Accident Analysis and Prevention 57 (2013) 55–66

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models

Ximiao Jiang[a], Baoshan Huang[a,*], Russell L. Zaretzki[b], Stephen Richards[a], Xuedong Yan[c], Hongwei Zhang[d]

[a] Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN 37996, United States
[b] Department of Statistics, Operations & Management Science, The University of Tennessee, Knoxville, TN 37996, United States
[c] MOE Key Laboratory for Transportation Complex Systems Theory and Technology, School of Traffic & Transportation, Beijing Jiaotong University, Beijing 100044, PR China
[d] Chang'an University, Xi'an, Shanxi, PR China

ARTICLE INFO

Article history:
Received 26 June 2012
Received in revised form 21 February 2013
Accepted 5 March 2013

Keywords:
Traffic injury severity
Zero-inflated ordered probit model
Single-vehicle crashes
Curb
Shoulder
Speed limit

ABSTRACT

The severity of traffic-related injuries has been studied by many researchers in recent decades. However, previous research has seldom accounted for the effects of curbed outside shoulders on traffic-related injury severity. This study applies the zero-inflated ordered probit (ZIOP) model to evaluate the influences of curbed outside shoulders, speed limit change, as well as other traditional factors on the injury severity of single-vehicle crashes. Crash data from 2003 to 2007 in the Illinois Highway Safety Database were employed in this study.

The ZIOP model assumes that injury severity comes from two distinct sources: injury propensity and injury severity when this crash falls into the *injury prone* category. The modeling results show that on one hand, single-vehicle crashes that occurring on roadways with curbed outside shoulders are more likely to be *injury prone*. On the other hand, the existence of a curb decreases the likelihood of severe injury if the crash was in the *injury prone* category. As a result, the marginal effect analysis implies that the presence of curbs is associated with a higher likelihood of no injury and minor injury involved crashes, but a lower likelihood of incapacitating injury and fatality involved crashes. In addition, in the presence of curbed outside shoulders, the change of speed limit adds no significant impact to the injury severity of single-vehicle crashes.

Moreover, the modeling results also highlight some interesting effects caused by vehicle type, light and weather conditions, and drivers' characteristics, as well as crash type and location. Through a comprehensive evaluation of the modeling results, the authors find that the ZIOP model performs well relative to the traditional ordered probit (OP) model, and can serve as an alternative in future studies of crash injury severity.

1. Introduction

Because of the suburbanization of city peripheries, the traffic volume of the original two-lane rural highways, typically with 55 mph speed limits, may exceed their capacities. To enhance driving quality and reduce traffic delay, many two-lane rural highways need to be widened to four-lanes to mitigate traffic pressure. In these highway improvement projects, curbs frequently are installed along the new four-lane highways to provide drainage and address issues such as access control, difficult terrain, and limited right-of-way.

According to AASHTO, curbs are used extensively on all types of urban highways but should be used cautiously on high-speed rural highways because they may cause vehicles to lose control or go airborne during impact (AASHTO, 2005). Therefore, the use of curbs is discouraged in many states in the U.S. on roadways with design speeds over 45 mph. When they are necessary for drainage control or other special functions, only certain types of curbs are permitted and must be used with specified placement requirements.

The authors conducted a survey of curb usage across the U.S. Each state's road design manuals and relevant guidelines, if any, were reviewed. The survey results show that 31 of 50 states have accessible regulations pertaining to the use of curbs. These guidelines reveal the effectiveness and performance of curbs and corresponding regulations on their use in high-speed highways.

* Corresponding author. Tel.: +1 865 789 5532; fax: +1 865 974 2669.
E-mail addresses: xjiang10@utk.edu (X. Jiang), bhuang@utk.edu (B. Huang), rzaretzk@utk.edu (R.L. Zaretzki), stever@utk.edu (S. Richards), yanxuedong@jtys.bjtu.edu.cn (X. Yan), zhwxm2002@yahoo.com.cn (H. Zhang).

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.03.018

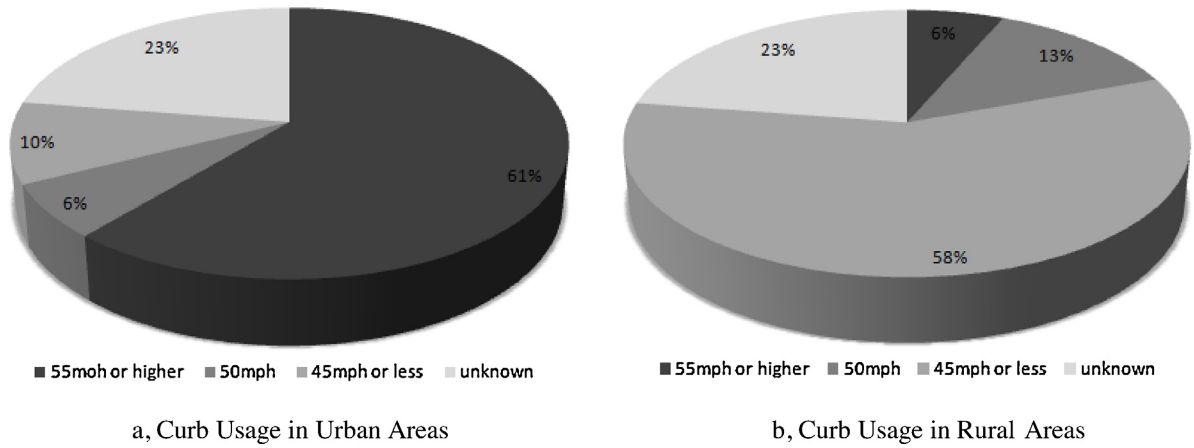

Fig. 1. States usage of curbs by speed limitation on highways.

The regulations vary among urban and rural highways in different states. Fig. 1 shows the percentage of states that discourage the use of curbs on highways exceeding specific speeds. The chart on the left depicts urban areas while the right-hand chart shows rural areas.

It can be observed from Fig. 1 that more than 60% of the US states allow the use of curbs on urban highways even at speed limits of 55 mph or above, while the corresponding figure is only 6% on rural highways. On the other hand, the percentage of states that only allow the use of curbs at the speed limit of 45 mph or less is 6% on urban highways, but more than 50% on rural highways. These results reveal to us that curbs are mostly discouraged on rural highways, especially at speed limits of 55 mph or above. According to the guidelines of many DOTs, the typical unfavorable effects that curbs produced are: the trend to influence drivers' behavior, which may increase the possibility to crash with the adjacent vehicles; the loss of vehicle control when impacting on curbs, even leading to roll over.

In recent years, a few researchers have particularly investigated the effects of curbs on the occurrences of crashes. For example, Baek et al. (2006) conducted a study to identify the mean crash rate differences between roadways with speed limits of 45 mph and those with speed limit of 55 mph, both with curbs next to the outside lanes in both directions. The results suggest that there is no significant difference on crash rate in both total and curb-involved crashes when speed limits change. This study was based on the data collected from 60 selected sites in North Carolina. The same authors (2008) studied 2274 collisions occurring on 191.85 miles of directional multilane highway segments in North Carolina from 2001 to 2003. The results indicate that multilane highways with curbed outside shoulders were associated with fewer total collisions and equal injury collisions when compared to no curbs, and segments with speed limit of 55 mph had more collisions than those with speed limits of 45 mph. Jiang et al. (2011) employed the Illinois Accident Data from 2003 to 2007 to analyze the effects of outside shoulder curbs on crash frequency. The results suggest that the presence of curbed outside shoulders on high-speed roads do not result in a significantly higher crash frequency when compared to other types of outside shoulder, and lowering speed limit from 55 mph to 45 mph in the presence of curbs is not necessary to reduce crash frequency.

However, the study of curbs as related to traffic injury severity is very limited. Plaxico et al. (2005) reported that single vehicle involved curb-related fatal crashes on roadways with speed limits of 40 mph (65 km/h) and above represent a very small percentage of total fatal crashes. Their successive research used a logistic regression model to compare the severity of all single-vehicle curb crashes and single-vehicle non-curb crashes under similar conditions. The results imply that in locations where curbs are located, single-vehicle crashes involving curbs were clearly no more severe than crashes involving other roadside objects. However, their study derived the effects of curbs only through the involvement of curbs in crashes; it did not account for the possibility that the presence of curbs may be a factor in non-curb-involved crashes. Moreover, the extent to which curbs may be involved in a crash might not be correctly identified in databases, owing to the complexity of crashes.

The objective of the present study was to develop a zero inflated ordered probit (ZIOP) model to analyze the effect of curbs along outsider shoulders on the injury severity of single-vehicle crashes. The ZIOP model is able to account for the preponderance of no injury crashes as well as to allow the effects of contributing factors to vary across injury levels. The traffic crash data were acquired from Illinois Highway Safety Database for the years 2003–2007. This research specifically focused on the analysis of crashes that occurred on non-freeway, non-intersection related, two-lane and four-lane road segments with posted speed limits of 45, 50, and 55 mph.

2. Literature review

The severity of traffic-related crashes is generally represented by a set of discrete categories describing the property loss and the injury of occupants, such as property damage only, minor injury, incapacitating injury and fatality. In the last few decades, a large variety of statistical models have been used to evaluate the influence of risk factors on the severity of traffic related crash injuries (Savolainen et al., 2011). These statistical models can be categorized into two general classes: nonparametric models and regression models.

2.1. Nonparametric models

Nonparametric models are known for their ability to explore the most significant contributing factors and achieve impressive goodness-of-fit. The major nonparametric models that have been applied in traffic injury studies include but are not limited to the decision tree based models (Chang and Wang, 2006; Berchialla et al., 2007), Neural Networks (Sohn and Shin, 2001; Sohn and Lee, 2003), and artificial neural networks (Abdelwahab and Abdel-Aty, 2001, 2002; Delen et al., 2006). Despite the great contribution of nonparametric models in either mining injury severity determinants or improving the goodness-of-fit, these models are not widely adopted because of the difficulty in interpreting the quantitative effect for each variable of interest.

2.2. Regression models

Due to the apparent shortcomings of nonparametric models, many researchers focus on regression based models to study injury severity in traffic accidents. One useful classification for regression type models that are utilized in this context considers two categories; Unordered-response models (UR) such as multinomial logit models (Agresti, 2007; Hensher et al., 2005; Train, 2009) are general models for multilevel responses but may ignore the natural ordering of the response while ordered-response (OR) models such as those based on the standard ordered probit model (SOP) (Cameron and Trivedi, 2009) and standard ordered logit (SOL), also known as the proportional odds model (Agresti, 2007), implicitly take this ordering into account.

2.2.1. Unordered-response models

Studies that have taken the unordered approach include binary logit and probit models (Al-Ghamdi, 2002; Winston et al., 2006; Lee and Abdel-Aty, 2008), multinomial logit (MNL) models (Shankar and Mannering, 1996; Khorashadi et al., 2005), nested MNL models (Shankar et al., 1996; Chang and Mannering, 1999; Lee and Mannering, 2002), mixed MNL models (Milton et al., 2008; Gkritza and Mannering, 2008), and markov switching MNL models (Malyshkina and Mannering, 2009). The nominal approach to the injury severity problem is more robust to under-reporting rates across injury severity levels of data in one category, e.g. non-injury events (Ye and Lord, 2011) and also provides additional flexibility in modeling variable effects across response categories (Savolainen et al., 2011; Castro et al., 2013). However, such approaches do not account for the ordering of injury data and thus important information may be lost.

2.2.2. Ordered-response models

As noted above, models that treat traffic injury severity as an ordered response variable are also popular in traffic injury studies. Standard ordered response models (SOR) include standard ordered logit (SOL) models (McCarthy and Madanat, 1994; O'Donnell and Connor, 1996; Pai and Saleh, 2008) and standard ordered probit (SOP) Models (Abdel-Aty, 2003; Shimamura et al., 2005; Rifaat and Chin, 2007). Unfortunately, these conventional ordered response models have been criticized due to their susceptibility to biased and inconsistent parameter estimation caused by the underreporting of crash-injury, particularly non-injury, data (Savolainen et al., 2011; Ye and Lord, 2011). Furthermore, these models restrict the effect of each variable to be monotonic across injury levels which may be unrealistic for some explanatory variables. Savolainen et al. (2011) illustrate this weakness by considering the effect of airbags on injury severity in a traffic crash. Holding other factors constant, for an accident with air bag deployment the SOR model will, due to its structure, estimate an increased probability of no injury and a diminished probability of fatal injury in comparison to a case without deployment. In fact, it is likely that the air bag will decrease the probability of no injury and fatal injury while increasing the probability of minor injury. The SOR approach does not have the flexibility to capture these effects.

In recent years, several researchers have dedicated their efforts to improve ordered response models from different perspectives. For example, Dissanayake and Lu (2002) employed sequential logit models to predict the crash severity outcome of single-vehicle fixed-object crashes involving young drivers. The sequential logit/probit generalizes the ordered logit/probit model by relaxing the restrictions of parameters imposed by standard ordered probability models and was shown to be robust in the presence of underreporting (Yamamoto and Shankar, 2004). One noted limitation is the fact that error terms associated with each level of injury severity are independent (Savolainen et al., 2011).

De Lapparent (2008) developed a bivariate ordered probit mode to analyze the decision to fasten the safety belt in a car and the resulting severity of accidents. This approach is closely related to econometric selection models and is designed account for overstating effects due to self-selection; see Savolainen et al. (2011) and Wooldridge (2002) for additional background on models for selection bias.

Wang and Abdel-Aty (2008) and Quddus et al. (2010) employed the generalized ordered logit (GOL) model in traffic injury studies. The GOL model, also known as the partial proportional odds model, relaxes the assumption of traditional ordered response models that the parameter estimates are constant across severity levels by allowing more flexibility. However, this model is often criticized for its potential to generate negative probability estimates (Greene and Hensher, 2010).

Eluru et al. (2008) developed a mixed generalized ordered response logit (MGORL) model to evaluate determinants of injury severity in pedestrian and bicycle involved crashes. The mixed (random) parameter model allows the parameters of the latent injury risk propensity to vary across observations, i.e., the effects of contributing factors on injury severity for each crash could be different. In addition, unlike traditional ordered response models that have fixed threshold values, the MGORL model allows for the observation specific thresholds that depend on explanatory variables. These features provide the ability to account for the unobserved heterogeneity among crashes and also resolve the problems with GOL models discussed above. In applying this model to accident severity of pedestrians and cyclists struck by automobiles, Eluru et al. (2008) reported a significant improvement in fit in comparison to the SOL model. However, in this context the mixed parameter approach did not offer any benefit beyond an approach that used only fixed parameters. Greene and Hensher (2010) applied a fixed parameter probit version of the model (HOPIT) in a social science application and showed that it also provides significant improvement over OP and generalized ordered probit (GOP) models. Castro et al. (2013) extended this approach to a spatial random coefficients generalized ordered response probit model. This model explicitly incorporates spatial dependency effects into the mixed generalized ordered response (MGOR) structure. Their case study indicates that the MGORP structure is superior to the traditional ordered-response structure in terms of capturing the variable effects of parameters across responses as well as overall fit. Moreover, this study clearly reveals the presence of spatial effects caused by site-specific unobserved factors that affect the latent injury propensity.

Finally, there are researchers who employed Bayesian inference to fit ordered response models (Xie et al., 2009; Jiang et al., 2012). This work concludes that Bayesian models can produce superior parameter estimates for small sample data than maximum likelihood estimation and can also provide the flexibility in model construction by considering the prior knowledge in the estimation contributing factors effects and thresholds.

2.3. Zero-inflated models

As noted above ordered response models have a number of strengths including their logical consistency in modeling ordered probabilities and interpretability. For this reason, the authors are motivated to use them despite the flexibility of nominal models. However, in addition to the weaknesses mentioned above, OP have difficulty fitting the abundance of *no injury* crashes which is a key feature of accident injury severity data. Therefore, researchers seek to modify the OP model attempting to maintain its interpretive features while avoiding bias due to a poor fit.

A growing body of literature has made efforts to address the so-called zero-inflated issues in crash frequency models. For example, Lord et al. (2005) employed the zero-inflated negative binomial (ZINB) and zero-inflated Poisson (ZIP) model to analyze motor vehicle crashes. Results indicate that ZINB models provide a better fit than conventional Poisson and negative binomial models. Zero-inflated models involve two states of modeling that allow the mixing of response counts with respect to zeros. These models are a special case of the finite mixture model. The latter one can be extended to many states and allows mixing with respect to both zeros and positives (Cameron and Trivedi, 1998). More recently, Malyshkina et al. (2010) developed the zero-state Markov switching negative binomial (MSNB) model to estimate five-year crash frequencies in Indiana. The MSNB model allows specific roadway segments to switch between the zero states and non-zero states over time. In their paper, the MSNB model was compared to the ZINB model. Results show that the zero-state MSNB model is a viable alternative and produces a superior fit relative to the ZINB model. Even though two-state models perform better than conventional count models, there is a dispute that the so-called zero state in crash frequency data may not physically exist (Lord et al., 2007).

Despite the considerable efforts described above to deal with the overabundance of zero counts in studies of numbers of accidents, very limited work has been done to adopt this methodology to account for the preponderance of non-injury observations in ordered response models. Harries and Zhao (2007) developed a zero-inflated ordered probit (ZIOP) model to estimate levels of tobacco consumption. The proposed ZIOP model is an extension of the basic ordered probit model that takes into account of the possibility that the zeros can arise from two distinct sources. In their research, the ZIOP model has a hierarchical set of two equations. In the upper level a probit selection equation is used to identify groups of individuals that will/will not consume tobacco while the lower level contains an OP model that estimates the probability of consuming a certain number of tobacco products given their potential to consume. Through application, Harries and Zhao (2007) demonstrate a few advantages of the ZIOP model over the conventional OP model. In particular the proportion of zeros that come from different sources may be controlled by one group of explanatory variables and, in addition, the effects of explanatory variables are allowed to vary across response classes.

Inspired by these advantages, the authors adopt the ZIOP model to study the effect of factors discussed above on injury severity. Although adoption of this approach may be justified by the fact that this mixture model improves the resulting fit, the authors attempt to rationalize this approach by claiming that the large proportion of zeros may come from two distinct populations. The first population is the group of *injury free accidents*, which we define as crashes that cannot result in an injury. Although this is an artificial construction, we might hypothesize that these would be low speed fender bender type accidents that cannot result in any type of reportable injury. Secondly, among accidents occurring with enough energy to cause harm to passengers which we refer to as *injury-prone*, the characteristics of vehicle, the driving style and the roadway will on average effect the severity of injury that occurs. Factors such as special apparatuses that protect drivers, pavement quality, and posted speed limits, and weather may be associated with injury severity. Similarly crashes occurring on highly congested roadways tend to be *injury free*, no matter what the posted speed limit is. Thus, it is likely that these two types of zeros are driven by alternative crash mechanisms. By explicitly considering both populations in the model, it will produce a superior model fit and less bias in our estimates.

3. Methodology

3.1. ZIOP model

The typical derivation of the standard OP model assumes that there is a latent continuous metric underlying the ordered responses whose stochastic component follows a normal distribution. The construction and fitting algorithm of the conventional OP model are described in Cameron and Trivedi (1998). In contrast, the data generation process assumed by the ZIOP has two levels: *Level 1* determines whether the accident is *injury-free* or *injury-prone* while *Level 2* determines the severity of injury-prone crashes through the OP mechanism just described.

Let s denote a binary variable indicating *injury free* ($s=0$) and *injury prone* ($s=1$). s is related to a latent variable γ^* through the criteria: $s=0$ for $\gamma^* < 0$ and $s=1$ for $\gamma^* > 0$. The latent variable γ^* represents the propensity of injury involvement and is given by

$$r_i^* = x_i^T\beta + \varepsilon_i \quad i = 1, \ldots, n \tag{1}$$

where $x_i = \left\{x_{i1}, \ldots, x_{ic}, \ldots, x_{ic}\right\}^T$ represents covariates in identifying injury propensity, and $x_i = \left\{\beta_1, \ldots, \beta_c, \ldots, \beta_c\right\}$ is the corresponding vector of parameters to be estimated. Accordingly, the probability of a crash being in the *injury prone* category is given by (Maddala, 1983)

$$P(s = 1|x) = P(r^* > 0|x) = \Phi(x_i^T\beta) \tag{2}$$

Conditional on $s=1$, the observed injury level $\widetilde{y} = \left\{\widetilde{y}_1, \ldots, \widetilde{y}_i, \ldots, \widetilde{y}_n\right\}$ can be connected to a latent variable y^* through a function $g(y_i^*)$.

$$y_i^* = w_i^T h + \mu_i \tag{3}$$

where $w_i = \left\{w_{i1}, \ldots, w_{id}, \ldots, w_{iD}\right\}^T$ represents explanatory variables in the second process of the ZIOP model; $h = \left\{h_1, \ldots, h_d, \ldots, h_D\right\}$ is associated vector of coefficients to be estimated; μ_i is a random error term (assumed to follow an independent and identically distributed standard normal distribution). The mapping between $\widetilde{y}_i$ and y_i^* is obtained by

$$\widetilde{y}_i = g(y_i^*) = \left\{\begin{array}{ll} 0 & \text{if } -\infty = \gamma_0 < y_i^* \le \gamma_1 \\ 1 & \text{if } \gamma_1 < y_i^* \le \gamma_2 \\ \vdots & \vdots \\ J & \text{if } \gamma_J < y_i^* \le \gamma_{J+1} = +\infty \end{array}\right\} \tag{4}$$

where $\gamma = \left\{\gamma_0, \gamma_1, \ldots, \gamma_J, \gamma_{J+1}\right\}$, is the threshold value for all categories, wherein $\gamma_0 = -\infty$, $\gamma_{J+1} = +\infty$, and the remaining threshold values are subjected to the constraint $\gamma_1 \le r_2 \le \cdots \le \gamma_J$. The function $g(\cdot)$ is taken to be non-decreasing, so that small and large values of y_i^* can be interpreted as corresponding to small and large values of $\widetilde{y}_i$.

In the current work, the response $\widetilde{y}$ has four categories: no injury, minor injury, incapacitating injury, and fatality. In addition, throughout this paper we assume that $\gamma_1 = 0$. Note that *Level 2* also allows for zero injury level, i.e., *no injury* and there is no requirement that $w_i = x_i$ so that separate explanatory factors can be used in both equations. Under the assumption that μ is standard Gaussian, *Level 2* creates the probability of each injury level as follows, conditional on $s=1$:

$$\begin{aligned} &P(\widetilde{y}_i = 0) = \Phi(-w_i^T h) \\ &P(\widetilde{y}_i = 1) = \Phi(\gamma_2 - w_i^T h) - \Phi(-w_i^T h) \\ &P(\widetilde{y}_i = J) = 1 - \Phi(\gamma_J - w_i^T h) \end{aligned} \tag{5}$$

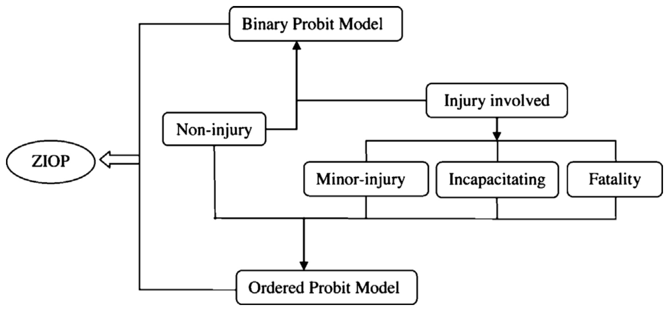

Fig. 2. Sketch of the ZIOP model.

While s and $\widetilde{y}$ are not individually observable in terms of the zeros, they are observable via the criterion

$$y = s \cdot \widetilde{y} \tag{6}$$

That is, to observe a $y = 0$ outcome, we require either that $s = 0$ (*injury free*) or jointly that $s = 1$ and $\widetilde{y} = 0$ (*injury prone* but no injury involved). To observe a positive y, we require jointly that the crash is *injury prone* ($s = 1$) and that the crash happened to be injury involved, $\widetilde{y} > 0$. Under the assumption that ε and μ identically and independently follow standard Gaussian distributions, the full probabilities for observed y are given by

$$P(y) = \left\{ \begin{array}{l} P(y=0|W,X) = P(s=0|X) + P(s=1|X)P(\widetilde{y}=0|W,s=1) \\ P()y=j|W,X = P(s=1|X)P(\widetilde{y}=j|W,s=1) \quad (j=1,\ldots,J) \end{array} \right\} \tag{7}$$

$$\left\{ \begin{array}{l} P(y=0|W,X) = \left[1-\Phi(X'\beta)+\Phi(X'\beta)\Phi(-W'h)\right] \\ P(y=j|W,X) = \Phi(X'\beta)\left[\Phi(u_{j+1}-W'h)\right] - \Phi(u_j - W'h) \quad (j=1,\ldots,J-1) \\ P(y=J|W,X) = \Phi(X'\beta)\left[1-\Phi(u_J - W'h)\right] \end{array} \right\}$$

It can be observed from this equation that the probability of a zero observation has been "inflated since the probability of no injury in a particular accident is a sum of the probability of the event being *injury free* and the probability of no injury given an *injury prone* event.

The parameters of the full model $\theta = (\beta', h', \gamma')'$ can be consistently and efficiently estimated using the maximum likelihood (ML) criteria. The log-likelihood function is given by

$$l(\theta) = \sum_{i=1}^{n}\sum_{j=0}^{J} l_{ij} \ln\left[P(y_i = j|x_i, w_i, \theta)\right] \tag{8}$$

where the indicator function l_{ij} is:

$$l_{ij} = \left\{ \begin{array}{ll} 1 & \text{if individual } i \text{ chooses outcome } j \\ 0 & \text{otherwise} \end{array} \right. \quad (i=1,\ldots,n;\ j=0,1,\ldots,J) \Bigg\}$$

The overall structure of the ZIOP is sketched in Fig. 2. The unconditional probability of positive injury is also a combination of the probability of being *injury prone* and the conditional probability of each injury level.

It may be useful at this point to contrast the ZIOP approach with the MGORL model which, to our knowledge, is the most sophisticated and flexible ordered response model that has been applied to injury severity data. It begins with a standard SOL/SOP latent variable formulation for the latent injury risk propensity. More importantly, an additive exponential form that is a function of accident dependent explanatory variables is proposed for the thresholds. Random parameters can be introduced at both levels to handle additional heterogeneity although these may be unnecessary (Eluru et al., 2008). In comparison, like standard SOL or SOP models, the ZIOP maintains fixed thresholds. The added flexibility in the ZIOP comes from the mixture distribution proposed for the non-injury category. Although the mathematical forms are quite different, the effect here is similar to a GOL model that allows one set of parameter coefficients for the first category while all more severe categories share a common set of coefficients. The advantage is that the mixture form ensures that the probabilities are strictly positive so that the model is logically consistent. The model improves upon the standard SOL/SOP formulation but falls short of the flexibility of the MGORL. To its advantage, however, it is computationally easier to implement and has a simpler interpretation due to fewer parameters. For example, applying the MGORL we might expect 13 parameters in the injury propensity equation as well as 8–13 for each of the 3-thresholds. In contrast the ZIOP uses 9 parameters in level 1, 13 parameters in level 2, and 2 for thresholds. A natural next step is to compare the models under a variety of conditions such as cases where data is censored or the effects depend upon severity level to gain a better perspective on relative strengths of both models.

3.2. Marginal effects and model evaluation

In order to evaluate the amount of change in the probability of certain injury severity with the change of each factor, marginal effects were computed. In the case under study, people may be interested in the marginal effects of each variable on the probability of being *injury free*, $P(s = 0)$ the probabilities of certain injury level given the crash is *injury prone*, $P(\widetilde{y} = j|s = 1)$, or on the overall probabilities of each injury level, $P(y = j)$.

The marginal effect of a dummy variable is defined as the difference in the predicted probability of each injury level with this dummy variable assigned to 0 and 1 in the predicted models, and holding all other covariates fixed. For example, the marginal effect of curbed outside shoulder type on minor injury, can be calculated from Eq. (9). The marginal effect of a continuous variable g can be calculated by $\partial P(\cdot)/\partial g$.

$$ME\,[P(y=1)] = \frac{1}{n}\sum_{i=1}^{n}[P(y_i = 1|x_{i1},\ldots,curb=1,\ldots,x_{iC},w_{i1},\ldots, curb=1,\ldots,w_{id}) - P(y_i=1|x_{i1},\ldots,curb=0,\ldots,x_{iC},w_{i1},\ldots curb=0,\ldots,w_{iD})] \tag{9}$$

Readers may refer to the work of Harries and Zhao (2007) for a detailed derivation.

3.3. Model evaluation

In this study, the authors also are interested in comparing the fit of the ZIOP model against traditional OP model. Therefore, the traditional OP model was also built with all factors included. Akaike's information criteria (AIC) (Akaike, 1973), Bayesian information criteria (BIC) (Sawa, 1978), and the Vuong test (Vuong, 1989; Greene, 1994) were employed to compare these two models. The model with the smaller AIC and BIC among all competing models is deemed the better model.

Vuong's test (Vuong, 1989) can outperform AIC in comparing models from different regression series. Given Eq. (10), define

$$m_i = \text{Log}\left(\frac{P_1(y_i\,|\,x_i, z_i)}{P_2(y_i\,|\,x_i, z_i)}\right) \tag{10}$$

where $p_N(y_i\,|x_i, z_i)$ is the estimated probability of the observed severity for event i using model N. Then, the Vuong statistic, V, is computed to test whether the two models are significantly different

Table 1
Variable definitions.

Variables	Description	Categories and classification criteria	
outshtp	Type of outside shoulder	Curb	Curbed shoulder
		Non-curb	None curb shoulder
medtype	Type of median	Flush	Flush median
		None	No median
		Raised	Raised median
no_lanes	Number of lanes	2	2 lane
		4	4 lane
spd_limt	Speed limit	45	45 mph (72 km/h)
		50	50 mph (80 km/h)
		55	55 mph (88 km/h)
rururb	Rural or urban	Rural	Rural area
		Urban	Urban area
light	Light condition	Daylight	Daylight
		Nighttime	Dark lighted and unlighted, dawn and dusk
weather	Weather condition	Clear	Clear
		Inclement	Rainy, foggy, snowy, etc.
phour	Peak hour	Yes	6:00–10:00 and 16:00–19:00
		No	10:00–16:00 and 19:00–6:00
vehtype	Vehicle type	Pass_car	Convertible, hatchback, sedan, etc.
		Light_trk	Pickup, SUV, minivan, small bus
drvage	Driver's age single-vehicle	Youngers	Driver's age in between 16 and 24
		Middle_aged	Driver's age in between 25 and 64
		Olders	driver's age ≥65
drvsex	Driver's gender single-vehicle	Male	Driver is male
		Female	Driver is female
crashtype	Type of crash	Animal	Vehicle collide with animals
		Overturn	Vehicle overturned in crashes
		Object	Vehicle impact with objects
surfyear	Year of surface construction	Pre_2000	Road surface constructed before 2000
		Post_2000	Road surface constructed after 2000
frst_loc	First involvement location	On_pave	Crash occurred on the roadway
		Off_pave	Crash occurred off the roadway
AADT		Annual average daily traffic in thousand	
Lanewid		Lane width in feet	
Outshwd		Outside shoulder width in feet	

in predicting the observed severity or not, which can be expressed as Eq. (11).

$$V = \frac{\sqrt{n}((1/n)\sum_{i=1}^{n} m_i)}{\sqrt{(1/n)\sum_{i=1}^{n}(m_i - \overline{m})^2}} \tag{11}$$

If $V > 1.96$, the first model is preferred with the second preferred if $V < 1.96$.

4. Data preparation

The Illinois Highway Safety Database (IHSD) was selected for this study. Illinois is one member of the Highway Safety Information System (HSIS), which is a multistate database that contains crash, roadway inventory, and traffic volume data for a select group of states. The HSIS dataset contains the subset of 1985–2007 crashes that occurred on the Illinois State-inventoried system and includes accident data, a roadlog file, a bridge (Structures) file, and a railroad (RR) grade crossing file (Council and Mohamedshah, 2009). The IHSD Accident Data from 2003 to 2007 and corresponding roadlog file were extracted and linked to each other using three common variables: county, route, and milepost. Overall there are 891,682 crashes extracted from the original database.

Explanatory factors considered in the present study include: outside shoulder type, median type, number of lanes, speed limit, rural or urban, light conditions, weather conditions, peak hour, vehicle type, driver's age, driver's gender, crash type, surface year, first involve location, annual average daily traffic (AADT), lane width, and outside shoulder width. Here, AADT was calculated through dividing the original AADT by 1000 so that a non-trivial marginal effect could be reported. The outside shoulder type specified in this paper is restricted to the shoulders adjacent to both directions of the travel lanes. All discrete factors were categorized into subclasses based on the original records as shown in Table 1.

The combined dataset was further cleaned according to the following criteria:

- Only single-vehicle crashes are studied in this paper;
- Freeway segments were excluded since the use of curbs on freeways is limited to special conditions in Illinois and specific features of curbs are required as well (IDOT, 2010);
- Intersection related crashes were excluded as the traffic and collision patterns are different from continuous segments;
- Speed limits of 45, 50 and 55 mph (72, 80 and 88 kph) were selected since this study focuses on high-speed roadways;
- The road segments with two lanes and four lanes were selected since they are the most common in sampled area;
- Pedestrian, pedal cyclist, motorcycle, large truck, and non-traditional vehicles involved accidents were screened out because they are minor and have different injury mechanisms;

Table 2
Injury severity distribution across key variables.

Distribution of injury severity across levels of categorical variables											
Variables	Categories	Injury severity								Total	
		0 (No injury)		1 (Minor injury)		2 (Incapacitating)		3 (Fatality)			
outshtp	Curb	2764	88.50%	296	9.48%	59	1.89%	4	0.13%	3123	6.77%
	Non_curb	38,491	89.44%	3514	8.17%	949	2.21%	82	0.19%	43,036	93.23%
medtype	Flush	4173	89.34%	389	8.33%	104	2.23%	5	0.11%	4671	10.12%
	None	33,286	89.55%	2980	8.02%	829	2.23%	76	0.20%	37,171	80.53%
	Raised	3796	87.93%	441	10.22%	75	1.74%	5	0.12%	4317	9.35%
no_lanes	2	32,741	89.59%	2923	8.00%	814	2.23%	69	0.19%	36,547	79.18%
	4	8514	88.58%	887	9.23%	194	2.02%	17	0.18%	9612	20.82%
spd_limt	45	5205	87.74%	589	9.93%	125	2.11%	13	0.22%	5932	12.85%
	50	3366	88.79%	341	8.99%	81	2.14%	3	0.08%	3791	8.21%
	55	32,684	89.70%	2880	7.90%	802	2.20%	70	0.19%	36,436	78.94%
rururb	Rural	27,641	90.15%	2280	7.44%	678	2.21%	62	0.20%	30,661	66.42%
	Urban	13,614	87.84%	1530	9.87%	330	2.13%	24	0.15%	15,498	33.58%
light	Daylight	10,720	81.78%	1827	13.94%	519	3.96%	42	0.32%	13,108	28.40%
	Nighttime	30,535	92.39%	1983	6.00%	489	1.48%	44	0.13%	33,051	71.60%
weather	Clear	33,385	90.45%	2694	7.30%	770	2.09%	61	0.17%	36,910	79.96%
	Inclement	7870	85.09%	1116	12.07%	238	2.57%	25	0.27%	9249	20.04%
phour	Yes	18,629	89.74%	1640	7.90%	455	2.19%	34	0.16%	20,758	44.97%
	No	22,626	89.08%	2170	8.54%	553	2.18%	52	0.20%	25,401	55.03%
vehtype	Pass_car	24,976	88.42%	2553	9.04%	665	2.35%	54	0.19%	28,248	61.20%
	Light_trk	16,279	90.89%	1257	7.02%	343	1.92%	32	0.18%	17911	38.80%
drvage	Youngers	9115	83.31%	1457	13.32%	342	3.13%	27	0.25%	10941	23.70%
	Middle_aged	28941	91.45%	2101	6.64%	565	1.79%	40	0.13%	31647	68.56%
	Olders	3199	89.58%	252	7.06%	101	2.83%	19	0.53%	3571	7.74%
drvsex	Male	24117	91.01%	1863	7.03%	465	1.75%	54	0.20%	26499	57.41%
	Female	17138	87.17%	1947	9.90%	543	2.76%	32	0.16%	19660	42.59%
crashtype	Animal	30023	96.73%	817	2.63%	194	0.63%	3	0.01%	31,037	67.24%
	Overturn	1078	49.13%	847	38.61%	265	12.08%	4	0.18%	2194	4.75%
	Object	10,154	78.54%	2146	16.60%	549	4.25%	79	0.61%	12,928	28.01%
surfyear	Pre_2000	30,089	89.17%	2849	8.44%	746	2.21%	59	0.17%	33,743	73.10%
	Post_2000	11,166	89.93%	961	7.74%	262	2.11%	27	0.22%	12,416	26.90%
frst_loc	On_pave	32,393	95.73%	1179	3.48%	258	0.76%	8	0.02%	33,838	73.31%
	Off_pave	8862	71.93%	2631	21.35%	750	6.09%	78	0.63%	12,321	26.69%

- Crashes with evidence of not using a seatbelt were excluded to minimize the effects of seatbelts on injury severity, hence seatbelt usage will not be taken as an explanatory variable;
- Crashes happening on "potentially standard" horizontal curves ($D \geq 2.5$) were removed to eliminate the inherent effect of horizontal curves on injury severity. In other words, all the crashes studied in this paper can be deemed to have happened on standard ($D < 2.5$) horizontal curves;
- All the crashes considered have drivers under normal conditions, i.e., the effects of alcohol, drug, fatigue and illness were eliminated.

The treated dataset applied in this paper includes 46,159 single-vehicle crashes that occurred on state routes and highways within the five years from 2003 to 2007. These crashes occurred on road segments having AADT (in thousands) ranges from 0.025 to 79.70, with an average value of 9.01; lane width ranges from 8 to 27 feet with an average of 11.85 feet; outside shoulder width ranges from 0 feet to 16 feet, with an average of 5.25 feet. In this study, the severity level of each crash was determined by the injury level of the most severely injured occupant in all of the vehicles involved. Four levels of injuries were identified from the original database: no injury, minor injury, incapacitating injury, and fatality. The average percentage of each injury level in ascending order is 89.38%, 8.25%, 2.18% and 0.19%, separately. Table 2 provides the distribution of injury severity by each explanatory variable.

It can be observed from Table 2 that a curbed shoulder was associated with lower percentages of no injury, incapacitating injury and fatality, but a higher percentage of minor injuries in the overall sampled crashes, in comparison with the non-curb shoulder type. A number of other variables also show noticeable differences in severity across levels. For example, night time was associated with a remarkably higher percentage of no injury crashes but with a lower percentage of minor injury crashes compared with day light condition; inclement weather conditions were more likely to result in more severe injuries in comparison with clear weather conditions; both younger drivers and older drivers were associated with a relatively higher percentage of injury involved crashes; more than half of "overturn" involved crashes resulted in injuries, and crashes that involved impacts with objects also resulted in a high likelihood of injury (around 23%). Collisions with animals had a much lower injury probability compared with the average event. Finally, crashes happening off pavement were more likely to result in injuries in comparison with on pavement crashes.

To quantitatively estimate the influences of these variables on the injury severity of single-vehicle crashes, the ZIOP model was fit and compared with the traditional OP model. In addition, the marginal effect of each explanatory condition was calculated in order to aid interpretation.

Table 3
Parameter estimates and model performance.

Parameter		Binary probit process			Ordered probit process		
		Estimate	SE	P value	Estimate	SE	P value
continuous variables	AADT	0.0066	0.0042	0.1167	−0.0065	0.0014	<0.0001
	Lanewid	–	–	–	−0.0991	0.0035	<0.0001
vehtype	Pass_car						
	Light_trk	−0.2202	0.0635	0.0005	0.0152	0.0308	0.6224
light	Daylight						
	Nighttime	−0.4166	0.0664	<0.0001	0.0330	0.0345	0.3392
weather	Clear						
	Inclement	0.4518	0.1170	0.0001	−0.3510	0.0331	<0.0001
drvsex	Female						
	Male	–	–	–	−0.2133	0.0205	<0.0001
drvage	Youngers						
	Middle_aged	−0.4809	0.0719	<0.0001	0.0771	0.0355	0.0301
	Olders	−0.8798	0.1020	<0.0001	0.4484	0.0780	<0.0001
outshtp	Non_curb						
	Curb	0.5208	0.1938	0.0072	−0.2167	0.0549	<0.0001
phour	Yes						
	No	–	–	–	0.0345	0.0186	0.0318
frst_loc	On_pave						
	Off_pave	1.0735	0.2086	<0.0001	0.1349	0.0599	0.0245
crashtype	Animal						
	Object	0.5424	0.1508	0.0003	0.4785	0.0773	<0.0001
	Overturn	2.1410	0.4797	<0.0001	0.9761	0.0765	<0.0001
u_1	1|2	–	–	–	0.9863	0.0200	<0.0001
u_2	2|3	–	–	–	2.0751	0.0443	<0.0001

Model comparison	ZIOP model	OP model
No. of observation	46,159	46,159
Log-likelihood	−15,333.23	−15,424.67
AIC	30,712.46	30,875.35
BIC	30,913.48	30,988.97
Vuong's test (ZIOP/OP)	5.6444	

Note: SE indicates standard error.

5. Results

5.1. Parameter estimates

In this section we review and interpret the parameters resulting from the fit of the ZIOP model. In the first step, variables identified as significant at the 95% confidence level by the traditional binary OP model were selected as candidates for the first level of the ZIOP model. All of the individual variables considered in this paper, as well as the two-way interacting variables spd_limt*outshtp, weather*outshtp and light*outshtp were selected as the candidates in the second process. After fitting the ZIOP model, a few variables that are not significant in both processes were removed and the ZIOP model was re-fitted. The final model includes all the variables that are significant in either process or both. These variables include: AADT, lane width, vehicle type, light conditions, weather conditions, driver's gender, driver's age, outside shoulder type, peak hour, first involved location, and crash type.

For comparison, a standard OP model with the same variables was also fit. The performance of each model is presented in Table 3. The results show that both the AIC and BIC values of the ZIOP model are significantly less, approximately 63 and 75 units respectively, than the OP model, indicating that the ZIOP model outperformed the OP model. For reference, an AIC difference of 2–3 units is usually considered significant. The result of the Vuong's test is 5.6444, which is much greater than the critical value 1.96, also suggesting that the ZIOP model is superior. In sum, in terms of AIC, BIC and the Vuong's test results, the ZIOP model appears to offer a clear improvement in the overall prediction of injury severity in comparison with the traditional OP model justifying our interest in this approach.

Parameter estimates from the ZIOP model are presented in Table 3. Parameter estimates provide the change in injury severity level in comparison to a reference attribute after holding fixed all of the other explanatory variables. Parameter estimates greater than 0 indicate that a particular cluster of a variable leads to a higher severity level, and vice versa.

To the authors' point of interest, the ZIOP model shows significant effects of curbed outside shoulders on traffic-related injury severity in single-vehicle crashes. In the first process, the coefficient for curb is significantly positive ($p = 0.0072$), indicating that crashes occurring on roadways with curbed outside shoulders were more likely to be *injury prone*. In the second process, curb is remarkably negative ($p < 0.0001$), which instead indicates that conditional on being *injury prone*, a curb is more likely to be associated with a lower level of injury in comparison to no curb. Using the discussion of energy above, curbs may be associated with higher energy accidents but may absorb more energy on average leading to lower event severity. The fact that crashes on roadways with curbs are less likely to result in "overturn" crash type (2.08%) as compared to crashes on non-curb roadways (4.95%) supports this argument. Since overturn has been proved to be associated with more severe injuries than other types of single vehicle crashes by both a previous study (Plaxico et al., 2005) and the findings in Table 3, it is expected that the presence of curbs is associated with less severe injuries. Alternatively, one may conjecture that drivers tend to drive more slowly when curbs present on shoulders. This assumption, if true, would also reduce injury severity.

Among other factors, the increase of AADT is not significant in injury propensity prediction, but is significantly associated with the lower injury severity level conditional on being in the *injury prone* group. Light trucks were associated with a significantly lower likelihood of *injury prone* crashes as compared to passenger cars, but did not show any significant difference on the conditional injury severity. The finding on injury propensity is to some extent consistent with the result reported by Krull et al. (2000), which indicates that injury severity increases with passenger cars as opposed to pickup trucks. The lower propensity of injury might be attributed to the higher rigidity of a light truck as opposed to a passenger car, or perhaps a difference in the driver population. Meanwhile, light truck vehicles are found to have a relatively higher percentage of “overturn” crashes (5.83%) in the sampled dataset in comparison with passenger cars (4.07%), which may cancel out the positive effect of light trucks and result in insignificant effect on the probability of conditional injury severity. The unstable characteristics of light truck vehicles have been discussed by many scholars. For example, NHTSA (1997) reported that light truck vehicles have a higher rate of rollover crashes as opposed to passenger cars.

The location of crashes was significant in both the injury propensity and injury severity processes. Off-pavement crashes were more likely to be *injury prone* and led to more severe injuries conditional on being in the *injury prone* category, as opposed to on-pavement crashes. Nighttime was associated with a significantly lower probability of *injury prone* crashes, but made no difference on the conditional injury severity as compared to day time. The less *injury prone* finding is consistent with many previous studies (Krull et al., 2000; Savolainen and Ghosh, 2008), since drivers are prone to drive more carefully at night. However, because traffic volume is lower at night, this may encourage drivers to maintain higher speeds, and visibility is poorer as well. These may cancel out the positive effect of driver’s attention and lead to an insignificant effect on the conditional injury severity. Inclement weather conditions significantly increased the likelihood of *injury prone* crashes as opposed to clear weather, but also notably decreased the chance of conditional severe injury. This is reasonable since inclement weather often results in slippery roads, spray and splashing and other unfavorable driving conditions that will interrupt driving maneuvers and vehicles trajectory in an emergency. On the other hand, in inclement weather, drivers may tend to pay more attention to roadways and other vehicles, and drive at lower speeds than they did in clear weather.

From the driver’s side, middle-aged and older drivers were less likely to be involved in *injury prone* crashes. This might be because they have more experience and drive less aggressively than younger drivers. On the contrary, old drivers were associated with a significantly higher probability of severe injury conditional on being *injury prone* than were younger drivers. This makes sense considering that older drivers may have poor health and are likely to react slowly in emergencies.

As for factors considered only in the second process of the ZIOP model, the increase of lane width significantly decreased the severity of injuries. Peak hour significantly positively affects injury severity conditional on being *injury prone* by decreasing the probability of severe injury in comparison to non-peak hour. This may be reasonable because peak hours may indicate lower driving speeds. Driver’s sex is extremely significant ($p < 0.0001$) in injury level estimation conditional on being in *injury prone* category. The result shows that crashes by male drivers are less likely to produce severe injuries than crashes by females. This may be caused by physiological and behavioral differences, and is supported by quite a few other scholars (Abdelwahab and Abdel-Aty, 2001; Evans, 1988, 2001). It is noteworthy that none of the proposed interacting variables is significant in the model, which to some extent answers our original question: is it necessary to drop the speed limit on roadways with curbs installed on outside shoulders? The insignificance of the interaction between outside should type and speed limit suggests that the change of speed limit does not add to the effects of curbs on injury severity.

5.2. Marginal effects

Because of the nonlinear nature of the OP model, it is difficult to directly interpret the parameters in a physically meaningful way. Therefore, marginal effects are computed in order to understand the impact of various factors on injury severity. Marginal effects are simply derivatives and must be evaluated at a particular point of the covariate space. There are two approaches to obtain the marginal effects: one is to evaluate the marginal effects at the joint mean of the covariates while the other option is to evaluate them separately for each observation and report the average across all observations. The authors choose to apply the second approach in this paper.

The marginal effects of each factor on the injury level probabilities are given in Tables 4 and 5. For comparison purposes, the authors also include the marginal effects computed from traditional OP model. Note that for variables appearing in both X and Z, the authors have combined the two parts of the marginal effects. In Table 4, the marginal effects on $P(y=0)$ using both ZIOP model and OP model are presented. For the ZIOP model, the authors also decompose the overall marginal into two parts: the effect on *injury free*, $P(s=0)$, and the effect on no injury conditional on being *injury prone*, $P(y=0|s=1)$. In Table 5, the marginal effects on the unconditional probabilities of all three positive injury levels are presented.

The marginal effects in Tables 4 and 5 highlight interesting differences from alternative models for some of the explanatory factors. A key example is the effect of outside shoulder. The ZIOP model shows that curbed outside shoulders brought about a 13.08% decrease in the probability of *injury free* crashes, but a 13.45% rise in the probability of no injury conditional on being *injury prone*. The former marginal effect shows that crashes that occurred on roadways with the presence of curbed outside shoulder were less likely to be firm safe as compared to non-curb outside shoulders. The latter marginal effect indicates that curbs were associated with a relatively higher likelihood of no injury compared with other outside shoulder types, conditional on being *injury prone*. As illustrated in the last a few sections, the ZIOP model is based on the assumption that zero observations come from two distinct sources. This allows flexibility in how the overall probability of a non-injury outcome is distributed. Combining the opposing effects of curb results in cancelation and an extremely small increase (0.36%) on the total probability of observing no injury. This is smaller than the parameter estimates from the OP model (1.44%).

In addition, the resulting marginal effect on the unconditional probabilities of each injury level also come from two sources. Table 5 reveals to us that curb as opposed to non-curb, were associated with 0.17% increase in the probability of minor injury, but 0.44% and 0.09% decrease in the probability of incapacitating injury and fatality, respectively. However, with only one latent variable, the conventional OP model shows a monotonically decreasing positive effect in the probabilities of three categories of injury involved crashes.

The marginal effects of other covariates presented in Tables 4 and 5 are not illustrated in detail because of the space

Table 4
Marginal effects for no injury category.

Parameter		OP	ZIOP		
		$P(y=0)$	$P(s=0)$	$P(y=0\|s=1)$	$P(y=0)$
continuous variables	AADT	0.0005	−0.0016	0.0022	0.0005
	Lanewid	0.0029	–	0.0126	0.0126
vehtype	Pass_car				
	Light_trk	0.0082	0.0553	−0.0472	0.0082
light	Daylight				
	Nighttime	0.0108	0.1046	−0.0898	0.0149
weather	Clear				
	Inclement	0.0317	−0.1135	0.1373	0.0238
drvsex	Female				
	Male	0.0285	–	0.0271	0.0271
drvage	Youngers				
	Middle_aged	0.0042	0.1208	−0.1086	0.0122
	Olders	−0.0150	0.2210	−0.2376	−0.0166
outshtp	Non_curb				
	Curb	0.0144	−0.1308	0.1345	0.0036
phour	Yes				
	No	−0.0066	–	−0.0044	−0.0044
frst_loc	On_pave				
	Off_pave	−0.0574	−0.2697	0.2034	−0.0662
crashtype	Animal				
	Object	−0.1179	−0.1363	0.0507	−0.0855
	Overturn	−0.2111	−0.5379	0.3161	−0.2218

Table 5
Marginal effects for minor injury, incapacitating and fatality categories.

Parameter		Minor injury, $P(y=1)$		Incapacitating, $P(y=2)$		Fatality, $P(y=3)$	
		OP	ZIOP	OP	ZIOP	OP	ZIOP
continuous variables	AADT	−0.0004	−0.0003	−0.0001	−0.0002	<0.0001	<0.0001
	Lanewid	−0.0020	−0.0082	−0.0008	−0.0039	−0.0001	−0.0005
vehtype	Pass_car						
	Light_trk	−0.0057	−0.0070	−0.0022	−0.0011	−0.0003	<0.0001
light	Daylight						
	Nighttime	−0.0075	−0.0129	−0.0029	−0.0019	−0.0004	<0.0001
weather	Clear						
	Inclement	−0.0219	−0.0119	−0.0086	−0.0102	−0.0012	−0.0017
drvsex	Female						
	Male	−0.0197	−0.0176	−0.0077	−0.0083	−0.0011	−0.0012
drvage	Youngers						
	Middle_aged	−0.0029	−0.0117	−0.0011	−0.0007	−0.0002	0.0002
	Olders	0.0103	0.0039	0.0041	0.0107	0.0006	0.0020
outshtp	Non_curb						
	Curb	−0.0099	0.0017	−0.0039	−0.0044	−0.0006	−0.0009
phour	Yes						
	No	0.0046	0.0028	0.0018	0.0013	0.0003	0.0002
Frst_loc	On_pave						
	Off_pave	0.0396	0.0515	0.0155	0.0135	0.0022	0.0013
crashtype	Animal						
	Object	0.0814	0.0598	0.0319	0.0228	0.0046	0.0029
	Overturn	0.1457	0.1610	0.0572	0.0544	0.0082	0.0064

limitations. We suggest that readers to use the results of Table 3 as a reference for the marginal effects.

6. Conclusions and discussion

There have been numerous efforts to investigate how the outcome of crashes relates to roadway design features, environmental conditions, drivers' characteristics, and traffic features. However, very few earlier studies have specifically estimated the effect on accident injury severity of curbed outside shoulders on high-speed roadways, as well as the possible effects that a speed limit change might produce given curbs have been installed. On the basis of crash data collected from Illinois Highway Safety Information database from 2003 to 2007, the influences of curbed outside shoulders, speed limit, as well as other traditional traffic safety factors, on the injury severity of single-vehicle crashes have been estimated using the ZIOP model.

The results suggest that on one hand, crashes occurring on roadways with curbed outside shoulders are more likely to be *injury prone*. On the other hand, a curb is more likely to be associated with lower levels of injury compared with no curb, conditional on being *injury prone*. In the other word, the presence of a curb negatively affects the propensity of injury, but positively affects the severity of injury given that the crash is injury prone. Overall, the presence of curbs along outside shoulders increased the probability of no injury and minor injury crashes, but decreased the likelihood of incapacitating injury and fatal crashes. The increase of speed limit does not produce significant impact on the severity of injury, given that a single-vehicle crash has occurred on roadways with curbs installed.

Other factors exhibiting statistically significant influence on injury severity include:

- Higher AADT and lane width, as well as the presence of peak hour were associate with lower level of injury severity given that crashes fall in the *injury prone* category;
- Light trucks as opposed to passenger cars and night time as opposed to day time were both associated with significantly lower likelihood of *injury prone* crashes, but did not bring about significant change to injury severity conditional on being *injury prone*;
- Inclement weather conditions significantly increased the likelihood of *injury prone* crashes as opposed to clear weather, but also notably decreased the chance of severe injury conditional on being *injury prone*;
- Male drivers were less likely to suffer severe injuries as opposed to females. Middle aged and older drivers were less likely to be involved in *injury prone* crashes, but older drivers had remarkably higher chances to be severely injured conditional on being *injury prone*;
- Off-pavement crashes versus on-pavement crashes, impacts with objects, and overturn crashes versus collision with animals crashes, were more likely to be *injury prone*, and result in more severe injuries conditional on being *injury prone*.

Methodologically, this paper has introduced the ZIOP model into the traffic safety literature for the first time. Implementation of the ZIOP model has resulted in a significantly superior statistical fit than the conventional OP model and has resulted in some significant relative differences in marginal effects. As intended, the improvement likely results from the improved fit to the non-injury category. Employing the ZIOP model as an alternative in future studies of traffic injury severity is strongly suggested.

Returning to the weaknesses of OP models with respect to underreporting of events further research must be performed. However, the flexibility provided by allowing the point mass estimate for the non-injury category may provide a significant advantage. The effectiveness of the method may also depend on the self-selection process involved in deciding not to report the event and how various explanatory factors may be related to this. Both simulation and analytical calculation will be useful in studying this issue.

Through the empirical study we have also seen that ZIOP has some ability to modify the effects of factors across the non-injury and injury severity levels. Currently, however, the model will not allow change in the directionality of effects within the injury prone population so that examples such as the effect of driver air bag on moderate and serious injury discussed in Savolainen et al. (2011) will still not be adequately addressed. Beyond comparisons between the ZIOP and MGORL discussed above, future research will focus on allowing random coefficients or variable thresholds as discussed in the MGORL model to provide the necessary flexibility. Because of the difficulty fitting the MGORL using conventional maximum likelihood methods, a Bayesian formulation of this model may also provide a useful stepping stone allowing the eventual development of a logically consistent and flexible model for ordered data.

Last but not the least, this paper is also limited by the incompleteness of the database so that a few typical variables were not accounted for in this research, such as type and geometric design of curbs, pavement conditions, vertical alignment and so on.

Acknowledgement

The authors would like to thank both the editor and reviewers for their patience and numerous thoughtful comments which improved the quality of the paper dramatically. The authors would also like to thank the Tennessee Department of Transportation for funding this research and HSIS for providing the crash data. The authors would also like to acknowledge support of Chinese National Natural Science Foundation (71210001).

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. , pp. 267–281.

Abdelwahab, H.T., Abdel-Aty, M.A., 2001. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transportation Research Record 1746, 6–13.

Abdelwahab, H.T., Abdel-Aty, M.A., 2002. Artificial neural networks and logit models for traffic safety analysis of toll plazas. Transportation Research Record 1784, 115–125.

Al-Ghamdi, A., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accident Analysis and Prevention 34 (6), 729–741.

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research 34 (5), 597–603.

AASHTO, 2005. A Policy on Geometric Design of Highways and Streets. Washington, DC.

Agresti, A., 2007. An Introduction to Categorical Data Analysis. A John Wiley & Sons, Inc., Hoboken, NJ.

Baek, J., Hummer, J., Williams, B.M., Cunningham, C.M., 2006. Reasonable speed limits on suburban multilane highways with curbs. Transportation Research Record 1969, 10–11.

Berchialla, P., Snidero, S., Stancu, A., Corradetti, R., Gregori, D., 2007. Predicting severity of foreign body injuries in children in upper airways: an approach based on regression trees. Risk Analysis 27 (5), 1255–1263.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.

Cameron, A.C., Trivedi, P.K., 2009. Microeconometrics: Methods and Applications. Cambridge University Press, Cambridge, UK.

Castro, M., Paleti, R., Bhat, C., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis and Prevention 52, 188–203.

Chang, L., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. Accident Analysis and Prevention 31, 579–592.

Council, F.M., Mohamedshah, Y.M., 2009. Guidebook for the Illinois State Datafiles. FHWA, U.S. Department of Transportation.

Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis and Prevention 38, 1019–1027.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis and Prevention 38 (3), 434–444.

Dissanayake, S., Lu, J., 2002. Analysis of severity of young driver crashes, sequential binary logistic regression modeling. Transportation Research Record 1784, 108–114.

De Lapparent, M., 2008. Willingness to use safety belt and levels of injury in car accidents. Accident Analysis and Prevention 40 (3), 1023–1032.

Evans, L., 1988. Risk of fatality from physical trauma versus sex and age. Journal of Trauma 28 (3), 368–378.

Evans, L., 2001. Female compared with male fatality risk from similar physical impacts. Journal of Trauma 50 (2), 281–288.

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. Accident Analysis and Prevention 40 (3), 1033–1054.

Gkritza, K., Mannering, F., 2008. Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles. Accident Analysis and Prevention 40 (2), 443–451.
Greene, W.H., 1994. Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working Paper, Department of Economics, New York University.
Greene, W.H., Hensher, D.A., 2010. Modeling Ordered Choices. Cambridge University Press, Cambridge, UK.
Hensher, D., Rose, J., Greene, D., 2005. Applied Choice Analysis. Cambridge University Press, Cambridge, UK.
Harries, M.N., Zhao, X.Y., 2007. A zero-inflated ordered probit model with an application to modeling tobacco consumption. Journal of Econometrics 141, 1073–1099.
IDOT, 2010, Illinois Department of Transportation, Bureau of design and environment manual-cross section elements 32-2.04.
Jiang, X.M., Yan, X.D., Huang, B.S., Richards, S., 2011. Influence of curbs on traffic crash frequency on high-speed roadways. Traffic Injury Prevention 12 (4), 412–421.
Jiang, X.M., Huang, B.S., Zaretzki, R., Richards, S., Yan, X.D., 2012. Two-vehicle injury severity models based on integration of pavement management and traffic engineering factors. Traffic Injury Prevention, http://dx.doi.org/10.1080/15389588.2012.731547.
Krull, K., Khattak, A., Council, F., 2000. Injury effects of rollovers and events sequence in single-vehicle crashes. In: Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, DC.
Khorashadi, A., Niemerier, D., Shankar, V., Mannering, F.L., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. Accident Analysis and Prevention 37 (5), 910–921.
Lee, J., Mannering, F.L., 2002. Impact of roadside features on the frequency and severity of run-off-roadside accidents: an empirical analysis. Accident Analysis and Prevention 34 (2), 149–161.
Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37 (1), 35–46.
Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero inflated models in highway safety. Accident Analysis and Prevention 39 (1), 53–57.
Lee, C., Abdel-Aty, M., 2008. Presence of passengers: does it increase or reduce driver's crash potential? Accident Analysis and Prevention 40 (5), 1703–1712.
Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, UK.
McCarthy, P.S., Madanat, S., 1994. Highway accident data analysis: alternative econometric methods. Transportation Research Record 1467, 44–49.
Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40 (1), 260–266.
Malyshkina, N.V., Mannering, F.L., 2009. Markov switching multinomial logit model: an application to accident-injury severities. Accident Analysis and Prevention 41, 829–838.
Malyshkina, N.V., Mannering, F.L., 2010. Zero-state Markov switching count-data models: an empirical assessment. Accident Analysis and Prevention 42 (1), 122–130.
National Highway Traffic Safety Administration (NHTSA), 1997. Traffic safety facts 1997. US Department of Transportation. National Highway Traffic Administration.
O'Donnell, C.J., Connor, D.H., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention 28 (6), 739–753.
Plaxico, C.A., Ray, M.H., Weir, J.A., Orengo, F., Tiso, P., 2005. Recommended guidelines for curb and curb-barrier installations. Transport. Res. Board, NCHRP Report 537.
Pai, C.W., Saleh, W., 2008. Modelling motorcyclist injury severity by various crash types at T-junctions in the UTK. Safety Science 46 (8), 1234–1247.
Quddus, M., Wang, C., Ison, S., 2010. Road traffic congestion and crash severity: an econometric analysis using ordered response models. Journal of Transportation Engineering 136, 424–435.
Rifaat, S.M., Chin, H.C., 2007. Accident severity analysis using ordered probit model. Journal of Advanced Transportation 41 (1), 91–114.
Sawa, T., 1978. Information criteria for discriminating among alternative regression models. Econometrica 46, 1273–1282.
Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of Safety Research 27 (3), 183–194.
Shankar, V., Mannering, F., Barfield, W., 1996. Statistical analysis of accident severity on rural freeways. Accident Analysis and Prevention 28 (3), 391–401.
Shimamura, M., Yamazaki, M., Fujita, G., 2005. Method to evaluate the effect of safety belt use by rear seat passengers on the injury severity of front seat occupants. Accident Analysis and Prevention 37, 5–17.
Savolainen, P.T., Ghosh, I., 2008. Examination of factors affecting driver injury severity in Michigan's single-vehicle-deer crashes. Transportation Research Record 2078, 17–25.
Sohn, S., Shin, H., 2001. Pattern recognition for a road traffic accident severity in Korea. Ergonomics 44 (1), 101–117.
Sohn, S.Y., Lee, S.H., 2003. Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea. Safety Science 41 (1), 1–14.
Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis and Prevention 43 (5), 1666–1676.
Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307–333.
Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, London, UK.
Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis and Prevention 40 (5), 1674–1682.
Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: The case of airbags and antilock brakes. Journal of Risk and Uncertainty 32 (2), 83–99.
Xie, Y.C., Zhang, Y.L., Liang, F.M., 2009. Crash injury severity analysis using Bayesian ordered probit models. Journal of Transportation Engineering 135 (1), 18–25.
Yamamoto, T., Shankar, V., 2004. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects. Accident Analysis and Prevention 36 (5), 869–876.
Ye, F., Lord, D., 2011. Investigating the effects of underreporting of crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit and mixed logit models. Transportation Research Record 2241, 51–58.