Accident Analysis and Prevention 129 (2019) 277–288

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Incorporating behavioral variables into crash count prediction by severity: A multivariate multiple risk source approach

Mohammad Razaur Rahman Shaon[a], Xiao Qin[a,*], Amir Pooyan Afghari[b], Simon Washington[c], Md Mazharul Haque[b]

[a] Department of Civil and Environmental Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI, 53201, USA
[b] School of Civil Engineering and Built Environment, Queensland University of Technology, Brisbane, QLD, 4001, Australia
[c] School of Civil Engineering, Faculty of Engineering, Architecture, and Information Technology, The University of Queensland, St Lucia, QLD, 4072, Australia

ARTICLE INFO

Keywords:
Crash causation mechanism
Multiple risk generating process
Driver behavioral factors
Multivariate analysis

ABSTRACT

The frequency and severity of traffic crashes have commonly been used as indicators of crash risk on transport networks. Comprehensive modeling of crash risk should account for both frequency and injury severity—capturing both the extent and intensity of transport risk for designing effective safety improvement programs. Previous research has revealed that crashes are correlated across severity categories because of the combined influence of risk factors, observed or unobserved. Moreover, crashes are the outcomes of a multitude of factors related to roadway design, traffic operations, pavement conditions, driver behavior, human factors, and environmental characteristics, or in more general terms: factors reflect both engineering and non-engineering risk sources. Perhaps not surprisingly, engineering risk sources have dominated the list of variables in the mainstream modeling of crashes whereas non-engineering sources, in particular, behavioral factors, are crucially omitted. It is plausible to assume that crash contributing factors from the same risk source affect crashes in a similar manner, but their influences vary across different risk sources. Conventional crash frequency modeling hypothesizes that the total crash count at any roadway site is well-approximated by a single risk source to which several explanatory variables contribute collaboratively. The conventional formulation is not capable of accounting for variations between risk sources; therefore, is unable to discriminate distinct impacts between engineering variables and non-engineering variables. To address this shortcoming, this study contributes to the development of multivariate multiple risk source regression, a robust modeling technique to model crash frequency and severity simultaneously.

The multivariate multiple risk source regression method applied in this study can effectively capture the correlation between severity levels of crash counts while identifyinging the varying effects of crash contributing factors originated from distinct sources. Using crashes on Wisconsin rural two-lane highways, two risk sources – engineering and behavioral – were employed to develop proposed models. The modeling results were compared with a single equation negative binomial (NB) model, and a univariate multiple risk source model. The results show that the multivariate multiple risk source model significantly outperforms the other models in terms of statistical fit across several measures. The study demonstrates a unique approach to explicitly incorporating behavioral factors into crash prediction models while taking crash severity into consideration. More importantly, the parameter estimates provide more insight into the distinct sources of crash risk, which can be used to further inform safety practitioners and guide roadway improvement programs.

1. Introduction

The frequency and severity of traffic crashes have been largely used in transportation safety as two indicators of crash risk (Washington et al., 2018). These two indicators form the overall risk at transport network locations and thus mitigating one without paying attention to the other one is incomplete and can be wrong. While highway agencies and departments of transportation aim to reduce the frequency and severity of traffic crashes, highway safety improvement programs are primarily focused on preventing severe and fatal crashes, as the cost per

* Corresponding author.
E-mail addresses: mshaon@uwm.edu (M.R. Rahman Shaon), qinx@uwm.edu (X. Qin), a.afghari@qut.edu.au (A.P. Afghari), hos@civil.uq.edu.au (S. Washington), m1.haque@qut.edu.au (M.M. Haque).

https://doi.org/10.1016/j.aap.2019.05.010
Received 1 March 2019; Received in revised form 2 May 2019; Accepted 10 May 2019
Available online 06 June 2019
0001-4575/

person of a fatal crash is almost 250 times higher than a non-injury crash (Harmon et al., 2018) –not to mention the number of persons involved in the crash and the total number of crashes across the transport network. The importance of crash severity levels is even more acute considering the collective social cost of crashes in the society which is substantially higher than the individual social costs (Tay, 2002). As a result, considering crash severity in conjunction with crash frequency is paramount in crash modeling and identification of high crash risk sites.

Crash prediction models have been widely used to study crash frequency and investigate crash contributing factors at transport network locations. These models have been traditionally applied to the crash frequencies aggregated over different crash severity levels. To incorporate crash severity into crash prediction models, numerous studies have modeled crash frequency of a particular severity level at a specific intersection or segment (Abdel-Aty and Radwan, 2000; Hauer et al., 1988; Lord and Persaud, 2000; Poch and Mannering, 1996; Shankar et al., 1995; Lyon et al., 2003; Qin et al., 2004; Abdel-Aty et al., 2005; Tarko et al., 2008; Geedipally et al., 2010; Geedipally and Lord, 2010). However, the independent (i.e., univariate) modeling of crash frequency with various injury severities may not be accurate because crash frequencies may be correlated across different crash severities due to the presence of shared effects from engineering, spatial, and unobserved factors (Wang et al., 2017). Neglecting such correlations may lead to biased parameter estimates and inaccurate inferences about crash contributing factors (Ma et al., 2008; Mannering and Bhat, 2014; Serhiyenko et al., 2016). Empirical evidence has shown that multivariate crash frequency models (e.g., multivariate Poisson lognormal model) can provide better predictive accuracy than its univariate counterparts (Ma and Kockelman, 2006; Wang et al., 2017). Hence, multivariate models have gained popularity, as they can model crash counts of different severities simultaneously and explore the effects of covariates in a more accurate fashion.

Research has established that traffic crashes are the results of chains of causal events that arise from a multitude of contributing factors associated with roadway design, traffic operations, pavement conditions, driver behavior, human factors, and environmental factors. These factors do not necessarily contribute equally to crashes at a site, though; therefore, it may be more plausible to consider traffic crashes at every site as the results of multiple risk sources, with each risk source playing either a primary or supporting role. However, conventional crash frequency models treat the crash count at a roadway site as the outcomes of a single risk source by using a single predictive equation estimated with Poisson or Negative Binomial (NB) distribution. While these single equation models are statistically sound and practically useful, their results may yield biased parameter estimates due to issues related with data overdispersion[1] (Shaon and Qin, 2016; Shaon et al., 2018b; Shirazi et al., 2016; Zou et al., 2015). Furthermore, single-equation models are incapable of assuming that crashes may have various risk sources, which could result in data heterogeneity. Not until recently have researchers developed the multiple risk source regression model to distinguish the distinct sources of crash contributing factors (Afghari et al., 2016; Washington and Haque, 2013). Multiple risk source regression modeling is a reasonable alternative to single equation predictive models for predicting risk-level crashes, considering that the contribution of explanatory variables originated from distinct risk sources to the outcome (i.e., predicted crash count at a site) may change. The feasibility of using a generalized structure for modeling crashes by multiple sources of risk has been investigated, and the models have been developed for univariate crash prediction (e.g., total crashes) (Afghari et al., 2016; Washington and Haque, 2013; Afghari et al., 2018b). However, these models ignored the possible correlation across crash counts of different crash severity levels, and thus the parameter estimates are prone to bias. Therefore, there is a need to incorporate crash severity into multiple risk source modeling of crashes.

Significant amount of research has been devoted to identifying and quantifying the effect of contributing factors on crash occurrence. Crash risk originated from driver behavior has been recognized as major crash contributors in highway safety literature (NHTSA, 2008; Rumar, 1985; Sabey and Staughton, 1975; Shaon et al., 2018a). Albeit of universal acceptance, incorporation of contributing factors originated from behavioral risk source into crash frequency modeling is limited due to data unavailability. There is no established method available to collect driver behavior related variables at crash sites. Alcohol-impaired, drug-impaired driving, distraction and speeding behaviors are frequently identified as contributing factors to crash occurrence (Rumar, 1985; Sabey and Staughton, 1975). The absence of these important pieces of behavioral information in crash data can cause unobserved heterogeneity and modeling result can yield biased parameter estimates (Mannering et al., 2016).

This study extends the idea of using a multiple risk source structure to develop a multivariate multiple risk source methodological approach to estimate both crash counts and severity, simultaneously. Similar to multivariate crash prediction models, it is hypothesized that the multivariate multiple risk source modeling approach will provide improved accuracy than a univariate model because it considers the correlation between crash counts of different severities and accommodates for unobserved heterogeneity which could result from the omission of multiple risk sources in modeling equations. In regard of multiple risk sources, the two risk sources - engineering and behavioral risk source related crash contributing factors were explored in the proposed model. Considering data limitation related to driver behaviors, a few behavioral variables that are uncorrelated with engineering factors and solely originated from a different source (e.g., physical and psychological characteristics) were incorporated as behavioral variables in this study. Furthermore, the risk-level predicted crashes from multiple risk source modeling could be useful in identifying sites for safety improvements and developing targeted and effective safety countermeasures.

2. Literature review

A large number of studies in the road safety literature estimated crash frequency by crash severity to evaluate the safety implications of contributing factors (Abdel-Aty and Radwan, 2000; Hauer et al., 1988; Lord and Persaud, 2000; Poch and Mannering, 1996; Shankar et al., 1995; Lyon et al., 2003; Qin et al., 2004; Abdel-Aty et al., 2005; Tarko et al., 2008; Geedipally et al., 2010; Geedipally and Lord, 2010; Qin et al., 2016). In this context, several previous studies noted that crash counts across different injury severity are likely to be correlated (Ma et al., 2008; Wang et al., 2017). Therefore, incorporating the correlation between crash counts of different injury severities is an important practice when estimating crash counts and severities, simultaneously. Such correlation can be effectively handled by multivariate regression models (Chiou and Fu, 2015; Pei et al., 2011; Wang et al., 2011; Ye et al., 2009; Zeng et al., 2016). Please refer to Mannering and Bhat (2014) and Mannering et al. (2016) for comprehensive list of literature on crash data modeling. Both multivariate Poisson and multivariate Poisson lognormal models are popular choices, but the latter is more effective for overdispersed data (Chib and Winkelmann, 2001; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007; Ye et al., 2008, 2009). The covariance structure used in the multivariate Poisson lognormal model allows for estimating model parameters with smaller standard errors while maintaining the core strength of the Poisson distribution. Studies have shown that this model of crashes outperforms

[1] Crash data are often characterized by the existence of a large sample variance compared with the sample mean. In a statistical term, the sample data is over-dispersed when the variance is greater than the mean. Data over-dispersion is often caused by unobserved data heterogeneity due to unobserved, unavailable, or unmeasurable variables that are important to explain model responses.

the univariate models in terms of statistical fit (Chib and Winkelmann, 2001; Ma et al., 2008; Park and Lord, 2007).

Substantial effort has been devoted to identify primary risk factors contributing to crashes at a site and quantify their effects on crash occurrences (Bahar et al., 2004; Garber and Ehrhart, 2000; Lee and Mannering, 2002; Miaou et al., 1992; Milton and Mannering, 1998; Persaud, 2001; Shaon, 2015; Tarko and Kanodia, 2004). Roadway design factors and traffic operational characteristics dominate this list of variables in the crash data modeling related literature (Abdel-Aty and Radwan, 2000; Anastasopoulos and Mannering, 2009; Chin and Quddus, 2003; El-Basyouny and Sayed, 2009; Fitzpatrick et al., 2010; Geedipally et al., 2012; Islam et al., 2014a, 2014b; Mitra and Washington, 2012; Montella and Imbriani, 2015; Oh et al., 2004; Qin et al., 2004, 2016; Quddus et al., 2001; Shankar et al., 1995; Shaon and Qin, 2016; Shaon et al., 2018b). The findings show that roadway geometric features such as lane width, shoulder width, and horizontal and vertical alignments are statistically significant in their correlation with crash occurrence. In addition, traffic operational variables such as Average Annual Daily Traffic (AADT), truck traffic and posted speed limit have been shown to have a significant influence on safety. Since these variables represent the engineering principles and practices in highway design and capacity analysis, they are often referred to as engineering variables. Understanding the safety performance of engineering variables is instrumental in identifying effective engineering solutions. Most proven safety countermeasures involve the modification and improvement of roadway and roadside design features as well as controlling traffic features on specific roadway sites (FHWA, 2017). The prevalence of studying these variables is also due to the availability and quality of data, as transportation agencies are required to collect and maintain them for highway performance monitoring, planning and program development, design, and operations, as well as maintenance activities.

Driver behavior variables, however, are not readily available even though they are considered universally as a major contributor to crashes (Afghari et al., 2018b; NHTSA, 2008; Rumar, 1985; Sabey and Staughton, 1975; Washington and Haque, 2013, Shaon et al., 2018a, 2018b). Standard procedures for collecting driver behavior data do not exist, as highway agencies are not obligated to collect such information for safety management systems. The behavior data collected from crash data represent a very small portion of driver activities in traffic events. The most relevant source for obtaining this information is perhaps the crash report where police officers may record information regarding driver's condition and his or her opinion of the possible contributing factors. This type of information, albeit extremely valuable, is often incomplete, underreported, and inconsistent. Reports show that risky driving behaviors such as distracted driving, impaired driving, speeding are often identified as major contributors to crash occurrences (Box, 2009; NHTSA, 2010; Redelmeier and Tibshirani, 1997). Such information, however, is usually available only for severe crashes in which thorough investigations are performed. One of the most exhaustive studies conducted so far is the National Motor Vehicle Crash Causation Survey administered by National Highway Traffic Safety Administration for which a group of experts reviewed a nationally representative sample of 5471 crashes during a 2.5-year period. Commonly used roadway or environmental conditions were found as the primary reason for only 135 crashes from this study – a mere 2.5 percent, which shows the necessity of incorporating driver behaviors into crash prediction models.

Although site-specific driver behavior variables may not be readily available, behavioral variables are sometimes collected at a larger geographic scale (e.g., county) to analyze the physical and psychological status of a community. Driver behavior is determined by drivers' commitment to the values and beliefs in safety, which is influenced by attitudes, social norms, and perceived risk. Social norms play an important role in driving behavior and risk perception (Carter et al., 2014). For example, some drivers may follow the behaviors of others in their community, regardless of roadway design or site characteristics (Schneider et al., 2018). Societal expectations of acceptable transportation risk can also influence risk-taking behavior (Moeckli and Lee, 2007). Proxy variables can be used to substitute driver risk factors in crash count modeling for measuring the effect of behavioral risk on crash occurrence, including total number of speeding offenses (Afghari et al., 2018a, b), operating while intoxicated citation count (Smith, 2000; Nagle, 2012), drug arrest count (Asbridge et al., 2012; Compton et al., 2009; Walsh et al., 2008), violent crime rate (Ando et al., 2018; Carter and Piza, 2017; Weiss, 2013), and liquor license rate (Lascala et al., 2000). For example, alcohol-impaired driving related fatalities are almost one-third of all fatalities that occur in the USA. Owusu-Ababio and Feng (2006) found liquor license related and liquor arrest related variables are significant in predicting alcohol-related crashes in Wisconsin. Drug use also affects the driver decision making capabilities. According to a meta-analysis of nine roadway safety studies conducted by Asbridge et al. (2012), the authors noted driving under the influence of cannabis was associated with a significantly increased risk of motor vehicle collisions compared with unimpaired driving with an odds ratio of 1.92. Authors also found that collision risk estimates are higher for fatal crashes for drugged driving. Although community safety and crashes are usually discussed separately, crime and crashes can interact with each other due to common factors. Ando et al. (2018) noted a positive relation between urban violence and crash count in Toyota city.

Understanding the effects of crash data generating mechanisms provides useful information about the sources of variance in crash data. Peng et al. (2014) used a generalized waring model to differentiate between different distinct sources of crash heterogeneity using different variance terms. The authors separated the observed variability into random errors; the proneness, which refers to the internal differences between observations, and the liability, which refers to the variance caused by unobserved exogenous variables. This new modeling structure has a better performance compared to the NB model and showed that a crash may originate from different sources through different processes, which contributes to additional variances. Explicitly incorporating the heterogeneous sources into crash modeling can be challenging, but one logical approach is to group contributing factors by risk source (e.g., environmental factors, roadway geometric design features, driver behavior) and assume that variables within the same risk source affect crash occurrence in a similar manner, but that sources contribute to crashes to different extents. This assumption resonates with what Lord, Washington, and Ivan have noted in their seminal work that concludes that over-dispersion arises from the actual nature of the crash process (Lord et al., 2005). In addition, crash risk generated from multiple risk source may have endogenous relation. For example, sometimes driver behaviors may be endogenous with roadway design elements. Several studies in the literature have discussed that the location-specific engineering factors such as roadway geometry promotes driver behavior (Afghari et al., 2018a; Chapman and Noyce, 2014). Afghari et al. (2018a) noted that less curved road segments or location specific rural environment result in more severe speeding among drivers. The NB distribution is therefore limited in that it assumes that only one underlying process affects the likelihood of crash frequency (Shankar et al., 1997).

Recently, researchers introduced a multiple risk generating process regression model in which crashes at a given site are assumed to have originated from distinct sources of risk, and their relationships are represented by multiple equations (Afghari et al., 2016; Afghari et al., 2018b; Washington and Haque, 2013). The authors argued that the single risk source assumption in traditional crash prediction models is statistically sound but cannot sufficiently address unobserved heterogeneity. The application of single risk source traditional models in blackspot identification centers on the assumption that operational causal factors such as roadway geometry or traffic factors operate in a single chain to form the total crash count. The result is that other risk sources such as driver behavioral factors, which are the cause of more than 50% of crashes, are neglected (NHTSA, 2008; Washington and Haque, 2013, Shaon et al., 2018a, b). A single risk source may attribute behavioral factors to operational factors, resulting in biased parameter estimates and erroneous model prediction. A multiple risk source regression model can add flexibility to estimate crashes based on their originating risk sources and provide meaningful parameter estimates. The empirical evidence shows that assumption of multiple risk sources in modeling equation provide improved model fit and can account for unobserved heterogeneity that results from ignoring risk sources (Afghari et al., 2016, 2018b).

3. Research hypothesis

To estimate crash counts and injury severity, simultaneously, a multivariate framework in regression modeling is needed to accommodate the correlation between crash counts of different injury severities. Equally important part in crash data modeling is to distinguish between the sources of crash risk. In spite of the importance of behavioral factors, a limited amount of research has directly incorporated these risks into crash prediction modeling because of the lack of site-specific driver behavior-related factors. Alternatively, the effect of site-specific variables, in combination with the influence of a broader safety culture represented by driver behavior, would provide many helpful insights.

The hypothesis of the methodological approach in this study is described below, and includes the theoretical support for this type of crash modeling:

- This study hypothesizes that a single risk source model (e.g., Poisson, NB) cannot sufficiently account for unobserved heterogeneity in crash data. Considering multiple underlying risk sources in crash data modeling may allow researchers to account for unobserved heterogeneity at each risk-generating source.
- This study hypothesizes that risk sources can be categorized based on distinct sources of data and their physical meaning. Two distinct risk sources, engineering, and behavioral risk sources are considered which simultaneously contribute to the crash occurrence on a roadway segment.
- Crash counts of different severities are correlated. Considering the correlation of crash severities in the modeling structure allows for the simultaneous estimation of crash frequency and severity and thus, reduces bias in the estimated model parameters. The multivariate structure is considered for two injury severities in this study: injury crashes and non-injury crashes.

Based on the above-mentioned hypothesis, the unstructured covariance matrix is used to define the correlation between injury severity levels, which contributes to the estimation of more precise model parameters. Multiple risk sources are considered to have varying contributions to crashes of all severities at each site and across sites. Site-specific risk-level weights (also vary between injury severity) are used to generate multiple proportions of total crashes. A bivariate (e.g., two risk sources) random error term at each risk level is used to account for unobserved heterogeneity and define the correlation between risk sources.

4. Methodology

Assuming observed crash count Y_i at location i, summed across underlying risk sources j, it can be hypothesized that each risk source is responsible for contributing to a proportion of the total observed crashes which are unobserved or latent at the crash location. To determine the latent probabilities of unobserved crash counts from different risk sources, let's assume the total observed crash count follows a Poisson distribution with a total predicted mean μ_i:

$$Y_{i\sim} \sim Poisson(\mu_i) \tag{1}$$

A latent mixture modeling approach can be used to link multiple risk-generating sources with the mean of Poisson distribution. The latent mixture approach requires the decomposition of the mean function of the Poisson distribution (μ_i) into multiple mixture components (Afghari et al., 2016; Afghari et al., 2018a, 2018b):

$$\mu_i = \sum_{j=1}^{J} \mu_{ji}\ and\ \mu_{ji} = w_{ji}\mu_i \tag{2}$$

Where, $\sum_{j=1}^{J} w_{ji} = 1$ and w_{ji} is the proportion (or weight) of the predicted crash count at site i attributed from latent risk source j and j is the total number of underlying risk sources. Assuming exponential functions for the decomposed means of the Poisson distribution, each of the above-mentioned predicted means is a function of a variety of contributing factors associated with unique risk sources:

$$\mu_{ji} = \alpha_{j0} F_i^{\alpha_1} exp\left(\sum \beta_j X_{ji} + \varepsilon_{ji}\right) \tag{3}$$

Where,
F_i = measure of exposure (shared between risk sources),
X_{ji} = explanatory variables for risk source j at site i,
α_j, β_j = estimated regression parameters,
ε_{ji} = model errors independent of all explanatory variables.

To account for unobserved heterogeneity arising from over-dispersion, error terms (ε_{ji}) are allowed to vary across observations. In addition, to account for the correlation between the underlying risk sources, the error terms are defined to follow a Multivariate Normal distribution which can be constructed as follows:

$$\boldsymbol{\varepsilon}_{\mathbf{i}} \sim Multivariate\ Normal\ (0,\ \boldsymbol{\Sigma}_{\mathbf{R}}) \tag{4}$$

Where,

$$\boldsymbol{\varepsilon}_{\mathbf{i}} = [\varepsilon_{1i}\ \varepsilon_{2i} ... \varepsilon_{Ji}]$$

$$and\ \boldsymbol{\Sigma}_{\mathbf{R}} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots\dots & \sigma_{1J}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots\dots & \sigma_{2J}^2 \\ \dots & \dots & \dots\dots & \dots \\ \sigma_{J1}^2 & \sigma_{K2}^2 & \dots\dots & \sigma_{JJ}^2 \end{bmatrix}$$

Please note that these error terms at the risk source level can also account for unobserved and/or unavailable factors that may have contributed to crash occurrence. The univariate risk source regression model, however, does not distinguish injury crashes from non-injury crashes. Thus, a multivariate modeling approach is needed for such a distinction where crashes by injury severity are modeled simultaneously. In a set of crash data at n roadway segments, let assume crashes are classified into K categories which represent K crash severities. Let $Y_{ki} = (y_{1i}, y_{2i}, \dots\dots\dots, y_{Ki})'$ be a K-dimensional vector that denotes the total crash count at i-th ($i = 1, 2, \dots, n$) roadway segment that belongs to k-th (k = 1, 2,….,K) injury severity. Assuming crash counts by crash severity follows the Poisson distribution with mean μ_{ki} for k = 1, 2, …., K and following a similar crash generating mechanism, crashes in each severity category are generated from multiple risk sources which are summed to obtain the total crash count at a location. The regression equation can be constructed as follows:

$$\begin{aligned} ln(\mu_{ki}) &= \ln(\mu_{ki}^{*}) + \varepsilon_{ki} \\ \mu_{ki}^{*} &= \sum_{j=1}^{J} \mu_{jki} \\ \mu_{jki} &= w_{jki}\mu_{ki} \end{aligned} \tag{5}$$

Where, w_{jki} is the proportion (or weight) of the total predicted crash counts for crash severity k at site i attributed from latent risk source j, and $\sum_{j=1}^{J} w_{jki} = 1$. The new error term $\boldsymbol{\varepsilon_{ki}}$ denotes the random effect which is uncorrelated with the explanatory variables and accounts for the unobserved heterogeneity arising from different crash severity levels. This new error term is also assumed to be multivariate normally distributed across crash counts of different severity levels. Let's assume $\boldsymbol{\varepsilon_{ki}} = (\varepsilon_{1i}, \varepsilon_{2i}, .. ,\varepsilon_{Ki})'$ represents a vector of random effects at each location i and it follows a K-dimensional normal distribution:

$$\boldsymbol{\varepsilon_{ki}} \sim Multivariate\ Normal\ (0,\ \boldsymbol{\Sigma_M})$$

Where, **0** is a K-dimensional zero vector and $\boldsymbol{\Sigma_M}$ is a J × J variance-covariance matrix. Following the above specification of the error term, it is equivalent to $\exp(\boldsymbol{\varepsilon_{ki}}) \sim Lognormal\ (0,\ \boldsymbol{\Sigma_M})$. The variance-covariance matrix Σ_M accounts for unstructured error and unobserved heterogeneous effects and can be formulated as follows:

$$\Sigma_M = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \ldots\ldots & \sigma_{1K}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \ldots\ldots & \sigma_{21}^2 \\ \ldots & \ldots & \ldots\ldots & \ldots \\ \sigma_{K1}^2 & \sigma_{K2}^2 & \ldots\ldots & \sigma_{KK}^2 \end{bmatrix} \tag{6}$$

The diagonal elements σ_{kk}^2 of the variance-covariance matrix represents the heterogeneous variance of ε_{ki}, and the off-diagonal elements σ_{rs}^2 represents the heterogeneous covariance between ε_{ri} and ε_{si} where $r \neq s$.

Following a similar concept of univariate modeling, the mean response from each risk source λ_{jti} can be expressed as follows:

$$\mu_{jki} = \alpha_{jk0} F_i^{\alpha_{t1}} exp(\sum \beta_{jk} X_{ji} + \varepsilon_{jki}) \tag{7}$$

Where, ε_{jki} is another error term used to account for unobserved heterogeneity and correlated between underlying risk-generating sources within each crash severity. This indicates the incorporation of additional K number of errors into the modeling structure. For example, let's assume we have crash data from 2 crash severity types (e.g., injury and non-injury) and there are 2 underlying risk sources (e.g., engineering and behavioral). Under multivariate multiple risk source modeling, $\varepsilon_{1i} = [\varepsilon_{11i}, \varepsilon_{12i}]$ and $\varepsilon_{2i} = [\varepsilon_{21i}, \varepsilon_{22i}]$ will account for unobserved heterogeneity and correlations between underlying risk-generating sources for crash severity 1 and crash severity 2, respectively. Because of multivariate modeling, there will be another error term $\varepsilon_i = [\varepsilon_{1i}, \varepsilon_{2i}]$ to account for unstructured errors and unobserved heterogeneous effects for each crash severity. The expected mean, variance, and covariance for multivariate multiple risk source regression model can be expressed as follows:

$$E[Y_{ki}] = \mu_{ki}^* \times exp(\frac{\sigma_{jj}}{2}) \tag{8}$$

$$Var[Y_{ki}] = E[Y_{ki}] + (E[Y_{ti}])^2 \times [\exp(\sigma_{ii}) - 1] \tag{9}$$

$$Cov[Y_{ri}, Y_{si}] = E[Y_{ri}] \times [\exp(\sigma_{rs}) - 1] \times E[Y_{si}] \tag{10}$$

The multivariate crash data modeling using Poisson-lognormal mixture can accommodate overdispersion in the data. From Eqs. (8) and (9), it can be noted that $Var[Y_{ki}] > E[Y_{ki}]$ since the diagonal elements of Σ, $e.g.\ \sigma_{kk}^2 > 0$. Additionally, the multivariate structure can incorporate the correlation among the components in a response vector as described in Eq. (10).

The proposed multivariate and univariate multiple risk source regression models are formulated and estimated in the Bayesian framework using OpenBUGS (Spiegelhalter et al., 2007). In the Bayesian framework, the regression parameters are estimated by maximizing the posterior which is a combination of the likelihood function and the defined prior information. It is necessary to specify a prior distribution for the parameters to obtain the Bayesian estimate. Prior distributions are meant to reflect prior knowledge about the parameters of interest. In the absence of solid prior information, uninformative priors can be assumed to estimate both univariate and multivariate multiple risk source models in the Bayesian framework:

$$\beta_{jk} \sim Normal(0,100)$$

$$w_{jki} \sim Uniform(a,\ b)$$

$$\Sigma_R^{-1} \sim Wishart(I_J,\ J)$$

$$\Sigma_M^{-1} \sim Wishart(I_K,\ K)$$

Where, I_J and I_K represents J × J and K × K dimensional identity matrix, respectively. Defining the value for lower limit a and upper limit b for the prior information of a risk-level weight should be approached with caution. Prior knowledge, if available, can be used to define these values (Afghari et al., 2016). The Markov chain Monte Carlo (MCMC) can also suffer from poor mixing, and the effective number of parameters of estimated models can be negative (it is guaranteed to be positive for properly defined and converged models) if there is a conflict between data and prior information. Several preliminary models were developed with different prior values for a and b to obtain good mixing between Monte Carlo chains and model convergence. The model performance of preliminary models was also compared to choose the optimal value. The final model is estimated with a = 0.45 and b = 0.95.

In addition, marginal effects are used to determine the impact of each covariate on the expected mean value of the dependent variable. The marginal effect represents the effect of a unit change in the independent variable on the expected mean of the dependent variable. For a multiple risk source model, the marginal effect for each explanatory variable can be estimated using the following equation (Washington et al., 2010):

$$E_{x_{jik}}^{\mu_i} = \frac{\delta\mu_i}{\mu_i} \times \frac{x_{jik}}{\delta x_{jik}} = w_j \beta_{jk} x_{jik} \tag{11}$$

In Eq. (11), the risk-level weights (w_j) will influence the estimated marginal effect of an explanatory variable in j-th risk source.

5. Data description

To empirically test the proposed methodology, the multivariate multiple risk source model is applied to traffic crashes along rural two-lane highway segments in Wisconsin, United States. Crash data for this network is available in "KABCO" scale containing a total count of crashes that occurred on Wisconsin state highways between 2011 and 2015. The KABCO injury codes presented in the dataset were consolidated into two levels in this study – injury crashes (K, A, B, and C) and non-injury crashes (O) to ensure that a sufficient number of observations was available in each crash severity level. A similar approach has been used by other researchers to ensure sufficient sample size for model estimation (Milton et al., 2008; Islam et al., 2014a, b; Uddin and Huynh, 2018):

Crash contributing factors were collected for two distinct risk sources (J = 2) including engineering and behavioral factors. The factors within engineering risk source include typical roadway geometric factors and traffic features for rural two-lane highways which were collected from MetaManager, a data management system developed and maintained by the Wisconsin Department of Transportation (St Clair, 2001). Typical roadway geometry related variables such as segment length, lane width, and shoulder width variables were collected from roadway inventory data table. The percent passing and posted speed limit variables were collected from the mobility data table. Please note that the segmentation of Wisconsin two-lane highways does not involve horizontal and vertical curves. This means the highway is not segmented at the starting or ending of a horizontal or vertical curve. To provide information on curves, no passing zone is generated from MARKINGview, an asset management tool used to capture and maintain traffic marking information and location of no passing zones. No passing zones are usually marked with a solid yellow line placed on hills or curves where you cannot see far enough ahead to pass safely. In the mobility data table, no passing zones are expressed in percentage of segment length. AADT and percent of truck on each segment were collected as traffic-related features for model development in this study.

To conform to the study hypothesis of multiple risk sources, the behavioral factors used in this study are solely generated from a different risk source than engineering risk factors. The behavioral factors used in this study were collected from the Uniform Crime Reporting (UCR) program in Wisconsin. The UCR program provides crime and arrest data from local law enforcement agencies to the Federal Bureau of Investigation and Bureau of Justice Information and Analysis (FBI, 2019). The collected behavioral factors include the operating a motor vehicle while intoxicated (OWI) rate, drug related arrest rate, and violent crime rates for each county between 2013 and 2014. In this context, the rate is defined as the total count per 10,000 people within a defined geographic area (e.g., county). Collected variables were averaged for each county over a two-year period to represent the behavior risk source. Community-level factors that heighten the risk of experiencing problems with alcohol include the per capita number of alcohol outlets in a community. The liquor license rate, which can be defined as the number of liquor outlet licenses per 500 people, was collected from the Wisconsin Department of Revenue for 2013 and 2014, and the average value was used (Wisconsin Department of Revenu, 2019).

Alcohol impairment of driving skill has been identified as a major traffic safety problem since early 20th century (Blomberg et al., 2005). National Survey on Drug Use and Health's (NSDUH) "State Estimates of Drunk and Drugged Driving" report released in 2012 indicates the prevalence of alcohol-impaired driving in Wisconsin are among the highest in the nation (Substance Abuse and Mental Health Services Administration, 2012). Besides alcohol impairment, drug-impaired driving has recently started raising government and public concerns in the USA as well as other countries (Asbridge et al., 2012; Walsh et al., 2008). In Wisconsin, the drug law violations are defined as the violation of laws prohibiting the production, distribution, and/or use of certain controlled substances and the equipment or devices utilized in their preparation and/or use. This includes the unlawful cultivation, manufacture, distribution, sale, purchase, use, possession, transportation, or importation of any controlled drug or narcotic substance. Based on above definition and observed consequences of impaired driving, both OWI and drug arrest rate can be considered as direct measures of crash risks the transportation network is exposed to in a community (e.g., county). Violent crimes used in this study includes murder and non-negligent manslaughter, rape, robbery, and aggravated assault.

The collected behavior variables were then linked to the roadway database using the spatial join tool in ArcMap, a geographic information system software package. The behavior variables were evenly assigned to the roadway segments within counties using their spatial coordinates. This means any roadway segment within a same county has same driver behavior related attribute. This approach has been used by multiple studies in literature to link aggregated level spatial variables such as weather conditions to roadway elements (Chang and Chen, 2005; Yu et al., 2015). Please note that although the driver behavior related variables collected in this study may change from one segment to another, it was hypothesized that use of aggregated level driver-behavior information may provide valuable inference about the impact of risky driving behavior on crash occurrence. The complete rural two-lane highway dataset contains 9605 segments extended over 8669 miles after cleaning for missed observations for target variables and very short segments. A sample of 6000 segments extended over 5400 miles from the complete dataset was used to evaluate the proposed methodology in this study. Table 1 provides the summary statistics of the variables used for this study.

6. Results and discussion

This section of the paper explains the application of the multivariate multiple risk source regression model to estimate crash count and injury severity simultaneously. The models developed for this study were designed to estimate injury and non-injury crashes (K = 2). The performance of the multivariate multiple risk source model was compared with the NB model, a univariate multiple risk source model, for both severity levels to determine whether the multivariate approach of multiple risk source model was theoretically sound and offered improved model performance. A total of two (2) MCMC chains were used to implement all models in the Bayesian framework. Model convergence was obtained through 130,000 iterations, and 30,000 samples were used as burn-in period. The Gelman–Rubin convergence statistics (G–R statistics) were reviewed to verify the model convergence (i.e., when the G–R statistic is less than 1.2) (Mitra and Washington, 2007).

Table 2 summarizes the modeling results for both injury and non-injury crashes in the study dataset. Table 2 shows that for the NB model, the estimated 95 percent posterior credible intervals for all coefficients in both injury and non-injury crashes did not include zero; hence, all coefficients are statistically significant at a 5 percent significance level. Though the drug arrest rate variable was statistically significant in predicting no injury crashes for the NB model, the variable was not statistically significant at 5 percent with both univariate and multivariate multiple risk source models. Afghari et al. (2016) found similar results when comparing the NB and multiple risk source model. It is found that six out of nine explanatory variables were not statistically significant in the multiple risk source model, but they were statistically significant in the single source NB model. The drug arrest rate was statistically significant in predicting injury crashes in all models, indicating driving under the influence of drug results in more injury crashes. All other posterior mean estimates of explanatory variables were statistically significant at a 5 percent significance level in both univariate and multivariate multiple risk source models.

Note that the posterior mean of the estimated parameters of explanatory variables cannot be directly compared between single risk and multiple risk source models, except for the exposure measure, because of the associated risk level weights of each variable. In non-injury crashes, the posterior mean of the estimated parameters for AADT ranges from 0.813 to 0.817 across three modeling approaches, indicating the multiple risk source modeling technique can maintain enough strength to estimate the Poisson mean while considering multiple risk sources. The posterior mean of the estimated parameter for AADT is positive for both severity levels implying that this variable has increasing effect on the number of injury and non-injury crashes.

In multiple risk source models, factors contributing to crashes are separated into two distinct sources: engineering and behavioral factors. The mean of posterior parameter estimate of risk-level weights indicates that on average, 70% of both injury and no injury crashes occur due to the engineering risk source, whereas behavioral risks contribute to 30% of the injury and no injury crashes in both univariate and multivariate multiple risk source models. The statistically significant covariates are similar across all models with regard to the engineering risk source for both injury and non-injury crashes. All parameter-mean estimates for explanatory variables in the engineering risk source have similar signs, indicating similar positive or negative effects on crash risk across modeling alternatives. The parameter-mean estimates for the engineering risk variable are similar when comparing between univariate and multivariate modeling approaches; this indicates that expanding the multiple risk source methodology to a multivariate structure can provide stable parameter estimates. The estimated standard deviation of most mean posterior parameter estimates in the multivariate multiple risk source model is smaller than the estimates for the univariate model; this indicates that more accurate parameters can be estimated using the multivariate structure, as noted in the literature (Park and Lord, 2007).

The posterior parameter estimates for the behavioral risk source variables paint a similar picture as the variables in the engineering risk source. The estimated parameter-mean values are mostly similar in both univariate and multivariate multiple risk source models. For instance, the mean of the posterior parameter estimates for OWI rate yielded a negative impact on crash risk, indicating that both injury and non-injury crash rates tend to decrease with an increase in OWI rate.

Table 1
Summary Statistics of Wisconsin Dataset.

Variables	Description	Mean	Standard Deviation	Minimum	Maximum
Crash Data					
"K + A" Crashes	Serious Injury Crashes	0.353	0.685	0	8
"B + C" Crashes	Minor Injury Crashes	1.511	2.786	0	67
"K + A+B + C" Crashes	Total Injury Crashes	1.864	3.067	0	70
"O" Crashes	Non-Injury Crashes	3.414	5.422	0	116
Exposure					
AADT	Annual Average Daily Traffic	3916.855	3106.853	80	66712
Engineering Risk Source					
Length	Segment Length in miles	0.900	0.423	0.050	2.600
LW	Lane Width in feet	12.020	0.875	9	20
SW	Shoulder Width in feet	6.771	2.797	0	15.500
Truck	Percentage of Heavy Truck	10.797	4.334	0	34.700
Speed	Posted Speed Limit in miles per hour	52.667	6.398	30	70
Passing	Percent Passing	0.464	0.270	0	1
Behavioral Risk Source					
OWI	Operating While Intoxicated citation rate per 10,000 population	46.864	24.653	6.100	335.850
Drug Arrest	Drug Arrest rate per 10,000 population	37.052	20.745	3.750	195.750
Violent Crime	Violent Crime rate per 10,000 population	13.715	11.922	0.700	95.100
Liquor License	Liquor license rate per 500 population	2.220	1.256	0.900	8.300

Table 2
Non-Injury and Injury Crash Modeling Results.

Parameters	Non-Injury NB Model	Univariate Multiple Risk Source Model	Multivariate Multiple Risk Source Model	Injury NB Model	Univariate Multiple Risk Source Model	Multivariate Multiple Risk Source Model
Exposure: AADT	0.814 (0.021)	0.813 (0.022)	0.817 (0.021)	0.780 (0.025)	0.778 (0.025)	0.780 (0.024)
Inverse-Dispersion	0.460 (0.016)			0.490 (0.022)		
Engineering Variables						
Constant	−4.162 (0.058)	−3.680 (0.339)	−3.578 (0.224)	−4.408 (0.195)	−3.901 (0.745)	−3.956 (0.532)
Length	0.930 (0.033)	1.219 (0.082)	1.252 (0.056)	1.010 (0.038)	1.146 (0.049)	1.184 (0.046)
Lane Width	−0.034 (0.014)	−0.045 (0.022)	−0.046 (0.021)	−0.074 (0.017)	−0.103 (0.027)	−0.095 (0.024)
Shoulder Width	−0.042 (0.005)	−0.041 (0.008)	−0.047 (0.007)	−0.040 (0.006)	−0.033 (0.008)	−0.035 (0.007)
Truck Percentage	−0.016 (0.003)	−0.020 (0.004)	−0.021 (0.004)	−0.015 (0.004)	−0.015 (0.004)	−0.016 (0.004)
Speed Limit	−0.023 (0.002)	−0.023 (0.004)	−0.025 (0.003)	−0.017 (0.003)	−0.010 (0.004)	−0.011 (0.004)
Percent Passing	0.215 (0.046)	0.432 (0.073)	0.402 (0.063)	0.154 (0.054)	0.305 (0.067)	0.280 (0.064)
Behavioral Variables						
Constant		−6.716 (1.110)	−6.355 (0.690)		−7.501 (1.294)	−7.010 (0.906)
OWI Rate	−0.004 (0.001)	−0.014 (0.004)	−0.015 (0.004)	−0.004 (0.001)	−0.016 (0.005)	−0.025 (0.005)
Drug Arrest Rate	0.002 (0.001)	***0.006 (0.003)***	***0.005 (0.004)***	0.003 (0.001)	0.017 (0.007)	0.022 (0.007)
Violent Crime Rate	0.007 (0.001)	0.017 (0.003)	0.022 (0.005)	0.009 (0.001)	0.018 (0.003)	0.023 (0.003)
Liquor License Rate	−0.136 (0.117)	−0.773 (0.206)	−0.726 (0.162)	−0.125 (0.014)	−0.985 (0.316)	−1.293 (0.29)
Risk-level Weights						
Engineering Risk		0.700 (0.116)	0.700 (0.132)		0.700 (0.121)	0.700 (0.138)
Behavioral Risk		0.300 (0.116)	0.300 (0.132)		0.300 (0.121)	0.300 (0.138)
Correlation between Risk Sources						
σ_{11}		0.539 (0.159)	0.214 (0.101)		0.524 (0.162)	0.198 (0.091)
σ_{22}		1.133 (0.497)	0.411 (0.240)		1.295 (0.610)	0.456 (0.304)
$\sigma_{12} = \sigma_{21}$		0.652 (0.226)	***−0.154 (0.134)***		0.697 (0.303)	***−0.066 (0.135)***

Note: 1) Parameter estimates presented in bold and italic font is not significant at 5 percent significance level; 2) The estimated standard error of mean parameter estimate is presented in parenthesis.

The mean of the posterior parameters estimated for the liquor license variable for both injury and non-injury crashes suggests a negative impact on crash risk. Both estimates from the data seem counter-intuitive if the OWI rate or the number of liquor licenses is regarded as the positive effect of liquor consumption on driving. OWI rate can be further regarded as a proxy of the number of drunk drivers who are more likely to be involved in a crash than sober ones. Unfortunately, such findings do not necessarily lead to a conclusive explanation as high OWI arrests may suggest intensive enforcement activities or more effective enforcement strategies. A meta-analysis shows that drink-driving checkpoints reduce alcohol-related crashes by 17% at a minimum and all crashes by 10–15% (Erke et al., 2009). In spite of an endeavor to collect information on enforcement, the data were incomplete and inconsistent and not helpful for this study. Another caveat in the UCR dataset is that liquor licenses are not separated by bar, restaurant, and off-premise liquor outlets, as several studies noted that crash risk increases with bar and off-premise liquor outlets but decreases for restaurants with a liquor license (Gruenewald and Johnson, 2010; Treno et al., 2007). Hence, the effects of these behavioral variables on crashes can be revealed and estimated via crash modeling but defining a cause-effect relationship requires additional information.

The mean posterior parameter estimates for the covariance matrix in the univariate multiple risk source model were found statistically significant at 5% significance level. This indicates that the two risk sources considered in the model are distinct and related. With the multivariate modeling approach, the posterior mean estimates of the covariance term between risk sources (mean: -0.154, std. dev. = 0.134) indicate that they are no longer statistically interrelated at a 5% significance level. Based on the posterior density of Σ_M, statistically significant positive correlations ($\sigma_{12} = \sigma_{21} = 0.629$)exist between crash counts at different levels of severity within a segment. The univariate risk source model is a special case of the multivariate multiple risk source model, with off-diagonal elements of Σ_M equal to zero. By incorporating a statistically significant correlation in the modeling structure, the correlation in injury severity counts was incorporated into the model framework.

Based on the above discussion on modeling results, it can be noted that a significant correlation exists between crash counts for different injury severity level. As described in the methodology section, this correlation influences the estimation of model parameters (Park and Lord, 2007). However, the posterior mean of the covariance matrix for the risk-level error term is no longer significant when the correlation between crash counts for different injury severity level is considered. The variance estimates indicate that the risk sources are indeed distinct for both injury and non-injury crashes; this suggests that the statistically significant correlation between engineering and behavioral risk sources can be a statistical artifact resulting from the absence of injury severity in the model. Hence, the multivariate multiple risk source regression model can provide informative parameter inferences with the existence of Σ_M and uncorrelated risk sources. A modified model was estimated with an uncorrelated error structure between risk sources. Modeling results show that parameter estimates for all covariates in both the engineering and behavioral risk sources yielded similar coefficients; thus, the results without correlation structure between risk sources were not presented here.

For the convenience of comparing the effect of individual factors from both engineering risk source and behavior risk source, marginal effects are estimated. The marginal effects of explanatory variables for both PDO and injury crashes are presented in Table 3.

According to Table 3, single risk models usually overestimate the contribution of a specific variable if this factor originates from a dominating source. For majority of the engineering risk factors, the estimated marginal effect is higher with a single risk source NB model than with a multiple risk source model. The estimated marginal effects of behavioral risk variables indicate that the single risk source NB model underestimates the effect of some variables related to the behavioral risk source; this may be why the effects of some variables related to the engineering risk source are overestimated. For example, the estimated marginal effect for OWI citation rate using single source NB model indicates that injury crash counts can decrease by 0.187 unit with a unit increase in OWI citation rate, whereas the multiple risk source model yielded a 0.525 and 0.722 unit increase in injury crash counts with an increase in OWI citation rate for the univariate and multivariate modeling approach, respectively. For injury crashes, the effect of shoulder width and speed limit may be overestimated in single risk models. Moreover, the marginal effect estimates indicate that single risk source models may underestimate the effect of OWI citation rate and liquor license rate. Comparing both posterior mean of parameter estimates and marginal effects, it can be noted that all variables have similar direction (positive or negative) in both single source and multiple source regression models. But the estimated marginal effects are significantly different for variables originated from supporting risk source such as behavioral risk variables in this study. Thus, a caution should be used while interpreting the parameter estimated from single source model if variables used for model development are generated from different risk sources.

7. Prediction accuracy

Table 4 provides the performance comparison for all models based on the Deviance Information Criterion (DIC). The DIC is a widely used GOF statistic for comparing models in a Bayesian framework (Spiegelhalter et al., 2002). The DIC consists of two components: (a) a measure of how well the model fits the data, Dbar ($D(\bar{\theta})$) and (b) a measure of model complexity (pD). Thus, DIC can provide a better comparison between models that are characterized by different complexities. The likelihood of a Bayesian model can be represented by Dbar ($D(\bar{\theta})$) and Dhat ($D(\hat{\theta})$). Dbar is the posterior mean of the deviance, whereas Dhat is a point estimate of the deviance. Mean Absolute Deviance (MAD) was estimated for each model to compare predictive accuracy. MAD can be calculated as follows:

$$MAD = \frac{1}{N}\sum |y_{it} - \hat{y}_{it}| \tag{13}$$

Where, N indicates the number of observations in the dataset.

A comparison of the DIC values between models illustrates that the multivariate multiple risk source regression model with uncorrelated error structure between risk sources performed better than other models. It is evident that excluding the correlation structure will result

Table 3
Average Marginal Effects for Non-Injury and Injury Crashes.

Variables	Non-Injury			Injury		
	Single Source NB Model	Univariate Multiple Risk Source Model	Multivariate Multiple Risk Source Model	Single Source NB Model	Univariate Multiple Risk Source Model	Multivariate Multiple Risk Source Model
Exposure: AADT	6.489	6.481	6.513	6.218	6.202	6.242
Engineering Risk						
Length	0.837	0.768	0.789	0.909	0.722	0.746
Lane Width	−0.408	−0.379	−0.387	−0.889	−0.866	−0.799
Shoulder Width	−0.284	−0.194	−0.223	−0.271	−0.156	−0.166
Truck Percentage	−0.173	−0.151	−0.159	−0.162	−0.113	−0.121
Speed Limit	−1.211	−0.848	−0.922	−0.895	−0.369	−0.405
Percent Passing	0.100	0.140	0.130	0.071	0.099	0.091
Behavioral Risk						
OWI Rate	−0.187	−0.197	−0.211	−0.187	−0.525	−0.722
Drug Arrest Rate	0.074	0.067	0.056	0.111	0.189	0.245
Violent Crime Rate	0.096	0.070	0.091	0.123	0.074	0.095
Liquor License Rate	−0.302	−0.515	−0.483	−0.277	−0.656	−0.861

Table 4
Comparison of Model Performance.

Methodology	Crash Type	$\bar{D}$	$\hat{D}$	DIC	pD	MAD
Single Source NB model	PDO	25040	25030	25060	14.13	2.325
	Injury	19880	19860	19890	13.08	1.413
	Total	**44920**	**44890**	**44950**	**27.21**	**1.869**
Univariate Multiple Risk Source model	PDO	20640	18900	22370	1734	2.286
	Injury	16910	15440	18370	1464	1.371
	Total	**37550**	**34340**	**40740**	**3198**	**1.829**
Multivariate Multiple Risk Source model w/ Correlated Error Structure	PDO	20360	18960	21750	1398	2.280
	Injury	16460	14910	18020	1308	1.353
	Total	**36820**	**33870**	**39770**	**2706**	**1.817**
Multivariate Multiple Risk Source model w/o Correlated Error Structure	PDO	20106	19554	20660	552.4	2.271
	Injury	16142	15057	16862	718.8	1.322
	Total	**36108**	**34766**	**37382**	**1271.2**	**1.797**

in a smaller effective number of parameter (pD) which will influence the estimation of the DIC value. The Dbar estimate indicates that the posterior mean of deviance is the smallest for the multivariate multiple risk source model without a correlated error structure compared with all other models. There is a significant improvement in DIC value with the multivariate modeling approach compared with the univariate multiple risk source model. The MAD estimates also indicate that multivariate risk source regression models can better predict both PDO and injury crashes compared with other models in this study.

8. Practical implications

One major benefit of the multiple risk source model over a single risk model is that risk-level predicted crash counts can be obtained from the former model which is not possible with the latter model. In the literature, driver error and engineering risk factors are identified as two major sources for crash occurrences (Shaon et al., 2018a, b). In Wisconsin, detailed crash report for each crash occurred on state trunk network is documented in Wisconsin Motor Vehicle Accident Reporting Form 4000 (MV4000) by investigating police officer(s) (WisDOT, 2019; Parker and Tao, 2006). The crash report for each crash includes fourteen specific driver-related factors and thirteen specific highway factors (e.g., geometry and pavement condition) that contribute to the occurrence of each crash. Table 5 describes the list of crash contributing circumstances listed in the MV4000 database.

A crash can be linked to behavior or engineering risk related crashes using these specific contributing factors noted in the MV4000 crash report. Using crash dataset from Wisconsin, Shaon and Qin found that 79% of total crashes are related to driver error (Shaon et al., 2018a, b). Please note that the engineering risk source variables used in this study were collected for each segment whereas the behavior variables are collected for each county and used as a proxy variable for behavioral risk in crash occurrence. Although the modeling results indicate that 30% of total crashes are generated due to behavioral risk source, this statement does not validate with the information listed by investigating police officer(s) for each crash. This may be because of the unavailability of important behavioral risk source variables that has a significant contribution to crash occurrence such as speeding behavior, fatigue or distracting driving, etc. Considering above-mentioned limitations, a comparison of predicted crashes between single and multiple risk source modeling was conducted to illustrate the strength of multiple risk source modeling. Table 6 described the predicted crash comparison for five rural two-lane sites based on observed non-injury crashes related to different risk sources.

In Table 6, the comparison sites were selected where at least one behavior crash was observed. From the overall comparison between single and multiple risk source models in Table 6, it can be noted that the latter model predicted more crashes compared to single source model. Afghari et al. also found similar information while identifying crash blackspots using multiple risk source model (Afghari et al., 2016).

The single risk source NB model can only predict total crash counts for a specific site. The decomposition of observed crashes described in Table 6 indicates that crashes may come from different risk sources. For example, For Site “A”, there were a total of 6 crashes observed on that segment which includes 1 and 5 crashes occurred due to behavior and engineering risk source, respectively. The predicted value from the single source model was 4 which indicates both engineering and behavior risk variables contributes to all 4 crashes. On the other hand, multiple risk source model predicts there are 4 crashes occurred due to engineering risk source and 1 crashes occurred due to behavioral risk source. The proposed model can also indicate limitation of important information by risk source. For Site “E”, 13 crashes were observed which include 7 and 6 crashes due to behavior and engineering risk, respectively. The predicted value from the single risk model was 5 indicating underestimation of crashes for that site. The multiple risk source model predicted 1 and 5 crashes occurred due to behavior and engineering risk, respectively. While multiple risk source model predicted engineering crashes near observed value (predicted 5 crashes out of 6 observed crashes), the behavior-related crashes were underestimated. This indicates important behavior information is needed to explain observed crashes. Considering the data limitation in this study, it can be noted that multiple risk source model is capable of predicting risk-level crashes which can help safety professionals to identify crash black-spots by risk source and design effective crash countermeasures.

9. Conclusions

Previous studies have explored many factors that could contribute to crash occurrence. Understanding crash-generating mechanisms, adopting appropriate hypotheses, and producing reliable parameter estimates from modeling crash data are challenging for researchers and traffic safety professionals. This study explored the influence of factors

Table 5
Possible Crash Contributing Circumstances listed in the MV4000 Database.

Driver-related Factors		Highway-related Factors	
• Driver condition	• Improper overtake	• Snow/ Ice/ Wet	• Other debris
• Physically disabled	• Improper turn	• Narrow shoulder	• Sign obscured/ missed
• Disregard traffic control	• Left of center	• Soft shoulder	• Narrow bridge
• Following too close	• Exceeding speed limit	• Loose gravel	• Construction zone
• Failure to yield	• Too fast for conditions	• Rough pavement	• Visibility obscured
• Failure to keep vehicle under control	• Unsafe braking	• Debris prior to accident	• Others
• In conflict	• Others		
• Inattentive driving			

Table 6
Comparison of Observed and Predicted Crash Counts between Single Source and Multiple Source Models.

Site ID	County	Total Observed Crashes	Observed Behavior Risk Related Crashes	Observed Engineering Risk Related Crashes	Predicted Crashes from Single Source NB Model	Predicted Behavior Crashes	Predicted Engineering Crashes
A	RICHLAND	6	1	5	4	1	4
B	GREEN	7	3	4	5	1	5
C	DODGE	7	2	5	2	1	2
D	JEFFERSON	7	2	5	9	3	6
E	JUNEAU	13	7	6	5	1	5

from distinct risk sources on crash occurrences while estimating crash frequency and injury severity simultaneously. While engineering risk factors were extensively utilized in crash modeling literature, use of behavioral risk factors are limited due to data unavailability. This study explored behavioral variables collected at larger geographic scale representing existing social norms that can influence driving behavior within a community. In association with engineering risk factors, these behavioral variables are considered to reflect crash risk which originates from a distinct source. The underlying hypothesis of the proposed modeling approach is that crash counts of different injury severities are correlated, and unobserved heterogeneity cannot be sufficiently captured using a single equation crash frequency model. While a large number of studies explored multivariate models to account for the correlation between injury severities, they did not distinguish between sources of crash risk. The complicated crash generation process can be addressed by considering multiple risk sources through the proposed method. Expanding univariate multiple risk source regression modeling to a multivariate framework enabled the incorporation of both injury count correlation and distinguish between crash risk from different sources.

The proposed models were applied to a crash count dataset from Wisconsin rural two-lane highways. Two distinct risk-generating sources – engineering and behavioral – were considered. The modeling results were compared with a single equation NB model and univariate multiple risk source model. The results showed that the multivariate multiple risk source regression model has the best prediction performance among all developed models, whilst also capturing more of the complexity in contributing crash sources. The model parameter estimates indicated that the multiple risk source modeling technique can maintain enough strength to estimate the Poisson mean while considering multiple risk sources in multivariate settings. The parameter estimates for behavioral risk source variables indicates both positive and negative effect of behavioral variables on crash occurrences. Parameter estimates of violent crime rate and drug arrest rate indicates high crime and arrest rates can be used to identify crash prone communities. The negative posterior mean for both OWI citation rate and liquor license rate may help in developing strategies to enforce locations with high liquor license rate in order to reduce crashes. A sample crash count comparison for five sites indicated that the proposed model can predict crashes from each risk source separately which cannot be obtained from single equation modeling. The study not only demonstrates a unique approach to explicitly incorporating behavioral factors into crash prediction models but also provides more insight into the sources of crash risk, which can be used to better inform safety practitioners and guide roadway improvement programs.

The proposed multivariate multiple risk source regression model was developed using the Bayesian framework. Despite the potential of the proposed methodology, the modeling framework may introduce computational complexity and data-specific effects. The risk-level weights used to link predicted crash risk from each risk-generating source to total crash count is solely data dependent. Future research should explore prior knowledge of risk distribution and use it as prior information in model development. Unobserved heterogeneity can be a major issue with crash datasets. It is also important to understand the source of unobserved heterogeneity so that appropriate caution can be taken during model development. Random parameters modeling is a well-accepted methodology to address unobserved data heterogeneity in crash datasets. Though there are no theoretical limitations with regard to implementing random parameters into the multiple risk source structure, the proposed multivariate models were assumed to have fixed parameters. The random parameters structure for covariates in each risk source can be explored in future studies to improve the accuracy of the proposed model and better understand the sources of heterogeneity in crash datasets. Furthermore, county-level behavior related variables were used in this study to represent risky driving behaviors. However, aggregated level behavior variables may not be available or usable to local agencies. Observable behaviors reflected by traffic citations such as speeding, operating while intoxicated, inattentive driving, etc. has long been considered as proxy variable for risky driving behavior. Traffic citation information can be considered to supplement behavioral risk source in future studies. One caveat of using location specific traffic citations is that the citation variables may be highly correlated among each other such as high speeding citation locations may be associated with high OWI citations. Caution should be used to account for possible endogeneity within risk source.

References

Abdel-Aty, M., Keller, J., Brady, P., 2005. Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. Transp. Res. Rec. 1908, 37–45.
Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32 (5), 633–642.
Afghari, A.P., Haque, M.M., Washington, S., 2018a. Applying fractional split model to examine the effects of roadway geometric and traffic characteristics on speeding behavior. Traffic Inj. Prev. 19 (8), 860–866.
Afghari, A.P., Haque, M.M., Washington, S., Smyth, T., 2016. Bayesian latent class safety performance function for identifying motor vehicle crash black spots. Transp. Res. Rec. 2601, 90–98.
Afghari, A.P., Washington, S., Haque, M.M., Li, Z., 2018b. A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk. Anal. Methods Accid. Res. 18, 1–14.
Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 (1), 153–159.
Ando, R., Higuchi, K., Mimura, Y., 2018. Data analysis on traffic accident and urban crime: a case study in Toyota city. Int. J. Transp. Sci. Technol. 7 (2), 103–113.
Asbridge, M., Hayden, J.A., Cartwright, J.L., 2012. Acute cannabis consumption and motor vehicle collision risk: systematic review of observational studies and meta-analysis. BMJ 344, e536.
Bahar, G., Erwin, T., Mackay, M., Smiley, A., Tighe, S., Zein, S., 2004. Canadian Guide to In-service Road Safety Reviews Transportation Association of Canada, Ottawa.
Blomberg, R.D., Peck, R.C., Moskowitz, H., Burns, M., Fiorentino, D., 2005. Crash Risk of Alcohol Involved Driving: a Case-control Study. Final Report, Contract DTNH22-94-C-05001.
Box, S., 2009. New Data From VTTI Provides Insight Into Cell Phone Use and Driving Distraction. Virginia Tech Transportation Institute, pp. 27.
Carter, J.G., Piza, E.L., 2017. Spatiotemporal convergence of crime and vehicle crash hotspots: additional consideration for policing places. Crime Delinq., 0011128717714793.
Carter, P.M., Bingham, C.R., Zakrajsek, J.S., Shope, J.T., Sayer, T.B., 2014. Social norms and risk perception: predictors of distracted driving behavior among novice adolescent drivers. J. Adolesc. Health 54 (5), S32–S41.
Chang, L.-Y., Chen, W.-C., 2005. Data mining of tree-based models to analyze freeway accident frequency. J. Safety Res. 36 (4), 365–375.
Chapman, J.R., Noyce, D.A., 2014. Influence of roadway geometric elements on driver behavior when overtaking bicycles on rural roads. J. Traffic Transp. Eng. 1 (1), 28–38.
Chib, S., Winkelmann, R., 2001. Markov chain monte carlo analysis of correlated count data. J. Bus. Econ. Stat. 19 (4), 428–435.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accid. Anal. Prev. 35 (2), 253–259.
Chiou, Y.-C., Fu, C., 2015. Modeling crash frequency and severity with spatiotemporal dependence. Anal. Methods Accid. Res. 5, 43–58.
Compton, R., Vegega, M., Smither, D., 2009. Drug-impaired Driving: Understanding the Problem and Ways to Reduce It: a Report to Congress.
El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accid. Anal. Prev. 41 (5), 1118–1123.
Erke, A., Goldenbeld, C., Vaa, T., 2009. The effects of drink-driving checkpoints on crashes—a meta-analysis. Accid. Anal. Prev. 41 (5), 914–923.
FBI, 2019. Uniform Crime Reporting (UCR) Program Federal Bureau of Investigation.
FHWA, 2017. Proven Safety Countermeasures. Office of Safety.
Fitzpatrick, K., Lord, D., Park, B.-J., 2010. Horizontal curve accident modification factor with consideration of driveway density on rural four-lane highways in texas. J. Transp. Eng. 136 (9), 827–835.
Garber, N., Ehrhart, A., 2000. Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. Transp. Res. Rec. 1717, 76–83.
Geedipally, S., Patil, S., Lord, D., 2010. Examination of methods to estimate crash counts by collision type. Transp. Res. Rec. 2165, 12–20.
Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of poisson–gamma models. Accid. Anal. Prev. 42 (4), 1273–1282.
Geedipally, S.R., Lord, D., Dhavala, S.S., 2012. The negative binomial-lindley generalized linear model: characteristics and application using crash data. Accid. Anal. Prev. 45, 258–265.
Gruenewald, P.J., Johnson, F.W., 2010. Drinking, driving, and crashing: a traffic-flow model of alcohol-related motor vehicle accidents. J. Stud. Alcohol Drugs 71 (2), 237–248.
Harmon, T., Bahar, G., Gross, F., 2018. Crash Costs for Highway Safety Analysis. FHWA-SA-17-071.
Hauer, E., Ng, J.C., Lovell, J., 1988. Estimation of safety at signalized intersections. Transp. Res. Rec. 1185 (1988), 48–61.
Islam, M.S., Ivan, J.N., Lownes, N.E., Ammar, R.A., Rajasekaran, S., 2014a. Developing safety performance function for freeways by considering interactions between speed limit and geometric variables. Transp. Res. Rec. 2435 (1), 72–81.
Islam, S., Jones, S.L., Dye, D., 2014b. Comprehensive analysis of single-and multi-vehicle large truck at-fault crashes on rural and urban roadways in alabama. Accid. Anal. Prev. 67, 148–158.
Lascala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accid. Anal. Prev. 32 (5), 651–658.
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 (2), 149–161.
Lord, D., Persaud, B., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transp. Res. Rec. 1717, 102–108.
Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37 (1), 35–46.
Lyon, C., Oh, J., Persaud, B., Washington, S., Bared, J., 2003. Empirical investigation of interactive highway safety design model accident prediction algorithm: rural intersections. Transp. Res. Rec. 1840, 78–86.
Ma, J., Kockelman, K.M., 2006. Bayesian multivariate poisson regression for models of injury count, by severity. Transp. Res. Rec. 1950 (1), 24–34.
Ma, J., Kockelman, K.M., 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.
Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.
Miaou, S.-P., Hu, P.S., Wright, T., Rathi, A.K., Davis, S.C., 1992. Relationship Between Truck Accidents and Highway Geometric Design: a Poisson Regression Approach. Transportation Research Record (1376).
Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395–413.
Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accid. Anal. Prev. 40 (1), 260–266.
Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accid. Anal. Prev. 39 (3), 459–468.
Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash modeling. Accid. Anal. Prev. 49, 439–448.
Moeckli, J., Lee, J.D., 2007. The making of driving cultures. Improving Traffic Safety Culture in the United States 38 (2), 185–192.
Montella, A., Imbriani, L.L., 2015. Safety performance functions incorporating design consistency variables. Accid. Anal. Prev. 74, 133–144.
Nagle, M., 2012. Indiana Traffic Safety Facts: Driver History and Crash Outcomes 2011.
NHTSA, 2008. National Motor Vehicle Crash Causation Survey: Report to Congress. National Highway Traffic Safety Administration Technical Report DOT HS. National Highway Traffic Safety Administration. pp. 059.
NHTSA, 2010. Traffic Safety Facts: Driver Electronic Device Use Observation Protocol. DOT HS 811. pp. 361.
Oh, J., Washington, S., Choi, K., 2004. Development of accident prediction models for rural highway intersections. Transp. Res. Rec. 1897, 18–27.
Owusu-Ababio, S., Feng, G., 2006. Predicting alcohol-related crash rate in wisconsin using neural networks. Intelligent Engineering Systems Through Artificial Neural Networks Volume 16 ASME Press.
Park, E., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. Transp. Res. Rec. 2019, 1–6.
Parker, S.T., Tao, Y., 2006. Wistransportal: a Wisconsin Traffic Operations Data Hub. Applications of Advanced Technology in Transportation. pp. 611–616.
Pei, X., Wong, S., Sze, N.-N., 2011. A joint-probability approach to crash prediction models. Accid. Anal. Prev. 43 (3), 1160–1166.
Peng, Y., Lord, D., Zou, Y., 2014. Applying the generalized waring model for investigating sources of variance in motor vehicle crash analysis. Accid. Anal. Prev. 73, 20–26.
Persaud, B.N., 2001. Statistical Methods in Highway Safety Analysis. Project No. 20-5 FY 1999.
Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. J. Transp. Eng. 122 (2), 105–113.
Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accid. Anal. Prev. 36 (2), 183–191.
Qin, X., Shaon, M.R.R., Chen, Z., 2016. Developing analytical procedures for calibrating the highway safety manual predictive methods. Transp. Res. Rec. 2583, 91–98.
Quddus, M., Chin, H., Wang, J., 2001. Motorcycle crash prediction model for signalised intersections. WIT Trans. Built Environ. 52.
Redelmeier, D.A., Tibshirani, R.J., 1997. Association between cellular-telephone calls and motor vehicle collisions. N. Engl. J. Med. 336 (7), 453–458.
Rumar, K., 1985. The Role of Perceptual and Cognitive Filters in Observed Behavior. Human Behavior and Traffic Safety. Springer, pp. 151–170.
Sabey, B.E., Staughton, G., 1975. Interacting roles of road environment vehicle and road user in accidents. Ceste I Mostovi.
Schneider, R.J., Sanatizadeh, A., Shaon, M.R.R., He, Z., Qin, X., 2018. Exploratory analysis of driver yielding at low-speed, uncontrolled crosswalks in milwaukee, wisconsin. Transp. Res. Rec., 0361198118782251.
Serhiyenko, V., Mamun, S.A., Ivan, J.N., Ravishanker, N., 2016. Fast bayesian inference for modeling multivariate crash counts. Anal. Methods Accid. Res. 9, 44–53.
Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.
Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29 (6), 829–837.
Shaon, M.R.R., Qin, X., 2016. Use of mixed distribution generalized linear models to quantify safety effects of rural roadway features. Transp. Res. Rec. 2583, 134–141.
Shaon, M.R.R., 2015. "Quantifying Rural Highway Safety Performance: Application of the Highway Safety Manual and Development of Mixed Distribution Statistical Models for Predicting Crash Frequency". Electronic Theses and Dissertations. https://openprairie.sdstate.edu/etd/1975.
Shaon, M.R.R., Qin, X., Chen, Z., Zhang, J., 2018a. Exploration of contributing factors related to driver errors on highway segments. Transp. Res. Rec., 0361198118790617.
Shaon, M.R.R., Qin, X., Shirazi, M., Lord, D., Geedipally, S.R., 2018b. Developing a random parameters negative binomial-lindley model to analyze highly over-dispersed crash count data. Anal. Methods Accid. Res. 18, 33–44.
Shirazi, M., Lord, D., Dhavala, S.S., Geedipally, S.R., 2016. A semiparametric negative binomial generalized linear model for modeling over-dispersed count data with a heavy tail: characteristics and applications to crash data. Accid. Anal. Prev. 91, 10–18.
Smith, D.J., 2000. Year. Human factors and traffic crashes. Proceedings of the Proceedings of 2000 Midwest Transportation Consortium Transportation Scholars Conference.
Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2007. Openbugs User Manual, Version 3.0. 2. MRC Biostatistics Unit, Cambridge.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Series B Stat. Methodol. 64 (4), 583–639.
St Clair, B., 2001. Year. Wisdot's meta-manager. Proceedings of the 4th National Transportation Asset Management Workshop.
Substance Abuse and Mental Health Services Administration, 2012. The Nsduh Report: State Estimates of Drunk and Drugged Driving.
Tarko, A., Inerowicz, M., Ramos, J., Li, W., 2008. Tool with road-level crash prediction for transportation safety planning. Transp. Res. Rec. 2083, 16–25.
Tarko, A.P., Kanodia, M., 2004. Hazard Elimination Program. Manual on Improving Safety of Indiana Road Intersections and Sections. Joint Transp. Res. Program. pp. 1–2.
Tay, R., 2002. Prioritising road safety initiatives: fatality versus social cost. Appl. Health Econ. Health Policy 1 (3), 129–134.
Treno, A.J., Johnson, F.W., Remer, L.G., Gruenewald, P.J., 2007. The impact of outlet densities on alcohol-related crashes: a spatial panel approach. Accid. Anal. Prev. 39 (5), 894–901.
Uddin, M., Huynh, N., 2018. Factors influencing injury severity of crashes involving hazmat trucks. Int. J. Transp. Sci. Technol. 7 (1), 1–9.
Walsh, J.M., Verstraete, A.G., Huestis, M.A., Mørland, J., 2008. Guidelines for research on drugged driving. Addiction 103 (8), 1258–1268.
Wang, C., Quddus, M.A., Ison, S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accid. Anal. Prev. 43 (6), 1979–1990.
Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accid. Anal. Prev. 99, 6–19.
Washington, S., Afghari, A.P., Haque, M.M., 2018. Detecting High-risk Accident Locations. Safe Mobility: Challenges, Methodology and Solutions. Emerald Publishing Limited, pp. 351–382.
Washington, S., Haque, M., 2013. On the commonly accepted assumptions regarding observed motor vehicle crash counts at transport system locations. 92nd Annual Meeting of Transportation Research Board (TRB).
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/ CRC press.
Weiss, A., 2013. Data-driven Approaches to Crime and Traffic Safety (DDACTS): an Historical Overview. No. DOT HS 809 689. pp. 2013.

Wisconsin Department of Revenu, 2019. Alcohol License Overview for Wisconsin. https://www.dhs.wisconsin.gov/publications/p0/p00778.pdf.
WisDOT, 2019. Wisconsin Crash Data User Guide. https://transportal.cee.wisc.edu/services/crash-data/.
Ye, X., Pendyala, R.M., Shankar, V.N., Konduri, K., 2008. Simultaneous Equations Model of Accident Frequency by Severity Level for Freeway Sections. No. 08-2347.
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. Saf. Sci. 47 (3), 443–452.
Yu, R., Xiong, Y., Abdel-Aty, M., 2015. A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. Transp. Res. Part C Emerg. Technol. 50, 68–77.
Zeng, Q., Huang, H., Pei, X., Wong, S., 2016. Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. Anal. Methods Accid. Res. 10, 12–25.
Zou, Y., Wu, L., Lord, D., 2015. Modeling over-dispersed crash data with a long tail: examining the accuracy of the dispersion parameter in negative binomial models. Anal. Methods Accid. Res. 5, 1–16.