ELSEVIER

Accident Analysis and Prevention 40 (2008) 45–54

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

# Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis

Huang Helai [a,*], Chin Hoong Chor [b,1], Md. Mazharul Haque [a,2]

[a] *Traffic Lab, Department of Civil Engineering, National University of Singapore, Engineering Drive 2, EW1, 04-02B, Singapore 117576, Singapore*

[b] *Department of Civil Engineering, National University of Singapore, Engineering Drive 2, Singapore 117576, Singapore*

Received 1 November 2006; received in revised form 3 April 2007; accepted 8 April 2007

**Abstract**

Most crash severity studies ignored severity correlations between driver–vehicle units involved in the same crashes. Models without accounting for these within-crash correlations will result in biased estimates in the factor effects. This study developed a Bayesian hierarchical binomial logistic model to identify the significant factors affecting the severity level of driver injury and vehicle damage in traffic crashes at signalized intersections. Crash data in Singapore were employed to calibrate the model. Model fitness assessment and comparison using intra-class correlation coefficient (ICC) and deviance information criterion (DIC) ensured the suitability of introducing the crash-level random effects. Crashes occurring in peak time and in good street-lighting condition as well as those involving pedestrian injuries tend to be less severe. But crashes that occur in night time, at T/Y type intersections, and on right-most lane, as well as those that occur in intersections where red light cameras are installed tend to be more severe. Moreover, heavy vehicles have a better resistance on severe crash and thus induce less severe injuries, while crashes involving two-wheel vehicles, young or aged drivers, and the involvement of offending party are more likely to result in severe injuries.


*Keywords:* Driver severity; Signalized intersection; Hierarchical logistic model; Bayesian analysis

## 1. Introduction

Signalized intersection is a hazardous location type on the road, which accounts for a substantial portion of traffic crashes. In order to develop cost-effective safety countermeasures, crash frequency and severity are two major concerns in understanding the relationship of crash occurrences and various risk factors. On the one hand, a large number of studies have focused on crash prediction models in examining the crash frequencies at intersections for different crash types (e.g. Kim and Washington, 2006; Kim et al., 2006). On the other hand, crash severity is also a safety concern of the traffic system. Before developing and implementing the traffic safety treatments, it would be very useful if a comprehensive understanding of the effects of risk factors on crash severity is available.

Analysis of crash severity can be conducted in different ways for various purposes. Some studies focused on the crash frequencies at specific traffic sites associated with different severity levels (e.g. fatal, serious, slight) to investigate how geometric, traffic, and environmental factors affect the crash severity. While this kind of studies normally take each crash as the subject unit, analysis can also be undertaken based on the driver–vehicle units involved in crashes to examine individual severity. Compared to the crash-based severity studies, individual severity analysis is promising and may yield a disaggregate understanding about severity levels of different driver–vehicle groups. This is especially useful when the severity levels of driver–vehicle units with different characteristics are desired (Hauer, 2006). This study focuses on examining the severity of driver injury and vehicle damage in traffic crashes at urban intersections.

Categorical data analysis techniques have generally been employed in most previous severity studies. While some (Hilakivi et al., 1989; Mannering and Grodsky, 1995; James and Kim, 1996; Shankar and Mannering, 1996; Mercier et al., 1997; Al-Ghamdi, 2002) have used binomial/multinomial logistic model to explore the significance of the risk factors by taking

* Corresponding author. Tel.: +65 65162255.
*E-mail addresses:* huanghelai@nus.edu.sg (H.L. Huang), cvechc@nus.edu.sg (H.C. Chin), mmh@nus.edu.sg (M.M. Haque).
[1] Tel.: +65 65162550.
[2] Tel.: +65 65162255.

0001-4575/$ – see front matter 
doi:10.1016/j.aap.2007.04.002

crash severity as a nominal, others (O'Donnell and Connor, 1996; Quddus et al., 2002; Rifaat and Chin, 2005; Abdel-Aty and Keller, 2005) have employed ordered logit/probit models to account for the ordered nature of severity levels.

However, since the techniques used in most past studies assumed independence between different observations, these techniques may not be adequate in modeling individual severity of driver injury and vehicle damage in the presence of potential correlations between those involved in the same multi-vehicle crashes. Actually, this correlation between samples has already been identified in some earlier studies; for example, Evans (1992, 1993) found that in a multiple vehicle crash, the risk of fatality was dependent on the characteristics of the other vehicles. Hence, the models without considering the covariance between individuals in the same crashes, especially when the covariance is significant, will result in inaccurate or biased estimates of factor effects.

A statistical modeling technique that allows hierarchical data structures to be easily specified and estimated is hierarchical models (see Snijders and Bosker, 2000; Goldstein, 2003). While several different terms have been used in the literature such as "multilevel models" and "random coefficient models", we use "hierarchical models" throughout this paper. Although the basic theories of hierarchical models have been developed and discussed for many years, it is only recently that many practical limitations on the use of hierarchical analysis have been overcome. A good number of applications of this modeling technique have been found in sociological research disciplines. In traffic safety research, Jones and Jorgensen (2003) presented a good exploration and discussion on the potential applications of the hierarchical models. Since then, the hierarchical modeling technique has been gaining an increasing amount of attention in accounting for the hierarchical data structure in road crash frequency and severity studies. For example, Jones and Jorgensen (2003) and Lenguerrand et al. (2006) developed hierarchical models to identify factors affecting crash severity, while Kim et al. (2007) employed the hierarchical crash prediction models for different crash types at rural intersections.

In the investigation of individual severity in crashes at signalized intersections in Singapore, a within-crash correlation was preliminarily identified, which will be shown in detail in Section 4 of model calibration and validation. Motivated by this correlation and inspired by the existing studies with hierarchical models, we proposed the use of a hierarchical binomial logistic (HBL) model to examine the significant risk factors related to severity of driver injury and vehicle damage in traffic crashes. In particular, crash was considered as cluster and there were a number of sub-clusters per cluster, i.e. driver–vehicle units involved in a crash. A full Bayesian method using Markov chain Monte Carlo (MCMC) algorithm was employed for model calibration to explicitly model the two-level data structure, i.e. crash-level and individual-level. Using the intra-class correlation coefficient (ICC) and deviance information criterion (DIC) in model assessment and comparison, the use of random effects on crash level in the model was further validated to be effective in this study in accounting for the within-crash correlation.

In the rest of this paper, a description of methodological framework, consisting of model development, assessment and comparison, is given first. Data collection and model calibration are then summarized to illustrate the proposed methodology and to understand the significant risk factors on individual severity. Summary and conclusion of this study, together with two potential extensions, are presented finally.

## 2. Methodology

### 2.1. Hierarchical binomial logistic model

In the presence of within-crash correlation of individual severity, models without appropriately considering the hierarchical data structure might yield inaccurate or biased parameter estimations. To account for this within-crash correlation, a HBL model with two-level specification was developed to estimate the effects of the selected covariates on severity level. Specifically, in the individual-level model (level 1), the response variable $Y$ for the $i$th driver–vehicle unit in $j$th crash only takes one of two values: $Y_{ij}=1$ in case of high severity, e.g. fatal or severe injury, while $Y_{ij}=0$ in case of low severity, e.g. slight or no injury. The probability of $Y_{ij}=1$ is denoted by $\pi_{ij}=\Pr(Y_{ij}=1)$, which follows a binomial distribution; hence

$$\operatorname{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} + \sum_{p=1}^{P}\beta_{pj}X_{pij} \tag{1}$$

where $X_{pij}$ is the $p$th covariate in the individual-level for $i$th driver–vehicle unit in $j$th crash, while $\beta_{0j}$and $\beta_{pj}$ are the intercept and the regression coefficients. In the context of the hierarchical model, the within-crash correlation is specified in the crash-level model (level 2) as:

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{Q}\gamma_{0q}Z_{qj} + u_{0j} \tag{2}$$

$$\beta_{pj} = \gamma_{p0} + \sum_{q=1}^{Q}\gamma_{pq}Z_{qj} + u_{pj} \tag{3}$$

In Eqs. (2) and (3), both intercept $\beta_{0j}$ and regression coefficients $\beta_{pj}$ in Eq. (1) vary with the different crashes. Specifically, two components are combined to decide the coefficient values. First, linear relationships are assumed for them with the crash-level covariates $Z_{qj}$, which is reasonable since the various crash features (e.g. street lighting, road surface condition) may result in different severity level. Second, besides the fixed parts which depend on the crash-level covariates $Z_{qj}$, random effects are also included to permit the potential random variations across the crashes ($u_{0j}$ and $u_{pj}$). These between-crash random effects vary across the different crashes only but are constant for all the driver–vehicle units within a same crash. This specification enables the model to account for the within-crash correlations (Jones and Jorgensen, 2003; Kim et al., 2007). Practically, the random effects are used to represent some unobservable variations between different crashes, which is the major difference between ordinary binomial logistic model (OBL) and HBL.

The full model with Eqs. (1–3) is academically named as random slope model (Snijders and Bosker, 2000). When the random effects are assumed only on the intercept, a simplified form can be obtained by dropping the crash-level covariate component $\sum_{q=1}^{Q}\gamma_{pq}Z_{qj}$ and the random part $u_{pj}$, which is referred to as random intercept model. The Eq. (3) is thus modified to be:

$$\beta_{pj} = \gamma_{p0} \tag{4}$$

In this study, to avoid excess complexity as the large set of covariates used, only the random intercept model was investigated. Hence, the combined model is yielded by substituting Eqs. (2) and (4) with Eq. (1) and is represented as follows:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \gamma_{00} + \sum_{p=1}^{P}\gamma_{p0}X_{pij} + \sum_{q=1}^{Q}\gamma_{0q}Z_{qj} + u_{0j} \tag{5}$$

The random effects $u_{0j}$ are generally assumed as a normal distribution with mean zero and variance $\tau_0^2$, as suggested by Snijders and Bosker (2000). The variance of outcome ($Y_{ij}$) therefore consists of two components: the variance of $u_{0j}$ ($\tau_0^2$) which captures the between-crash variability (level 2), and the variance associated with logistic distribution which captures the within-crash variability (level 1).

In interpreting the fixed effect part of coefficient estimation, a similar way can be followed as with the OBL, the exponential of effect coefficients, i.e. exp(γ), can be calculated to obtain odds ratio (O.R.) estimates in HBL model. This provides a basic interpretation for the magnitude of $\gamma$: if O.R. is less than 1.0, a unit increase in the variable $X_{pij}$ or $Z_{qj}$ will reduce the odds of being severe by a multiplicative effect of $\exp(\gamma)$ and vice versa. For the categorical covariates in the model where dummy variables are applied, $\exp(\gamma_a–\gamma_b)$ represents the odds ratios between these two categorical variables, *a* and *b*. In this case, the parameter or its estimate makes sense only by comparing one category with another.

### 2.2. Bayesian inference

There are several methods available for model calibration in hierarchical models (see Goldstein, 2003). Instead of using likelihood-based estimation, this study employed Bayesian inference to calibrate the proposed two-level model (Gelman et al., 2003). Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations. Specifically, in Bayesian models, given model assumptions and parameters, the likelihood of the observed data is used to modify the prior beliefs of the unknowns, resulting in the updated knowledge summarized in posterior densities. Hence, the distinctions between fixed and random effects disappear since all effects are now considered random and the hierarchical structure is explicitly accounted for. Several studies have demonstrated the advantages of Bayesian inference over classical estimation methods in philosophical aspect as well in practical aspect in transportation applications (e.g. Washington et al., 2005; Mitra and Washington, 2007).

In the absence of strong prior information for the model unknowns, uninformative priors were assumed for all regression coefficients ($\gamma_{00}$,$\gamma_{\text{p0}}$ and $\gamma_{\text{0q}}$) with normal distributions (0, 1000), and the variance $\tau_0^2$ of the normal distributed random effects $\mu_{0j}$ with inverse Gamma distribution (0.001, 0.001). The model was computed via the Gibbs sampler, a MCMC technique (Gilks et al., 1995), which was implemented using WinBUGS software (Spiegelhalter et al., 2003a). The 95% Bayesian credible interval (95% BCI) was used to examine the significance of covariates, which provides probability interpretations with normality assumption on unknowns and confidence interval estimations (Gelman et al., 2003). Specifically, those coefficient estimations were identified as significant, whose 95% BCIs do not cover “0”, i.e., the 95% BCIs of O.R. do not cover “1”. Besides, engineering and intuitive judgment should be able to confirm the validity and practicality of the sign of each covariate and the rough magnitude of each estimated coefficient.

### 2.3. Assessment of random effects using intra-class correlation coefficient

An intra-class correlation coefficient $\rho$ (ICC) is normally defined to examine the proportion of specific crash-level variance (level 2) in overall residual variance (Jones and Jorgensen, 2003; Kim et al., 2007). Since the logistic distribution for the individual-level (level 1) residual implies a variance of $\pi^2/3 = 3.29$, this implies that for a two-level logistic random intercept model with an intercept variance of $\tau_0^2$, the ICC for between-crash residual is

$$\rho = \frac{\tau_0^2}{\tau_0^2 + \pi^2/3}$$

The ICC is an indicator of the magnitude of the within-crash correlation. A value of $\rho$ close to zero means that there is a very small variation between the different crashes, indicating that OBL model may be adequate for the data. On the other hand, a relative large value of $\rho$ implies a favor for hierarchical model, e.g. HBL model in this study.

### 2.4. Model comparison using deviance information criterion

To further ensure the advantage of employing HBL over OBL, an OBL model with the same covariates and dataset can also be estimated to compare with the calibrated HBL model. The OBL model was obtained by dropping random effects $u_{0j}$, which means ignoring the severity correlations between driver–vehicle units within the same crashes. So the Eq. (5) changes to:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \gamma_{00} + \sum_{p=1}^{P}\gamma_{p0}X_{pij} + \sum_{q=1}^{Q}\gamma_{0q}Z_{qj} \quad (6)$$

For model comparison, deviance information criterion (DIC), proposed by Speigelhalter et al. (2003b) was used. In complex hierarchical models where parameters may outnumber observations, DIC provides a Bayesian measure of model complexity and fit that can be combined to compare models of arbitrary structure (Speigelhalter et al., 2003b). This can overcome the problems of classical criteria, such as Akaike information criterion (AIC) and Bayesian information criterion (BIC). These classical criteria require the specification of the number of parameters in each model. Specifically, DIC is defined as:

$$\text{DIC} = D(\bar{\gamma}) + 2p_{\text{D}} = \overline{D(\gamma)} + p_{\text{D}}$$

where $D(\bar{\gamma})$is the deviance evaluated at the posterior means of estimated unknowns ($\bar{\gamma}$), and posterior mean deviance $\overline{D(\gamma)}$can be taken as a Bayesian measure of fit or "adequacy". $p_{\text{D}}$ is motivated as a complexity measure for the effective number of parameters in a model, as the difference between $\overline{D(\gamma)}$ and $D(\bar{\gamma})$, i.e., mean deviance minus the deviance of the means. As a generalization of AIC, DIC can thus been considered a Bayesian measure of fit or adequacy, penalized by an additional complexity term $p_{\text{D}}$. As with AIC, models with lower DIC values are preferred.

## 3. Dataset for analysis

For this study, crash data in Singapore from 2003 to 2005 were used. Singapore is a heavily urbanized island country with an area of about 700 km$^2$ and 3235 km of roads (in 2005). Of the total of 19832 reported crashes in this period, 4095 cases occurring at signalized intersections were extracted and used in the model. In these, 7840 driver–vehicle units were involved, resulting in an average involvement rate of 1.91 individuals per crash.

In the dataset, each observation is associated with a driver–vehicle unit involved in the crashes at intersections. Two categorical severity indicators are of interest, which are driver injury severity: (a) fatal or serious injury, DI(A), (b) slight or no injury, DI(B); and vehicle damage severity: (a) extensive damage, VD(A), (b) slight or no damage, VD(B). To yield a net effect estimate of each potential factor on individual severity, a binary dependent variable was defined by combining the two severity indicators: (a) DI(A) or/and VD(A), denoted as IS(A), representing high individual severity (b) otherwise is low individual severity denoted as IS(B). A summary of severity statistics is given for years in Table 1.

In addition to severity levels, a record of crash IP number, geometric features, traffic conditions, driver and vehicle characteristics was also reported. There are a total of 25 variables coded for each intersection crash in the dataset. A number of variables like location code, vehicle registration number, nature of vehicle registration, etc. were excluded as they were irrelevant to the analytical purpose. A correlation matrix for those remaining variables, which were hypothesized to relate to the severity levels, was checked to avoid multi-collinearity as well as wrong signs or implausible magnitudes in the estimated coefficients. For the highly correlated variables, only the most significant variable was retained in the analysis; for example, weather condition was excluded because of its high correlation with road surface. Finally, a total of 10 covariates in the crash-level were used, i.e. Day of Week, Time of Day, Intersection Type, Nature of Lane, Road Surface, Street Lighting, Road Speed Limit, Vehicle Movement, Presence of Red Light Camera, and Pedestrian Involved. In addition, to explore how differently the various driver–vehicle characteristics affected the severity levels, five covariates in the individual-level, i.e. driver–vehicle level, were selected, i.e. Vehicle Type, Driver Age, Driver Gender, Involvement of Offending Party, Passenger Involved. Unfortunately, several vehicle safety features such as airbags, and anti-lock brakes, are not included in the crash dataset. But although those variables may be important to affect the individual severity, they are not so useful in Singapore since most vehicles are less than 6 years old and are hence equipped with the latest protective features in modern cars. Moreover, the stringent compulsory annual inspection on all vehicles to ensure they are road worthy means that these features are in serviceable conditions.

The definitions of the selected covariates, together with their mean and standard deviation (S.D.), are presented in Table 2. For convenience of analysis, all these variables were split as groups of dummy variables based on the engineering experiences or existing findings in previous studies. For example, Vehicle Type was categorized as three groups of two-wheel vehicle, light vehicle and heavy vehicle, since the vehicle weight had been identified relevant to injury severity (Evans and Frick, 1994).

Table 1
Summary of crash severity at signalized intersection by years

| Year | DI(A) | DI(B) | % of DI(A) | VD(A) | VD(B) | % of VD(A) | IS(A) | IS(B) | % of IS(A) |
|---|---|---|---|---|---|---|---|---|---|
| 2003 | 39 | 2622 | 1.49 | 491 | 2170 | 22.63 | 508 | 2153 | 23.59 |
| 2004 | 37 | 2885 | 1.28 | 398 | 2524 | 15.77 | 412 | 2510 | 16.41 |
| 2005 | 36 | 2221 | 1.62 | 173 | 2084 | 8.30 | 192 | 2065 | 9.30 |
| Total | 112 | 7728 | 1.45 | 1062 | 6778 | 15.67 | 1112 | 6728 | 16.53 |

*Note*: DI(A): driver with fatal/serious injury; VD(A): vehicle with extensive damage; IS(A): DI(A) or/and VD(A); DI(B): driver with slight or no injury; VD(B): vehicle with slight or no injury; IS(B): otherwise.

Table 2
Covariates used in the model

| Covariates | Description of the variables | Mean | S.D. |
|---|---|---|---|
| Day of week | If crash at weekend = 1, otherwise = 0 | 0.164 | 0.370 |
| Time of day | | | |
| Day time | If crash in 10 a.m.–5 p.m. = 1, otherwise = 0 | 0.289 | 0.453 |
| Night time | If crash at 8 p.m.–7 a.m. = 1, otherwise = 0 | 0.434 | 0.496 |
| Peak time | If crash at 7 a.m.–10 a.m. or 5 p.m.–8 p.m. = 1, otherwise = 0 | 0.278 | 0.448 |
| Intersection type | | | |
| X intersection | If crash at X type intersection = 1, otherwise = 0 | 0.014 | 0.115 |
| T/Y intersection | If crash at T/Y type intersection = 1, otherwise = 0 | 0.232 | 0.422 |
| Other types | If crash at other type intersection = 1, otherwise = 0 | 0.755 | 0.430 |
| Nature of lane | | | |
| Single lane | If crash on single lane = 1, otherwise = 0 | 0.025 | 0.155 |
| Left-most lane | If crash on Left-most lane = 1, otherwise = 0 | 0.163 | 0.369 |
| Right-most lane | If crash on right-most lane = 1, otherwise = 0 | 0.256 | 0.437 |
| Centre lane | If crash on centre lane = 1, otherwise = 0 | 0.556 | 0.497 |
| Road surface | If road surface is dry = 0, otherwise = 1 | 0.129 | 0.335 |
| Weather condition | If weather condition is fine = 0, otherwise = 1 | 0.098 | 0.297 |
| Street lighting | If street lighting is fine = 0, otherwise = 1 | 0.338 | 0.473 |
| Road speed limit (km/h) | | | |
| 40 | If road speed limit is 40 km/h = 1, otherwise = 0 | 0.005 | 0.068 |
| 50 | If road speed limit is 50 km/h = 1, otherwise = 0 | 0.891 | 0.311 |
| 60 | If road speed limit is 60 km/h = 1, otherwise = 0 | 0.072 | 0.258 |
| 70 | If road speed limit is 70 km/h = 1, otherwise = 0 | 0.032 | 0.176 |
| Vehicle movement | | | |
| Single vehicle self-skidded | If single vehicle self-skidded = 1, otherwise = 0 | 0.031 | 0.172 |
| Single vehicle against stationary object or pedestrian | If single vehicle against stationary object or pedestrian = 1, otherwise = 0 | 0.029 | 0.169 |
| Between moving vehicle and stationary vehicle | If between moving vehicle and stationary vehicle = 1, otherwise = 0 | 0.882 | 0.323 |
| Between moving vehicles | If between moving vehicles = 1, otherwise = 0 | 0.053 | 0.223 |
| Other movements | If other movements = 1, otherwise = 0 | 0.006 | 0.076 |
| Presence of red light camera | If a red light camera is present = 1, otherwise = 0 | 0.072 | 0.258 |
| Pedestrian involved | If passengers involved = 1, otherwise = 0 | 0.051 | 0.220 |
| Vehicle type | | | |
| Two-wheel vehicle | If vehicle type is motor scooter or motorcycle = 1, otherwise = 0 | 0.304 | 0.460 |
| Light vehicle | If vehicle type is motorcar, station wagon, goods can, pick-up or minibus = 1, otherwise = 0 | 0.572 | 0.495 |
| Heavy vehicle | If vehicle type is Bus, bendy, lorry, tip truck, trailer, crane or other heavy vehicles = 1, otherwise = 0 | 0.124 | 0.329 |
| Driver age | | | |
| $\leq 25$ | If driver age < = 25 = 1, otherwise = 0 | 0.162 | 0.368 |
| 26–45 | If driver age within 26–45 = 1, otherwise = 0 | 0.480 | 0.500 |
| 46–65 | If driver age within 46–65 = 1, otherwise = 0 | 0.326 | 0.469 |
| >65 | If driver age >65 = 1, otherwise = 0 | 0.033 | 0.178 |
| Driver gender | If driver is female = 1, otherwise = 0 | 0.104 | 0.305 |
| Involvement of offending party | If driver is likely at = faul = 1, otherwise = 0 | 0.627 | 0.484 |
| Passenger involved | If with passengers on board = 1, otherwise = 0 | 0.170 | 0.376 |

## 4. Model calibration and validation

A preliminary examination of potential within-crash covariance in the collected dataset identified a significant correlation between individuals involved in same multi-vehicle crashes, which represent 83.5% of all crashes at signalized intersections in Singapore. In particular, in a multi-vehicle crash, if the severity of driver–vehicle unit was IS(A), then the others had a probability of 31% also to be in IS(A). On the other hand, if a driver–vehicle unit was in IS(B), then the others had only 12% chance to be in IS(A). This significantly lower ratio clearly implies that the correlation among the individual severities in a multi-vehicle crash may exist. Hence, the proposed HBL model may be more appropriate in modeling the data than OBL model. The results for model calibration as well as quantitative assessment are presented in this section.

In the model calibration, beginning with the 15 covariates in the dataset, each variable was tested for the statistical sig-

Table 3
Posterior summaries of parameter estimates

| Parameters | Effect estimate | | Odds ratio | 95% BCI of odds ratio | |
|---|---|---|---|---|---|
| | Mean | S.D. | | 2.5% | 97.5% |
| Fixed effects | | | | | |
| Time of day | | | | | |
| Day time[a] | 0 | 0 | 1.00 | 1.00 | 1.00 |
| Night time | 0.17 | 0.09 | 1.19 | 1.04 | 1.39 |
| Peak time | −0.89 | 0.36 | 0.41 | 0.12 | 0.85 |
| Intersection type | | | | | |
| X intersection | −0.72 | 1.27 | 0.49 | 0.07 | 5.38 |
| T/Y intersection | 0.18 | 0.06 | 1.20 | 1.02 | 1.36 |
| Other types[a] | 0 | 0 | 1.00 | 1.00 | 1.00 |
| Nature of lane | | | | | |
| Single lane | −1.05 | 0.98 | 0.35 | 0.07 | 2.27 |
| Left-most lane | −0.37 | 0.42 | 0.69 | 0.33 | 1.50 |
| Right-most lane | 0.23 | 0.08 | 1.26 | 1.07 | 1.83 |
| Centre lane[a] | 0 | 0 | 1.00 | 1.00 | 1.00 |
| Street lighting | −1.17 | 0.34 | 0.31 | 0.14 | 0.59 |
| Presence of red light camera | 0.73 | 0.12 | 2.08 | 1.68 | 2.53 |
| Pedestrian involved | −0.96 | 0.46 | 0.38 | 0.14 | 0.92 |
| Vehicle type | | | | | |
| Two-wheel vehicle | 1.29 | 0.21 | 3.63 | 2.53 | 5.75 |
| Light vehicle[a] | 0 | 0 | 1.00 | 1.00 | 1.00 |
| Heavy vehicle | −2.07 | 0.36 | 0.13 | 0.11 | 0.23 |
| Driver age | | | | | |
| ≤25 | 0.15 | 0.13 | 1.16 | 1.02 | 1.43 |
| 26–45[a] | 0 | 0 | 1.00 | 1.00 | 1.00 |
| 46–65 | −0.16 | 0.19 | 0.85 | 0.61 | 1.19 |
| >65 | 0.53 | 0.28 | 1.70 | 1.03 | 3.74 |
| Involvement of offending party | 0.49 | 0.13 | 1.63 | 1.21 | 2.14 |
| Random effects | | | | | |
| Between-crash variance ($\tau_0^2$) | 1.34 | 0.87 | | 0.56 | 2.29 |
| Within-crash variance | 3.29 | | | | |
| ICC | 0.289 | | | | |

[a] Represents the reference category used in the model for the multinomial variable.

nificance and the insignificant ones were eliminated. In the final model, three chains of 20,000 iterations each produced trace plots with a good degree of mixing, and Brooks, Gelman and Rubin convergence diagnostics (Brooks and Gelman, 1998) using Bayesian output analysis (BOA) program (Smith, 2001) indicated convergence. Particularly, after discarding 10,000 burn-in samples and thinning to retain every fifth sample to reduce autocorrelation (leaving a total of 6000 posterior samples), the 0.975 quantiles of the corrected scale reduction factor (CSRF) for the parameters were each 1.2 or less. Posterior distributions were all uni-modal. The means, standard deviations and associated 95% BCI of estimated random effects and regression coefficients were monitored and listed in the Table 3.

To check the model adequacy, underlying assumptions for the HBL model in Eq. (5) were assessed. Posterior samples of the crash-level random effects ($u_{0j}$) can be thought of as residuals, and thus can be examined with usual model diagnostics. In the MCMC simulation, 200 random effects $u_{0j}$ were randomly sampled, and the fact that they averaged very close to zero was reassured. Normal probability plots, revealing no strong abnormalities, also validate the normality and exchangeability assumptions.

As shown in Table 3, the variance of $u_{0j}(\tau_0^2)$, indicating the magnitude of the between-crash variance, is 1.34. Hence, the ICC is calculated by:

$$\rho = \frac{1.34}{1.34 + \pi^2/3} = 28.9\%$$

This means that 28.9% of unexplained variations in individual severity were resulted from between-crash variance, which strongly suggests the usefulness of the model specification of hierarchical structure. If an OBL mode was implemented without considering the random effects between crashes, the results will be biased and inaccurate.

Model comparison using DIC further strengthened this argument. DIC values for fitted OBL model Eq. (6) and HBL model Eq. (5) are given in Table 4. Results show that $\overline{D(\gamma)}$ of HBL model (1984.5) is less than one third of that obtained in OBL model (6165.5). After penalized by $p_D$, the DIC value for HBL

Table 4
Results of model comparison using DIC

| | $\overline{D(\gamma)}$ | $D(\bar{\gamma})$ | $p_D$ | DIC |
|---|---|---|---|---|
| Ordinary logistic model | 6165.5 | 6139.1 | 26.4 | 6191.9 |
| Hierarchical logistic model | 1984.5 | 901.1 | 1083.4 | 3067.9 |

model (3067.9) is also hugely less than that in OBL model (6191.9). This further proves that the use of crash-level random effects in HBL model can substantially improve the model fit.

## 5. Discussions on significant risk factors

Summary statistics for the posterior samples of fixed effects of significant covariates are presented in Table 3. In the final HBL model, nine variables were identified as significant judged by 95% BCI. They are: (1) Time of Day, (2) Intersection Type, (3) Nature of Lane, (4) Street Lighting, (5) Presence of Red Light Camera, (6) Pedestrian Involved, (7) Vehicle Type, (8) Driver Age, (9) Involvement of Offending Party. The detailed interpretations for these significant risk factors are offered in the following.

### 5.1. Time of day

The time of crash occurrence was classified into three periods, i.e., day time (10 a.m.–5 p.m.), night time (8 p.m.–7 a.m.), and peak time (7 a.m.–10 a.m. or 5 p.m.–8 p.m.). Compared with crash occurrences during day time, crashes which occur at night time have 19% higher odds of high severity (IS(A)) (O.R. 1.19, 95% BCI (1.04, 1.39)). This finding is consistent with Simoncic (2001) who found crashes at night were more serious than those during daytime. This may be expected since speeding and alcohol use resulting in higher crash severity are more likely in these hours. Moreover, at night the effect of street lighting comes into play and this was also found to be significant in this study. The high probability of IS(A) in night time is consistent with previous studies for severities of motorcycle crashes (Quddus et al., 2002) and single vehicle crashes (Rifaat and Chin, 2005) in Singapore. Furthermore, individuals involved at crashes in peak time (O.R. 0.41, 95% BCI (0.12, 0.85)) were also found to have reduced odds of being IS(A) by 60%. It can be reasoned that due to the higher traffic volume, the vehicle speeds during peak time are substantially reduced compared to off-peak time, hence resulting in lower crash severity. This is consistent with Zhang et al. (2000), in which the odds of fatality in crashes that occurred in 70–90 kph zones were almost six times more than those in crashes occurring in zones with slower speeds.

### 5.2. Intersection type

It was found that crashes occurring at T/Y type intersections (O.R. 1.20, 95% BCI (1.02, 1.36)) increase the odds of being IS(A) by 20%, in contrast to other type of intersections. Results indicate that, though insignificant, X type intersections may have an averagely positive effect on reducing the crash severity. Vehicles on the minor road at T/Y type intersections, merging into the major road, have a higher probability to be seriously collided by the going-through vehicles on the major road. This is similar to the right-turn traffic (left-driving) at X type intersections. In addition, a shorter sight distance, commonly associated with a T/Y type intersections, may also be a factor causing more severe crashes.

### 5.3. Nature of lane

Another significant geometric factor is Nature of Lane, where the right-most (left driving) lane (O.R. 1.26, 95% BCI (1.07, 1.83)) was identified to be significant on increasing the odds of severe crashes by 26%, compared with central lane. This result is consistent with the Khorashadi et al. (2005) who found that for right driving, if the location of collision is on the left lane, the likelihood of injury severity increased by 268.1%. The higher severity risk may be caused by higher speed on right-most lane than on other lanes. According to Bedard et al. (2002), traveling at speeds exceeding 112 kph was independently associated with a 164% increase in the odds of a fatality compared with speeds less than 56 kph.

### 5.4. Street lighting

Street lighting was identified as a significant factor (O.R. 0.31, 95% BCI (0.14, 0.59)). The odds ratio value indicates that a bad street lighting condition can increase the odds of severe crash by about 69%. This result is generally expected because drivers may have more reaction time and better perception ability on crash risk in good street lighting environments. Yau (2004) also found that street lighting condition affects the crash severity for the single vehicle crashes in Hongkong. This finding implies that improving the street lighting can substantially improve the safety condition at intersections.

### 5.5. Presence of red light camera

Results show that among the highly significant risk factors, Presence of Red Light Camera (O.R. 2.08, 95% BCI (1.68, 2.53)) is associated negatively with crash severity. In other words, the presence of red light camera is associated with higher severity level. In the sites with red light camera, the odds of being IS(A) increase by 108%. This may seem surprising compared to findings in many studies in which the red light camera has been proved to be useful in reducing the violation and crash frequencies, as well as relieving the crash severity. In a recent driver behavior study in Singapore, Huang et al. (2006) have found that the presence of a red light camera is effective in curbing the red light running as well as reducing crash risk in angle crashes. Although red light camera itself may not increase the risk of severe crashes, it is associated with high risk sites. Specifically, intersections with red light camera may have already been placed in sites with more severe crashes since traffic authorities always install cameras at extraordinarily hazardous sites. Moreover, this reinforces the findings by Chin and Quddus (2003), where the presence of a surveillance camera was found to be associated with an increase in the total crash frequency at intersections. These results imply that, keeping other covariates unchanged, some unmeasured factors may have effects on the relative severity.

### 5.6. Pedestrian involved

The variable Pedestrian Involved is a significant factor affecting driver severity (O.R. 0.38, 95% BCI (0.14, 0.92)). The involvement of pedestrians substantially reduces the odds of being IS(A) by about 62%. This is intuitively reasonable since pedestrians, rather than the drivers, are much easier to be injured seriously in the collisions. It is also supported by Chang and Wang (2006), who found that pedestrians were more likely to have higher risks of being injured than other types of vehicle drivers in traffic crash. Crash severity statistic also confirms this finding that of driver–vehicle units involved in the crashes of “vehicle against pedestrian” type, only 3.4% were injured severely and/or damaged extensively, compared with the overall rate of 16.5% as shown in Table 1.

### 5.7. Vehicle type

Vehicle type was classified as three categories in this study, i.e., two-wheel vehicle, light vehicle, and heavy vehicle. By taking the most common light vehicle as reference, the other two dummy variables for two-wheel vehicle (O.R. 3.63, 95% BCI (2.53, 5.75)) and heavy vehicle (O.R. 0.13, 95% BCI (0.11, 0.23)) were all found to have significant effects on individual severity. Compared with light vehicle, two-wheel vehicle increased the odds of being IS(A) by 263%, representing the most significant factor in the model. The severity risk in two-wheel vehicle (e.g. motorcycles) is expected as two-wheel riders do not have the facility of safety protections that are available in light vehicle (e.g. cars), such as seatbelt, airbag etc. Again the two-wheel riders may be thrown off from the vehicle at the time of collision while in the case of car crashes this may rarely happen. Kocklelman and Kweon (2002) found that riding a motorcycle is causing more severe injury than driving a car. Again heavy vehicle reduces the odds of being IS(A) by 87%. It is not surprising that as the vehicle weight increases, the risks of being injured or damaged decrease substantially, even though other driver–vehicle units involved in the same crash may be more vulnerable to be injured or damaged. This finding is also supported by Levine et al. (1999), who reported that every 454 kg (1000 lbs) increase in vehicle weight was equivalent to the driver’s ability to withstand front impact crashes of 10 more kph (6 mph) before being fatally injured. However, it is interesting to notice that as found in Rifaat and Chin (2005), the truck crashes in single vehicle crashes are more likely to result in serious injuries and fatalities. This contradiction can be explained by the different collision types between intersection crash and single vehicle crash. In contrast to intersection crash, more severe crashes may be caused by higher energy exchange for trucks with roadside objects in single vehicle crashes. Moreover, as found in Rifaat and Chin (2005), the higher relative fatality risk was associated with truck crashes mainly on high speed roads such as expressway rather than other highway types where signalized intersections are located.

### 5.8. Driver age

The demographic variable, driver age, was found to be significant on individual severity, in which both young group (O.R. 1.16, 95% BCI (1.02, 1.43)) and aged group (O.R. 1.70, 95% BCI (1.03, 3.74)) were identified to have effects on increasing the odds of being IS(A). Odds ratios indicate that a 16% increase of the IS(A) odds is associated with young drivers while 70% for aged drivers. It is likely because young drivers drive more recklessly (Rifaat and Chin, 2005; Kocklelman and Kweon, 2002) while aged drivers have relatively weak risk detecting and reacting abilities. Again Hilakivi et al. (1989) also showed that young drivers as well as older drivers are more at risk of being involved in severe crashes. Another reason for young drivers to be involved with severe crashes may be that they represent a large proportion of riders of two-wheel vehicles, which have been proven to be associated with a higher risk of being involved in more severe crashes (Rifaat and Chin, 2005; Quddus et al., 2002). Furthermore, as indicated by Rifaat and Chin (2005), decrease of visual power, deterioration of muscle strength and reaction time may be responsible for the aged drivers to be involved in severe crashes.

### 5.9. Involvement of offending party

Involvement of offending party affects crash severity significantly (O.R. 1.63, 95% BCI (1.21, 2.14)). The at-fault driver–vehicle unit has 63% higher odds to be IS(A) than the not-at-fault party. This provides a more convincible evidence for educating drivers to keep away from risk-taking maneuvers.

## 6. Conclusions and recommendations

This study developed a Bayesian HBL model to identify the risk factors on individual severity of driver injury and vehicle damage at urban intersections. It is helpful to account for the severity correlation of driver–vehicle units involved in the same multi-vehicle crashes. The estimation of random effects using ICC showed that 28.9% of unexplained variation in severity level was resulted from between-crash variance. Model comparison with ordinary logistic model using DIC further ensured the suitability and model-improving effectiveness of introducing the crash-level random effects. This means, if ordinary logistic model was used, 28.9% residual variance could not be explained by the model, which might result in inaccurate coefficient estimates of risk factors. The Bayesian hierarchical modeling approach also showed flexibilites to explicitly explore the hierarchical data structure in traffic safety field.

Of the covariates including various geometric features, traffic conditions, and driver–vehicle characteristics, nine variables were identified as significant using 95% BCI. Among these, the crash-level significant factors are Time of Day, Intersection Type, Nature of Lane, Street Lighting, Presence of Red Light Camera, and Pedestrian Involved. In particular, it was found that crashes occurring in peak time, in good street-lighting condition, and in the case of pedestrians involved are associated with lower severity, while those occurring in night time, at T/Y type intersections, on right-most lane, and in the presence of red light cameras have larger odds of being severe. Vehicle type, Driver Age and Involvement of Offending Party were also found to affect severities of driver injury and vehicle damage significantly. Specifically, results indicated that heavy vehicles have a better resistance on serious injury or extensive damage, while two-wheel vehicles, young or aged drivers, with the involvement of offending party have a higher risk of being high severity.

This study has a great potential in traffic safety discipline, especially when the correlation exists in the dataset. This study illustrated a way to analyze the potential within-crash correlations in severity study using the hierarchical modeling technique. It also proved and emphasized the importance of accounting for this kind of within-cluster correlation in yielding reliable and accurate effect estimates for various risk factors.

Two important extensions of this research can be proposed. Firstly, while this study only considered the random intercept in the regression equations, the random effects on the covariate coefficients can also be examined with careful specifications, resulting in random slope model Eqs. (1–3). In the random slope model, the cross-level interaction between covariates could be appropriately specified and estimated. Secondly, the hierarchical data structure in traffic data is not only limited in crash-specific correlation in severity analysis. A more general form can also be proposed for traffic safety study to be a five-level hierarchy, i.e., geographic region–traffic site–crash–vehicle–occupant. The involvement and emphasis for different sub-groups of these levels depend on different research purposes and also rely on the heterogeneity examination on crash data employed. The Bayesian approach provides us with a flexible and reliable model calibration and assessment measure for these potential explorations and applications.

## References


Abdel-Aty, M., Keller, J., 2005. Exploring the overall and specific crash severity levels at signalized intersections. Accid. Anal. Prev. 37 (3), 417–425.

Al-Ghamdi, A.S., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accid. Anal. Prev. 34 (6), 729–741.

Bedard, M., Guyatt, G.H., Stones, M.J., Hirdes, J.P., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accid. Anal. Prev. 34 (6), 717–727.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulation. J. Comput. Graph. Stat. 7 (4), 434–455.

Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accid. Anal. Prev. 38 (5), 1019–1027.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accid. Anal. Prev. 35 (2), 253–259.

Evans, L., 1992. Car size or car mass: which has greater influence on fatality risk? Am. J. Pub. Health 82 (8), 1105–1112.

Evans, L., 1993. Mass ratio and relative driver fatality risk in two-vehicle crashes. Accid. Anal. Prev. 25 (2), 213–224.

Evans, L., Frick, M., 1994. Car mass and fatality risk: has the relationship changed? Am. J. Public Health 84 (1), 33–36.

Gelman, A., Carlin, J.B., Stern, H.S., 2003. Bayesian Data Analysis, 2nd ed. Chapman & Hall, New York.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1995. Markov Chain Monte Carlo methods in Practice. Chapman & Hall, New York.

Goldstein, H., 2003. Multilevel Statistical Models, 3nd ed. Edward Arnold, London.

Hauer, E., 2006. The frequency-severity indeterminacy. Accid. Anal. Prev. 38 (1), 78–83.

Hilakivi, I., Veilahti, J., Asplund, P., Sinivuo, J., Laitinen, L., Koskenvuo, K., 1989. A sixteen-factor personality test for predicting automobile driving accidents of young drivers. Accid. Anal. Prev. 21 (5), 413–418.

Huang, H.L., Chin, H.C., Heng, H.H., 2006. Effect of red light camera on accident risk at intersections. In: TRB 2006 Annual Meeting CD-ROM, Transportation Research Board, National Research Council, Washington, DC.

James, J.L., Kim, K.E., 1996. Restraint use by children involved in crashes in Hawaii, 1986–1991. Transportation Research Record 1560, 8–11.

Jones, A.P., Jorgensen, S.H., 2003. The use of multilevel models for the prediction of road accident outcomes. Accid. Anal. Prev. 35 (1), 59–69.

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: An explanatory analysis. Accid. Anal. Prev. 37 (5), 910–921.

Kim, D.G., Washington, S., 2006. The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. Accid. Anal. Prev. 38 (6), 1094–1100.

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

Kim, D.G., Lee, Y., Washington, S., Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: application of hierarchical binomial logistic models. Accid. Anal. Prev. 39 (1), 125–134.

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

Lenguerrand, E., Martin, J.L., Laumon, B., 2006. Modeling the hierarchical structure of road crash data: application to severity analysis. Accid. Anal. Prev. 38 (1), 43–53.

Levine, E., Bedard, M., Molloy, D.W., Basilevsky, A., 1999. Determinants of driver fatality risk in front impact fixed object collisions. Mature Med. Canada 2, 239–242.

Mannering, F.L., Grodsky, L.L., 1995. Statistical analysis of motorcyclists' perceived accident risk. Accid. Anal. Prev. 27 (1), 21–31.

Mercier, C.R., Shelley, M.C., Rimkus, J., Mercier, J.M., 1997. Age and gender as predictors of injury severity in head-on highway vehicular collisions. Transp. Res. Rec. 1581, 37–46.

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

O'Donnell, C.J., Connor, D.H., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accid. Anal. Prev. 28 (6), 739–753.

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 (4), 445–462.

Rifaat, S.M., Chin, H.C., 2005. Analysis of severity of single-vehicle crashes in Singapore. In: TRB 2005 Annual Meeting CD-ROM, Transportation Research Board, National Research Council, Washington, DC.

Shankar, V., Mannering, F., 1996. An exploratory multinomial Logit analysis of single-vehicle motorcycle accident severity. J. Saf. Res. 27 (3), 183–194.

Simoncic, M., 2001. Road fatalities in Slovenia involving a pedestrian, cyclist or motorcyclist and a car. Accid. Anal. Prev. 33 (2), 147–156.

Smith, B.J., 2001. Bayesian Output Analysis Program (BOA). Version 1.0.0 for S-PLUS and R. available at http://www.public-health.uiowa.edu/boa.

Snijders, A.B., Bosker, R.J., 2000. Multilevel Analysis, An Introduction to Basic and Advanced Multilevel Modeling. SAGE Publications, London.

Speigelhalter, D.J., Best, N.G., Carlin, B.P., Linde, V.D., 2003b. Bayesian measures of model complexity and fit (with discussion). J. Roy. Stat. Soc., Series B 64 (4), 583–616.

Spiegelhalter, D.J., Thomas, A., Best, N.G., Lunn, D., 2003a. WinBUGS version 1.4.1 User Manual. MRC Biostatistics Unit, Cambridge, UK.

Washington, S., Congdon, P., Karlaftis, M., Mannering, G., 2005. Bayesian multinomial logit models: exploratory assessment of transportation applications. In: TRB 2005 Annual Meeting CD-ROM, Transportation Research Board, National Research Council, Washington, DC.

Yau, K., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accid. Anal. Prev. 36 (3), 333–340.

Zhang, J., Lindsay, J., Clarke, K., Robbins, G., Mao, Y., 2000. Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. Accid. Anal. Prev. 32 (1), 117–125.