.23
Median: (1: if present, 0 if not present)	−1.108***	0.317	−0.865***	0.329	−0.25	0.381
Side friction (1: high interaction, 0: low interaction)	−0.393*	0.229	0.542**	0.248	1.433***	0.308
Summary statistics			527			
No of observation			0.294***			
ρ (Std. Error)			−436.871			
Log-likelihood at convergence			465.682***			
$\chi^2 = 2(LL_\beta - LL_0)$			0.376			
Pseudo R^2			(0.081)			

Note: The base category is "no injury". "R" stands for reference category. ρ is the ratio of the random-effect variance to the total variance. One, two, and three asterisks denote the significance at the 10%, 5%, and 1% level, respectively.

increase in the HO crash frequency with all other variables being constant.

The natural logarithm of heavy-vehicle traffic was found to have a positive relationship with HO crashes, which agrees with expectations. The reason for this finding is that heavy vehicles travel at low speeds, so they increase the risk of overtaking manoeuvres made by vehicles behind them; this conclusion is intuitively reasonable for this study where approximately 88% of the study segments are two-lane undivided roadways. In these circumstances, fast-moving vehicles must cross the road centreline to pass slow-moving vehicles; thus, they are more likely to encounter a vehicle approaching from the opposite direction. The IRR for heavy vehicle traffic is 1.65, which implies that a one unit increase in log of heavy vehicle traffic will result in a 65% increase in HO crashes, with the remaining predictor values held constant. Horizontal curvature were found to have a positive effect on the HO crash frequency, which is consistent with intuitive expectations because sharp curves limit drivers' sight distance to see oncoming vehicles and make it difficult for them to control the vehicle, especially at high speeds, and hence increase the risk of HO crashes (Gårder, 2006). From the IRR value, a 1 km^{-1} sharper curve is associated with a 7.1% increase in the crash frequency. Both paved and unpaved shoulder widths had a negative relationship with HO collisions. A head-on crash occurs when a vehicle crosses the road centreline either intentionally (e.g., manoeuvres relevant to turning, overtaking, lane changing, etc.) or unintentionally (e.g., losing control, driver distraction, drowsiness, inattention, etc.). In such conditions, a wider shoulder width could provide more room for a vehicle in the opposing lane to avoid colliding with the errant vehicle (Gårder, 2006). According to the IRR, widening 1 meter shoulder in each direction will correspond to a nearly 0.40% reduction in the HO crash frequency.

With regard to terrain type, HO crashes were more likely to occur on undulating terrains, which is expected because drivers' sight distance is reduced in such areas, so that they may not have sufficient reaction time to avoid colliding with an errant oncoming vehicle. In addition, the majority of the study segments are two-lane undivided roadways. In such locations, drivers typically attempt to make an overtaking manoeuvre when they are following a slow-moving vehicle (e.g., truck, bus, etc.). Therefore, they are more likely to experience HO crashes while driving on undulating terrain. This finding is also consistent with that of the study by Zhang and Ivan (2005). The IRR for terrain type is 1.36, indicating undulating terrain are 36% more likely to experience HO crashes relative to flat terrain. Speed limit was found to be negatively associated with the crash frequency. A possible reason for this finding is that lower speed limits are generally assigned to road segments with poor safety conditions. Consistent results can be found in past studies (Milton and Mannering, 1998; Qin et al., 2005; Zhang and Ivan, 2005). However, a higher speed limit should not be interpreted as a measure to reduce both the frequency of HO crashes.

4.2. Analysis of head-on crash severity

Table 6 lists the results of REGOPM for the injury severity of HO crashes. From the table, the χ^2 statistic, as twice the differences in the log-likelihoods of the full model and the null model, yielded a value of 465, which rejects the null hypothesis that all the model parameters are zero at the 0.001 level. Moreover, a value of 0.376 for the pseudo-R^2 also indicates a reasonable goodness of fit for the model. The LR test for the parallel slope assumption turned out to be highly significant, rejecting the null hypothesis that each explanatory variable has the same effect in each severity category. This result favours the generalized ordered probit model over the standard ordered probit model (OPM). Moreover, the pseudo-R^2 for the REGOPM (0.376) is greater than that of the standard model (0.277), which also supports the appropriateness of the REGOPM. The term ρ, which represents the ratio of random-effects variance component to the total variance, was statistically significant at the 0.1% level (Z-value = 3.63), supporting the validity of random-effect specification against the pooled model. The magnitude of ρ was 0.294, meaning that about 30% of the total variance was explained by unobservable segment-specific effects. This implies a significant evidence of intra-segment correlation (i.e., the presence of heterogeneity among segments) in the crash data. As a result of the above discussion, both observed and unobserved heterogeneity was captured by the REGOPM in which the former was accounted for by allowing the model coefficients to vary across the severity levels, while the later was controlled for using random-effects specification.

Table 7
Marginal effects for REGOPM.

Variable	No injury	Slight injury	Serious injury	Fatal injury
Horizontal curvature	−0.08	−0.095	0.088	0.087
Paved shoulder width	−0.329	−0.06	0.272	0.116
Access points	0.034	0.106	−0.026	−0.115
Land use: (1) No activity	–	–	–	–
(2) Low activity	0.259	−0.311	0.134	−0.082
(3) High activity	0.527	−0.25	−0.194	−0.084
Terrain type (1: undulating terrain, 0: flat terrain)	0.459	−0.404	−0.172	0.117
Median: (1: if present, 0 if not present)	0.371	−0.084	−0.227	−0.059
Side friction (1: high interaction, 0: low interaction)	0.131	−0.311	−0.159	0.339

In total, 7 variables were found to be significantly associated with the injury severity in the model: horizontal curvature, paved shoulder width, access points, land use, terrain type, presence of median, and side friction. To better reflect the effect of significant variables on the probability of different severity categories, the marginal effects of each variable were calculated and are presented in Table 7.

With respect to horizontal curvature, HO crashes occurring on curved segments were expected to result in more severe injury. This finding is not surprising because sharp curves restrict drivers' sight distance to see the traffic ahead, so that they may not have sufficient time to take proper action while encountering an errant vehicle approaching from the opposite direction. This result can also be derived from the marginal effects: the probabilities of severe injuries (serious & fatal) increase as the horizontal curvature increases ($ME_{serious}$ = 0.088, ME_{fatal} = 0.087), while the probabilities of non-severe injuries decrease ($ME_{no\text{-}injury}$ = -0.08, ME_{slight} = -0.095). A similar result was found in previous studies (Abdel-Aty and Abdelwahab, 2004; Ma et al., 2008; Wang et al., 2009, 2011; Xie et al., 2009; Christoforou et al., 2010).

The variable paved shoulder width was found to have a positive effect on the severity of HO crashes. This finding is somewhat counterintuitive since shoulder width had a negative effect on the crash frequency. An explanation for this result is that wider shoulders might encourage drivers to travel at high speeds, so that they may not have sufficient time to avoid a collision with an errant vehicle crossing the centreline in the opposing traffic, thus increasing the chance of more severe injuries (Gårder, 2006). Haleem and Abdel-Aty (2010) pointed out that wider shoulders increase the risk of sideswipe crashes by encouraging inappropriate use of the shoulder for merging or lane-changing manoeuvres. As a consequence, vehicles travelling in the traffic lane are forced to cross the centreline to avoid a sideswipe collision with such vehicles, leading to more severe injuries in HO crashes. According to the marginal effects, wider shoulders reduce the probabilities of no injury and slight injury by 33% and 6%, respectively, and increase the probabilities of serious and fatal injuries decrease by 27.2% and 11.6%, respectively. Access point was found to be negatively associated with injury severity. As mentioned in the previous section, the increase in the number of access points leads more conflicts among traffic movements, which increases the frequency of HO crashes. However, under such conditions, drivers tend to travel more carefully and slowly. Therefore, they are less likely to be involved in severe injuries once a head-on occurs. From the marginal effects, the risk of having severe injuries decreases as the number of access points increases ($ME_{\text{no-injury}} = 0.034$, $ME_{\text{fatal}} = -0.115$).

With regard to land use, HO crashes that occurred on segments with roadside activities (e.g., commercial, residential, industrial, etc.) were found to be less severe than those that occurred on segments with no activity (e.g., agricultural or forested areas). This finding seems intuitively reasonable because drivers are more cautious and travel at lower speeds while driving on roadways with high roadside activities. Therefore, they are expected to be involved in less severe crashes in such areas. On the other hand, on road segments with no or very few roadside activities, drivers tend to drive at high speeds, increasing the likelihood of more severe injuries when a HO crash occurs. This finding is consistent with those in past research (Zajac and Ivan, 2003; Lee and Abdel-Aty, 2005; Boufous et al., 2008; Pai, 2009; Wang et al., 2009). The marginal effects show that there are 25.9% and 52.7% probabilities of non-injury crashes for road segments located in areas with low and high activities, respectively.

Regarding terrain type, HO crashes occurring on undulating terrains were more likely to result in severe injuries. This is due to the restriction of drivers' sight distance especially at crests. In such situations, an oncoming vehicle could be suddenly appeared in the driver's vision field, so that he or she may not have sufficient reaction time to avoid colliding with that vehicle. In addition, as noted earlier, much of the study segments are two-lane undivided roadways. In such locations, drivers tend to pass slow-moving vehicles. Therefore, such dangerous manoeuvres increase the risk of severe injuries in HO crashes while driving on undulating terrain. Based on the marginal effects, the risk of severe injury outcomes (serious and fatal) compared to slight injury category increases for non-flat terrain.

As expected, the presence of median was associated with less severe injuries. In fact, road medians had a positive correlation with non-severe crashes by separating the two opposing traffic flows. A recent study by Bham et al. (2012) also confirmed that undivided highways are greatly associated with the head-on crash severity. According to the marginal effects, the presence of a median increases the probability of non-injury outcomes by 37%, while the probabilities of slight, serious, and fatal injuries decrease by 8.4%, 22.7%, and 5.9%, respectively. Side friction was positively associated with the injury severity. The reason for this finding is that on road segments with high interaction with roadside activities (e.g., parking, bus stopping, loading, etc.), vehicles travelling on the through traffic are more likely to be faced with unexpected events such as people or other vehicles involved in the roadside activity, so that they may be forced to suddenly change their position or even cross centerline to avoid a collision. The marginal effects show that there is about 34% probability of fatal injury crashes for road segments where roadside activities were not protected from through traffic.

By focusing in Tables 5 and 6, three variables including horizontal curvature, paved shoulder width, access points, and terrain type were found to be significantly associated with both the frequency and severity of HO crashes. The horizontal curvature and terrain type had the same effect on these two responses, while the paved shoulder width and access points had the opposite effect. In conclusion, a specific explanatory variable affecting a certain crash outcome, either the crash frequency or severity, does not necessarily contribute to another (e.g., speed limit, median, side friction, and land use) or may even have the opposite effect (paved shoulder width and access points). An advantage of modelling crash frequency and severity separately is that it allows a different set of variables to be independently associated with the crash outcomes. This helps better understand the real impact of contributing variables on the crash incidence to develop safety countermeasures in a more effective manner. As an example, by focusing solely on the crash frequency, "access points" was found to be the most influential variable affecting the HO frequency. At first glance, "access points" should be prioritised for remedial action; however, due to its inverse effect on the injury severity, "access points" may not be the most critical factor that needs to be addressed to reduce HO crashes.

Based on the results of the severity model, HO crashes are more likely to result in severe injuries on undivided road segments with no or few roadside activities, wider shoulder widths, and fewer access points. In such circumstances, drivers tend to travel at excessive speeds, which lead to more severe outcomes. Consequently, road safety plans should be implemented to reduce such dangerous driving behaviours. The installation of centreline median barriers on undivided roadways seems to be the most effective countermeasure to keep vehicles from encroaching into the opposing lane. However, due to the great expenses associated with the installation and maintenance of continuous median barriers, road segments with a higher risk of HO crashes should be considered for such countermeasures. For other locations, centreline rumble strips could be effectively installed to reduce crossover collisions (Gårder, 2006). Table 8 presents some potential countermeasures proposed for the variables that affect HO crash frequency and severity. For example, according to the table, the use of speed enforcement cameras in areas with high speed violations could be an effective countermeasure to reduce HO crashes resulted from excessive speeds. Head-on crashes occurring on curved segments could be reduced by proposing a wide range of actions, such as straightening sharp curves, installing median barriers, widening the pavement width, improving superelevation. These measures could be prioritised and implemented in terms of their cost effectiveness. Further research should be considered to investigate the efficiency of different measures in reducing the HO crash outcomes.

According to the findings of other studies (Gårder, 2006; Radun et al., 2009; Tijerina et al., 2010; Mohamed et al., 2012; Merat and Jamson, 2013), driver error (e.g., inattention, fatigue, distraction, etc.) is one of the main causes of crashes related to unintentional lane departure (e.g., run-off-road, sideswipe, head-on crashes). To reduce head-on crashes caused by unintentional crossings of the centreline, in-vehicle warning technologies such as lane departure warning system, lane keeping system, and driver fatigue warning system could be effectively applied. The first two systems alert the driver when the vehicle is departing its lane, while a fatigue warning system alerts the driver of drowsiness while driving the vehicle. Driver fatigue is responsible for a considerable number of traffic accidents around the world. For example, fatigue was found to cause nearly 20% of road accidents in the UK and as many as 40% in Australia and North America (Merat and Jamson, 2013). In addition, providing car-parking bays and rest facilities, especially in rural areas, could also minimize fatigue-related crashes.

5. Conclusion and recommendations for future studies

This study provided insight into the effects of roadway geometric design, environmental and traffic characteristics on the frequency and severity of HO crashes occurred on 448 segments

Table 8
Causes of head-on crashes and corresponding potential countermeasures.

Causes of head-on crashes	Potential countermeasures
Undivided roadways (absence of median)	• Install median barrier (e.g., wire rope safety barrier, raised median, etc.) • Install rumble strips at the centreline • Improve centreline delineation
Excessive speeds	• Install advisory speed signs (e.g., radar speed sign) • Speed limit enforcement (e.g., speed cameras, radar speed guns, etc.) • Intelligent Speed Adaptation (ISA)
Horizontal curvature	• Straighten sharp horizontal curves • Install median barrier along adverse curves • Improve superelevation • Install centreline rumble strips through moderate curves • Improve sight distance (e.g., removal of roadside vegetation) • Widen road pavement through curves • Enhance delineation (guideposts, warning signs, etc.) • Speed limit enforcement
High heavy-vehicle traffic	• Provide passing lanes • Require heavy-vehicles to keep left except to turn right or pass
Undulating terrains	• Soften sharp vertical alignments • Install median barrier or rumble strips through adverse vertical alignments • Provide climbing lanes
Access points: Large volume of merging traffic	• Provide acceleration lanes • Widen & lengthen existing acceleration lanes • Signalize left turns
Large volume of right turns	• Implement channelization • Create right-turn lanes (e.g., flush median, right turn bay, etc.) • Provide right-turn signal phase • Prohibit right turns
Shoulder width	• Widen shoulder width • Install shoulder rumble strips • Improve delineation

References: Chen et al. (2002), Neuman et al. (2003), Mccartt et al. (2004), Persaud et al. (2004), Zhang and Ivan (2005), Gårder (2006), Ma and Kockelman (2006), Ma et al. (2008), Neuman (2008), Neuman et al. (2009).

of Malaysian federal roads over the 4-year period between 2007 and 2010. Seven count-data models, including Poisson, NB, RENB, HP, HNB, ZIP, and ZINB models were developed and compared to identify the factors that contribute to the frequency of HO crashes. The comparison results showed that the RENB model was preferred over the others to fit the crash frequency. Based on the results of the RENB model, the variables horizontal curvature, terrain type, heavy-vehicle traffic, and access points were found to be positively related to the frequency of HO crashes, while speed limit and shoulder width had a negative impact on the outcome. At the next step, the REGOPM was applied to model the injury severity of HO crashes. The aim was to identify the contributions of site-specific variables to the injury outcomes, given a HO crash had occurred. The results of the REGOPM indicated that horizontal curvature, paved shoulder width, terrain type, and side friction were associated with more severe crashes, whereas access points, land use, and the presence of median reduced the probability of severe crashes. The advantage of modelling the crash frequency and severity independently is that it allows for different sets of variables to influence the outcomes separately. For example, while curvature was positively associated with both the frequency and severity of head-on crashes, access points and shoulder width had opposite effects on these outcomes. Such asymmetric results cannot be uncovered by focusing merely on either of these outcomes.

Based on the results of this study, some potential countermeasures were proposed to minimize the risk of vehicles being involved in HO crashes. Further research in this domain should be performed to evaluate the efficiency of these measures in reducing the likelihood of such crossover collisions. One limitation of this study was that it did not consider the effects of driver, weather, and vehicle characteristics on the severity analysis, which is typically conducted at individual crash level. As such, a REGOPM was adopted to relax potential biases due to such omitted variables. To recommend, an injury severity model should be developed in the future by using a more comprehensive database to gain a better understanding of various factors associated with the injury severity. In addition, this study is one of the first attempts to investigate the applicability of REGOPM in modelling crash injury severity. Since the results showed that the model had more accurate performance over the standard ordered probit model, future research in this domain should be conducted to explore whether REGOPM is applicable to other crash contexts.

Acknowledgements

This study was supported financially by Univeristi Sains Malaysia Research University grant (1001/814123). The authors would like to acknowledge the support. The authors also wish to thank the Malaysia Institute of Safety Research and the Highway Planning Unit for the data used in this work.

References

Abdel-Aty, M.A., Abdelwahab, H.T., 2004. Predicting injury severity levels in traffic crashes: a modeling comparison. J. Transport. Eng. 130 (2), 204–210.
Abdelwahab, H., Abdel-Aty, M., 2004. Investigating the effect of light truck vehicle percentages on head-on fatal traffic crashes. J. Transport. Eng. 130 (4), 429–437.
Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate poisson lognormal models for crash severity modeling and site ranking. Transport. Res. Rec.: J. Transport. Res. Board 2136 (1), 82–91.
Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannering, F.L., 2012. A multivariate tobit analysis of highway accident-injury-severity rates. Accid. Anal. Prevent. 45, 110–119.
Baba, N., 2009. Increased presence of foreign investors and dividend policy of japanese firms. Pacific-Basin Fin. J. 17 (2), 163–174.
Bham, G.H., Javvadi, B.S., Manepalli, U.R., 2012. Multinomial logistic regression model for single-vehicle and multivehicle collisions on urban us highways in arkansas. J. Transport. Eng. 138 (6), 786–797.
Boes, S., Winkelmann, R., 2004. Income and happiness: new results from generalised thresholds and sequential models. IZA Discussion paper, No 1175.
Boes, S., Winkelmann, R., 2006. The Effect of Income on Positive and Negative Subjective Well-being. University of Zurich, Socioeconomic Institute, Manuscript, IZA Discussion Paper (1175).
Boes, S., Winkelmann, R., 2010. The effect of income on general life satisfaction and dissatisfaction. Social Indicators Res. 95 (1), 111–128.
Boucher, J.P., Santolino, M., 2010. Discrete distributions when modeling the disability severity score of motor victims. Accid. Anal. Prevent. 42 (6), 2041–2049.
Boufous, S., Finch, C., Hayen, A., Williamson, A., 2008. The impact of environmental, vehicle and driver characteristics on injury severity in older drivers hospitalized as a result of a traffic crash. J. Safety Res. 39 (1), 65–72.
Brown, H., Roberts, J., 2011. Exercising choice: the economic determinants of physical activity behaviour of an employed population. Soc. Sci. Med. 73 (3), 383–390.
Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accid. Anal. Prevent. 33 (1), 99–109.
Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accid. Anal. Prevent. 52, 188–203.
Cerrelli, E.C., 1997. Fatal Crash Involvements: What Are the Odds? US Department of Transportation, National Highway Traffic Safety Administration.
Chen, G., Meckle, W., Wilson, J., 2002. Speed and safety effect of photo radar enforcement on a highway corridor in british columbia. Accid. Anal. Prevent. 34 (2), 129–138.
Chen, Y., Persaud, B., Sacchi, E., Bassani, M., 2012. Investigation of models for relating roundabout safety to predicted speed. Accid. Anal. Prevent..

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accid. Anal. Prevent. 35 (2), 253.

Christoforou, Z., Cohen, S., Karlaftis, M.G., 2010. Vehicle occupant injury severity on highways: an empirical investigation. Accid. Anal. Prevent. 42 (6), 1606–1620.

Clark, A., Etilé, F., Postel-Vinay, F., Senik, C., Van Der Straeten, K., 2005. Heterogeneity in reported well-being: evidence from twelve european countries*. Econ. J. 115 (502), C118–C132.

Clifton, K.J., Burnier, C.V., Akar, G., 2009. Severity of injury resulting from pedestrian–vehicle crashes: what can we learn from examining the built environment? Transport. Res. Part D: Transport Environ. 14 (6), 425–436.

Conroy, C., Tominaga, G.T., Erwin, S., Pacyna, S., Velky, T., Kennedy, F., Sise, M., Coimbra, R., 2008. The influence of vehicle damage on injury severity of drivers in head-on motor vehicle crashes. Accid. Anal. Prevent. 40 (4), 1589–1594.

Daniel, J., Chien, S.I.-J., 2004. Truck safety factors on urban arterials. J. Transport. Eng. 130 (6), 742–752.

Easa, S.M., You, Q.C., 2009. Collision prediction models for three-dimensional two-lane highways. Transport. Res. Rec.: J. Transport. Res. Board 2092 (1), 48–56.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate poisson-lognormal regression. Accid. Anal. Prevent. 41 (4), 820–828.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accid. Anal. Prevent. 40 (3), 1033–1054.

Gårder, P., 2006. Segment characteristics and severity of head-on crashes on two-lane rural highways in maine. Accid. Anal. Prevent. 38 (4), 652–661.

Geedipally, S.R., Patil, S., Lord, D., 2010. Examination of methods to estimate crash counts by collision type. Transport. Res. Rec.: J. Transport. Res. Board 2165 (−1), 12–20.

Haleem, K., Abdel-Aty, M., 2010. Examining traffic crash injury severity at unsignalized intersections. J. Safety Res. 41 (4), 347–357.

Hauer, E., Council, F.M., Mohammedshah, Y., 2004. Safety models for urban four-lane undivided road segments. Transport. Res. Rec.: J. Transport. Res. Board 1897 (−1), 96–105.

Hausman, J.A., Hall, B.H., Griliches, Z., 1984. Econometric Models for Count Data with an Application to the Patents-r&d Relationship. National Bureau of Economic Research Cambridge, Mass., USA.

Hilbe, J.M., 2011. Negative Binomial Regression. Cambridge Univ Pr.

Hosseinpour, M., Prasetijo, J., Yahaya, A.S., Ghadiri, S.M.R., 2013. A comparative study of count models: application to pedestrian-vehicle crashes along Malaysia federal roads. Traffic Injury Prevent. 14 (6), 630–638.

Huang, H., Chin, H.C., 2010. Modeling road traffic crashes with zero-inflation and site-specific random effects. Statist. Methods Appl. 19 (3), 445–462.

Irtad, 2011. Road Safety Annual Report 2011. OECD/ITF.

Isgin, T., Bilgic, A., Forster, D.L., Batte, M.T., 2008. Using count data models to determine the factors affecting farmers' quantity decisions of precision farming technology adoption. Comput. Electron. Agric. 62 (2), 231–242.

Jonsson, T., Ivan, J.N., Zhang, C., 2007. Crash prediction models for intersections on rural multilane highways: differences by collision type. Transport. Res. Record: J. Transportati. Res. Board 2019 (−1), 91–98.

Kaplan, S., Prato, C.G., 2012. Risk factors associated with bus accident severity in the united states: a generalized ordered logit model. J. Safety Res. 43 (3), 171–180.

Khan, A., Ullah, S., Nitz, J., 2011. Statistical modelling of falls count data with excess zeros. Injury Prevent. 17 (4), 266–270.

Kim, D.-G., Washington, S., Oh, J., 2006. Modeling crash types: new insights into the effects of covariates on crashes at rural intersections. J. Transport. Eng. 132 (4), 282–292.

Kockelman, K.M., Kweon, Y.-J., 2002. Driver injury severity: an application of ordered probit models. Accid. Anal. Prevent. 34 (3), 313–321.

Kumara, S., Chin, H., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Injury Prevent. 4 (1), 53–57.

Kumara, S.S., Chin, H.C., Weerakoon, W.S., 2003. Identification of accident causal factors and prediction of hazardousness of intersection approaches. Transport. Res. Rec.: J. Transport. Res. Board 1840 (1), 116–122.

Kweon, Y.-J., Kockelman, K.M., 2005. Safety effects of speed limit changes: use of panel models, including speed, use, and design variables. Transport. Res. Rec.: J. Transport. Res. Board 1908 (1), 148–158.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in florida. Accid. Anal. Prevent. 37 (4), 775–786.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid. Anal. Prevent. 34 (2), 149–161.

Lord, D., 2006. Modeling motor vehicle crashes using poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accid. Anal. Prevent. 38 (4), 751–766.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transport. Res. Part A: Policy Practice 44 (5), 291–305.

Ma, J., Kockelman, K.M., 2006. Bayesian multivariate poisson regression for models of injury count, by severity. Transport. Res. Rec.: J. Transport. Res. Board 1950 (−1), 24–34.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods. Accid. Anal. Prevent. 40 (3), 964–975.

Mccartt, A.T., Northrup, V.S., Retting, R.A., 2004. Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in northern virginia. J. Safety Res. 35 (1), 107.

Merat, N., Jamson, A.H., 2013. The effect of three low-cost engineering treatments on driver fatigue: a driving simulator study. Accid. Anal. Prevent. 50, 8–15.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395–413.

Miranda-Moreno, L.F., Fu, L., 2006. A comparative study of alternative model structures and criteria for ranking locations for safety improvements. Networks Spat. Econ. 6 (2), 97–110.

Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accid. Anal. Prevent. 39 (3), 459–468.

Mohamed, N., Mohd-Yusoff, M.-F., Othman, I., Zulkipli, Z.-H., Osman, M.R., Voon, W.S., 2012. Fatigue-related crashes involving express buses in malaysia: will the proposed policy of banning the early-hour operation reduce fatigue-related crashes and benefit overall road safety? Accid. Anal. Prevent. 45, 45–49.

Neuman, T., 2008. Guidance for implementation of the aashto strategic highway safety plan: a guide for reducing head-on crashes on freeways. Transport. Res. Board.

Neuman, T., Pfefer, R., Slack, K., Hardy, K., Mcgee, H., Prothe, L., Eccles, K., Council, F., 2003. Nchrp Report 500 Volume 4: A Guide For Addressing Head-on Collisions. Transportation Research Board of the National Academies, Washington, DC.

Neuman, T., Slack, K., Hardy, K., Bond, V., Foss, R., Goodwin, A., Sohn, J., Torbic, D., Harwood, D., Potts, I., 2009. Nchrp Report 500: Guidance for Implementation of the Aashto Strategic Highway Safety Plan: Volume 23: A Guide for Reducing Speeding-related Crashes. Transportation Research Board of the National Academies, Washington, DC.

Nhtsa, 2007. Traffic Safety Facts 2005: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. U. S. Department of Transportation, Washington DC.

Noland, R.B., Quddus, M.A., 2005. Congestion and safety: a spatial analysis of london. Transport. Res. Part A: Policy Practice 39 (7), 737–754.

Obeng, K., 2008. Injury severity, vehicle safety features, and intersection crashes. Traffic Injury Prevent. 9 (3), 268–276.

Pai, C.-W., 2009. Motorcyclist injury severity in angle crashes at t-junctions: identifying significant factors and analysing what made motorists fail to yield to motorcycles. Safety Sci. 47 (8), 1097–1106.

Pai, C.-W., Saleh, W., 2008. Exploring motorcyclist injury severity in approach-turn collisions at t-junctions: focusing on the effects of driver's failure to yield and junction control measures. Accid. Anal. Prevent. 40 (2), 479–486.

Park, E.S., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. Transport. Res. Rec.: J. Transport. Res. Board 2019 (−1), 1–6.

Persaud, B.N., Retting, R.A., Lyon, C.A., 2004. Crash reduction following installation of centerline rumble strips on rural two-lane roads. Accid. Anal. Prevent. 36 (6), 1073.

Peterson, B., Harrell Jr., F.E., 1990. Partial proportional odds models for ordinal response variables. Appl. Statist., 205–217.

Qi, Y., Srinivasan, R., Teng, H., Baker, R., 2013. Analysis of the frequency and severity of rear-end crashes in work zones. Traffic Injury Prevent. 14 (1), 61–72.

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. J. Transport. Eng. 131 (5), 345–351.

Quddus, M.A., 2013. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and gis. J. Transport. Saf. Secur. 5 (1), 27–45.

Quddus, M.A., Noland, R.B., Chin, H.C., 2002. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. J. Safety Res. 33 (4), 445–462.

Quddus, M.A., Wang, C., Ison, S.G., 2009. The impact of road traffic congestion on crash severity using ordered response models. In: Proceedings of the Paper Presented to the 88th Annual Meeting of the Transportation Research Board, Washington, DC.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. Open access available at: http://www.R-project.org

Radun, I., Ohisalo, J., Radun, J.E., Summala, H., Tolvanen, M., 2009. Fell asleep and caused a fatal head-on crash? A case study of multidisciplinary in-depth analysis vs. the court. Traffic Injury Prevent. 10 (1), 76–83.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociol. Methodol. 25, 111–164.

Rich, J., Prato, C.G., Hels, T., Lyckegaard, A., Kristensen, N.B., 2013. Analyzing the relationship between car generation and severity of motor-vehicle crashes in denmark. Accid. Anal. Prevent. 54 (0), 81–89.

Rifaat, S.M., Tay, R., De Barros, A., 2012. Urban street pattern and pedestrian traffic safety. J. Urban Des. 17 (3), 337–352.

Rmp, 2011. Statistical Report on Road Accidents in Malaysia 2010. Royal Malaysia Police, Traffic Branch, Royal Malaysia Police, Bukit Aman, Kuala Lumpur.

Rodgers, G.B., Leland, E.W., 2005. An evaluation of the effectiveness of a baby walker safety standard to prevent stair-fall injuries. J. Safety Res. 36 (4), 327–332.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prevent. 27 (3), 371–389.

Son, H., 2011. Development of crash prediction models with individual vehicular data. Transport. Res. Part C: Emerg. Technol..

Statacorp, 2009. Stata Statistical Software: Release 11.0. College Station, TX, USA.
Tijerina, L., Blommer, M., Curry, R., Greenberg, J., Kochhar, D., Simonds, C., Watson, D., 2010. Effects of adaptive lane departure warning system on driver response to a surprise event. Transport. Res. Rec.: J. Transport. Res. Board 2185 (1), 1–7.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: J. Econometr. Soc., 307–333.
Wang, C., Quddus, M.A., Ison, S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accid. Anal. Prevent. 43 (6), 1979–1990.
Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accid. Anal. Prevent. 40 (5), 1674–1682.
Wang, Z., Chen, H., Lu, J.J., 2009. Exploring impacts of factors contributing to injury severity at freeway diverge areas. Transport. Res. Rec.: J. Transport. Res. Board 2102 (−1), 43–52.
Wegman, F., 2004. Fewer Crashes and Fewer Casualties by Safer Roads. SWOV Institute for Road Safety Research.
Xie, Y., Zhang, Y., Liang, F., 2009. Crash injury severity analysis using bayesian ordered probit models. J. Transport. Eng. 135 (1), 18–25.
Yasmin, S., Eluru, N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accid. Anal. Prevent..
Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safety Sci. 47 (3), 443–452.
Zajac, S.S., Ivan, J.N., 2003. Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural connecticut. Accid. Anal. Prevent. 35 (3), 369–379.
Zhang, C., Ivan, J.N., 2005. Effects of geometric characteristics on head-on crash incidence on two-lane roads in connecticut. Transport. Res. Rec.: J. Transport. Res. Board 1908 (−1), 159–164.