




# Injury severity analysis: comparison of multilevel logistic regression models and effects of collision data aggregation

Taimur Usman<sup>1</sup>  · Liping Fu<sup>1,2</sup> · Luis F. Miranda-Moreno<sup>3</sup>

Received: 6 May 2015 / Revised: 27 January 2016 / Accepted: 28 January 2016 / Published online: 23 February 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** This paper describes an empirical study aiming at identifying the main differences between different logistic regression models and collision data aggregation methods that are commonly applied in road safety literature for modeling collision severity. In particular, the research compares three popular multilevel logistic models (i.e., sequential binary logit models, ordered logit models, and multinomial logit models) as well as three data aggregation methods (i.e., occupant based, vehicle based, and collision based). Six years of collision data (2001–2006) from 31 highway routes from across the province of Ontario, Canada were used for this analysis. It was found that a multilevel multinomial logit model has the best fit to the data than the other two models while the results obtained from occupant-based data are more reliable than those from vehicle- and collision-based data. More importantly, while generally consistent in terms of factors that were found to be significant between different models and data aggregation methods, the effect size of each factor differ substantially, which could have significant implications for

evaluating the effects of different safety-related policies and countermeasures.

**Keywords** Injury severity · Multilevel logistic regression models · Collision data aggregation

## 1 Introduction

The outcome of a collision is polytomous in nature such as no injury (NI), minimal injury, minor injury, major (incapacitating) injury, and fatal injury. This type of data is mostly modeled using logistic regression models. Most of the models are extensions of the multinomial logit models based on the assumption of independent severity classes [1–11]. Although different modeling methodologies are available from literature to examine collision severity as related to various influencing factors, little is known on the relative merits of these alternatives. The first objective of this research is therefore to compare three most widely used logistic regression models, namely, sequential binary logit models, ordered logit models, and multinomial logit models in a multilevel framework for injury severity analysis.

Some of the issues related to injury severity analysis are within-crash correlation, hierarchical nature of collision data, misclassification, underreporting, endogeneity, sample size, and spatial correlation [5, 11–21]. While a number of recent studies have been devoted to addressing some of these issues, the issue pertaining to the hierarchical nature of collision data has not been addressed adequately. Collision data is hierarchical in nature with possible correlation at the occupant or vehicle level. Ignoring such correlation (intra-class correlation) could lead to false estimation of standard errors and undermine the true

---

✉ Taimur Usman  
taimurus@gmail.com

Liping Fu  
lfu@uwaterloo.ca

Luis F. Miranda-Moreno  
luis.miranda-moreno@mcgill.ca

<sup>1</sup> Department of Civil & Environmental Engineering,  
University of Waterloo, Waterloo, ON N2L 3G1, Canada

<sup>2</sup> School of Transportation and Logistics, Southwest Jiaotong  
University, Chengdu 610031, China

<sup>3</sup> Department of Civil Engineering & Applied Mechanics,  
McGill University, Montreal, QC H3A 2K6, Canada

significance of parameter estimates [22]. However, little work has been done to account for the multilevel structure of the collision data. Jones and Jørgensen [17] and Lenguerrand et al. [20] were among the first, as identified in Usman et al. [23], to recognize the need to consider the hierarchical crash-car-occupant structure of collision data for crash severity modeling. They discussed the potential issues of ignoring the clustering nature of data and the correlation within the clusters, such as erroneous estimates of model coefficients and understated standard errors and confidence intervals for the effects. They, however, did not discuss the effects of data aggregation. Their conclusions were similar to those from other disciplines such as epidemiology, social research, and political science [24–27]. The second objective of this research is therefore to evaluate the effects of data aggregation through an empirical investigation using three levels of aggregations, i.e., occupant level, vehicle level, and collision level.

This paper contributes to the literature by generating new knowledge about the implications of different modeling alternatives and data aggregation methods for collision severity analysis. The paper first describes the data used in the empirical investigation, including study sites, data sources, and data processing and integration. The three

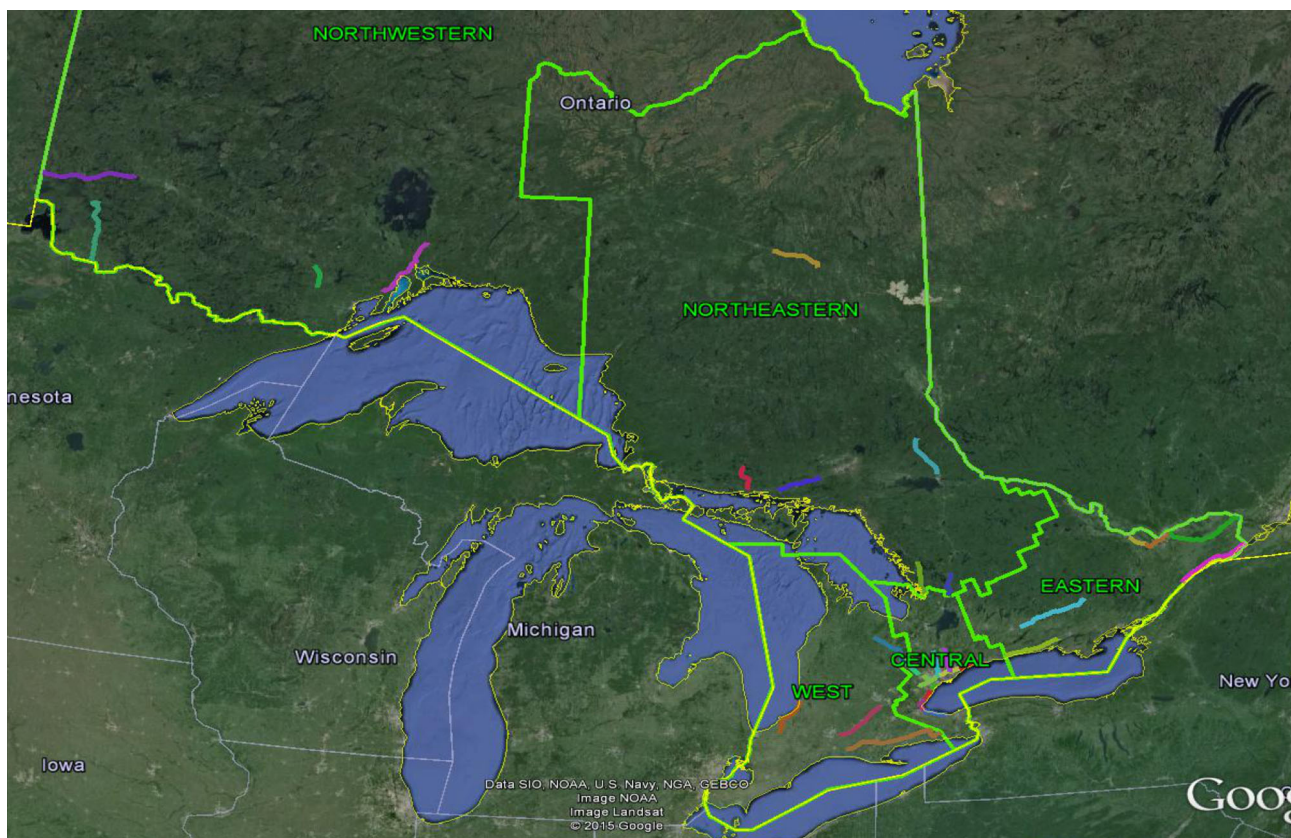
different data aggregation methods are discussed in details. Then, an overview of the three logistic regression models in the construct of the multilevel framework is provided, followed by a discussion on the model calibration process and the results. Finally, the main findings are summarized, focussing particularly on the differences from different approaches.

## 2 Data description

This research makes use of a collision database prepared in our previous effort [23, 28, 29]. This dataset is unique in several aspects, including reliable observations on traffic and environmental conditions when the collision occurred, and extensive spatial and temporal coverage, as described in the following section.

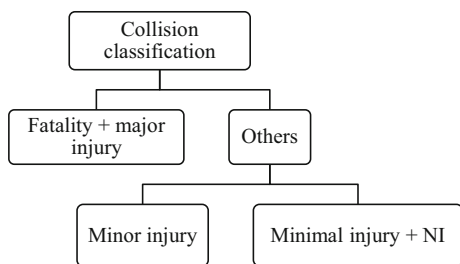
### 2.1 Study sites

A total of 31 patrol routes, each representing a highway section covered by a single maintenance unit (yard), from different regions of Ontario, Canada, were selected for this analysis as shown in Fig. 1. These sites were selected based



**Fig. 1** Study sites





**Fig. 3** Data classification scheme (vehicle- and collision-based data)

4. Major injury, where the victim was admitted to the hospital either for treatment or observation;
5. Fatality, where the victim died within 30 days of collision or on site.

Minimal injury and NI collisions were grouped together into one category because they are similar in terms of consequence. Similarly, major injuries and fatalities were also grouped into a single category. This merging of categories will also take care of the possible correlation that could exist between those closely related outcomes of a collision severity [12, 30]. The hierarchic structure of collision data is shown in Fig. 2, which shows that for a given collision, vehicles are nested within the collision and persons are nested within vehicles and each person could have a given level of severity.

Data from other sources such as weather and traffic were merged with the person-based collision data based on date, time, and location for the 31 patrol routes. A stepwise aggregation process was followed to convert the data from occupant-based records to vehicle-based, and finally to collision-based records. Three datasets were thus formed for this analysis: occupant-based dataset with three levels (occupant—vehicle—collision), vehicle-based dataset with two levels (vehicle—collision), and collision-based dataset with a single level. For the vehicle- and collision-based data, collision severity levels were assigned to the respective vehicles and collisions as per the classification scheme shown in Fig. 3. Note that this classification scheme was not used for occupant-based data as each person has a unique injury severity level.

### 3 Model development

Different approaches can be used for collision severity analysis: (a) incorporating severity into the collision frequency models by modeling collisions classified by severity types [31–34]; and (b) modeling the conditional probability of each severity level for a given collision [14, 15, 17, 35, 36]. In this research, we adopted the second approach for three reasons: (i) different factors could have

different effects on collision occurrence and severity (e.g., seat belt use has nothing to do with collision occurrence, but is an important factor in severity analysis); (ii) data that could be used for joint models are limited in nature because most of the data are collected after the collision has happened [12]; and, (iii) consequence outcomes and injury data are at the individual, vehicle, or accident level. Three different model structures were considered for the conditional probability of a collision for each of the three datasets discussed previously.

Multilevel framework is used to account for the correlation between vehicles in a collision or persons in a vehicle. In a multilevel setting, correlation at a sub-level is taken care of by inclusion of random parameters which are constant within the sub-level but are allowed to vary at the upper levels [18, 20, 37].

#### 3.1 Multilevel logistic regression models

The first modeling structure considered is the multilevel multinomial logit model. In this model, a base category is selected out of the different outcomes and other categories are estimated with respect to the base category. Many researchers have used multinomial logit models for accident severity analysis [1–10]. If the three severity levels are represented by 0, 1, and 2 with 0 as the reference or base category then the model structure for a three-level data structure (occupant-based data) is given by Eq. (1). The resulting models are called multilevel multinomial logit models (MML).

$$\ln \left[ \frac{P(Y = 1/X)}{P(Y = 0/X)} \right] = \beta_{10} + \sum_{n=1}^N \beta_{1n} X_{1ijn} + U_{jk} + V_k, \left. \begin{array}{l} \ln \left[ \frac{P(Y = 2/X)}{P(Y = 0/X)} \right] = \beta_{20} + \sum_{n=1}^N \beta_{2n} X_{2ijn} + U_{jk} + V_k, \end{array} \right\} \quad (1)$$

where  $P$  represents the probability of severity level (either 0, 1 or 2);  $i$ ,  $j$ , and  $k$  represent occupant, vehicle, and collision levels, respectively;  $U_{jk}$  and  $V_k$  denote the second level (vehicle) and the third level (collision) random effect factors which are assumed to follow a logistic distribution;  $\beta$  is a model coefficient to be estimated; and  $X_{ijk}$  represents a set of explanatory variables at the individual level.  $U_{jk}$  remains constant for occupants within a vehicle but varies across vehicles and collisions. Similarly,  $V_k$  is constant for vehicles in a collision but varies across collisions.  $U_{jk}$  and  $V_k$  are obtained by considering the intercept as a random parameter.

The second modeling structure is the sequential binary logistic model. Collision data were divided into two mutually exclusive injury outcomes for a given collision at a given level, and binary logit models were specified at

**Table 1** List of variables used in the analysis

| Category                           | Variable   | Definition  |
|------------------------------------|--|---|
| Road characteristics               | Road classification                                | Freeway = 1, kings highway multilane = 2, kings highway two lane <sup>c</sup> = 3               |
|                                    | Road alignment                                     | Straight on level <sup>c</sup> = 1, straight on hill = 2, curve on level = 3, curve on hill = 4 |
|                                    | Number of lanes                                    | Number of lanes   |
|                                    | Collision location                                 | Intersections = 1, segment <sup>c</sup> = 2, bridges/underpasses = 3                            |
|                                    | Speed limit  | km/h  |
| Weather and environment conditions | Light  | Light + dawn = 1, dark + dusk <sup>c</sup> = 2  |
|                                    | Precipitation type                                 | Other <sup>c</sup> = 0, freezing rain/snow = 1  |
|                                    | Hourly precipitation                               | Precipitation intensity in “cm”   |
|                                    | Temperature  | Measured in °C  |
|                                    | Wind speed   | km/h  |
|                                    | Visibility   | km  |
|                                    | Road surface condition                             | Road surface condition in winter i.e., icy, snow covered, etc. represented by RSI               |
| Vehicle                            | Day  | Weekdays = 0, weekends <sup>c</sup> = 1   |
|                                    | Vehicle type <sup>b</sup>                          | SUVs/car/station wagon <sup>c</sup> = 1, van = 2, large trucks etc. = 3                         |
|                                    | Vehicle condition <sup>b</sup>                     | Otherwise = 0, defective <sup>c</sup> = 1   |
|                                    | Vehicle age <sup>b</sup>                           | In years  |
| Driver/person                      | Driver age <sup>b</sup>                            | In years  |
|                                    | Driver sex <sup>b</sup>                            | Male = 1, female <sup>c</sup> = 2   |
|                                    | Driver condition at time of collision <sup>b</sup> | Otherwise <sup>c</sup> = 0, normal = 1  |
|                                    | Position in vehicle <sup>a</sup>                   | Front = 1, rear <sup>c</sup> = 2  |
| Traffic                            | Safety equipment used <sup>a</sup>                 | Used safety device = 0, not used or bad use <sup>c</sup> = 1                                    |
|                                    | Hourly traffic volume                              | ln(Hourly traffic volume)   |

<sup>a</sup> Used for occupant-based data only

<sup>b</sup> Used for vehicle- and occupant-based data only

<sup>c</sup> Base category

**Table 2** Collision count by severity and percent change in collision severity distribution due to data aggregation

| All sites   | Minimal injury + NI | Minor injury | Fatality + major injury | Total  |
|---|---------------------|--------------|-------------------------|--------|
| Collision count by severity   |                     |              |                         |        |
| Occupant-based data   | 30,246              | 7,733        | 1,585                   | 39,564 |
| Vehicle-based data  | 12,698              | 5,905        | 1,032                   | 19,635 |
| Collision-based data  | 7,349               | 5,531        | 895                     | 13,775 |
| Percent change in collision severity distribution due to data aggregation |                     |              |                         |        |
| Occupant-based data   | 76.45 %             | 19.55 %      | 4.00 %                  |        |
| Vehicle-based data  | 64.67 %             | 30.07 %      | 5.26 %                  |        |
| Collision-based data  | 53.35 %             | 40.15 %      | 6.50 %                  |        |

each level such as shown in Fig. 2. Many researchers have used binary logit models for accident severity analysis [5, 11, 13–21].

For multilevel data, the resulting model is called the multilevel sequential binary logit model (MBL). The mathematical form of the model for a three-level data structure (occupant-based data) is shown in Eq. (2):

$$\ln \left[ \frac{P(Y = 1)}{P(Y = 0)} \right] = \beta_0 + \sum_{n=1}^N \beta_n X_{ijkn} + U_{jk} + V_k, \quad (2)$$

where  $P$  represents the probability of severity level (either 0, or 1).

The third modeling structure considered in this research is multilevel ordered logit model. Ordered logit models are

extensions of multinomial logit models to account for the inherent ordering of severity levels in collisions, such as, from no injury to injury and to fatal [10, 38–44]. The mathematical form of a multilevel ordered logit model (MOL) for a three-level data structure (occupant-based data) is shown in Eq. (3):

$$\log \left[ \frac{S_{ijk}^s}{S_{ijk}^r} \right] = \beta_0 + \beta_1 X_{ijk} + U_{jk} + V_k, \quad (3)$$

where severity (represented by “S”) with superscript “r” represents the base severity against which other severity levels, denoted by superscript “s,” are compared at the occupant level. The reference category could be either the least or most severe one. If  $Y$  denotes the observed severity level,  $Y^*$  the unobserved injury severity level from Eq. (3), and  $\mu_1, \mu_2, \dots, \mu_j$  the cut-off points or threshold values for the injury severity levels, then

$$\begin{aligned} Y &= 1 \text{ if } Y^* \leq \mu_1, \\ Y &= 2 \text{ if } \mu_1 < Y^* \leq \mu_2, \\ &\vdots \\ Y &= j \text{ if } \mu_{j-1} < Y^*. \end{aligned} \quad (4)$$

The probability of a particular injury severity level  $Y = j$  can be estimated using Eq. (5) [45]:

$$\begin{aligned} P(Y = j) &= P(\mu_{j-1} < Y^* < \mu_j) \\ &= \frac{\exp(\mu_j - \sum \beta_k x_k)}{1 + \exp(\mu_j - \sum \beta_k x_k)} - \frac{\exp(\mu_{j-1} - \sum \beta_k x_k)}{1 + \exp(\mu_{j-1} - \sum \beta_k x_k)}, \end{aligned} \quad (5)$$

where  $\beta_k$  are model coefficients to be estimated and  $\{X_1, X_2, \dots, X_k\}$  represents a set of explanatory variables. An important aspect of ordered logit models is the proportional odds (or parallel slopes) assumption, where the variables are assumed to have the same slope across all levels of severity/outcome [46–48] with the exception of the intercept [49]. Results of ordered logit models are therefore unidirectional (show either an increase or decrease in severity) and are thus very easy to interpret. This unidirectional effect can sometimes lead to undesirable effects where a variable could cause the probability of high or low severity collision to increase at the cost of the other [38].

The presence of correlation is confirmed by calculating the intra-class correlation (correlation among observations within the same cluster). Intra-class correlation, denoted by  $\rho$ , is a coefficient with values ranging from 0 to 1 and is calculated as the ratio of the variance at the sub-level to the total variance [23, 50, 51] as given in Eq. (6):

$$\rho = \frac{\sigma_{\text{sub-level}}^2}{\sum \sigma^2}. \quad (6)$$

The higher the value of  $\rho$ , the greater the correlation is and the higher the consequences of ignoring it will be [30]. For details on how  $\rho$  can be calculated, readers are referred to e.g., Jones and Jørgenson [18].

### 3.2 Exploratory data analysis

There are a large number of factors that influence the severity of collisions under winter conditions [52, 53]. The main factors can be grouped into three categories, namely road driving conditions, vehicle characteristics, and driver attributes. Road driving conditions include road geometry, environment, and pavement surface conditions. The latter are affected by weather and maintenance operations. Different sets of variables were considered in analyzing the three datasets as listed in Table 1.

Table 2 provides a summary of collision counts by severity for the different datasets and the changes in the proportions of different types of injury severity levels due to aggregation at each step.

As shown in Fig. 2, a collision may involve several vehicles and the occupants of an involving vehicle may experience different levels of injury severity. As a result, modeling the collision severity at the collision level will result in a loss of information and misrepresentation of certain severity levels, as show in Table 2. For example, if we aggregate data for a collision with three fatal injuries and two vehicles involved, the fatality count for occupant-, vehicle-, and collision-based datasets will be three (03), two (02), and one (01), respectively.

## 4 Model calibration and results

MLwin<sup>1</sup> was used to calibrate the three alternative models discussed in Sect. 3. Tables 3 through 5 provide the calibration results for collision-based data, vehicle-based data, and occupant-based data. MLWin uses Quasi-likelihood for models with discrete dependent variables and thus the reported likelihood estimates are only approximate leading to unreliable likelihood ratio tests [54]. A positive sign is used as an indicator of increase in severity level with respect to the associated variable. Results from all the models are consistent in terms of the direction of their effect on severity; however, effect of the size of coefficient varies across different models and aggregation levels. For

<sup>1</sup> Rasbash, J., Charlton, C., Browne, W.J., Healy, M. and Cameron, B. (2005) *MLwin Version 2.22*. Centre for Multilevel Modeling, University of Bristol.

**Table 3** Modeling results for collision-based data

| Categories                   | Variable                         | MBL fatal |          | MBL minor |        | MOL      |        | MML—fatality versus NI |       | MML—minor versus NI |       |
|------------------------------|----------------------------------|-----------|----------|-----------|--------|----------|--------|------------------------|-------|---------------------|-------|
|                              |                                  | Coeff.    | Sig.     | Coeff.    | Sig.   | Coeff.   | Sig.   | Coeff.                 | Sig.  | Coeff.              | Sig.  |
| Road type                    | Constant                         | -0.856    | 0.021    | 0.020     | 0.917  |          |        | 0.034                  | 0.926 | 0.016               | 0.931 |
|                              | Freeways                         | -0.114    | 0.476    | 0.017     | 0.875  | -0.159   | 0.101  | -0.101                 | 0.523 | 0.027               | 0.789 |
|                              | Multilane kings                  | 0.287     | 0.089    | -0.038    | 0.772  | -0.044   | 0.704  | 0.262                  | 0.123 | -0.032              | 0.791 |
| Other                        | 2 Lane kings                     | 0.000     |          | 0.000     |        | 0.000    |        | 0.000                  |       | 0.000               |       |
|                              | Light/dawn                       | -0.207    | 0.004    |           |        |          |        | -0.181                 | 0.009 |                     |       |
|                              | Dark/dusk                        | 0.000     |          |           |        |          |        | 0.000                  |       |                     |       |
|                              | Weekdays                         | -0.197    | 0.009    | -0.097    | 0.018  | -0.137   | 0.000  | -0.247                 | 0.001 | -0.096              | 0.014 |
| Road related                 | Weekends                         | 0.000     |          | 0.000     |        | 0.000    |        | 0.000                  |       | 0.000               |       |
|                              | Road alignment—straight on hill  |           |          | 0.186     | 0.004  | 0.180    | 0.003  |                        |       | 0.172               | 0.004 |
|                              | Road alignment—curve on level    |           |          | -0.004    | 0.955  | 0.014    | 0.832  |                        |       | -0.008              | 0.902 |
|                              | Road alignment—curve on hill     |           |          | -0.021    | 0.795  | -0.038   | 0.617  |                        |       | -0.014              | 0.852 |
|                              | Road alignment—straight on level |           |          | 0.000     |        | 0.000    |        |                        |       | 0.000               |       |
|                              | Speed limit                      | 0.013     | 0.001    | 0.008     | 0.000  | 0.010    | 0.000  | 0.014                  | 0.000 | 0.008               | 0.000 |
| Weather                      | Number of lanes                  | -0.103    | 0.000    | -0.083    | 0.000  | -0.087   | 0.000  | -0.127                 | 0.000 | -0.082              | 0.000 |
|                              | RSI                              | 0.317     | 0.018    |           |        |          |        | 0.309                  | 0.018 |                     |       |
|                              | Weather—freezing rain, snow      |           |          |           |        | -0.085   | 0.048  |                        |       |                     |       |
| Traffic                      | Weather—other                    |           |          |           |        | 0.000    |        |                        |       |                     |       |
|                              | Wind speed (km/h)                |           |          | -0.005    | 0.012  | -0.005   | 0.012  |                        |       | -0.005              | 0.012 |
|                              | ln(Traffic)                      | -0.320    | 0.000    | -0.056    | 0.005  | -0.135   | 0.000  | -0.339                 | 0.000 | -0.054              | 0.004 |
|                              | $\mu_1$                          |           |          |           |        | -0.881   | 0.000  |                        |       |                     |       |
|                              | $\mu_2$                          |           |          |           |        | 1.742    | 0.000  |                        |       |                     |       |
|                              | -2 × log likelihood (null)       | 508.443   |          | 18,437.2  |        | 19,301.4 |        | 19,301.4               |       |                     |       |
| -2 × log likelihood (full)   | -3,340.050                       |           | 17,892.1 |           | 15,804 |          | 14,944 |                        |       |                     |       |
| Collision level observations | 13,775                           |           | 12,880   |           | 13,775 |          | 13,775 |                        |       |                     |       |

*MBL* = multilevel binary logit models  
*MOL* = multilevel ordered logit models  
*MML* = multilevel multinomial logit models

**Table 4** Modeling results for vehicle-based data

| Categories                  | Variable                                  | MBL fatal |       | MBL minor |       | MOL      |       | MML—fatality versus NI |       | MML—minor versus NI |       |
|-----------------------------|---|-----------|-------|-----------|-------|----------|-------|------------------------|-------|---------------------|-------|
|                             |   | Coeff.    | Sig.  | Coeff.    | Sig.  | Coeff.   | Sig.  | Coeff.                 | Sig.  | Coeff.              | Sig.  |
| Road type                   | Constant                                  | -0.488    | 0.260 | 0.397     | 0.042 |          |       | 0.487                  | 0.248 | 0.397               | 0.037 |
|                             | Freeways                                  | 0.029     | 0.868 | 0.165     | 0.102 | -0.026   | 0.780 | 0.111                  | 0.519 | 0.179               | 0.062 |
|                             | Multilane kings                           | 0.297     | 0.112 | 0.030     | 0.804 | -0.011   | 0.920 | 0.318                  | 0.087 | 0.037               | 0.746 |
|                             | 2 Lane kings                              | 0.000     |       | 0.000     |       | 0.000    |       | 0.000                  |       | 0.000               |       |
| Other                       | Light/dawn                                | -0.175    | 0.021 |           |       |          |       | -0.184                 | 0.012 |                     |       |
|                             | Dark/dusk                                 | 0.000     |       |           |       |          |       | 0.000                  |       |                     |       |
|                             | Weekdays                                  | -0.159    | 0.047 |           |       |          |       | -0.163                 | 0.034 |                     |       |
|                             | Weekends                                  | 0.000     |       |           |       |          |       | 0.000                  |       |                     |       |
| Road related                | Collision location—intersections          | -0.148    | 0.294 |           |       | 0.134    | 0.031 | -0.227                 | 0.100 |                     |       |
|                             | Collision location—bridges/underpasses    | 0.665     | 0.021 |           |       | -0.240   | 0.124 | 0.707                  | 0.009 |                     |       |
|                             | Collision location—segment                | 0.000     |       |           |       | 0.000    |       | 0.000                  |       |                     |       |
|                             | Road alignment—straight on hill           |           |       | 0.202     | 0.000 | -0.193   | 0.001 |                        |       | 0.198               | 0.000 |
| Driver                      | Road alignment—curve on level             |           |       | 0.146     | 0.032 | -0.159   | 0.013 |                        |       | 0.137               | 0.032 |
|                             | Road alignment—curve on hill              |           |       | 0.125     | 0.105 | -0.098   | 0.185 |                        |       | 0.129               | 0.077 |
|                             | Road alignment—straight on level          |           |       | 0.000     |       | 0.000    |       |                        |       | 0.000               |       |
|                             | Speed limit                               | 0.014     | 0.000 | 0.008     | 0.000 | -0.008   | 0.000 | 0.015                  | 0.000 | 0.008               | 0.000 |
| Vehicle                     | Number of lanes                           | -0.109    | 0.000 | -0.080    | 0.000 | 0.085    | 0.000 | -0.132                 | 0.000 | -0.080              | 0.000 |
|                             | RSI                                       |           |       | -0.322    | 0.000 | 0.266    | 0.000 |                        |       | -0.312              | 0.000 |
|                             | Driver age (years)                        | 0.007     | 0.020 | 0.002     | 0.046 | -0.003   | 0.003 | 0.008                  | 0.008 | 0.003               | 0.003 |
|                             | Driver—male                               | 0.139     | 0.078 | -0.241    | 0.000 | 0.175    | 0.000 |                        |       | -0.242              | 0.000 |
| Weather                     | Driver—female                             | 0.000     |       | 0.000     |       | 0.000    |       |                        |       | 0.000               |       |
|                             | Driver condition—normal                   | -0.605    | 0.000 | -0.170    | 0.000 | 0.285    | 0.000 | -0.711                 | 0.000 | -0.177              | 0.000 |
|                             | Driver condition—other (drinking etc.)    | 0.000     |       | 0.000     |       | 0.000    |       | 0.000                  |       | 0.000               |       |
|                             | Vehicle age (years)                       |           |       | 0.013     | 0.001 | -0.012   | 0.000 |                        |       | 0.013               | 0.000 |
| Traffic                     | Vehicle type—vans                         | 0.192     | 0.021 | 0.116     | 0.003 | -0.129   | 0.000 | 0.217                  | 0.007 | 0.112               | 0.003 |
|                             | Vehicle type—large trucks etc             | 0.042     | 0.771 | -0.612    | 0.000 | 0.495    | 0.000 | -0.089                 | 0.519 | -0.595              | 0.000 |
|                             | Vehicle type—car/station wagon            | 0.000     |       | 0.000     |       | 0.000    |       | 0.000                  |       | 0.000               |       |
|                             | Wind speed (km/h)                         |           |       | -0.006    | 0.003 | 0.006    | 0.003 |                        |       | -0.006              | 0.003 |
| In(Traffic)                 | ln(Traffic)                               | -0.384    | 0.000 | -0.139    | 0.000 | 0.201    | 0.000 | -0.447                 | 0.000 | -0.141              | 0.000 |
|                             | $\mu_1$                                   |           |       |           |       | -1.195   | 0.000 |                        |       |                     |       |
|                             | $\mu_2$                                   |           |       | 1.230     | 0.000 |          |       |                        |       |                     |       |
|                             | $-2 \times \log \text{likelihood (null)}$ | -3,834.8  |       | 24,302.0  |       | 20,777.2 |       | 20,777.2               |       |                     |       |
| Variance at collision level | $-2 \times \log \text{likelihood (full)}$ | -11,816.4 |       | 22,859.8  |       | 13,911.2 |       | 11,398.3               |       |                     |       |
|                             | Variance at collision level               | 2.352     | 0.000 | 0.199     | 0.000 | 2.198    | 0     | 2.248                  | 0.000 | 0.000               | 0.000 |



Table 4 continued

| Categories | Variable                     | MBL fatal |      | MBL minor |      | MOL      |      | MML—fatality versus NI |      | MML—minor versus NI |      |
|------------|------------------------------|-----------|------|-----------|------|----------|------|------------------------|------|---------------------|------|
|            |                              | Coeff.    | Sig. | Coeff.    | Sig. | Coeff.   | Sig. | Coeff.                 | Sig. | Coeff.              | Sig. |
|            | Variance at vehicle level    | 3.29      | 0    | 3.29      | 0    | 3.29     | 0    | 3.29                   |      | 0                   |      |
|            | Collision level observations | 13,775.0  |      | 13,060.0  |      | 13,775.0 |      | 13,775                 |      |                     |      |
|            | Vehicle level observations   | 19,635.0  |      | 18,603.0  |      | 19,635.0 |      | 19,635                 |      |                     |      |

evaluating the effect of individual factors, their elasticities are calculated and given in Table 6. For a continuous variable  $X_{ki}$ , elasticity for a particular collision severity outcome “ $i$ ” is computed as

$$E_{X_{ki}}^{P(i)} = [1 - P(i)]\beta_{ki}X_{ki}, \tag{7}$$

where  $P(i)$  is the probability of collision severity outcome “ $i$ ,” and  $\beta_{ki}$  is the coefficient associated with variable  $X_{ki}$ . For categorical variables elasticity is calculated as  $E = (\exp \beta - 1)/\exp \beta$  [3, 7, 8]. Table 7 gives values predicted from the models and the observed severity ratios.

### 4.1 Comparison of quality of fitting

As explained in the previous section, likelihood estimates from MLWin are approximate and the usual goodness of fit criterion such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) could not be applied [54]. AIC [55], defined as  $-2LL + 2p$ , is a test statistics used to identify the best fit model from a set of models. The term  $LL$  is the log likelihood of a fitted model and  $p$  the number of parameters, which is included to penalize models with higher number of parameters. A model with smaller AIC value represents a better overall fit. Similarly, Bayesian Information Criterion (BIC) [56], defined as  $-2LL + p\ln(n)$ , which is another test statistics and a variation of AIC, is used to identify the best fit model from a set of models. The term “ $n$ ” represents the number of observations used to calibrate the model. A model with smaller BIC value represents a better overall fit. Alternatively, results from the models were compared to the actual observations and it was found MML models have a better prediction performance compared to MOL models except for collision-based fatalities where MOL has a slightly better prediction. Similarly, MML models have better prediction results compared to MBL models for occupant- and vehicle-based data. For collision-based data, MBL results are slightly better for NI + minimal injury and minor injury collisions, whereas for fatality collisions, MML results are closer to the observed severity ratios. Based on the discussion in this section, MML is found to perform better as a whole than MBL and MOL.

### 4.2 Effects of data aggregation and correlation

If the collision data are used at a disaggregated level of analysis such as occupant based or vehicle based, then efforts should be made to account for the correlation that exists between occupants in a vehicle or vehicles in a collision such as shown by the variance terms in Tables 4 and 5. Occupant-based results (Table 5) show that around 79 % of the variation ( $\rho = 0.79$ ) is accounted for at the

Table 5 Modeling results for occupant-based data

| Categories   | Variable                               | MBL fatal |       | MBL minor |       | MOL    |       | MML—fatality versus NI |       | MML—minor versus NI |       |
|--------------|--|-----------|-------|-----------|-------|--------|-------|------------------------|-------|---------------------|-------|
|              |  | Coef.     | Sig.  | Coef.     | Sig.  | Coef.  | Sig.  | Coef.                  | Sig.  | Coef.               | Sig.  |
| Road type    | Constant                               | 0.341     | 0.504 | 0.559     | 0.003 |        |       | 1.349                  | 0.009 | 0.561               | 0.002 |
|              | Freeways                               | 0.054     | 0.791 | 0.105     | 0.259 | -0.012 | 0.892 | 0.048                  | 0.812 | 0.130               | 0.149 |
|              | Multilane kings                        | 0.340     | 0.134 | 0.103     | 0.353 | -0.085 | 0.409 | 0.297                  | 0.187 | 0.094               | 0.380 |
|              | 2 Lane kings                           | 0.000     |       | 0.000     |       | 0.000  |       | 0.000                  |       | 0.000               |       |
| Other        | Light/dawn                             | -0.159    | 0.058 |           |       | 0.057  | 0.075 | -0.166                 | 0.043 |                     |       |
|              | Dark/dusk                              | 0.000     |       | 0.000     |       | 0.000  |       | 0.000                  |       |                     |       |
|              | Collision location—intersections       | -0.151    | 0.336 |           |       | 0.133  | 0.024 | -0.189                 | 0.223 |                     |       |
|              | Collision location—bridges/underpasses | 0.823     | 0.015 |           |       | -0.247 | 0.100 | 0.873                  | 0.007 |                     |       |
| Road related | Collision location—segment             | 0.000     |       |           |       | 0.000  |       | 0.000                  |       |                     |       |
|              | Road alignment—straight on hill        |           |       | 0.131     | 0.017 | -0.131 | 0.013 |                        |       | 0.133               | 0.012 |
|              | Road alignment—curve on level          |           |       | 0.180     | 0.004 | -0.193 | 0.002 |                        |       | 0.165               | 0.007 |
|              | Road alignment—curve on hill           |           |       | 0.198     | 0.007 | -0.168 | 0.018 |                        |       | 0.190               | 0.007 |
|              | Road alignment—straight on level       |           |       | 0.000     |       | 0.000  |       |                        |       | 0.000               |       |
|              | Speed limit                            | 0.016     | 0.001 | 0.009     | 0.000 | -0.008 | 0.000 | 0.018                  | 0.000 | 0.009               | 0.000 |
|              | Number of lanes                        | -0.122    | 0.000 | -0.079    | 0.000 | 0.084  | 0.000 | -0.141                 | 0.000 | -0.078              | 0.000 |
|              | RSI                                    |           |       | -0.295    | 0.000 | 0.248  | 0.000 |                        |       | -0.306              | 0.000 |
|              | Driver age (years)                     | 0.007     | 0.000 | 0.004     | 0.000 | -0.004 | 0.000 | 0.008                  | 0.000 | 0.004               | 0.000 |
|              | Driver—male                            |           |       | -0.377    | 0.000 | 0.329  | 0.000 |                        |       | -0.381              | 0.000 |
| Vehicle      | Driver—female                          |           |       | 0.000     |       | 0.000  |       |                        |       | 0.000               |       |
|              | Driver condition—normal                | -0.561    | 0.000 |           |       | 0.147  | 0.000 | -0.590                 | 0.000 |                     |       |
|              | Driver condition—other (drinking etc.) | 0.000     |       |           |       | 0.000  |       | 0.000                  |       | 0.000               |       |
|              | Vehicle age (years)                    |           |       | 0.018     | 0.000 | -0.016 | 0.000 |                        |       | 0.017               | 0.000 |
|              | Vehicle type—vans                      |           |       | -0.287    | 0.000 | 0.263  | 0.000 | -0.190                 | 0.037 | -0.288              | 0.000 |
|              | Vehicle type—large trucks etc          |           |       | -0.910    | 0.000 | 0.790  | 0.000 | -0.442                 | 0.001 | -0.886              | 0.000 |
|              | Vehicle type—car/station wagon         |           |       | 0.000     |       | 0.000  |       | 0.000                  |       | 0.000               |       |
|              | Vehicle condition—non-defective        | -0.411    | 0.031 |           |       | 0.156  | 0.063 | -0.419                 | 0.024 |                     |       |
|              | Vehicle condition—defective            | 0.000     |       |           |       | 0.000  |       | 0.000                  |       |                     |       |
|              | Position in vehicle—front              |           |       | 0.241     | 0.000 | -0.215 | 0.000 | 0.166                  | 0.062 | 0.244               | 0.000 |
| Person       | Position in vehicle—rear               |           |       | 0.000     |       | 0.000  |       | 0.000                  |       | 0.000               |       |
|              | Safety equipment—used                  | -0.826    | 0.000 | -0.733    | 0.000 | 0.836  | 0.000 | -1.070                 | 0.000 | -0.660              | 0.000 |
|              | Safety equipment—not or bad used       | 0.000     |       | 0.000     |       | 0.000  |       | 0.000                  |       | 0.000               |       |

Table 5 continued

| Categories                  | Variable                           | MBL fatal |        | MBL minor |        | MOL       |        | MML—fatality versus NI |       | MML—minor versus NI |       |
|-----------------------------|------------------------------------|-----------|--------|-----------|--------|-----------|--------|------------------------|-------|---------------------|-------|
|                             |                                    | Coef.     | Sig.   | Coef.     | Sig.   | Coef.     | Sig.   | Coef.                  | Sig.  | Coef.               | Sig.  |
| Weather                     | Wind speed (km/h)                  |           |        | -0.006    | 0.003  | 0.005     | 0.012  |                        |       | -0.006              | 0.003 |
|                             | Visibility (km)                    |           |        | -0.004    | 0.046  | 0.006     | 0.003  |                        |       | -0.004              | 0.046 |
| Traffic                     | Hourly precipitation (cm/h)        |           |        |           |        | 0.169     | 0.039  |                        |       |                     |       |
|                             | ln(Traffic)                        | -0.418    | 0.000  | -0.166    | 0.000  | 0.208     | 0.000  | -0.474                 | 0.000 | -0.169              | 0.000 |
|                             | $\mu_1$                            |           |        |           |        | -1.566    | 0.000  |                        |       |                     |       |
|                             | $\mu_2$                            |           |        |           |        | 0.702     | 0.001  |                        |       |                     |       |
|                             | $-2 \times \log$ likelihood (null) | -24,094.8 |        | 37,806.7  |        | 13,896.6  |        | 13,896.6               |       |                     |       |
|                             | $-2 \times \log$ likelihood (full) | -44,556.4 |        | 33,338.5  |        | -4,166.57 |        | -11,450.9              |       |                     |       |
|                             | Variance at collision level        | 8.179     | 0.000  | 0.782     | 0.000  | 7.561     | 0      | 7.832                  | 0     | 0                   | 0     |
|                             | Variance at vehicle level          | 0         | 0      | 0.1       | 0      | 0         | 0      | 0                      | 0     | 0                   | 0     |
|                             | Variance at occupant level         | 3.29      | 0      | 3.29      | 0      | 3.29      | 0      | 3.29                   | 0     | 0                   | 0     |
|                             | Collision level observations       | 13,775    |        | 13,375    |        | 13,775    |        | 13,775                 |       |                     |       |
| Vehicle level observations  | 19,635                             |           | 19,046 |           | 19,635 |           | 19,635 |                        |       |                     |       |
| Occupant level observations | 39,564                             |           | 37,979 |           | 39,564 |           | 39,564 |                        |       |                     |       |

occupant level, whereas the collision level accounts for 19 % of the variation ( $\rho = 0.19$ ) and vehicle level for 2 % ( $\rho = 0.02$ ). Similarly, vehicle-based results (Table 4) show that around 94 % of the variation ( $\rho = 0.94$ ) is accounted for at the vehicle level, whereas the collision level accounts for 0.06 % of the variation ( $\rho = 0.06$ ). This flexibility offered by multilevel modeling improves the reliability of the modeling results obtained with such models as compared to single-level models [57–59].

Data used in a collision level severity analysis are, however, aggregated to the level of a collision. This takes care of the correlation within the data but can result in some immediate problems: (i) loss of information by reducing the number of observations, (ii) miss-specification of collision attributes resulting in erroneous share of high severity levels (Table 2), and (iii) the incapability to analyze different variables related to individual persons or vehicles at aggregate level such as seat belt use, position in the vehicle, vehicle age and type, etc. These could result in biased parameter estimates (see e.g., Mensah and Hauer [60] for some of these issues in collision frequency modeling). In this research, we utilized the multilevel framework to account for the correlation between occupants in the same vehicle and vehicles in the same collision. Treating occupant-based data results as the base case we compare modeling results from MML models for the three datasets.

The percent change in parameter estimates for fatality and major injury collisions show a difference ranging from -131 % to 214 % (average = 13 %) between occupant-based (as the base case) and vehicle-based and -9 % to 310 % (average = 62 %) between occupant-based and collision-based data. The difference between vehicle-based data (as the base case) and collision-based data is -52 % to 191 % (average reduction in size of the parameter estimate = 28 %). For minor injuries the difference is from -49 % to 139 % (average = 20 %) between occupant-based and vehicle-based data and from -29 % to 134 % (average = 54 %) between occupant-based and collision-based data, whereas for vehicle-based data (as the base case) and collision-based data this difference is from -3 % to 186 % (average = 64 %). This shows that aggregating the data results in underestimation of the parameters estimates. This could be of grave consequences if the purpose of the analysis is to evaluate the effects of some policies through some variables, in which case precise estimation of the magnitude of the parameter for the variable is of great importance. Besides data aggregation, another reason for this is the model setting (Table 1) where it can be seen that not all the variables used in the occupant-based data model are used for the other two level of aggregation. This will also result in parameter estimates for the remaining variables to be rescaled. This is evident from the results as well

**Table 6** Elasticities for the three datasets

| Variable                               | MBL fatal | MBL minor | MOL    | MML—fatality versus NI | MML—minor versus NI |
|--|-----------|-----------|--------|------------------------|---------------------|
| <b>Occupant-based model</b>            |           |           |        |                        |                     |
| Freeways                               | 0.053     | 0.1       | 0.012  | 0.047                  | 0.122               |
| Multilane kings                        | 0.288     | 0.098     | 0.089  | 0.257                  | 0.09                |
| Light/dawn                             | -0.172    |           | -0.055 | -0.181                 |                     |
| Accident location—intersections        | -0.163    |           | -0.125 | -0.208                 |                     |
| Accident location—bridges/underpasses  | 0.561     |           | 0.28   | 0.582                  |                     |
| Road alignment—straight on hill        |           | 0.123     | 0.14   |                        | 0.125               |
| Road alignment—curve on level          |           | 0.165     | 0.213  |                        | 0.152               |
| Road alignment—curve on hill           |           | 0.18      | 0.183  |                        | 0.173               |
| Driver age (years)                     | 0.261     | 0.122     | 0.122  | 0.297                  | 0.121               |
| Driver—male                            |           | -0.458    | -0.28  |                        | -0.464              |
| Driver condition—normal                | -0.752    |           | -0.137 | -0.804                 |                     |
| Vehicle age (years)                    |           | 0.072     | 0.064  |                        | 0.068               |
| Vehicle type—vans                      |           | -0.332    | -0.231 | -0.209                 | -0.334              |
| Vehicle type—large trucks etc          |           | -1.484    | -0.546 | -0.556                 | -1.425              |
| Vehicle condition—non-defective        | -0.508    |           | -0.144 | -0.52                  |                     |
| Position in vehicle—front              |           | 0.214     | 0.24   | 0.153                  | 0.217               |
| Safety equipment—used                  | -1.284    | -1.081    | -0.567 | -1.915                 | -0.935              |
| Speed limit                            | 1.492     | 0.685     | 0.608  | 1.668                  | 0.68                |
| Number of lanes                        | -0.838    | -0.443    | -0.47  | -0.963                 | -0.434              |
| RSI                                    |           | -0.193    | -0.162 |                        | -0.198              |
| Wind speed (km/h)                      |           | -0.079    | -0.066 |                        | -0.079              |
| Visibility (km)                        |           | -0.052    | -0.078 |                        | -0.052              |
| Hourly precipitation (cm/h)            |           |           | -0.01  |                        |                     |
| ln(traffic)                            | -3.282    | -1.064    | -1.331 | -3.699                 | -1.075              |
| <b>Vehicle-based model</b>             |           |           |        |                        |                     |
| Freeways                               | 0.029     | 0.152     | 0.026  | 0.105                  | 0.164               |
| Multilane kings                        | 0.257     | 0.03      | 0.011  | 0.272                  | 0.036               |
| Light/dawn                             | -0.191    |           |        | -0.202                 |                     |
| Weekdays                               | -0.172    |           |        | -0.177                 |                     |
| Collision location—intersections       | -0.16     |           | -0.125 | -0.255                 |                     |
| Collision location—bridges/underpasses | 0.486     |           | 0.271  | 0.507                  |                     |
| Road alignment—straight on hill        |           | 0.183     | 0.213  |                        | 0.18                |
| Road alignment—curve on level          |           | 0.136     | 0.172  |                        | 0.128               |
| Road alignment—curve on hill           |           | 0.118     | 0.103  |                        | 0.121               |
| Driver age (years)                     | 0.253     | 0.053     | 0.079  | 0.29                   | 0.08                |
| Driver—male                            | 0.13      | -0.273    | -0.161 |                        | -0.274              |
| Driver condition—normal                | -0.831    | -0.185    | -0.248 | -1.036                 | -0.194              |
| Vehicle age (years)                    |           | 0.054     | 0.05   |                        | 0.054               |
| Vehicle type—vans                      | 0.175     | 0.11      | 0.138  | 0.195                  | 0.106               |
| Vehicle type—large trucks etc          | 0.041     | -0.844    | -0.39  | -0.093                 | -0.813              |
| Speed limit                            | 1.27      | 0.527     | 0.529  | 1.362                  | 0.533               |
| Number of lanes                        | -0.722    | -0.385    | -0.411 | -0.876                 | -0.389              |
| RSI                                    |           | -0.18     | -0.149 |                        | -0.176              |
| Wind speed (km/h)                      |           | -0.068    | -0.068 |                        | -0.069              |
| ln(Traffic)                            | -2.909    | -0.765    | -1.111 | -3.388                 | -0.784              |
| <b>Collision-based model</b>           |           |           |        |                        |                     |
| Freeways                               | -0.121    | 0.017     | -0.147 | -0.106                 | 0.027               |

**Table 6** continued

| Variable                        | MBL fatal | MBL minor | MOL    | MML—fatality versus NI | MML—minor versus NI |
|---------------------------------|-----------|-----------|--------|------------------------|---------------------|
| Multilane kings                 | 0.249     | -0.039    | -0.043 | 0.23                   | -0.033              |
| Light/dawn                      | -0.23     |           |        | -0.198                 |                     |
| Weekdays                        | -0.218    | -0.102    | -0.128 | -0.28                  | -0.101              |
| Road alignment—straight on hill |           | 0.17      | 0.197  |                        | 0.158               |
| Road alignment—curve on level   |           | -0.004    | 0.014  |                        | -0.008              |
| Road alignment—curve on hill    |           | -0.021    | -0.037 |                        | -0.014              |
| Weather—freezing rain, snow     |           |           | -0.081 |                        |                     |
| Speed limit                     | 1.19      | 0.454     | 0.544  | 1.275                  | 0.463               |
| Number of lanes                 | -0.69     | -0.345    | -0.346 | -0.847                 | -0.348              |
| RSI                             | 0.242     |           |        | 0.235                  |                     |
| Wind speed (km/h)               |           | -0.036    | -0.035 |                        | -0.037              |
| ln(Traffic)                     | -2.412    | -0.262    | -0.604 | -2.542                 | -0.257              |

**Table 7** Prediction results from models versus observed results

| Severity type                | MBL (%) | MML (%) | MOL (%) | Observed (%) |
|------------------------------|---------|---------|---------|--------------|
| <b>Occupant-based model</b>  |         |         |         |              |
| NI + minimal injury          | 78.3    | 77      | 80.0    | 76.4         |
| Minor injury                 | 19.9    | 20.5    | 17.5    | 19.5         |
| Fatal + major injury         | 1.8     | 2.5     | 2.5     | 4.0          |
| <b>Vehicle-based model</b>   |         |         |         |              |
| NI + minimal injury          | 63.5    | 64.3    | 69.1    | 64.7         |
| Minor injury                 | 31.2    | 30.4    | 27.1    | 30.1         |
| Fatal + major injury         | 5.3     | 5.3     | 3.8     | 5.3          |
| <b>Collision-based model</b> |         |         |         |              |
| NI + minimal injury          | 55.1    | 55.9    | 56.9    | 53.4         |
| Minor injury                 | 40.7    | 39.4    | 37.9    | 40.2         |
| Fatal + major injury         | 4.2     | 4.7     | 5.2     | 6.5          |

where the range is wider for the difference between occupant- and collision-based data than those from occupant- and vehicle-based data.

### 4.3 Comparison of significant factors

Despite different in quality of fitting and effect sizes of various safety factors from different models and data aggregation methods, there were consistent results in terms of the factors that were found to have statistically significant effect on collision severity. This section discusses the main findings on the contributing factors and the magnitude of their effects (Tables 3 through 6).

#### 4.3.1 Driver characteristics and accident impact type

One percent change in driver age will cause an average increase of 0.297 in the probability of suffering a fatal/major injury and 0.121 increases in the probability of having minor injuries. For male drivers, the probability of suffering minor injuries are 0.46 less compared to female drivers. Alcohol can increase the probability of fatality/major injuries by 0.80. Collisions on bridges increase the probability of fatality/major injuries by 0.58, whereas those occurring at intersections reduce it by 0.21.

#### 4.3.2 Road characteristics

Multilane-divided highways increase the probabilities of fatality/major injuries by 0.26 and minor injuries by 0.09, whereas for freeways these figures are 0.05 and 0.12 compared to undivided two-lane highways. Improvement in road surface condition causes the probability of minor injuries to reduce by 0.20. The presence of curves or hilly terrain increases the probability of minor injuries from 0.12 to 0.17. Increase in number of lanes decreases the probability of fatal/major injuries by 0.96 and minor injuries by 0.43. Increase in speed limit increases the probability of fatality/major injuries by 1.67 and minor injuries by 0.68.

#### 4.3.3 Vehicle and individual

Heavy weight and non-defective vehicles decrease the probability of fatal/major injuries from 0.21 to 0.56 and minor injuries by 0.33–1.43. Increase in the age of a vehicle increases the chances of minor injuries by 0.07. Front position increases the chances of fatal/major injuries by 0.15 and minor injuries by 0.22, whereas the use of safety devices decreases the chances of fatal/major injuries by 1.92 and minor injuries by 0.94.

#### 4.3.4 Weather and environment

Increase in wind speed and visibility decreases the probability of minor injuries by 0.08 and 0.05. The presence of lighting conditions reduces the chances of fatality/major injuries by 0.18.

#### 4.3.5 Traffic volume

Traffic volume is the most influential factor of all and an increase in traffic volume decreases the probability of fatal/major injuries by 3.70 and minor injuries by 1.08. Intuitively, a higher traffic volume will lead to more congestion resulting in lower speeds.

### 5 Conclusions and future research

Three alternative logistic regression models, namely multinomial logit model, sequential binary logit model, and ordered logit model applied in a multilevel framework, were compared and evaluated for their performance for predicting the conditional probabilities of different severity levels of a given collision. These models were applied to collision data aggregated at three levels—occupant level, vehicle level, and collision level. These three levels were used to evaluate the effects of data aggregation and correlation on collision severity analysis. Collision data from six winter seasons (2,000–2,006) and 31 sites containing 13,775 collisions, involving 39,564 individuals and 19,635 vehicles was used for this analysis. Based on the modeling results, it was found that multilevel multinomial logit (MML) has the best overall fit to the data, and occupant-based data results are more reliable than vehicle- and collision-based data.

Moreover, it was found that data aggregation affects the parameter estimates, on the average, by as much as 13 % for vehicle-based aggregated data and 62 % collision-based aggregated data compared to occupant-based data. Similarly, from correlation perspective, around 79 % of the variation is accounted for when using occupant-based data compared to the 19 % variation accounted for by collision-based data. This could have significant implications for evaluating the effects of different safety-related policies and countermeasures when using, showing the importance of data analysis at a disaggregate level.

Our future efforts will be directed toward the comparison of data compiled from winter seasons and snow storm events using the results from this research. Moreover, other modeling types such as latent class models will also be evaluated and compared to the modeling results from this analysis.

**Acknowledgments** This research was supported by MTO in part through the Highway Infrastructure and Innovations Funding Program

(HIIFP). The authors wish to acknowledge in particular the assistance of Max Perchanok from MTO.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### References

- Shankar V, Mannering F (1996) An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *J Saf Res* 27(3):183–194
- Lee, J. and F. Mannering (1999). Analysis of roadside accident frequency and severity and roadside safety management. Final Research Report. Research Project T9903, Task 97. Report Number WA-RD 475.1
- Lee J, Mannering F (2002) Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accid Anal Prev* 34:149–161
- Ulfarsson GF, Mannering FL (2004) Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. *Accid Anal Prev* 36:135–147
- Holdridge MJ, Shankar VN, Ulfarsson GF (2005) The crash severity impacts of fixed roadside objects. *J Saf Res* 36:139–147
- Khorashadi A, Niemeier D, Shankar V, Mannering F (2005) Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. *Accid Anal Prev* 37:910–921
- Ulfarsson GF, Kim S, Lentz ET (2006) Factors affecting common vehicle-to-vehicle collision types road safety priorities in an aging society. *Transp Res Rec* 1980:70–78
- Malyshkina NV, Mannering F (2008) Effect of increases in speed limits on severities of injuries in accidents. *Transp Res Rec* 2083:122–127
- Miranda-Moreno, Luis F., Liping Fu, Satish Ukkusuri, and Dominique Lord (2009). How to incorporate accident severity and vehicle occupancy into the hotspot identification process? 88th Annual Meeting of the Transportation Research Board. Paper No. 09 -2824
- Mergia WY (2010) Exploring factors contributing to injury severity at freeway merging and diverging areas. MSc Thesis, University of Dayton
- Lee C, Abdel-Aty M (2008) Two-level nested logit model to identify traffic flow parameters affecting crash occurrence on freeway ramps. *Transp Res Rec* 2083:145–152
- Savolainen PT, Mannering FL, Lord D, Quddus MA (2011) The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accid Anal Prev* 43(5):1666–1676
- Nassar SA, Saccomanno FF, Shortreed JH (1994) Disaggregate analysis of road accident severities. *Int J Impact Eng* 15(6):815–826
- Saccomanno FF, Nassar SA, Shortreed JH (1996) Reliability of statistical road accident injury severity models. *Transp Res Rec* 1542:14–23
- Shankar V, Mannering F, Barfield W (1996) Statistical analysis of accident severity on rural freeways. *Accid Anal Prev* 28(3):391–401
- Carson J, Mannering F (2001) The effect of ice warning signs on ice-accident frequencies and Severities. *Accid Anal Prev* 33:99–109

17. Dissanayake S, Lu J (2002). Analysis of severity of young driver crashes sequential binary logistic regression modeling. *Transportation Research Record* 1784, 108–114. Paper No. 02-2302
18. Jones AP, Jørgensen SH (2003) The use of multilevel models for the prediction of road accident outcomes. *Accid Anal Prev* 35:59–69
19. Donnell ET, Mason JM Jr (2004) Predicting the severity of median-related crashes in pennsylvania by using logistic regression. *Transp Res Rec* 1897:55–63
20. Lenguerrand E, Martin JL, Laumon B (2006) Modelling the hierarchical structure of road crash data—Application to severity analysis. *Accid Anal Prev* 38:43–53
21. Milton JC, Shankar VN, Mannering FL (2008) Highway accident severities and the mixed Logit model: an exploratory empirical analysis. *Accid Anal Prev* 40:260–266
22. Gibbons RD, Hedeker D (1997) Random effects probit and logistic regression models for three-level data. *Biometrics* 53:1527–1537
23. Usman T, Fu L, Miranda-Moreno L (2011) Accident prediction models for winter road safety: does temporal aggregation of data matters? *Transp Res Rec* 2237:144–151
24. Ronald H. H., Thomas and S. Loring (2000). “An Introduction to multilevel modeling techniques quantitative methodology series”
25. Newsom, J.T., and Nishishiba, M., (2002). Hierarchical linear modeling of dyadic data. nonconvergence and sample bias in hierarchical linear modeling of dyadic data, 2004, <http://www.upa.pdx.edu/IOA/newsom/mlrlyad4.doc> Accessed 29 Mar 2010
26. Schreiber JB, Griffin BW (2004) Review of multilevel modeling and multilevel studies in *The Journal of Educational Research* (1992–2002). *J Educ Res* 98:24–33
27. Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York
28. Usman T, Fu L, Miranda-Moreno Luis F (2010) Quantifying safety benefit of winter road maintenance: accident frequency modeling. *Accid Anal Prev* 42(6):1878–1887
29. Usman T, Fu L, Miranda-Moreno LF (2012) A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level. *J Accid Anal Prev* 48:368–378
30. Hutchings C, Knight S, Reading JC (2003) The use of generalized estimating equations in the analysis of motor vehicle crash data. *Accident Anal. Prev.* 35(1):3–8
31. Bijleveld FD (2005) The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. *Accid Anal Prev* 37(4):591–600
32. Ma J, Kockelman KM (2006) Bayesian multivariate Poisson regression for models of injury count, by severity. *Transp Res Rec* 1950:24–34
33. Ma J, Kockelman KM, Damien P (2008) A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accid Anal Prev* 40(3):964–975
34. Park ES, Lord D (2007) Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transp Res Rec* 2019:1–6
35. Wong J, Chung Y (2008) Comparison of methodology approach to identify causal factors of accident severity. *Transp Res Rec* 2083:190–198
36. Yau KKW (2004) Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accid Anal Prev* 36(2004):333–340
37. Rasbash J, Steele F, Browne WJ, Goldstein H (2009) *A User’s Guide to MLwiN*, version 2.10 Centre for Multilevel Modelling. University of Bristol, London
38. Savolainen P, Mannering F (2007) Probabilistic models of motorcyclists’ injury severities in single- and multi-vehicle crashes. *Accid Anal Prev* 39:955–963
39. O’Donnell CJ, Connor DH (1996) Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice. *Accid Anal Prev* 28(6):739–753
40. Khattak A, Kantor P, Council FM (1998) Role of adverse weather in key crash types on limited access: roadways implications for advanced weather systems. *Transp Res Rec* 1621:10–19
41. Quddus MA, Wang C, Ison SG (2010) Road traffic congestion and crash severity: econometric analysis using ordered response models. *J Transp Eng* 136(5):424–435
42. Srinivasan, K. K. (2002). Injury severity analysis with variable and correlated thresholds ordered mixed logit formulation. *Transportation Research Record* 1784 Paper No. 02-3805
43. Wang X, Kockelman KM (2005) Occupant injury severity using a heteroscedastic ordered logit model: distinguishing the effects of vehicle weight and type. *Transp Res Rec* 1908:195–204
44. Zhang H (2010). Identifying and quantifying factors affecting traffic crash severity in louisiana. Ph.D. dissertation, Louisiana State University
45. Train KE (2009) *Discrete choice methods with simulation*, 2nd edn. Cambridge University Press, New York
46. Kosmelj K and Vadnal K (2003). Comparison of two generalized logistic regression models; a case study. In: 25th international conference on information technology interfaces ITI 2003, June 16-19, 2003, Cavtat, Croatia
47. Kamarudin MNBC, Ahmad I, Zaharim A, Abdullah S, Kamarudin H (2007). A comparison on two generalized logistic regression models: a case study on failure mode for multiple reflow effect on ball grid array (BGA) application. In: regional conference on engineering mathematics, mechanics, manufacturing & architecture
48. Dissanayake S (2004) Comparison of severity affecting factors between young and older drivers involved in single vehicle crashes. *IATSS Res* 28(2):48–54
49. Jung S, Qin X, Noyce DA (2010) Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accid Anal Prev* 42(2010):213–224
50. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1(1):30–46
51. Newsom JT, Nishishiba M (2002). Hierarchical linear modeling of dyadic data. Non-convergence and sample bias in hierarchical linear modeling of dyadic data, <http://www.upa.pdx.edu/IOA/newsom/mlrlyad4.doc> Accessed 29 Mar 2010
52. Miaou Shaw-pin, Song Joon Jin, Mallick Bani K (2003) Roadway traffic crash mapping: a space-time modeling approach. *J Transp Stat* 6(1):33–57
53. Andrew V, Bared J (1998). Accident models for two-lane rural segments and intersections. *TRR* 1635, Paper No. 98-0294
54. Pickery J, Loosveldt G (2002) A multilevel multinomial analysis of interviewer effects on various components of unit non-response. *Qual Quant* 36:427–437
55. Akaike H (1974) A new look at the statistical model of identification. *IEEE Trans Autom Control* 19:716–723
56. Schwarz GE (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
57. Washington SP, Karlaftis MG, Mannering FL (2010) *Statistical and econometric methods for transportation data analysis*, 2nd edn. Chapman Hall/CRC, Boca Raton
58. Aguero-Valverde J, Jovanis PP (2008) Analysis of road crash frequency with spatial models. *Transp Res Rec* 2061:55–63
59. Lord D, Mannering F (2010) the statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transp Res Part A* 44(2010):291–305
60. Mensah A, Hauer E (1998) Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. *Transp Res Rec* 1635:37–43