Accident Analysis and Prevention 42 (2010) 1556–1565

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

# Multilevel data and Bayesian analysis in traffic safety

Helai Huang*, Mohamed Abdel-Aty[1]

*Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816, USA*

ARTICLE INFO

*Article history:*
Received 13 November 2009
Received in revised form 19 March 2010
Accepted 24 March 2010

*Keywords:*
Road safety
Crash prediction models
Bayesian hierarchical models
Multilevel data
Spatiotemporal data

ABSTRACT

*Background:* Traditional crash prediction models, such as generalized linear regression model, are incapable of taking into account multilevel data structure. Therefore they suffer from a common underlying limitation that each observation (e.g. a crash or a vehicle involvement) in the estimation procedure corresponds to an individual situation in which the residuals exhibit independence.
*Problem:* However, this "independence" assumption may often not hold true since multilevel data structures exist extensively because of the traffic data collection and clustering process. Disregarding the possible within-group correlations may lead to production of models with unreliable parameter estimates and statistical inferences.
*Proposed theory:* In this paper, a 5 × ST-level hierarchy is proposed to represent the general framework of multilevel data structures in traffic safety, i.e. [Geographic region level – Traffic site level – Traffic crash level – Driver-vehicle unit level – Occupant level] × Spatiotemporal level. 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. To properly accommodate the potential cross-group heterogeneity and spatiotemporal correlation due to the multilevel data structure, a Bayesian hierarchical approach that explicitly specifies multilevel structure and reliably yields parameter estimates is introduced and recommended.
*Case studies:* Using Bayesian hierarchical models, the results from several case studies are highlighted to show the improvements on model fitting and predictive performance over traditional models by appropriately accounting for the multilevel data structure.


## 1. Introduction

Road safety is a socio-economic concern, resulting in tremendous life and property loss. To improve safety, it is challenging to obtain a comprehensive understanding of traffic system safety because road traffic is such a complicated system, which may be affected by a diversity of risk factors representing environmental, road geometric, traffic, and driver-vehicle characteristics. Moreover, the understanding of traffic system safety may be further obscured since crash occurrences are necessarily discrete, often sporadic and random events. Hence, obtaining unbiased and relatively accurate estimation and prediction of traffic system safety has become a central concern in road safety management.

Crash prediction model (also called safety performance function) is one of the most important techniques in investigating the relationship between crash occurrence and risk factors associated with various traffic entities. These risk factors are assumed to provide information on the behavior of crash occurrence, which is commonly measured by crash frequency with various degrees of crash severity (Hauer, 1992). Appropriate probabilistic forms and statistically significant factors are identified based on the examination of crash occurrence mechanism and model fitting performance to the historical crash data. Hence, the safety variability associated with traffic entities is modeled by risk factors identified and disturbance/error terms used to account for hidden or unobserved safety-related features. Generally, the better the underlying disturbances are accounted for, the better the model performance and consequently the better the safety estimation and prediction. As a result, apart from scrutinizing risk factors, to address model disturbance is a major challenge for safety modelers.

In terms of disturbance source, two sorts of disturbances could be defined, i.e. unstructured and structured. In this sense, model performance is subject to (a) an adequate understanding and accommodation of structured disturbance; and (b) a rational selection of probabilistic distributions to account for unstructured disturbance. Unfortunately in current road safety research, while considerable research has been made on the latter, effort is significantly insufficient towards a better understanding on structured disturbances. These structured disturbances are closely related to specific safety data structure. Among these, the most typical is

* Corresponding author. Tel.: +1 407 823 4902.
*E-mail addresses:* huanghelai@alumni.nus.edu.sg (H. Huang), mabdel@mail.ucf.edu (M. Abdel-Aty).
[1] Tel.: +1 407 823 5657; fax: +1 407 823 3315.

0001-4575/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2010.03.013

hierarchically and spatiotemporally structured safety data, i.e. multilevel data structure.

This paper aims at a comprehensive inspection of typical structured data used in traffic safety evaluation, so as to raise the awareness of safety analysts and stimulate the necessary concern on structured safety data. Specifically, a conceptual model with a $5 \times$ ST-level hierarchy is proposed as an innovative approach to represent the general framework of multilevel data structures in traffic safety. To properly model the potential cross-group and/or spatiotemporal heterogeneities due to the multilevel data structure, Bayesian hierarchical approach that explicitly specifies multilevel structure and reliably yields parameter estimates is introduced and recommended. Finally, several recently accomplished studies by the current authors and their colleagues are summarized as case study to illustrate the proposed methodology with empirical evaluation.

## 2. Review on crash prediction models

The typical structure of crash prediction model could be expressed as a general form as follows:

$$Y|\boldsymbol{\theta} \sim Dist(\boldsymbol{\theta}) \quad \text{with} \quad \boldsymbol{\theta} = f(\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\varepsilon}) \tag{1}$$

where $Y$: dependent variable(s) of interest, e.g. crash frequency or severity; $Dist(\boldsymbol{\theta})$: adapted distribution for $Y|\boldsymbol{\theta}$ and its parameter(s); $\mathbf{X}$: covariates representing various exposure/risk factors to crash occurrence; $\boldsymbol{\beta}$: coefficients, i.e. factor effects of $\mathbf{X}$ on $\boldsymbol{Y}$; $f(\cdot)$: link function relating $\mathbf{X}$ and $Y$; $\boldsymbol{\varepsilon}$: disturbance/error terms in the model.

The dependent variable $Y$ is assumed to follow some distribution with parameter(s) $\boldsymbol{\theta}$, which is further modeled as a link function $f(\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\varepsilon})$. The selection of the distribution depends on the nature of crash features of interest. Particularly, in predicting crash frequency, Poisson distribution is traditionally employed to model the count data (e.g. Jovanis and Chang, 1986; Joshua and Garber, 1990; Jones et al., 1991; Miaou and Lum, 1993). In contrast, when crash severity is concerned, discrete outcome distributions are generally used, such as those in nominal models (e.g. Mannering and Grodsky, 1995; Shankar and Mannering, 1996; Mercier et al., 1997; Simoncic, 2001; Al-Ghamdi, 2002) or ordered discrete models (O'Donnell and Connor, 1996; Quddus et al., 2002; Rifaat and Chin, 2005; Abdel-Aty and Keller, 2005). The distribution parameters ($\boldsymbol{\theta}$) are then related to the risk factors using a link function which, in a conceptual sense, consists of three components:

(i) a suitable transformation function for $\boldsymbol{\theta}$ based on the nature of data type, for example, a logistic function for binary data or exponential function for count data;
(ii) an expression combining $\mathbf{X}$ and $\boldsymbol{\beta}$, typically assuming a linear combination of $\mathbf{X}$ or their transforms, i.e. $\mathbf{X}\boldsymbol{\beta}$, and
(iii) the term $\boldsymbol{\varepsilon}$ to represent various disturbance/error terms assumed in the model.

Considerable efforts have been made to establish the suitability of various crash prediction models for both crash frequency and severity. Traditionally, generalized linear regression models (GLMs), such as Poisson model, Logit or Probit Models are broadly applied to build probabilistic formulations on the relationship of the crash occurrence with a variety of possible covariates. In most of these classical models, the disturbance term $\boldsymbol{\varepsilon}$ is inherently determined by the adapted distribution, resulting in some constraints for the mean and the variance of the model, e.g., 'variance = mean' in Poisson model, or 'variance = mean $\times$ (1 − mean)' in Binomial Logit model. Hence, they may not be adequate to account for some over-dispersed data, which are commonly found in crash data. To overcome the over-dispersion problem in count data, some over-dispersed Poisson models have proved to be useful by relaxing the condition of 'variance = mean' in standard Poisson model. An additional stochastic component $\varepsilon$ is introduced to the link function. By respectively assuming $\exp(\varepsilon)$ a Gamma distribution or a Lognormal distribution, Poisson-Gamma model (also called Negative Binomial model, NB) (e.g. Miaou, 1994; Kulmala, 1995; Shankar et al., 1995; Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000) or Poisson-Lognormal distribution (Lord and Miranda-Moreno, 2008; Huang et al., 2009; Haque et al., 2010) is typically employed. It is worth noting that the source of over-dispersion has not been explicitly distinguished in these models. In other words, the model disturbance is modeled unstructuredly.

Furthermore, these GLMs suffer from a common underlying limitation that each observation (e.g. a crash or a vehicle involvement) in the estimation procedure corresponds to an individual situation. Hence, the residuals from the model exhibit independence. However, this "independence" assumption may often not hold true since multilevel data structures especially spatiotemporal structures exist extensively because of the traffic data collection and clustering process. Disregarding the possible cross-group heterogeneities or spatiotemporal correlations may lead to production of models with unreliable parameter estimates and statistical inferences.

In recent years, several advanced statistical techniques have been explored to more appropriately represent the nature of crash data. For example, a number of studies have employed zero-inflated models to take into account the excess zero observations in crash data (e.g., Miaou, 1994; Lee et al., 2002; Lee and Mannering, 2002; Shankar et al., 2003; Wang et al., 2003; Kumara and Chin, 2003; Qin et al., 2004; Lord et al., 2005, 2007), whereas Lord et al. (2005, 2007) have questioned the validity of the basic zero-state assumption in these models. In this regard, Markov switching models (Malyshkina et al., 2009; Malyshkina and Mannering, 2009, 2010) and finite mixture models (Park and Lord, 2009) have been newly tested. Especially, the Markov switching models allow specific road entities to switch between multiple crash states over time. Furthermore, the use of variable dispersion parameters in negative binomial models have been reported useful to improve the model-fitting (Heydecker and Wu, 2001; Miaou and Lord, 2003; Miranda-Moreno et al., 2005; El-Basyouny and Sayed, 2006; Mitra and Washington, 2007; Lord and Park, 2008). Multivariate count models have also been applied to jointly model crash frequency at different levels of injury severity (Park and Lord, 2007; Ma et al., 2008; Ye et al., 2009; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009). More recently, a more flexible random parameter modeling approach, including random intercept and/or random slope, is emerging in the literature, in which model parameters are allowed to vary from site to site (Li et al., 2008; Anastasopoulos and Mannering, 2009; Huang et al., 2008, 2009; El-Basyouny and Sayed, 2009; Huang and Chin, in press). The surge of aggregate crash prediction models in response to the safety conscious planning has boosted the exploration of spatiotemporal models to account for the unmeasured confounders and spatiotemporal autocorrelations among adjacent geographic units (e.g., counties, TAZs) (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., in press). It is noted that most of these models have been accomplished in the methodological framework of Bayesian hierarchical approach.

Clearly, these recent advancements have significantly improved the analytical capability over traditional crash prediction models. However, till now, there is no systematic and consistent examination found in the literature on the multilevel data characteristics in general traffic safety research. This issue is comprehensively examined in the next section.

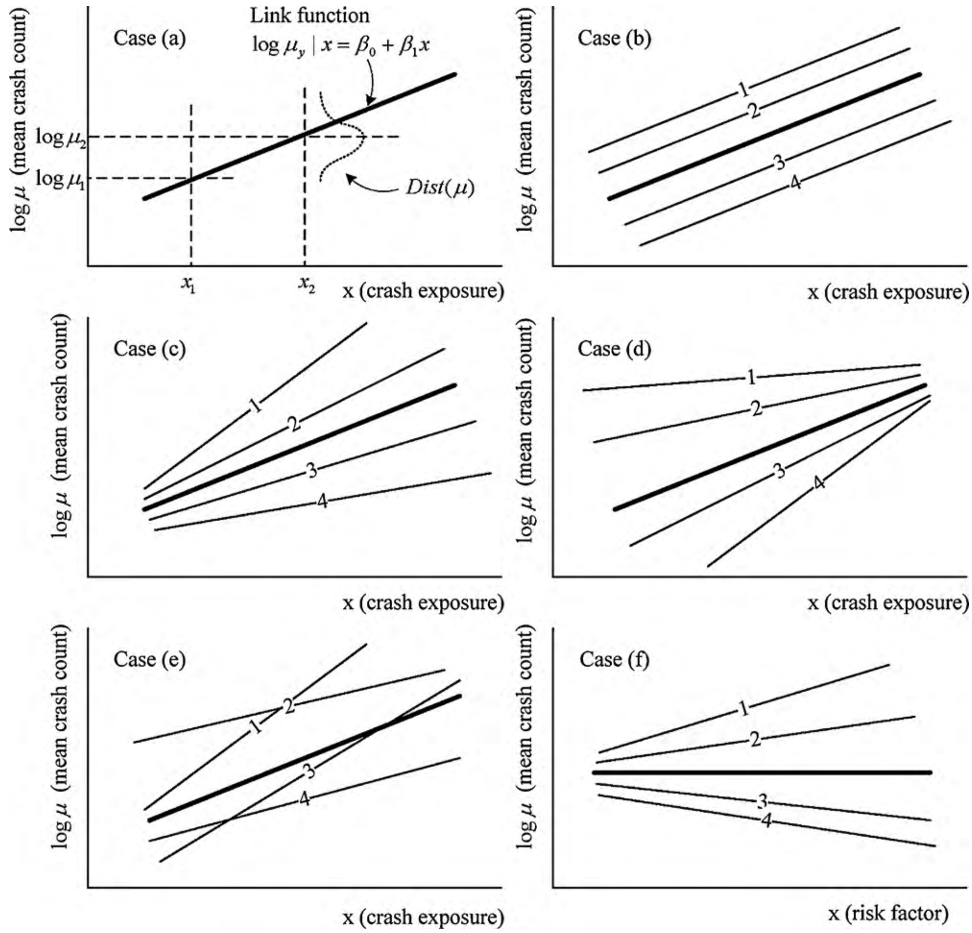


**Fig. 1.** Possible relationship between crash occurrence and risk factors.

## 3. Multilevel data structure in traffic safety

### 3.1. A common neglect in existing models: cross-group heterogeneity

To clearly explain the underlying limitation of 'independence' assumption in GLMs, we take an example of a simple regression relationship between crash frequency and crash exposure. The crash exposure is defined as the amount of opportunities for crashes of a certain type which drivers or the traffic system experience. In this example, the crash exposure is assumed to have a linear relationship with the logarithm of the mean crash count ($\log \mu$): higher exposure is associated with more crashes. A standard GLM may generate the relationship as shown in Case (a) of Fig. 1. Given the crash exposure ($x$), the variation between different observations ($y$) is restricted by distribution adapted ($Dist(\mu)$). Put it in another way, the only stochastic component of variation is introduced by $Dist(\mu)$.

In particular, standard Poisson model assumes a fixed variance for different observations given $\mu$, which exactly equals to $\mu$. Hence the variation of $y$ is only determined through observed heterogeneity, i.e. crash exposure in this example. In over-dispersed Poisson models, by adding an additional disturbance ($\varepsilon$) to relax the constraint of 'variance = mean', the new mean crash count ($\tilde{\mu}$) is subject both to the deterministic variation associated with crash exposure but also to the unobserved heterogeneity introduced by $\varepsilon$. For a given crash exposure ($x$), there is a distribution of $\tilde{\mu}$'s rather than a single value for the mean crash count $\mu$. Nevertheless, over-dispersed Poisson models only take an overall same distribution on the disturbance among individual observations. Hence different observations are still independent with each other. The potential structural improvement of an over-dispersed Poisson model over a standard Poisson model is only the capability to account for unobserved cross-individual heterogeneity in addition to the observed variations.

However, the "independence" assumption may often not hold true since multilevel data structure exists extensively, either intrinsically in traffic data or extrinsically resulting from the manner data are collected or clustered. For example, to study the relationship of crash count and exposure, a number of selected road segments may be nested in several areas of interest (e.g. cities). Moreover, for each selected road segment, there may be several observations from different time periods. In this case, some cross-group heterogeneities, either observed or unobserved, may exist due to spatial and temporal effects of crash data. Indeed, some characteristic variations may necessarily exist between different areas or between road segments.

For instance, suppose the data in the above example are collected from four different areas, in each of which a number of road segments are involved in the study. The Cases (b)–(f) in Fig. 1 illustrate various potential relationships between crash count and exposure. As discussed previously, if a standard GLM model is used on the aggregate dataset (Case (a)), the area-level context in which the road segments belong to is completely ignored: the same single straight line relationship is held to exist everywhere. In effect the model has explained "everything in general and nothing in particular". However, given the different features among the areas such as the socio-economic, demographic and transportation facility characteristics, there may be varying crash count/exposure relationships. Also, traffic enforcement may actually play a role, since different police departments might be reporting crashes differently, e.g., ignoring minor crashes. One possible result shown in the Case (b) of Fig. 1, is the varying-intercept pattern. Here each of four areas (line nos. 1–4) has their own crash count/exposure relation represented by a separate line. The single thicker line represents the general relationship across all four areas. The parallel lines imply that, while the crash count/exposure relation in each area is the same, some areas have uniformly higher crash frequency than others. In Cases (c) and (d), the situations are more complicated as the steepness of the lines varies from area to area. In Case (c), the pattern is such that areas make very little difference for the relatively low exposure roads but there is a high degree of between-area variation in crash count for higher exposure roads. In contrast, Case (d) shows large area-specific differentials exist for the road segment with lower exposure. In Case (e), there is a complex interaction between crash count and exposure. In some areas it is the lower exposure roads which have relatively high crash frequency, whereas in others it is the higher exposure roads. While the final plot, Case (f), is unlikely to occur in terms of the current example, it can be expected for some other risk factors. Across all the areas there is no relationship between the crash count and the risk factor (the single thicker line is horizontal) but in specific areas there are distinctive relationships. This situation is similar to Case (c) but here the differences result from some areas having high crash frequency for high value of the risk factor, while in others they have the lowest frequency.

The cross-area variations of slopes and intercepts could be caused by various area-specific heterogeneities. For those observable heterogeneities, it is theoretically possible to factorize and then account for them by using some classical techniques such as GLM with consideration of interactions, ANOVA, or ANOCOVA. But traffic crash is a complex event with a large number of factors involved. Ideally, all of the relevant factors in different levels (e.g. road segment level and area level) should be considered in the model. In practice, however, some of the factors may not be available or even uncollectable for study. A model may only consider the most important factors and omit the others. It assumes that similar groups (i.e. with same selected observable factors) have the same pattern of crash occurrence. In the real world, however, similar groups (e.g. area) may be different in omitted factors and thus may have different means. These unobservable or omitted heterogeneities introduce additional variance to the data and cause the over-dispersion. Consequently, without appropriately accounting for the cross-group heterogeneities, the estimates of the standard error in the regression coefficients may be underestimated. Moreover, in the presence of spatiotemporal correlation, the accommodation of these specific data features would be valued.

The patterns shown in the above example exist almost everywhere in traffic safety studies since most crash datasets are collected with an inherent hierarchical and spatiotemporal structure. For example, in predicting crash severities, it is reasonable to assume that the characteristics of the vehicles within which casualties are traveling will affect their probability of survival. If this is the case, then casualties within the same vehicle would tend to have more similar severity than casualties in different vehicles, rendering the assumption of residual independence invalid. The same argument may be extended to encompass the effect of similarities between different crashes, traffic sites, or geographical regions.

### *3.2. A 5 × ST-level hierarchy in traffic safety data*

For the purpose of systematic inspection, a 5 × ST-level hierarchy, as shown in Fig. 2, is proposed to represent the general framework of multilevel data structures in traffic safety. Along the vertical of this triangular prism is a five-level hierarchy representing various traffic entity with spatial distribution including from macroscopic to the microscopic levels, Geographic region level–Traffic site level–Traffic crash level–Driver-vehicle unit level–Occupant level. All these traffic entity levels are structured along a horizontal time axis, defined as Time level, thus resulting in a "**5 × ST**"-level hierarchy, i.e. **5** entity levels × SpatioTemporal level. The involvement and emphasis for different sub-groups of these levels depend on different research purposes and also rely on the heterogeneity examination on the crash data employed. Generally, macro-analysis focus on the top three levels, i.e. Geographic region level, Traffic site level, and Traffic crash level, while micro-analysis concern the bottom three levels, i.e. Traffic crash level, Driver-vehicle unit level, and Occupant level.

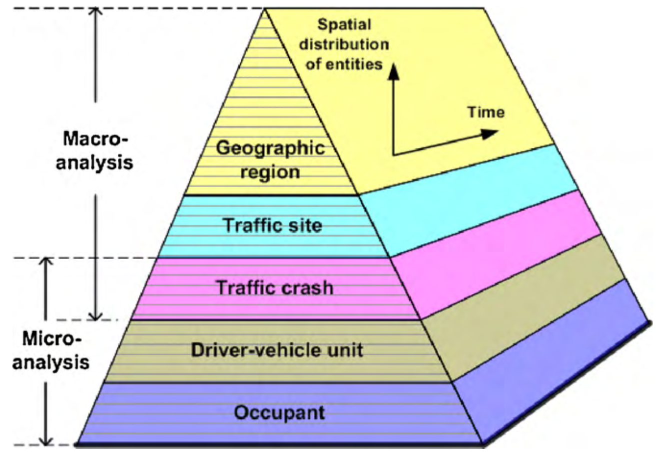


**Fig. 2.** A 5 × ST-level hierarchy in traffic safety data.

Specifically, the Geographic regional level could be a number of regions, countries, states, counties or cities, etc. Inter-regional studies generally include the traffic data collected from the regions of interest. This level is normally associated with a number of contextual factors potentially affecting the traffic safety situation such as driving regulations, road density, spatial features, population and other socio-economic features. Nested under Geographic regional level is Traffic site level, which is of greatest interest in many traffic safety studies. It consists of what constitute the basic road network, namely road segments (link) and road junctions (node). A corridor could also be considered as within Traffic site level in case that the safety situation of a whole corridor is of concern. Various collective or comparative safety studies are conducted regarding road design, operation and assessment. Next, while traffic sites necessarily reside in some geographic region, traffic crashes of various types occur at different traffic sites. Traffic crash level has been the most direct and thus most used criterion in monitoring the safety situation for traffic sites. It is intuitively reasonable that characteristics of crashes occurring at a same site should be correlated due to the same context in terms of geometric, traffic, and regulatory control factors. Measures such as crash severity, collision type and possible crash causes are used to characterize the traffic crashes. Driver-vehicle unit level is the most concerned entity in traffic safety as it directly relates to the driver behavior and vehicle maneuver leading to crashes as well as overall vehicle crash worthiness. Individual severity of driver injury or vehicle damage may potentially show a strong correlation between those involved in the same multi-vehicle crashes. Various driver and vehicle characteristics are factors distinguishing different involved units in this level. The lowest level in the hierarchy is vehicle occupants involved in crashes, which refer to both drivers and passengers. This level is commonly concerned in studies on body injury worthiness related to age and gender, as well as in-vehicle protection measures for different seats. Finally, traffic data in any entity level are necessarily marked by a time scale (horizontal axis in the prism), with which the interest of studies may be on the time serial correlations of traffic safety situation.

In the proposed conceptual model, it seems a bit confusing for that drivers are included in both Driver-vehicle unit level and Occupant level. Note that drivers in Driver-vehicle unit level represent the persons who control the vehicle and as such the emphasis is on their driving behavior related characteristics. Whereas, drivers in Occupant level are seen as one of the vehicle occupants in which their body injury worthiness and in-vehicle protection facilities are of most concern.

Following the framework, typical data-clustering designs in traffic safety research could vary depending on the research purposes. For example, in some inter-regional studies, with Geographic region level as the higher level, study subjects in the lower level could be safety performance of various traffic sites, drivers or vehicles. In these cases, two-entity-level design could be used to explicitly examine the safety effects of risk factors in both individual and contextual levels. The two-entity-level design can be naturally extended to reflect three-entity-level data structure, for example, Geographic region level–Traffic site level–Traffic crash level. It is also worth noting that for the levels of geographic region and/or traffic site, spatial effects could be present among adjacent entities of analysis. The geographic distribution of the regions or sites should be taken into account as closer traffic entities may have certain similarity in crash occurrence. Moreover, when time series are considered, panel data design or repeated cross-sectional design could be used. In panel data design, a set of sites within regions are pre-selected on which repeated measures along time are conducted, whereas repeated cross-sectional design consider a number of time periods, in each of which selected sites may be different. In the scope of micro-analysis, crash risk and severity are mostly concerned for the Driver-vehicle unit level and Occupant level. By controlling for covariates at Traffic crash level, taking Driver-vehicle units as observation unit may reveal crash propensity variation among different drivers and vehicles. For the Occupant level, research could be conducted to investigate the injury severity variation among different seats by taking each individual occupant as an observation unit, whereas covariates at the upper levels such as Driver-vehicle unit level and Traffic crash level could be controlled for.

## 4. Bayesian hierarchical method on multilevel crash data

The previous sections show that an appropriate method is needed to account for the multilevel data structure in the traffic safety discipline. In this section, a methodological framework using Bayesian hierarchical modeling is established to properly model the potential heterogeneities due to the multilevel data structure. A number of advantages of this method ensure its great potential of extensive applications in traffic safety analysis.

### 4.1. Hierarchical model

To model the multilevel data structure, several potential solutions have been found in the literature. For example, some researchers have employed the artificial intelligent models (AI) in crash prediction such as the most widely used neural networks (NN) and Bayesian NN (Mussone et al., 1999; Abdelwahab and Abdel-Aty, 2001; Riviere et al., 2006; Xie et al., 2007). But the NN has been criticized for being black boxes incapable of generating explicit functional relationships and statistically interpretable results. Another useful technique in accounting for correlated data is the generalized estimating equations (GEE), which is regarded as an extension of GLM (Lord and Persaud, 2000; Abdel-Aty and Abdalla, 2004; Wang and Abdel-Aty, 2006). GEE is also called as 'marginal' model, as distinguished from 'subject specific' model, such as hierarchical model in this paper. When dealing with multilevel data structure, GEE aims to provide estimates with acceptable properties only for the fixed parameters in the model, while treating the existence of any random parameters as a necessary 'nuisance'. Hence, the GEE may merely be superior in the case where the exact form of the multilevel data structure is unknown.

Another way to distinctly address the multilevel data structure is to use hierarchical models (also called as multilevel model, random effect model or random parameter model). Hierarchical modeling is a statistical technique that allows multilevel data structures to be properly specified and estimated (see Gelman and Hill, 2007). Although the basic theories of hierarchical model have been developed and discussed for many years, it is only recent that many practical limitations on the use of hierarchical analysis have been overcome. Currently, hierarchical models have become commonplace in research in a variety of other disciplines such as sociology, education, political science, and public health. In employing the hierarchical model in the first application in a traffic crash study, Shankar et al. (1998) showed that the inclusion of site-specific random effects and time indicators into the NB regression model can significantly improve the explanatory power of crash models. 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 multilevel data structure in crash prediction. For example, while some researchers (Mitra and Washington, 2007; Chin and Quddus, 2003; Yang and MacNab, 2003; Kim et al., 2007; Li et al., 2008; Quddus, 2008; Huang et al., 2009; Haque et al., 2010) employed the hierarchical models for predicting crash frequency, others (Jones and Jorgensen, 2003; Lenguerrand et al., 2006; Huang et al., 2008) developed hierarchical models to identify factors affecting crash severity.

As defined by Gelman and Hill (2007), a multilevel/hierarchical model is a regression (a linear or generalized linear model) in which the parameters – the regression coefficients – are given a probability model. Hence, this higher level model has parameters of its own – the hyperparameters of the model – which are also estimated from the data. In the context of GLM, the hierarchical modeling (also called hierarchical GLMs) is mainly working on the link function: disturbance terms are added to the model corresponding to different sources of variation in the multilevel data.

Specifically, recall the general expression of statistical modeling in Eq. (1), while the first part of the expression ($Dist(\boldsymbol{\theta})$) remains to represent different characteristics of the crash features of interest, it is the disturbance term $\boldsymbol{\varepsilon}$ which differs the hierarchical modeling to classical statistical models. It should be noted that here the $\boldsymbol{\varepsilon}$ represents a general concept for the disturbances. In fact, it could consist of many components, with some of which working on the intercept, others on the slopes in the link function.

A two-level hierarchical model is used to mathematically interpret how the method works on the multilevel data. As with most practices, a basic linear combination of $\mathbf{X}$ and $\boldsymbol{\beta}$ is assumed to simplify the interpretation. Furthermore, the covariate vector $\mathbf{X}$ is divided into three components, $c(1, \mathbf{X}^{L1}, \mathbf{X}^{L2})$, to respectively represent the factors associated with intercept, individual level (level 1) and group level (level 2). Correspondingly, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are also divided into different components to serve different functions with the bold symbol representing vector or matrix. Hence, the link function becomes the combination of models in terms of two levels,

$$\begin{aligned}&\text{Level 1 model}: \quad f^{-1}(\boldsymbol{\theta}) = \beta_0^{L1} + \boldsymbol{X}^{L1}\boldsymbol{\beta}^{L1} + \varepsilon^{L1}\\ &\text{Level 2 model}: \quad \beta_0^{L1} = \beta_{00}^{L2} + \boldsymbol{X}^{L2}\boldsymbol{\beta}_0^{L2} + \varepsilon_0^{L2}\\ &\boldsymbol{\beta}^{L1} = \boldsymbol{\beta}_{01}^{L2} + \boldsymbol{X}^{L2}\boldsymbol{\beta}_1^{L2} + \boldsymbol{\varepsilon}_1^{L2}\end{aligned} \tag{2}$$

The combined model is obtained by substituting the level 2 model into level 1 model,

$$\begin{aligned}f^{-1}(\boldsymbol{\theta}) = {}&(\beta_{00}^{L2} + \boldsymbol{X}^{L1}\boldsymbol{\beta}_{01}^{L2} + \boldsymbol{X}^{L2}\boldsymbol{\beta}_0^{L2} + \boldsymbol{X}^{L1}\boldsymbol{X}^{L2}\boldsymbol{\beta}_1^{L2})\\ &+ (\varepsilon^{L1} + \varepsilon_0^{L2} + \boldsymbol{X}^{L1}\boldsymbol{\varepsilon}_1^{L2})\end{aligned} \tag{3}$$

It is clear that now the link function consists of two parts: fixed part and random part. The fixed part means a deterministic relationship fully depending on covariate $\mathbf{X}$, while random part is stochastically determined by a number of disturbance terms. The components in both the two parts are interpreted as follows.

*Fixed part*:

(1) $\beta_{00}^{L2}$: The intercept, which is the main effect with all covariates equal zero. By centering all covariates on their mean, this term represents the main effect on the average values of covariates.
(2) $\boldsymbol{X}^{L1}\boldsymbol{\beta}_{01}^{L2}$: $\boldsymbol{\beta}_{01}^{L2}$ is the mean of the main-effect coefficient of level 1 covariates $\boldsymbol{X}^{L1}$ on the dependent variable.
(3) $\boldsymbol{X}^{L2}\boldsymbol{\beta}_0^{L2}$: $\boldsymbol{\beta}_0^{L2}$ is the main-effect coefficient of level 2 covariates $\boldsymbol{X}^{L2}$ on the dependent variable.
(4) $\boldsymbol{X}^{L1}\boldsymbol{X}^{L2}\boldsymbol{\beta}_1^{L2}$: $\boldsymbol{\beta}_1^{L2}$ is the interactive-effect coefficient of $\boldsymbol{X}^{L1}$ and $\boldsymbol{X}^{L2}$. This component make it possible to in-depth understanding of how contextual factor (level 2 covariates) could affect the individual factor (level 1 covariates).

*Random part*:

(1) $\varepsilon^{L1}$: The disturbance term associated with level 1 analysis. Normally, it is assumed to be identical independent distributed (IID) among individuals with mean zero and variance to be estimated. The associated unknown variance structure of this term facilitates the estimation of unobservable or omitted between-individual heterogeneity. The additional disturbance in over-dispersed Poisson models is a typical example, in which with Gamma distribution assumption on $\exp(\varepsilon)$ resulting in Poisson-Gamma model, and Lognormal distribution assumption in Poisson-Lognormal model.
(2) $\varepsilon_0^{L2}$: The disturbance term associated with level 2 analysis. It is also common to assume IID among groups (level 2) with mean zero and variance to be estimated. With this term, those individuals (level 1) belonging to a same group (level 2) share a same variance component, thus resulting in a within-group covariance. As a result, the model intercept now consists of three parts: $\beta_{00}^{L2} + \varepsilon^{L1} + \varepsilon_0^{L2}$ and is hence variable by between-individual (or within-group) variation $\varepsilon^{L1}$ as well as between-group variation $\varepsilon_0^{L2}$.
(3) $\boldsymbol{X}^{L1}\boldsymbol{\varepsilon}_1^{L2}$: $\boldsymbol{\varepsilon}_1^{L2}$ is the disturbance vector on the slope of level 1 covariates $\boldsymbol{X}^{L1}$ associated with level 2. $\boldsymbol{\varepsilon}_1^{L2}$ makes the slope of $\boldsymbol{X}^{L1}$ variable according to the data clustering. In other words, individuals in a same group share with a same variance on the slope. As a result, the slope of $\boldsymbol{X}^{L1}$ consists of two parts: $\boldsymbol{\beta}_{01}^{L2} + \boldsymbol{\varepsilon}_1^{L2}$ and is hence variable by between-group variation $\boldsymbol{\varepsilon}_1^{L2}$. Note that while $\boldsymbol{X}^{L1}$ has varying main-effect slope, the main-effect slope for $\boldsymbol{X}^{L2}$ is fixed. A higher level analysis, for example in a three-level model, could make this level 2 slope varying.

It is clear that $\varepsilon_0^{L2}$ and $\boldsymbol{\varepsilon}_1^{L2}$ are the unique features of hierarchical models while all of the rest components could be included and estimated in classical models. It is just these two stochastic terms making it possible to account for the unobservable or omitted heterogeneity in level 2 model.

In the framework of hierarchical modeling, the two-level model shown in Eq. (3) is also called as varying-intercept and varying-slope model. Obviously, this full-version model could be simplified by taking account of either varying intercept or varying slope, resulting in varying-intercept model and varying-slope model.

Varying-intercept model:

$$f^{-1}(\boldsymbol{\theta}) = (\beta_{00}^{L2} + \boldsymbol{X}^{L1}\boldsymbol{\beta}_{01}^{L2} + \boldsymbol{X}^{L2}\boldsymbol{\beta}_0^{L2} + \boldsymbol{X}^{L1}\boldsymbol{X}^{L2}\boldsymbol{\beta}_1^{L2}) + (\varepsilon^{L1} + \varepsilon_0^{L2}) \tag{4}$$

Varying-slope model:

$$f^{-1}(\boldsymbol{\theta}) = (\beta_{00}^{L2} + \boldsymbol{X}^{L1}\boldsymbol{\beta}_{01}^{L2} + \boldsymbol{X}^{L2}\boldsymbol{\beta}_0^{L2} + \boldsymbol{X}^{L1}\boldsymbol{X}^{L2}\boldsymbol{\beta}_1^{L2}) + (\varepsilon^{L1} + \boldsymbol{X}^{L1}\boldsymbol{\varepsilon}_1^{L2}) \tag{5}$$

Clearly, all these models could be expanded to accord with more complicated designs. The above derivative also shows that the hierarchical modeling provides a rather flexible technique to account for various study purposes and different extent of model complexity such as within-level or between-level interactions, varying intercept, and varying slopes. Moreover, spatiotemporal effects could also be incorporated into the hierarchical models by specifying the hypothesized features on the $\varepsilon_0^{L2}$ and $\boldsymbol{\varepsilon}_1^{L2}$, such as spatial or temporal autoregressive models (see Banerjee et al., 2003). Recently, several safety studies have successfully applied the spatial models (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., in press).

Fitting hierarchical models, as well as displaying, checking, analyzing the model results necessarily get much more complicated than classical models. Given the increased 'costs' of using hierarchical models, a number of major advantages are identified in the traffic safety research context. First, hierarchical modeling provides a coherent model that simultaneously incorporates both individual-level and group-level models. Second, it is more efficient in inference for parameters. Compared to complete pooling across all groups and no-pooling, the modeling paradigm of hierarchical analysis represents a preferred partial pooling, i.e. a compromise between the two extremes. Moreover, as hierarchical modeling combines information from multilevel variations, it is feasible to use all the data to perform inference for groups with small sample size. The last but not the least, is the capability to provide more efficient crash prediction for new observation or unit.

### 4.2. Bayesian inference

Bayesian approach is a prevailing way to explicitly model the hierarchical structure. With the recent development of computing capacity and Bayesian analysis techniques, a good number of researchers have been working on estimating the models in a Bayesian framework. Bayesian inference (BI) 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. Instead of giving "maximum likelihood" estimates for the studied unknowns totally based on the sample data in MLE inference, the essential characteristic of Bayesian methods is its explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis. Specifically, the ultimate aim of Bayesian data analysis is to obtain the marginal posterior distribution of all unknowns, and then integrate this distribution over the unknowns that are not of immediate interest to obtain the desired marginal distribution.

As indicated from a large number of theoretical studies and applications, BI shows numerous theoretical and practical advantages over the "classical" likelihood-based inference methods (also

**Table 1**
Model comparison: hierarchical model vs. GLM model.

| | Deviance information criterion (DIC) | Mean absolute deviance (MAD) | Mean squared predictive error (MSPE) |
|---|---|---|---|
| Negative Binomial | 7661 | 0.84 | 1.42 |
| Poisson-Lognormal | 7655 | 0.82 | 1.34 |
| Hierarchical Poisson | 7344 | 0.63 | 1.12 |
| Hierarchical Poisson (AR-1)[a] | 7339 | 0.51 | 1.03 |

*Note*: results retrieved from Table 1 in Huang et al. (2009).
[a] Serial correlation coefficient = 0.775, Bayesian credible interval (0.72, 0.80).

called frequentist methods). Several major advantages in the traffic safety context are identified as follows.

(1) Crash data is difficult to collect and gradually available, e.g. year by year. There is also possible variation for crash prediction models as the outcome of changes of some influential factors, e.g. the installation of red light camera, or the adjustment of amber interval time. The Bayesian algorithm provides a quite flexible and reliable measure to realize the updating requirement. In Bayesian context, the previous model, any engineering experiences or justified previous findings could be used as the prior knowledge of the updated model.
(2) Missing data occur very commonly in crash records. In Bayesian method, missing data are automatically modeled as latent variables in a manner that takes into account the information contained in other observed data.
(3) Bayesian posterior distributions for parameters are perfectly valid for any size of sample. One of the most important strengths of BI is the capability to handle small size data. The extensive application of empirical Bayesian approach in observational before-after study of safety treatment evaluation is a good supportive example of this statement (Hauer, 1997).
(4) Regarding model comparison, frequentist hypothesis tests require that only two models are compared, and these models must be nested. In a Bayesian setting, any number of non-nested models may be compared.

The well-known computing approach for BI is Markov chain Monte Carlo (MCMC) methods (Gelman et al., 2003). MCMC is a general method based on drawing values of unknowns from approximate distributions and then correcting those draws to better approximate the target posterior distribution. Gibbs sampler and the Metropolis–Hastings algorithm are the most widely used simulation algorithms in MCMC. In complex hierarchical models where parameters may outnumber observations, the recently developed deviance information criterion (DIC), a Bayesian generalization of AIC, could be used to measure the model complexity and fit (Speigelhalter et al., 2003a). Gelman–Rubin convergence statistics (Brooks and Gelman, 1998) could be employed to monitor the convergence of MCMC simulation chains. BUGS modeling language (Bayesian Inference using Gibbs Sampling) is a prevailed tool to allow the computation using MCMC algorithms for all sorts of Bayesian models, including most of the hierarchical models applied. WinBUGS package (Spiegelhalter et al., 2003b) provides a flexible and simplified platform to modeling with the BUGS programs. In particular, since specification of the full conditional densities is not necessary in WinBUGS, small changes in program code can achieve a wide variation in modeling options and thus facilitating sensitivity analysis and prior assumptions.

## 5. Illustrative examples

In this section, several studies recently accomplished by the current authors and their colleagues are deliberately selected to exemplify the proposed conceptual model of the 5 × ST-level hierarchy with Bayesian hierarchical estimation approach. Levels covered in these examples vary from levels of Geographic region, traffic site, traffic crash, driver-vehicle unit to spatial and temporal levels. While Examples 1–3 address crash frequency analysis, Examples 4–5 deal with crash severity analysis.

### 5.1. *Example 1: Crash frequency ∼ Intersection level × Time level*

To evaluate alternative approaches in identifying crash hotspots, Huang et al. (2009) developed a set of Bayesian hierarchical models to predict the expected crash rate for 582 intersections in Singapore. Crash historical data in 10 years (1997–2006) were employed to accomplish the analysis. Specifically, a two-level design was employed to accommodate the site-specific effect for multiple yearly observations for specific sites, i.e. Crash frequency ∼ Intersection level × Time level. The potential correlation among multiple observations at a specific site was modeled in two ways, one with a site-specific random effect, and the other with a time serial autoregressive lag-1 (AR-1) specification.

Table 1 cites the model comparison results presented in Huang et al. (2009). The results showed that both the hierarchical models (Hierarchical Poisson model, HP and Hierarchical Poisson (AR-1) model, HP(AR-1)) with accommodation for site-specific effect and serial correlation have better goodness-of-fit than non-hierarchical models (NB model and Poisson-Lognormal model, PLN). Specifically, judged by the MAD and MSPE values, the HP and HP(AR-1) models significantly outperform NB and PLN models with respect to model-fitting. Likewise, by using DIC, the model comparison shows that the accommodation of multilevel data structure can substantially improve the model performance. Given the better model fitting results, it is also not surprising to find that the Bayesian hierarchical models outperform the results based on traditional crash prediction models in correctly predicting crash hotspots. The evaluation of predictive performance was conducted by using 3-year data to predict the 10-year average. This study indicates that the flexibility in model specification of the Bayesian modeling approach can generate great potentials to improve the crash prediction models and subsequent applications by explicitly accounting for various structured heterogeneities in crash data.

### 5.2. *Example 2: Crash frequency ∼ [County level – Corridor level – Intersection level] × Spatial effect*

In an exploratory study of modeling signalized intersection safety (Guo et al., 2009), a total of 170 four-legged signalized intersections along 25 principle and minor arterials crossing two Central Florida counties were investigated. By a preliminary analysis, a three-level design was formulated to accommodate the multilevel data structure as well as the spatial effect, i.e. Crash frequency ∼ [County level – Corridor level – Intersection level] × Spatial effect. In the model, as discussed in Guo et al. (2009), there are potentially three levels of correlation for the crash frequency at an individual intersection during the observation period (2000–2005), i.e. County level, Corridor level, and Intersection level. Specifically, crashes that occurred within a same county are possibly more "similar" comparing to those in other counties while intersections within the same corridor may also be correlated with each other due to some omitted corridor-specific characteristics. In the analysis, the county level was accounted for by fixed effects with two dummy variables; and random effect stochastic terms were used to specify the potential corridor-specific heterogeneity. Moreover, a point-referenced conditional autoregressive (CAR) prior was also incorporated to test the spatial correlation among intersections with varying distance. In other words, intersections closer to each other were hypothesized to be more similar than those far apart.

The Bayesian inference was employed to estimate the proposed models. Results showed that the within-corridor correlation is substantial with significant CAR effects in NB CAR model (16.8, Bayesian credible interval, BCI (2.54, 62.77)) and HP CAR model (0.79, BCI (0.58, 1.04)). As reflected by the model-fitting results and DIC values, it found that the HP-CAR model (DIC = 1378) outperformed all the other models including standard Poisson model (DIC = 3209), NB model (DIC = 1378), NB mixed effect model (DIC = 1626) and NB CAR model (DIC = 1628). The study implied that the hierarchical spatial models provide a better representation of the stochastic processes underneath the observed safety data with multilevel structures. This result also confirmed the well-known fact that the NB can be generated by a HP model and implied that HP model can be valuable in modeling the safety data with multilevel data structure.

### 5.3. Example 3: Crash frequency ~ County level × Spatial effect

The necessity and capability of spatial model to account for spatial data as shown in Example 2 has also been explored in an aggregate crash prediction model (Huang et al., in press). In the study, crash statistics in a 5-year period (2003–2007) were investigated for 67 Florida counties. By using the Moran's I statistics in a GIS platform, positive spatial correlation was detected among adjacent counties with most of the $Z$ scores higher than 1.68 which represents a 90% confidence level. Hence, a Bayesian area-referenced CAR model was developed to accommodate for the potential spatial effects. The spatial model was assumed to be able to account for various confounding factors which are not observed or unobservable in the analysis. Furthermore, in the context of area-referenced CAR model, those confounding factors are supposed to be spatially correlated among adjacent counties and their effects on crash risks are homogeneous.

Estimations of the marginal standard deviations of state-wide heterogeneity and clustering effects among adjacent counties were used to calculate the proportion of the variability in the random effects that is due to clustering. The results showed that variations accounted by spatial clustering are substantial for all the all-crash and severe-crash risk models, specifically, 51.7% and 42.4%, respectively for models with daily VMT as exposure, and 25.9% and 26.4%, respectively for models with population as exposure.

### 5.4. Example 4: Severity of driver injury and vehicle damage ~ Traffic crash level – Driver-vehicle unit level

The Examples 1 through 3 have illustrated the application of Bayesian hierarchical models in addressing crash frequency prediction models with multilevel and spatiotemporal data structure. Examples 4 and 5 will exemplify how to deal with multilevel data in crash severity analysis.

In Huang et al. (2008), a Bayesian hierarchical binomial logistic model was developed 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 from 2003 to 2005 were used. It was found that in the 4095 signalized intersection related crashes, 7840 driver-vehicle units were involved, resulting in an average involvement rate of 1.91 individuals per crash. A preliminary inspection on the crash severity of these data revealed a significant correlation between individuals involved in the same multi-vehicle crashes, which represent 83.5% of all crashes. Specifically, in a multi-vehicle crash, if a driver and/or vehicle was injured and/or damaged severely, then the others had a probability of 31% also to be so. On the other hand, if there was no severe damage and injury to a driver-vehicle unit, the chance for the others to be severe is only 12%. This motivated a two-level hierarchical design of the model, i.e. Crash Severity ~ Traffic crash level – Driver-vehicle level, so as to take into account the potential cross-crash heterogeneity. For the purpose of comparison, standard logistic model has also been developed to justify the use of the more 'expensive' two-level model.

Model estimation results showed that the variance of the crash-specific random effects is 1.34, which is statistically significant with BCI (0.56, 2.29). By use of intra-class correlation coefficient (ICC), this means the proportion of unexplained variations in individual severity resulted from cross-crash variance is about 28.9%. Thus, it is not surprising to hypothesize that accounting for the multilevel data structure in the proposed hierarchical model would lead to a better model performance. By the model-fitting component in DIC, the goodness-of-fit of the proposed two-level model improved significantly over the standard logistic model in which all driver-vehicle units were treated independently. After penalized by the effective number of parameters in the model, the DIC value for the hierarchical model (DIC = 3067.9) is also considerably less than that in the standard logistic model (DIC = 6191.9). This further showed that the use of crash-level random effects in the hierarchical model can substantially improve the model performance. The improved model performance is useful to reduce the variability in estimating the safety effects of risk factors at both crash level and driver-vehicle unit level.

### 5.5. Example 5: Crash severity ~ Road segment level – Traffic crash level

A study was recently conducted to examine crash injury severity related to visibility obstruction due to fog/smoke (Huang et al., 2010). A total of 994 fog/smoke related crashes were identified from the Florida state-road crash database for the period of 2003–2007. These crashes are sparsely distributed within the Florida road network. Based on the spatial locations of these crashes, a total of 597 road segments were defined, which have largely uniform road characteristics. The lengths of these segments range from 2 to 5 miles. A two-level ordered logistic model, i.e. Crash severity ~ Road segment level – Traffic crash level, was developed to account for the five ordered crash severity levels (from no injury to fatality) and the potential cross-road segment heterogeneity. Specifically, while an ordinal response model normally takes fixed thresholds to define the boundary between the intervals corresponding to observed severity outcomes, the proposed multilevel model specifies a set of variable thresholds for individual segments. The crash-level covariates (e.g. ADT, driver age, etc.) were used to explain the different severity outcomes, and the segment-level covariates were incorporated into the determination of variable threshold values. Segment-specific random effect has also been utilized to estimate the omitted confounding factors associated with road segments. Results showed that the precision parameter of the random effects (65.82) is significant judged by the BCI (10.06, 145.9). Furthermore, by using the DIC, the proposed multilevel crash severity prediction model (DIC = 2844.7) was also found to be better comparing to ordinary ordered models (DIC = 4237.2).

In summary, the studies as presented in this section exemplified the potential benefit of accounting for various multilevel safety data structures. Compared to conventional models, Bayesian hierarchical approach can yield more accurate and efficient safety estimation for entities at different levels. Most importantly, it should be emphasized that benefits of these efforts were arisen from rational examination on realistic data structures, rather than a blind treatment for disturbance.

## 6. Conclusion and future study

The past decade has witnessed a revolutionary advancement in statistical modeling techniques, and multilevel modeling approach is among the most notable. Given the hierarchically and spatiotemporally structured safety data, it is promising to advance traffic safety modeling by adopting this emerging technique. Towards this end, this study takes a substantial step to regularize typical safety data structures and the corresponding analytical approach related to the development of crash prediction model. The proposed $5 \times$ ST-level hierarchy represents a conceptual framework with the up-to-date understanding on safety data structure. The state-of-the-art technique, i.e. Bayesian hierarchical approach, has shown good potential to explicitly account for the specific multilevel data structure as represented in the $5 \times$ ST-level hierarchy.

Although many successes have been reported in the literature, much work is still needed toward the wide application of the proposed theory and method in the traffic safety field. The first issue in need of future effort is the reliability test of the hierarchical models in yielding better results in traffic safety. Although these types of model have been well coordinated in many other fields such as economics and sociology, it is only less than 10 years when traffic safety analysts first reported relevant results in well-known publications. Especially, while most applications are limited to varying-intercept model, varying-slope model deserves more effort because of the complexity in model estimation and result interpretation. There is also a need to test the applicability, robustness and transferability of these emerging techniques. Another issue related to application concern is the cost-benefit rate. The current methods available in calibrating hierarchical models especially in a context of the Bayesian approach, are relatively computing and intellectual intensive. The cost-benefit rate for applying those techniques to industry should be carefully evaluated.

## Acknowledgements

We would like to sincerely thank our colleagues for their invaluable supervision and collaboration to fulfill the studies which are used as examples in this paper. They are Prof. Chin Hoong Chor, Dr. Md. Mazharul Haque at National University of Singapore, Dr. Xuesong Wang at Tongji University, Dr. Feng Guo at Virginia Tech Transportation Institute, and our former students at the University of Central Florida, Ali Lotfi Darwiche and Al-Ahad Ekram, etc.

## References

Abdel-Aty, M.A., Abdalla, M.F., 2004. Linking roadway geometrics and real-time traffic characteristics to model daytime freeway crashes: generalized estimating equations for correlated data. Transportation Research Record 1879, 106–115.

Abdel-Aty, M., Keller, J., 2005. Exploring the overall and specific crash severity levels at signalized intersections. Accident Analysis and Prevention 37, 417–425.

Abdel-Aty, M., Radwan, E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32, 633–642.

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

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38, 618–625.

Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. In: TRB 2009 Annual Meeting CD-ROM, Transportation Research Board. National Research Council, Washington, D.C.

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

Anastasopoulos, P.Ch., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41, 153–159.

Banerjee, S., Carlin, B., Gelfand, A., 2003. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall, CRC Press.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulation. Journal of Computational and Graphical Statistics 7 (4), 434–455.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis and Prevention 35, 253–259.

El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transportation Research Record 1950, 9–16.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41, 820–828.

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

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Guo, F., Wang, X., Abdel-Aty, M., 2009. Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis and Prevention 42 (1), 84–92.

Haque, M.M., Chin, H.C., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis & Prevention 42, 203–212.

Hauer, E., 1992. Traffic conflicts and exposure. Accident Analysis and Prevention 14, 359–364.

Hauer, E., 1997. Observational Before–After Studies in Road Safety. Pergamon/Elsevier Science, Inc., Tarrytown, New York.

Heydecker, B.G., Wu, J., 2001. Identification of sites for accident remedial work by Bayesian statistical methods: an example of uncertain inference. Advances in Engineering Software 32, 859–869.

Huang, H., Abdel-Aty, M., Darwiche, A.L., in press. County-level crash risk analysis in Florida: Bayesian spatial modeling. Transportation Research Record.

Huang, H., Abdel-Aty, M., Ekram, A., Oloufa, A., Chen, Y., Morrow, R., 2010. Fog related crashes in Florida: identifying crash characteristics, spatial distribution and injury severity. In: TRB 2010 Annual Meeting CD-ROM, Transportation Research Board. National Research Council, Washington, D.C.

Huang, H., Chin, H.C., in press. Modeling road traffic crashes with zero-inflation and site-specific random effects. Statistical Methods & Applications, doi:10.1007/s10260-010-0136-x.

Huang, H., Chin, H.C., Haque, M.M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. Accident Analysis and Prevention 40, 45–54.

Huang, H., Chin, H.C., Haque, M.M., 2009. Empirical evaluation of alternative approaches in identifying crash hotspots: naive ranking, empirical Bayes and full Bayes. Transportation Research Record 2103, 32–41.

Jones, B., Jansen, L., Mannering, F.L., 1991. Analysis of the frequency and duration of the freeway accidents in Seattle. Accident Analysis and Prevention 23, 239–255.

Jones, A.P., Jorgensen, S.H., 2003. The use of multilevel models for the prediction of road accident outcomes. Accident Analysis and Prevention 35, 59–69.

Joshua, S.C., Garber, N.J., 1990. Estimating truck accidents rate and involvements using linear and Poisson regression models. Transportation Planning and Technology 15 (1), 41–58.

Jovanis, P., Chang, H., 1986. Modeling the relationship of accident to mile traveled. Transportation Research Record 1068, 42–51.

Kim, D.G., Lee, Y., Washington, S., Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: application of hierarchical binomial logistic models. Accident Analysis and Prevention 39, 125–134.

Kulmala, R., 1995. Safety at Rural Three- and Four-arm Junctions: Development and Application of Accident Prediction Models. Technical Research Center at Finland, VTT Publications, Espoo.

Kumara, S.S.P., Chin, H.C., 2003. Modeling accident occurrence at signalized T-intersections with special emphasis on excess zeros. Traffic Injury Prevention 3 (4), 53–57.

Lee, A.H., Stevenson, M.R., Wang, K., Yau, K.K.W., 2002. Modeling young driver motor vehicle crashes: data with extra zeros. Accident Analysis and Prevention 34, 515–521.

Lee, J., Mannering, F.L., 2002. Impact of roadside features on the frequency and severity of run-off-road accidents: an empirical analysis. Accident Analysis and Prevention 34, 349–361.

Lenguerrand, E., Martin, J.L., Laumon, B., 2006. Modeling the hierarchical structure of road crash data: application to severity analysis. Accident Analysis and Prevention 38, 43–53.

Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008. The choice of statistical models in road safety countermeasures effectiveness studies in Iowa. Accident Analysis and Prevention 40, 1531–1542.

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46 (5), 751–770.

Lord, D., Park, Y.J., 2008. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. Accident Analysis and Prevention 40, 1441–1457.
Lord, D., Persaud, B.N., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transportation Research Record 1717, 102–108.
Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models for motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37, 35–46.
Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero-inflated models in highway safety. Accident Analysis and Prevention 39, 53–57.
Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40, 964–975.
Malyshkina, N.V., Mannering, F.L., 2009. Markov switching multinomial logit model: an application to accident-injury severities. Accident Analysis and Prevention 41, 829–838.
Malyshkina, N.V., Mannering, F.L., 2010. Zero-state Markov switching count-data models: an empirical assessment. Accident Analysis and Prevention 42, 122–130.
Malyshkina, N.V., Mannering, F.L., Tarko, A.P., 2009. Markov switching negative binomial models, an application to vehicle accident frequencies. Accident Analysis and Prevention 41 (2), 217–226.
Mannering, F.L., Grodsky, L.L., 1995. Statistical analysis of motorcyclists' perceived accident risk. Accident Analysis and Prevention 27, 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. Transportation Research Record 1581, 37–46.
Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road section: Poisson versus negative binomial regression. Accident Analysis and Prevention 26, 471–482.
Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transportation Research Record 1840, 31–40.
Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention 25, 689–709.
Miaou, S., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space–time modeling approach. Journal of Transportation and Statistics 6 (1), 33–57.
Miranda-Moreno, L.F., Fu, L., Saccomanno, F.F., Labbe, A., 2005. Alternative risk models for ranking locations for safety improvement. Transportation Research Record 1908, 1–8.
Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accident Analysis and Prevention 39, 459–468.
Mussone, L., Ferrari, A., Oneta, M., 1999. An analysis of urban collisions using an artificial intelligence model. Accident Analysis and Prevention 31, 705–718.
O'Donnell, C.J., Connor, D.H., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention 28, 739–753.
Park, E.S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019, 1–6.
Park, B.J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention 41, 683–691.
Poch, M., Mannering, F.L., 1996. Negative binomial analysis of intersection accident frequencies. Journal of Transportation Engineering 122 (2), 105–113.
Qin, X., John, N.I., Nalini, R., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis and Prevention 36, 183–191.
Quddus, M.A., 2008. Modeling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis & Prevention 40, 1486–1497.
Quddus, M.A., Noland, R.B., Chin, H.C., 2002. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research 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. D.C.
Riviere, C., Lauret, P., Ramsamy, J.F.M., Page, Y., 2006. A Bayesian neural network approach to estimating the energy equivalent speed. Accident Analysis and Prevention 38, 248–259.
Shankar, V.N., Albin, R.B., Milton, J.C., Mannering, F.L., 1998. Evaluation of median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effect negative binomial model. Transportation Research Record 1635, 44–48.
Shankar, V.N., Mannering, F.L., Barfield, W., 1995. Effect of roadway geometric and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention 27, 371–389.
Shankar, V.N., Mannering, F., 1996. An exploratory multinomial Logit analysis of single-vehicle motorcycle accident severity. Journal of Safety Research 27 (3), 183–194.
Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Neberagal, M.B., 2003. Modeling crashes involving pedestrians and motorized traffic. Safety Science 41 (7), 627–640.
Simoncic, M., 2001. Road fatalities in Slovenia involving a pedestrian, cyclist or motorcyclist and a car. Accident Analysis and Prevention 33, 147–156.
Speigelhalter, D.J., Best, N.G., Carlin, B.P., Linde, V.D., 2003a. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, Series B 64 (4), 583–616.
Spiegelhalter, D.J., Thomas, A., Best, N.G., Lunn, D., 2003b. WinBUGS version 1.4.1 User Manual. MRC Biostatistics Unit, Cambridge, UK.
Wang, K., Lee, A.H., Yau, K.K.W., Philip, J.W., 2003. A bivariate zero-inflated Poisson regression model to analyze occupational injuries. Accident Analysis and Prevention 35, 625–629.
Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38 (6), 1137–1150.
Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis. Accident Analysis and Prevention 39, 922–933.
Yang, C., MacNab, 2003. A Bayesian hierarchical model for accident and injury surveillance. Accident Analysis and Prevention 35, 91–102.
Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safety Science 47 (3), 443–452.