Accident Analysis and Prevention 146 (2020) 105730

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

An energy loss-based vehicular injury severity model

Ang Ji*, David Levinson

The University of Sydney, School of Civil Engineering, Sydney, Australia

ARTICLE INFO

Keywords:
Injury severity
Regression model
Vehicle crashes
Energy absorption

ABSTRACT

How crashes translate into physical injuries remains controversial. Previous studies recommended a predictor, Delta-V, to describe the crash consequences in terms of mass and impact speed of vehicles in crashes. This study adopts a new factor, energy loss-based vehicular injury severity (ELVIS), to explain the effects of the energy absorption of two vehicles in a collision. This calibrated variable, which is fitted with regression-based and machine learning models, is compared with the widely-used Delta-V predictor. A multivariate ordered logistic regression with multiple classes is then estimated. The results align with the observation that heavy vehicles are more likely to have inherent protection and rigid structures, especially in the side direction, and so suffer less impact.

1. Introduction

The prevention of traffic crashes requires an in-depth analysis of the severity of crashes. The term 'severity' in the traffic safety literature may refer to the probability of crashes, the damage magnitude, or both (Shelby, 2011). The degree of crash harm depends on the relationship between physical injuries and crash mechanisms, but understanding is often limited by complicated crash mechanics (Carlson, 1979). Therefore, statistical models, which exploit previous crash data, are used to predict injury severity with several explanatory variables. For example, factors in most vehicular injury analyses include crash mechanical properties, driver characteristics (such as age, gender, and alcohol or drug use), safety equipment, vehicle use, and surrounding conditions (Massie et al., 1995; Abdel-Aty et al., 1998, 2011; Yau, 2004; Chang and Wang, 2006; Keall et al., 2004; Smink et al., 2005).

There is no consensus on how to evaluate severity. To assess multiple discrete severity outcomes, the Abbreviated Injury Score (AIS) and Maximum Abbreviated Injury Score (MAIS) are frequently used to classify the extent of injuries (Carlson and Kaplan, 1975; Carlson, 1977; Augenstein et al., 2003; Digges and Dalmotas, 2001; Conroy et al., 2008). They have seven classifications in general: 0 – Property Damage Only (PDO), 1 – Minor Injury, 2 – Moderate Injury, 3 – Serious Injury, 4 – Severe Injury, 5 – Major Injury, and 6^+ – Fatality (Untreatable cases). Similarly, there are also some indices like Injury Severity Score (ISS) evaluating one injury severity score derived from three different body regions (Greenspan et al., 1985; Ehrlich et al., 2006). This paper uses MAIS for the severity evaluation, as it was recorded per single vehicle and was commonly provided by the datasets employed here.

To specify the crash severity, Mackay (1968), applying momentum conservation theory, first defined Delta-V as the change of speed before and after the collision. The probability and magnitude of injuries and fatalities are often based on the Delta-V of involved vehicles. Delta-V has been applied for the measurement of vehicular crash severity for several decades (Joksch, 1993; Roberts and Compton, 1993; Evans, 1994). Evans (1994) evaluated this model as one of the best analytical methods when real crash testing is unavailable. The fundamental mechanical meaning of Delta-V is related to the forces that occupants receive inside the crashing vehicles (Carlson, 1979). In reaction to the imposed force after the collision, occupants strike the surfaces in the front of the vehicle, which is the main reason for injuries in most cases. However, the recent application of Delta-V depends on four crash modes (Front, Near side, Far side, and Rear detected by the event data recorder), which may not be comparable so that different directions yield different conclusions (Andricevic et al., 2018). Delta-V fails to consider the influence of long crash pulse on injury (Ydenius, 2010; Tsoi and Gabler, 2015). The limitations of Delta-V have prompted the discussion of alternative crash indicators.

Injury severity studies typically establish the statistical relationship between the dependent variable, injury severity, and several independent variables. Ordered logistic regression (OLR) models and their variants, with different levels of severity assessments, perform well (Yasmin et al., 2015; Ferreira et al., 2017; Quddus et al., 2002; Duncan et al., 1998; Kockelman and Kweon, 2002; Ye and Lord, 2014; Bogue et al., 2017; Abay et al., 2013). Table 1 summarizes severity indices, crash indicators, and regression models in previous automobile crash severity studies.

* Corresponding author.
E-mail address: ang.ji@sydney.edu.au (A. Ji).

https://doi.org/10.1016/j.aap.2020.105730
Received 25 October 2019; Received in revised form 23 June 2020; Accepted 5 August 2020
Available online 21 August 2020

Acronyms

AIC	Akaike Information Criterion
AIS	Abbreviated Injury Score
CDS	Crashworthiness Data System
CI	confidence interval
CIREN	Crash Injury Research and Engineering Network
CRASH	Calspan Reconstruction of Accident Speeds on Highways
ISS	Injury Severity Score
MAIS	Maximum Abbreviated Injury Score
MICE	Multiple Imputations by Chain Equations
MNLR	multinomial logistic regression
NASS	National Automotive Sampling System
oEES	Objective Energy Equivalent Speed
OLR	ordered logistic regression
OLS	ordinary least squares
PDOF	principal direction of force
POLR	proportional odds ordered logistic regression
RD	Residual Deviance
RFE	Recursive Feature Elimination
ELVIS	Model of Energy loss-based vehicular injury severity
Delta-V	Model of probability and magnitude of injuries and fatalities

Table 1
Model types of previous injury severity studies with Delta-V.

Model	Severity index	Main crash indicator	Regression model
Carlson (1979)	Average AIS	Longitudinal ΔV *intervals*	Ordinary least squares (OLS)
Joksch (1993)	Probability of fatality	Longitudinal $(\Delta V)^4$	OLS
Evans (1994)	Probability of injury and fatality	Longitudinal $(\Delta V)^{3.54}$ (fatalities); $(\Delta V)^{2.22}$ (injuries)	OLS
Ydenius (2010)	MAIS (≥ 3, < 3)	Mean and peak acceleration	Binomial logistic
Tsoi and Gabler (2015)	MAIS (≥ 3, < 3)	Occupant impact velocity, acceleration severity index, and vehicle pulse index	Binomial logistic
Augenstein et al. (2003), Kononen et al. (2011) and Nishimoto et al. (2017)	MAIS (≥ 3, < 3) or ISS (≥ 15, < 15)	ΔV in four crash modes	Binomial logistic
Andricevic et al. (2018)	MAIS (≥ 2, < 2) and (≥ 3, < 3)	Objective Energy Equivalent Speed (oEES)	Binomial logistic

However, few logistic models have been estimated to explain injury severity due to impact energy. This study uncovers the correlation between the injury severity of occupant and a multi-directional energy-related predictor.

This paper considers dynamic crash mechanisms and then proposes an energy loss-based vehicular injury severity (ELVIS) evaluation indicator. The model is first calibrated by a dataset with separate two types of collision, considering the longitudinal and lateral impacts simultaneously. It aims to include both magnitude and direction elements in its expression. By comparing with the Delta-V, we explain the occupant injury severity due to energy absorption in a specific collision. Additionally, because of the discrete and ordinal nature of the dependent variable, the ordered logistic regression model is used to estimate model coefficients. The rest of the paper is structured as followed: Section 2 introduces how data get processed in this study, and Section 3 provides a theoretical framework for the ELVIS function. In Section 4, we evaluate the regression model by the calibrated parameter and different probability distributions; meanwhile, comparing the Delta-V model with the new model with their statistical fitness and predicted accuracy. The discussion (Section 5) interprets the findings from the ELVIS model. A brief conclusion (Section 6) summarizes the outcomes.

2. Materials

The injury severity analysis in this paper first investigates crashes with the two-vehicle straight-line and side-impact collision only and then extends to all types of crashes. In these cases, we apply energy loss theory to explain the injury mechanisms. MAIS, which evaluates the most severe injury in each collided vehicle, is the injury severity assessment in following procedures.

The National Automotive Sampling System (NASS)/ Crashworthiness Data System (CDS), which contains a sample of representative crash information from the NASS database collected across the US, is employed for this research. It collects weighted data at 24 Primary Sampling Units (PSUs) and aims to produce statistical relationships between biomedical and engineering evidence in police-

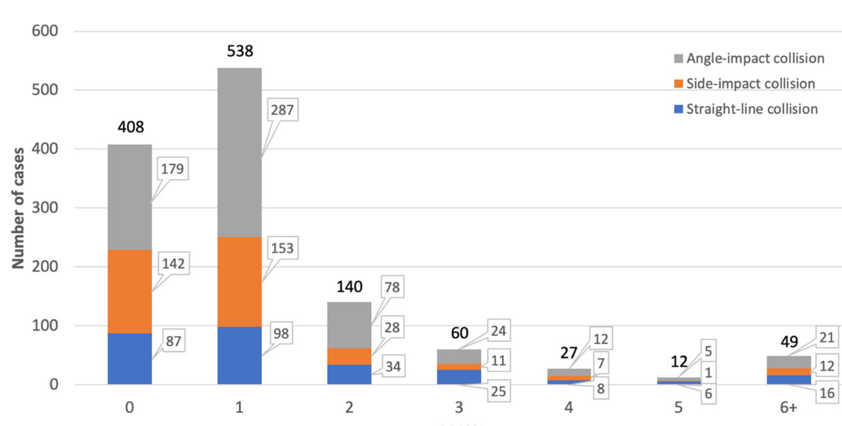


Fig. 1. MAIS distribution in the NASS/CDS dataset with two-vehicle crashes only.

reported automobile crash investigation (Radja, 2016). We applied two-vehicle crashes for the years 2009–2015 (coding methodology changed in 2008) that mainly recorded the information about occupants, vehicles, and crash scenarios. Crash events with passenger cars only are examined. The MAIS distribution of this dataset is shown in Fig. 1, which reveals the higher frequency of lower severity levels (MAIS 0 and 1).

This paper estimates the injury severity at the vehicle level, which means each observation represents a vehicle involved in crashes. This vehicle-based study considers factors that describe the general information of collided vehicles and their external environment conditions. A variable selection procedure (as illustrated in Fig. 2), employing the Recursive Feature Elimination (RFE) method and the elimination of highly-correlated attributes, is able to reduce redundancy and over-fitting. RFE is a backward selection method by establishing a model (random forest in this study) on all features, calculating importance scores, and removing the least important ones (Guyon et al., 2002). We then removed cases with omitted items in necessary variables (MAIS, car weight, and computed Delta-V) and preserved only two-vehicle involved cases. Missing values of other variables are replaced by an algorithm called Multiple Imputations by Chain Equations (MICE). A 1234-case all-type dataset (including 274 injuries in straight-line crashes and 354 in side-impact crashes) results. Table 2 describes the selected variables for the ELVIS model.

Table 2
Description of variables applied in this analysis.

Variable	Variable description	Frequency	Share	Total
Dependent variables				
MAIS	0 – Property Damage Only	408	33.06%	1234
	1 – Minor Injury	538	43.60%	
	2 – Moderate Injury	140	11.35%	
	3 – Serious Injury	60	4.86%	
	4 – Severe Injury	27	2.19%	
	5 – Major Injury	12	0.97%	
	6^+ – Fatality (Untreatable)	49	3.97%	
Independent variables (essential variables &other top five attributes)				
Delta-V (with angle)	Values of Delta-V (in m/s)	–	–	1234
W_L (ELVIS)	Energy loss from the specific crash (in joules)	–	–	1234
Number of occupants	The number of occupants seated in this vehicle (1–6)	–	–	1234
Restraint type	1 – Airbags front only	382	30.96%	1234
	2 – Airbags front &side	847	68.64%	
	3 – Passive belts or unknown	5	0.40%	
Road alignment	1 – Straight	1045	84.68%	1234
	2 – Curve right	102	8.27%	
	3 – Curve left	87	7.05%	
Road profile	1 – Level	898	72.77%	1234
	2 – Uphill	166	13.45%	
	3 – Crest	21	1.70%	
	4 – Downhill	147	11.91%	
	5 – Sag	2	0.17%	
Surface type	1 – Asphalt	1088	88.17%	1234
	0 – Others	146	11.83%	

3. Methodology

3.1. Newtonian mechanics in crashes

In typical two-vehicle crashes, the destructive energy is dissipated first in the deformation of designed protective structures and second in decelerating the uncrashed structures (Wood, 1997). Then, the acceleration is imposed on vehicle bodies and on unrestrained or restrained occupants as well. Occupants retain their forward motion due to the acceleration imposed on their bodies and finally collide with the interior surface of vehicles or are restrained by safety belts (they collide with the belt, which is usually less injurious). For instance, the overall injury severity index (*IS*) is given by IS = $\bar{a}_b^{2.5}$ in Gadd (1966).

According to the theory of Newtonian mechanics, the average acceleration imposed on occupant-vehicle compartment can be expressed as,

$$\bar{a}_b = \frac{E_b}{\bar{m}_b \cdot d_b} \tag{1}$$

in which $\bar{a}_b$ is the average imposed acceleration imposed on car occupant compartment; E_b is the energy absorption for decelerating the uncrashed compartment; $\bar{m}_b$ is the mass at the gravity center of the uncrashed vehicle portions; d_b is the crash displacement of the compartment.

When it comes to a two-vehicle collision, the acceleration ratio of the subject vehicle s and the partner vehicle p is computed as,

$$\frac{\bar{a}_s}{\bar{a}_p} = \frac{E_s}{E_p} \cdot \frac{\bar{m}_p}{\bar{m}_s} \cdot \frac{d_p}{d_s} \tag{2}$$

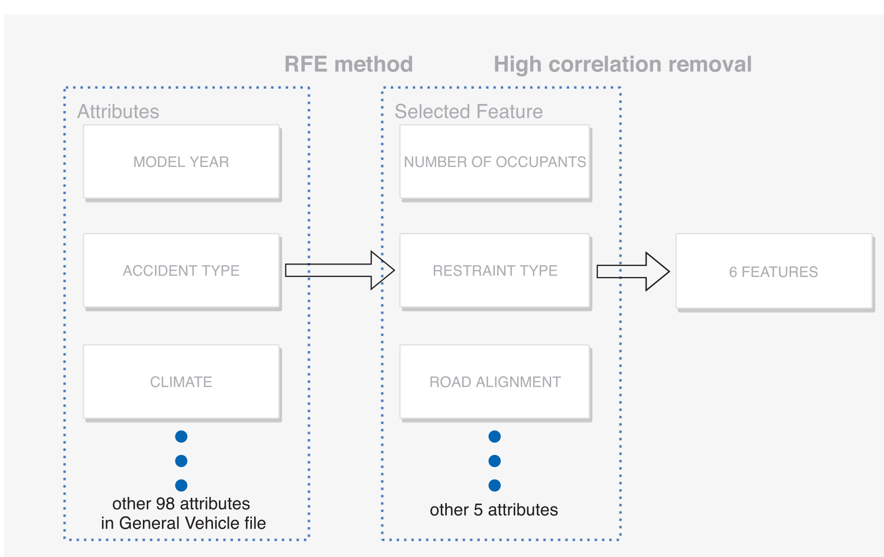


Fig. 2. The feature selection process of data for the prediction model.

Eq. (2) shows that the energy absorption of the uncrashed vehicle body is proportional to the average imposed acceleration, and thereby positively correlated to the occupant's injury severity. That indicates that besides the protection structures, occupants in the uncrashed compartment that absorbs less energy suffer less from crash impacts.

Two main factors (related to the last two terms in Eq. (2)) that influence the absorption are discussed in this paper: protective structure (which absorbs destructive energy) and stiffness (which reduces vehicle deformation). High stiffness and extra structure both increase vehicle mass. In the next section, we investigate the effect of vehicle mass on the energy absorption of vehicle bodies.

3.2. The two-vehicle straight-line energy loss function

Thus, we establish the two-vehicle energy absorption function in two parts: energy loss (W_L) and energy absorption distribution factor. The computation of energy loss combines the energy conservation equation and momentum conservation equation when preserving the assumption of inelastic collision. Employing the coefficient of restitution e (in Eq. (3)) might improve model fit by allowing for elastic collisions, but we assume $e = 0$ here for simplicity:

$$e = -\frac{\text{relative velocity after impact}}{\text{relative velocity before impact}} = -\frac{u_1 - u_2}{v_1 - v_2} = -\frac{v - v}{v_1 - v_2} = 0 \tag{3}$$

Where v_1 and v_2 are the speeds before impact, u_1 and u_2 are the speeds after impact, assuming two vehicles stick to each other with the same speed ($u_1 = u_2 = v$).

We hypothesize this inelasticity simplification makes ELVIS easier to obtain and a preliminary predictor requiring little data input. We then derive the equations in the form of Delta-V (directly provided by the NASS/CDS dataset) for two vehicles, respectively. A simplified computation of the straight-line energy absorption follows:

To begin with, we first define the general form of Delta-V in the following equation:

$$\begin{aligned}\Delta V_1 &= v - v_1 = \frac{m_2}{m_1 + m_2}\sqrt{(v_1^2 + v_2^2 + 2v_1v_2\cos\gamma)}\\ \Delta V_2 &= v - v_2 = \frac{m_1}{m_1 + m_2}\sqrt{(v_1^2 + v_2^2 + 2v_1v_2\cos\gamma)}\end{aligned} \tag{4}$$

in which m_1 and m_2 are the mass of vehicle 1 and 2; ΔV_1 and ΔV_2 are the speed changes (before and after the collision) for vehicle 1 and 2; γ is the angle between two initial speeds.

For straight-line collisions, Eq. (4) simplifies to Eq. (5):

$$\begin{aligned}\Delta V_1 &= \frac{m_2(v_2 - v_1)}{m_1 + m_2}\\ \Delta V_2 &= \frac{m_1(v_1 - v_2)}{m_1 + m_2}\end{aligned} \tag{5}$$

Note that ΔV_1 and ΔV_2 could be in the same or different direction, so their angle of collision speed should be determined before application.

The energy loss function includes the square of Delta-V and the coefficient of restitution e in its form. The relationship between W_L and e was derived in Daily et al. (2006) for possible future extensions, but here we ignore restitution ($e = 0$) in the following equations. Eq. (6) describes the energy loss (W_L) in a specific crash as:

$$\begin{aligned}W_L &= 0.5\cdot\frac{m_1m_2(v_1 - v_2)^2}{(m_1 + m_2)}\cdot(1 - e^2)\\ &= 0.5\cdot\frac{m_1}{m_2}(m_1 + m_2)\Delta V_1^2\cdot(1 - e^2)\\ &= 0.5\cdot\frac{m_2}{m_1}(m_1 + m_2)\Delta V_2^2\cdot(1 - e^2)\end{aligned} \tag{6}$$

Moreover, the power value of the mass ratio (α), unveils the relationship between the mass ratio and the energy absorption in Eq. (7) and is estimated or calibrated from real data. The more massive vehicle is initially hypothesized to absorb less impact energy within its uncrashed part. For that reason, the value of α is expected to be negative:

$$\frac{W_{L1}}{W_{L2}} = \left(\frac{m_1}{m_2}\right)^{\alpha} \tag{7}$$

Finally, the overall energy absorption for each vehicle is presented in Eq. (8). If the weights of two collided vehicles are equal, the occupants receive the same impact. Details of the computation are in Appendix A:

$$\begin{aligned}W_{L1} &= 0.5\cdot\frac{\left(\frac{m_1}{m_2}\right)^{\alpha+1}}{1 + \left(\frac{m_1}{m_2}\right)^{\alpha}}(m_1 + m_2)\Delta V_1^2\\ W_{L2} &= 0.5\cdot\frac{(\frac{m_2}{m_1})^{\alpha+1}}{1 + (\frac{m_2}{m_1})^{\alpha}}(m_1 + m_2)\Delta V_2^2\end{aligned} \tag{8}$$

3.3. Two-vehicle all-type crashes energy loss function

According to the longitudinal straight-line energy loss function, the energy absorption due to the lateral movement can be similarly computed by Eqs. (6) and (8) but with a different α. That is because vehicles have different structures and designs longitudinally and laterally (in general, vehicles' longitudinal reinforcements are better). The total energy absorption is then the summation of lateral and longitudinal energy loss after decomposing the Delta-V vector into two directions:

$$\begin{aligned}W_{L1} &= W_{L1}(\alpha_{\text{long}}, \Delta V_{1,\text{long}}) + W_{L1}(\alpha_{\text{lat}}, \Delta V_{1,\text{lat}})\\ W_{L2} &= W_{L2}(\alpha_{\text{long}}, \Delta V_{2,\text{long}}) + W_{L2}(\alpha_{\text{lat}}, \Delta V_{2,\text{lat}})\end{aligned} \tag{9}$$

It respectively computes the energy absorption in two directions that do not influence each other. After the calibration of α for two different collision types (straight-line and side-impact), the regression process tests different forms to find the best fit for each types.

3.4. Regression model specification

Due to the different magnitudes of two explanatory variables, W_L (10^6) and Delta-V (10^1), their values should be rescaled to compare them with the same magnitude. The Z-score method is applied to standardize W_L and Delta-V without any impacts on regression outcomes. We regard the dependent variable as an ordered factor.

3.4.1. Proportional odds ordered logistic regression model

The proportional odds ordered logistic regression (POLR) builds the relationship between the explanatory variable(s) and the tendency of the dependent variable being in higher categories compared to the lower one. One of the main features of this model is that it keeps the fixed coefficient across each split. The link functions or Cumulative Density Functions (CDF) can differ from various pre-defined assumptions of the distribution.

Regardless of choices of link functions, the general form of the model specification is given as:

$$Y_{\text{in}}^* = \beta_{0,i} + \beta_i X_{\text{in}} + \varepsilon_i \tag{10}$$

where Y_{in}^* is the intermediate parameter of injury severity MAIS for observation n at the level of i, X_{in} denotes the multivariate explanatory variable, $\beta_{0,i}$ is the intercept of this level, and β_i is the coefficient of variable X_{in}.

During the fitting process, this model predicts the values of the latent variable and coefficients by yielding the Principle of Maximum Likelihood. The values of MAIS serve as the thresholds to determine the segments. For example, the predicted probability of each MAIS level with the logit link function is calculated by:

$$
\begin{aligned}
P(0 < Y \leq 1) &= P(\text{MAIS} = 1) = \frac{1}{1 + \exp(-Y_1)} \\
P(1 < Y \leq 2) &= P(\text{MAIS} = 2) = \frac{1}{1 + \exp(-Y_2)} - \frac{1}{1 + \exp(-Y_1)} \\
&\vdots \\
P(5 < Y \leq 6+) &= P(\text{MAIS} = 6+) = 1 - \frac{1}{1 + \exp(-Y_5)}
\end{aligned}
\tag{11}
$$

Finally, the maximum probability of an event and its corresponding response are selected for the prediction outcome. With the interpretations of the models mentioned above, the values of the uncertain coefficient α can be determined by comparing their Residual Deviances (RD). In addition, we compare models with different link functions or different variables by the Akaike Information Criterion (AIC). All the features above are solved in *R* packages (Venables and Ripley, 2002; Christensen, 2019).

In this paper, the regression model expressions for Delta-V and ELVIS are shown in Eq. (12). Note that the Delta-V is expressed by both its modulus and its angle from the vertical (represented by $\tan\theta = \frac{\Delta V_{\text{lat}}}{\Delta V_{\text{long}}}$).

$$
\begin{aligned}
\text{MAIS} &\sim \Delta V + \tan\theta + \text{Number of Occupants} + \text{Restraint Type} \\
&\quad + \text{Road Alignment} + \text{Road Profile} + \text{Surface Type} \\
\text{MAIS} &\sim \text{ELVIS} + \text{Number of Occupants} + \text{Restraint Type} \\
&\quad + \text{Road Alignment} + \text{Road Profile} + \text{Surface Type}
\end{aligned}
\tag{12}
$$

3.5. Machine learning model

Machine learning algorithms can provide a credible prediction by generating classifiers without any pre-defined distribution assumptions. Random Forest, a widely used machine learning tool in traffic safety research, is employed in this examination (Breiman, 2001).

First, two divisions randomly separate the whole dataset with 80% for training and 20% for testing. Then, the training part produces several classifiers, and the classifiers predict the outcomes of the test part by voting for the best description. The accuracy (the number of correct predictions divided by the total number of cases) is one of the measures that evaluates the quality of prediction models. Note the setting of hyper parameters is tuned for model performance criteria through an n-fold cross-validation technique, which repeats the steps above n times.

4. Model results

4.1. The proportional ordered logistic regression

The POLR model sets MAIS as a categorized and ordered factor measuring the most severe injury of occupants. With this regression model, straight-line and side-impact crashes are separately involved. The calibration results are shown in Fig. 3.

The values of α with the minimum RD are -0.8 for head-on and rear-end crashes and -5.0 for side-impact crashes; hence, the initial hypothesis of negative α value is corroborated. In terms of absolute values of α, it is much larger in side-impact crashes rather than straight-line scenarios.

This is consistent with the fact that, due to their rigid structures and protections, the bodies of heavier vehicles tend to absorb less energy from both types of crashes. Occupants in heavier vehicles may thereby receive less impact that forces them to collide the interior.

After the calibration, the analytical capability of the model is examined by fitting the data. To explore the influence of different link functions, we conduct two more experiments to compare with the logit function. Table 3 presents the results for three models.

In the results of varying link functions, the logit-function model holds the minimum AIC among three models, and its confidence level is above 99.9%. Logit is chosen as the best-fit POLR model and is applied in the following process.

Following the calibrated parameters and the link function, we further check the statistical significance of variable parameter estimation in Table 4. 2.5% and 97.5% confidence intervals (CI) are provided.

The coefficients of the explanatory variables reveal the relationship between two factors. The results demonstrate that the parameters of energy absorption and the number of occupants are statistically significant. For W_L, with the increase of energy absorbed by vehicles, the positive coefficient indicates a higher likelihood of being severely injured. Similarly, the larger the number of occupants sitting in vehicles, the higher the likelihood that occupants get serious injuries. In summary, occupants of most heavy vehicles receive less impact from crashes, especially from the side-impact collision, due to their reinforcement and rigid structure. With a specific α value and the logit link function, the energy absorption and the number of occupants have a significant impact on the injury severity of occupants.

4.2. Comparison of energy loss and Delta-V models

In order to further examine the statistical power of the proposed model to describe the injury severity, we compare the ELVIS and the Delta-V models in Table 5. The POLR model with the logit link function is deployed to Delta-V as well to estimate seven classifications of MAIS.

The results demonstrate that the ELVIS model fits the data better with a lower AIC. The difference in AIC between two models is more than 10, which strongly favors the one with the minimum AIC value over another (Burnham and Anderson, 2004). We believe the reason is that ELVIS considers the effect of the protection that the bodies of heavier vehicles absorb less impact energy. It distributes energy absorption between vehicles according to their mass ratio and the distribution parameters of two directions.

4.3. Random forest algorithm tests for accuracy

After checking the energy loss-based model's goodness-fit statistically, we want to ensure that the result is robust to methodology for parameter estimation. The random forest machine learning algorithms is deployed to test the predictive capability of models.

Since most cases in the NASS/CDS dataset have less severe injuries, we need to consider this imbalance in the analysis. One popular solution transforms multi-class tasks into two-class tasks (Tax and Duin, 2002; Delen et al., 2006; Li et al., 2012). Therefore, we reduce the number of classifications into two categories (MAIS 3− and MAIS 3 & 3+). After that, we also conduct a 10-fold cross-validation process to tune the hyper parameter (number of trees 'mtry' in Random Forest) and further mitigate the bias problem. Table 6 shows the settings and results of machine learning models.

Both predictors are assessed through this process. Machine learning models maintain the same functions with the regression expressions in

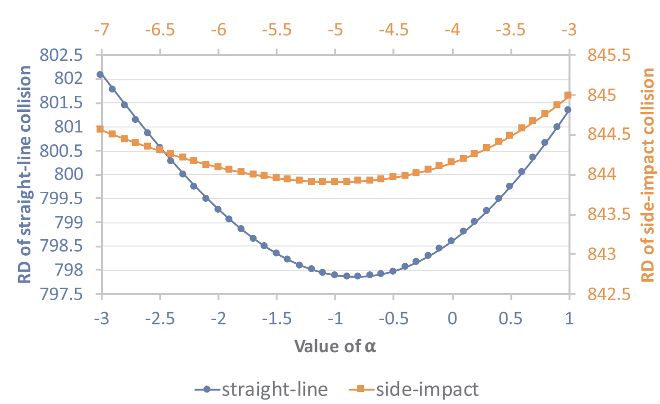


Fig. 3. Calibration for the parameter α in two directions.

Table 3
Interpretations of the proportional ordered logistic regression model with three link functions.

Link function	Class	Estimate	Std. error	Pr(> |z|)	AIC
logit	0|1	−0.9672	0.1935	5.8022e−07***	3279.49
	1|2	1.0994	0.1947	1.6257e−08***	
	2|3	2.0216	0.2042	4.2166e−23***	
	3|4	2.6846	0.2181	7.8422e−35***	
	4|5	3.1155	0.2311	1.9941e−41***	
	5|6+	3.3499	0.2398	2.3004e−44***	
probit	0|1	−0.5910	0.1107	9.2735e−08***	3297.87
	1|2	0.6565	0.1109	3.2451e−09***	
	2|3	1.1536	0.1136	3.1812e−24***	
	3|4	1.4778	0.1173	2.0658e−36***	
	4|5	1.6745	0.1205	6.7617e−44***	
	5|6+	1.7768	0.1226	1.2719e−47***	
cloglog	0|1	−0.7734	0.1640	2.0615e−07***	3346.38
	1|2	0.8490	0.1667	2.4236e−08***	
	2|3	1.0406	0.1675	1.8827e−23***	
	3|4	1.2867	0.1694	7.0965e−35***	
	4|5	1.3032	0.1708	4.0434e−43***	
	5|6+	1.4639	0.1717	2.2472e−45***	

*** P ≤ 0.001.

Table 4
Significance tests for model variables.

Variable	Coefficient	Std. error	*P*-value	2.5% CI	97.5% CI
Alignment 1|2	−0.1330	0.2003	5.0655e−01	−0.5277	0.2584
Alignment 2|3	−0.0735	0.2090	7.2507e−01	−0.4851	0.3353
Occupants	0.1749	0.0515	6.8916e−04***	0.0739	0.2761
Profile 1|2	−0.0970	0.1599	5.4420e−01	−0.4114	0.2159
Profile 2|3	0.2384	0.3872	5.3810e−01	−0.5253	0.9983
Profile 3|4	0.0099	0.1687	9.5330e−01	−0.3216	0.3400
Profile 4|5	0.6453	1.1591	5.7772e−01	−1.7196	3.0108
Restrain type 1|2	0.0300	0.1149	7.9426e−01	−0.1951	0.2554
Restrain type 2|3	−1.8623	1.1217	9.6856e−02*	−4.8407	0.0429
Surface type 0|1	−0.1994	0.1714	2.4475e−01	−0.5351	0.1373
ELVIS (W_L)	0.6380	0.0574	1.1403e−28***	0.5277	0.7530

* P ≤ 0.05 *** P ≤ 0.001.

Table 5
Comparison of ELVIS and Delta-V in POLR model with logit-link function.

Model	Class	Estimate	Std. error	Pr(> |z|)	AIC
ELVIS (W_L)	0|1	−0.9672	0.1935	5.8022e−07***	3279.49
	1|2	1.0994	0.1947	1.6257e−08***	
	2|3	2.0216	0.2042	4.2166e−23***	
	3|4	2.6846	0.2181	7.8422e−35***	
	4|5	3.1155	0.2311	1.9941e−41***	
	5|6+	3.3499	0.2398	2.3004e−44***	
Delta−V	0|1	−0.8913	0.1925	3.6431e−06***	3291.05
	1|2	1.1801	0.1943	1.2474e−09***	
	2|3	2.0720	0.2035	2.4319e−24***	
	3|4	2.7034	0.2163	7.7909e−36***	
	4|5	3.1202	0.2287	2.1759e−42***	
	5|6+	3.3539	0.2373	2.4597e−45***	

*** P ≤ 0.001.

Eq. (12). Fig. 4 illustrates the confusion matrix (in a heatmap form) for each model. In each confusion matrix, its diagonal represents the cases that classifiers successfully predict the actual outcome, while off-diagonal values are wrong predictions. The values of accuracy in two models for the 'Test' set are 82.59% for the ELVIS model and 78.54% for the Delta-V model respectively. Because of the overlapping confidence intervals of accuracy between the two models, we conduct a two-sample *t*-test to check whether the difference is significant ($t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx 3.50$). The computed *P*-value is less than 0.005 to reject the null hypothesis. In other words, the ELVIS model predicts better than Delta-V with a probability of more than 99.5%.

Table 6
Comparison of ELVIS and Delta-V in machine learning 'Random Forest' model.

Indicator	mtry	TrainSet	Accuracy	TestSet	Accuracy	95% CI
ELVIS (W_L)	4	3−: 873	81.36%	3−: 213	82.59%	(77.28%, 87.11%)
		3+: 114		3+: 34		
Delta-V	4	3−: 873	76.94%	3−: 213	78.54%	(72.89%, 83.49%)
		3+: 114		3+: 34		

5. Discussion

In the sections above, we have revealed the statistical and predictive capability of the ELVIS model in two-vehicle crashes. The mass ratio, which distributes the energy absorption, is the core ELVIS model component that is worth discussion. The calibrated power values for the mass ratio in longitudinal and lateral directions are both negative, which indicates that occupants in heavy vehicles absorb less impact from the crashes, suffering less significant injuries. This result, however, leads to the 'bullying phenomenon' in reality, in which people desire to purchase large mass cars like SUVs and pickup trucks to take advantage of drivers in compact or light cars. Even heavy-vehicle drivers who take less risk than others may behave more aggressively more often (Monfort and Nolan, 2019). To mitigate this effect, the policy that uniformly reduces the mass range is recommended. For instance, a mass range reduction of 20% was found to lower the fatality rate by 3% in Sweden (Buzeman et al., 1998). Another finding is that the absolute power value of the mass ratio in the lateral direction is more than six times greater than in the longitudinal direction. That indicates heavy vehicles generally contain more effective masses in their side structures, while light vehicles are designed to reduce such masses for fuel economy, cost reduction, or other purposes. Due to the limited side protective ability of light passenger vehicles, car manufactures are supposed to implement reinforcing side impact protection designs. Side airbags and enforced door and B-pillar structures have reduced injury severity for passenger cars on real roads (Stigson and Kullgren, 2011). The improvements of side protection induce a more than 70 percent overall injury reduction, as reported by Volvo (Jakobsson et al., 2008). Therefore, it is expected that future side protection designs could be more effective in mitigating the weakness of light passenger vehicles.

The number of vehicle occupants is positively correlated to the most severe injuries in one vehicle, consistent with previous studies (Chang and Mannering, 1999; Christoforou et al., 2010). That means the high vehicle occupancy, all else equal, increases likelihood that some of the occupants are seriously injured.

6. Conclusions

Injury reduction requires analysis of the impact suffered by occupants in vehicular crashes. Building relationships between physical injuries and crash mechanisms in the real world is complicated, so statistical models are used to handle this problem.

While the assessment criteria for injury severity has many forms in literature, this paper uses the common MAIS system to compare two different models: the widely-used Delta-V approach and our proposed energy-loss vehicle injury severity (ELVIS) method.

In brief, the ELVIS model considers energy absorption from

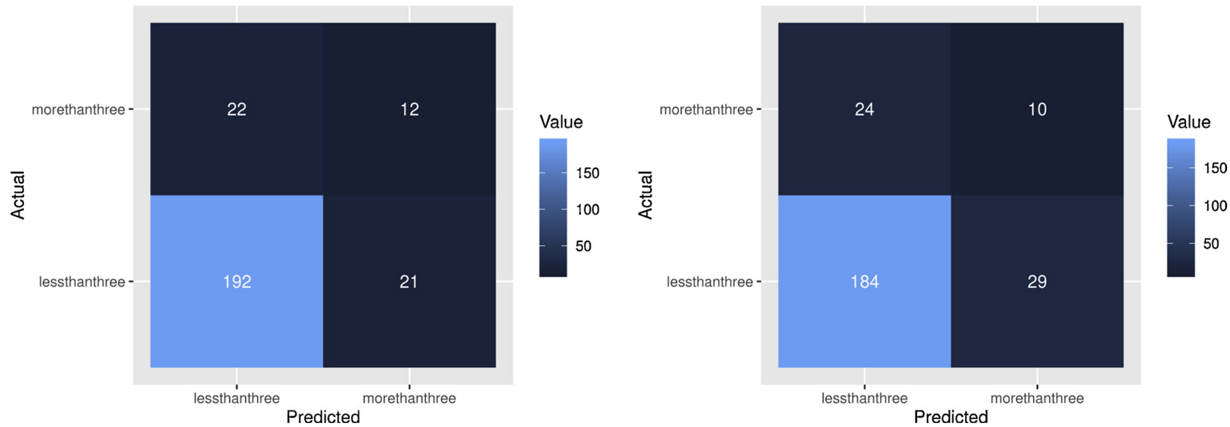


Fig. 4. The confusion matrix of prediction by model with the ELVIS predictor (left) and Delta-V predictor (right).

longitudinal and lateral directions when two vehicles with different masses collide to estimate injury. Lighter vehicles generally suffer more severely from crashes, the protective effect reflects that heavier vehicles have better-designed structures and is embedded in this new method.

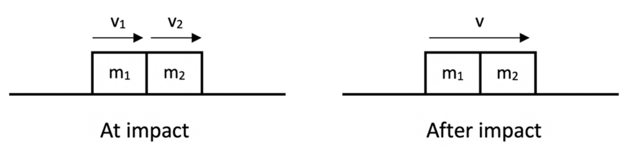
Different regression models and various values of the parameter α are calibrated to trace the statistical best fit of the ELVIS model with the observed crash severity data. From the results, we find that the proportional ordered logistic regression (POLR) with the logit link fits well. The power values of the mass ratio α in this model are negative, which corroborates the hypothesis about the protective effect. The protective effect of heavy vehicles is more significant in the lateral direction. That suggests car manufacturers to focus more on the design of lateral structures for compact cars. The ELVIS and Delta-V models are compared with both the POLR model and with a Random Forest machine learning model, for both approaches ELVIS outperforms Delta-V in most situations.

Future research could extend the model by studying more crashes with different collision angles and establishing the relationships between crash types and their respective α values. It may also depend on whether fragile or weak structures of vehicles receive the crash impact. Other factors that significantly influence the energy absorption by vehicles are also expected to improve estimation outcomes. Extensions to consider elastic collisions and restitution coefficients may provide additional useful insights for realistic crash studies.

Authors' contribution

Ang Ji: conceptualization, methodology, software, validation, formal analysis, investigation, writing – original draft. David Levinson: writing – review & editing, supervision.

Appendix A. The computation of energy loss in one direction



The process of crashes can be divided into two states: at impact and after impact. When comparing the two states, momentum conservation and energy conservation should keep valid in the whole process. If assuming the collision is inelastic, two vehicles stick to each other and continue moving for a distance. The momentum conservation of this process is shown below:

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2) v \tag{13}$$

The speeds after impact of two vehicles are the same, which can be then transformed into:

$$v = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2} \tag{14}$$

At the same time, the energy should also be conserved in this process as well. Therefore, the kinetic energy loss can be expressed as:

$$\begin{aligned} W_L &= 0.5 \cdot (m_1 + m_2) v^2 - 0.5 \cdot m_1 v_1^2 - 0.5 \cdot m_2 v_2^2 \\ &= 0.5 \cdot \frac{m_1 m_2}{m_1 + m_2} (v_1 - v_2)^2 \end{aligned} \tag{15}$$

After obtaining the total energy loss, it is hypothesized to be distributed according to their mass ratio, as displayed in Eq. (7). Then, their respective energy loss can be described as:

$$W_{L1} = \left(\frac{m_1}{m_2}\right)^{\alpha} W_{L2} = \left(\frac{m_1}{m_2}\right)^{\alpha}(W_L - W_{L1}) \Rightarrow W_{L1} = \frac{\left(\frac{m_1}{m_2}\right)^{\alpha}}{1+\left(\frac{m_1}{m_2}\right)^{\alpha}} W_L$$
$$W_{L2} = \left(\frac{m_2}{m_1}\right)^{\alpha} W_{L1} = \left(\frac{m_2}{m_1}\right)^{\alpha}(W_L - W_{L2}) \Rightarrow W_{L2} = \frac{\left(\frac{m_2}{m_1}\right)^{\alpha}}{1+\left(\frac{m_2}{m_1}\right)^{\alpha}} W_L \tag{16}$$

By considering Eqs. (6) and (16) simultaneously, we can finally get the energy absorption function as Eq. (8). If the weight of two vehicles is the same (m), Eq. (8) is transformed as:

$$W_{L1} = 0.5\cdot\frac{\left(\frac{m}{m}\right)^{\alpha+1}}{1+\left(\frac{m}{m}\right)^{\alpha}}(m+m)\Delta V_1^2 = 0.5\cdot m\cdot\Delta V_1^2$$
$$W_{L2} = 0.5\cdot\frac{\left(\frac{m}{m}\right)^{\alpha+1}}{1+\left(\frac{m}{m}\right)^{\alpha}}(m+m)\Delta V_2^2 = 0.5\cdot m\cdot\Delta V_2^2 \tag{17}$$

Substituting Eq. (17) into Eq. (5), we establish the relationship between the energy absorption and collision speeds of two vehicles as,

$$W_{L1} = 0.5\cdot m\cdot\Delta V_1^2 = 0.5\cdot m\cdot[0.5*(V_2 - V_1)]^2 = 0.125\cdot m\cdot(V_1 - V_2)^2$$
$$W_{L2} = 0.5\cdot m\cdot\Delta V_2^2 = 0.5\cdot m\cdot[0.5*(V_1 - V_2)]^2 = 0.125\cdot m\cdot(V_1 - V_2)^2 \tag{18}$$

Finally, it can be concluded that the energy absorption by occupants in the two vehicles is also equal when meeting the above conditions. That is, the protective effects are eliminated due to the same weight of vehicles.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.aap.2020.105730.

References

Abay, K.A., Paleti, R., Bhat, C.R., 2013. The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. Transp. Res. Part B: Methodol. 50, 74–89.

Abdel-Aty, M., Ekram, A.A., Huang, H., Choi, K., 2011. A study on crashes related to visibility obstruction due to fog and smoke. Accid. Anal. Prev. 43, 1730–1737.

Abdel-Aty, M.A., Chen, C.L., Schott, J.R., 1998. An assessment of the effect of driver age on traffic accident involvement using log-linear models. Accid. Anal. Prev. 30, 851–861.

Andricevic, N., Junge, M., Krampe, J., 2018. Injury risk functions for frontal oblique collisions. Traff. Inj. Prev. 19, 518–522.

Augenstein, J., Perdeck, E., Stratton, J., Digges, K., Bahouth, G., 2003. Characteristics of crashes that increase the risk of serious injuries. Annual Proceedings/Association for the Advancement of Automotive Medicine 561.

Bogue, S., Paleti, R., Balan, L., 2017. A modified rank ordered logit model to analyze injury severity of occupants in multivehicle crashes. Anal. Methods Accid. Res. 14, 22–40.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33, 261–304.

Buzeman, D.G., Viano, D.C., Lövsund, P., 1998. Car occupant safety in frontal crashes: a parameter study of vehicle mass, impact speed, and inherent vehicle protection. Accid. Anal. Prev. 30, 713–722.

Carlson, W.L., 1977. Crash injury loss: the effect of speed, weight and crash configuration. Accid. Anal. Prev. 9, 55–68.

Carlson, W.L., 1979. Crash injury prediction model. Accid. Anal. Prev. 11, 137–153.

Carlson, W.L., Kaplan, R.J., 1975. Case studies considered as retroactive experiments. Accid. Anal. Prev. 7, 73–80.

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

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

Christensen, R.H.B., 2019. ordinal-Regression Models for Ordinal Data. R Package Version 2019.4-25. http://www.cran.r-project.org/package=ordinal/.

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

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. Prev. 40, 1589–1594.

Daily, J., Strickland, R., Daily, J., 2006. Crush analysis with under-rides and the coefficient of restitution. Proceedings of the 24th Annual Special Problems in Traffic Crash Reconstruction.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accid. Anal. Prev. 38, 434–444.

Digges, K., Dalmotas, D., 2001. Injuries to Restrained Occupants in Far-Side Crashes. Technical Report. SAE Technical Paper.

Duncan, C.S., Khattak, A.J., Council, F.M., 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Transp. Res. Rec. 1635, 63–71.

Ehrlich, P.F., Brown, J.K., Sochor, M.R., Wang, S.C., Eichelberger, M.E., 2006. Factors influencing pediatric injury severity score and glasgow coma scale in pediatric automobile crashes: results from the crash injury research engineering network. J. Pediatr. Surg. 41, 1854–1858.

Evans, L., 1994. Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian mechanics. Accid. Anal. Prev. 26, 609–616.

Ferreira, S., Amorim, M., Couto, A., 2017. Risk factors affecting injury severity determined by the MAIS score. Traff. Inj. Prev. 18, 515–520.

Gadd, C.W., 1966. Use of a Weighted-Impulse Criterion for Estimating Injury Hazard. Technical Report. SAE Technical Paper.

Greenspan, L., McLELLAN, B.A., Greig, H., 1985. Abbreviated injury scale and injury severity score: a scoring chart. J. Trauma 25, 60–64.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.

Jakobsson, L., Isaksson-hellman, I., Lindman, M., 2008. Whips (volvo cars' whiplash protection system) – the development and real-world performance. Traff. Inj. Prev. 9, 600–605. https://doi.org/10.1080/15389580802435636.

Jokseh, H.C., 1993. Velocity change and fatality risk in a crash – a rule of thumb. Accid. Anal. Prev. 25.

Keall, M.D., Frith, W.J., Patterson, T.L., 2004. The influence of alcohol, age and number of passengers on the night-time risk of driver fatal injury in New Zealand. Accid. Anal. Prev. 36, 49–61.

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

Kononen, D.W., Flannagan, C.A., Wang, S.C., 2011. Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accid. Anal. Prev. 43, 112–122.

Li, Z., Liu, P., Wang, W., Xu, C., 2012. Using support vector machine models for crash injury severity analysis. Accid. Anal. Prev. 45, 478–486.

Mackay, G.M., 1968. Injury and collision severity. SAE Trans. 3108–3113.

Massie, D.L., Campbell, K.L., Williams, A.F., 1995. Traffic accident involvement rates by driver age and gender. Accid. Anal. Prev. 27, 73–87.

Monfort, S.S., Nolan, J.M., 2019. Trends in aggressivity and driver risk for cars, suvs, and pickups: vehicle incompatibility from 1989 to 2016. Traff. Inj. Prev. 20, S92–S96. https://doi.org/10.1080/15389588.2019.1632442.

Nishimoto, T., Mukaigawa, K., Tominaga, S., Lubbe, N., Kiuchi, T., Motomura, T., Matsumoto, H., 2017. Serious injury prediction algorithm based on large-scale data and under-triage control. Accid. Anal. Prev. 98, 266–276.

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

Radja, G.A., 2016. National Automotive Sampling System-Crashworthiness Data System, 2015 Analytical User's Manual. Technical Report DOT HS 812 321.

Roberts, V.L., Compton, C.P., 1993. The Relationship Between Delta V and Injury. Technical Report. SAE Technical Paper.

Shelby, S.G., 2011. Delta-V as a measure of traffic conflict severity. International conference on road safety and simulation, 3rd Indianapolis, USA.

Smink, B.E., Ruiter, B., Lusthof, K.J., De Gier, J., Uges, D.R., Egberts, A.C., 2005. Drug use and the severity of a traffic accident. Accid. Anal. Prev. 37, 427–433.

Stigson, H., Kullgren, A., 2011. Effect of side impact protection in reducing injuries. 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV). http://www-esv.nhtsa.dot.gov/Proceedings/22/isv7/main.htm.

Tax, D.M., Duin, R.P., 2002. Using two-class classifiers for multiclass classification. Object Recognition Supported by User Interaction for Service Robots 124–127.

Tsoi, A.H., Gabler, H.C., 2015. Evaluation of vehicle-based crash severity metrics. Traff. Inj. Prev. 16, S132–S139.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, 4th ed. Springer, New York ISBN 0-387-95457-0.

Wood, D.P., 1997. Safety and the car size effect: a fundamental explanation. Accid. Anal. Prev. 29, 139–151.

Yasmin, S., Eluru, N., Pinjari, A.R., 2015. Analyzing the continuum of fatal crashes: a generalized ordered approach. Anal. Methods Accid. Res. 7, 1–15.

Yau, K.K., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accid. Anal. Prev. 36, 333–340.

Ydenius, A., 2010. Influence of crash pulse characteristics on injury risk in frontal impacts based on real-life crashes. Traff. Inj. Prev. 11, 526–534.

Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Anal. Methods Accid. Res. 1, 72–85.