Investigating the Effect of Modeling Single-Vehicle and Multi-Vehicle Crashes Separately on Confidence Intervals of Poisson-gamma Models

Srinivas Reddy Geedipally
Engineering Research Associate
Texas Transportation Institute
Texas A&M University
3135 TAMU
College Station, TX 77843-3135
Tel. (979) 862-1651
Fax. (979) 845-6006
Email: srinivas8@tamu.edu

Dominique Lord
Assistant Professor
Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 458-3949
Fax. (979) 845-6481
Email: d-lord@tamu.edu

October 14, 2009

1Corresponding author
ABSTRACT

Crash prediction models still constitute one of the primary tools for estimating traffic safety. These statistical models play a vital role in various types of safety studies. With a few exceptions, they have often been employed to estimate the number of crashes per unit of time for an entire highway segment or intersection, without distinguishing the influence different sub-groups have on crash risk. The two most important sub-groups that have been identified in the literature are single- and multi-vehicle crashes. Recently, some researchers have noted that developing two distinct models for these two categories of crashes provides better predicting performance than developing models combining both crash categories together. Thus, there is a need to determine whether a significant difference exists for the computation of confidence intervals when a single model is applied rather than two distinct models for single- and multi-vehicle crashes. Building confidence intervals have many important applications in highway safety.

This paper investigates the effect of modeling single- and multi-vehicle (head-on and rear-end only) crashes separately versus modeling them together on the prediction of confidence intervals of Poisson-gamma models. Confidence intervals were calculated for total (all severities) crash models and fatal and severe injury crash models. The data used for the comparison analysis were collected on Texas multilane undivided highways for the years 1997-2001. This study shows that modeling single- and multi-vehicle crashes separately predicts larger confidence intervals than modeling them together as a single model. This difference is much larger for fatal and injury crash models than for models for all severity levels. Furthermore, it is found that the single- and multi-vehicle crashes are not independent. Thus, a joint (bivariate) model which accounts for correlation between single- and multi-vehicle crashes is developed and it predicts wider confidence intervals than a univariate model for all severities. Although joint models predict wider confidence intervals, this research still supports previous studies that recommended modeling single- and multi-vehicle crashes separately for analyzing highway segments.

Keywords: Single-vehicle crashes, multi-vehicle crashes, confidence intervals
INTRODUCTION

Crash prediction models still constitute one of the primary tools for estimating traffic safety. These statistical models play a vital role in various types of safety studies, but they are most often used for estimating the safety performance of various transportation elements or entities (Persaud and Nguyen, 1998; Lord, 2000; Ivan et al., 2000; Lyon et al., 2003; Tarko et al., 2008). In this context, these models have been employed to estimate the number of crashes per unit of time for an entire highway segment or intersection, without distinguishing the influence different sub-groups have on crash risk. A few exceptions have been noted in the literature however. For instance, some researchers have developed distinct predictive models to estimate the safety performance as a function of different categories of vehicles, such as passenger vehicles and truck crashes (Jovanis and Chang, 1986; Miller et al., 1998; Lee and Abdel-Aty, 2005), different time periods (Cercarelli et al., 1992; Mensah and Hauer, 1998; Zador, 1985) or the number of vehicles involved in each crash, i.e., single-vehicle (SV) and multi-vehicle (MV) crashes (Qin et al., 2004; Lord et al., 2005; Griffith, 1999) among others.

Given the differences observed in the characteristics associated with SV and MV crashes, some transportation safety analysts have proposed that distinct crash prediction models should be developed for these two categories of crashes when the objective of the study consists of estimating the safety performance of highway segments (Mensah and Hauer, 1998; Ivan, 2004; Lord et al., 2005; Jonsson et al., 2007; Hardwood et al., 2007; Bonneson et al., 2007). These researchers noted that developing two distinct models provides better predicting performance than developing models combining both crash categories together. In most cases, the motivation for separating models by the number of vehicles involved in the crash is based on shape of the functional form linking both crash types to the traffic flow variable that has been found to be very different from one another (Ivan, 2004; Lord et al., 2005).

Since these two categories of models have been used to estimate or predict the number of crashes on highway segments (Hardwood et al., 2007; Bonneson et al., 2007; Lord et al., 2008), there is a need to determine whether there exists differences in the computation of confidence intervals when a single model is applied rather than two distinct models. Confidence intervals can play a major role in the selection of various highway design alternatives (Lord, 2008) and the identification of hazardous sites (Hauer, 1997) among others. The primary objective of this paper is to investigate if there is an important difference in the prediction of confidence intervals when a unique model is estimated compared to a distinct model for SV and MV crashes. The secondary objective is to examine if there is any difference in the prediction of confidence intervals when a bivariate negative binomial (BNB) model is used rather than a univariate negative binomial (UNB) model for analyzing SV and MV crashes. Confidence intervals were calculated for the Poisson mean, gamma mean and predicted response, respectively (Wood, 2005). Crash data collected on Texas undivided roads from 1997-2001 were used for this comparison analysis.
This paper is divided into five sections. The first section provides a discussion on the modeling prospects related to SV and MV crashes. The second section gives a brief description of data used in this study. The third section describes the methodology utilized for the comparison analysis. The fourth section presents the results of this analysis. The last section provides a summary of the work carried out in this research and recommendation for further work.

BACKGROUND

Several researchers have examined the characteristics and the differences associated with SV and MV crashes (direct comparison). For instance, Ostrom and Eriksson (1993) were the first to examine crash characteristics as a function of the number of vehicles involved in a crash. They studied factors influencing crashes involving intoxicated drivers in northern Sweden. These authors reported that the driver’s blood alcohol content was more significantly related to SV crash fatalities than those associated with MV crash fatalities.

Shankar et al. (1995) analyzed the safety effects of highway design features, weather and other seasonal variables on different crash types. Using data collected on a 61-km section near Seattle, WA, they analyzed several variables, such as the number of horizontal curves and their spacing, rainfall amount, snowing conditions, and their relationship to different types of SV and MV crashes. The authors concluded that models predicting crashes for different crash types had a greater explanatory power than a single model that combined all crash types together.

Mensah and Hauer (1998) examined the effects of aggregated and disaggregated traffic flow variables on the estimation of predictive models. They used data on two-lane rural highways published in Persaud and Mucsi (1995) for their analysis. Using exploratory data analyses and regression methods, they found that SV and MV crashes have significant different characteristics. They reported that using an aggregated model that combine both SV and MV crashes predicts fewer crashes than combining the output of two separate models for the same two categories of crashes.

Griffith (1999) studied SV and MV crashes caused by alcohol and drug-impaired drivers. This author found that SV run-off-the-road crashes on freeways resulted in a higher number of nonfatal injuries than for MV crashes involving an impaired driver. In contrast, MV crashes on freeways resulted in a higher number of fatal injuries than for SV run-off-the-road crashes.

Ivan et al. (1999) investigated differences in causality factors for SV and MV crashes on two-lane rural highways in Connecticut. They found that contributing factors were different for each category of crashes. For example, SV crashes were negatively associated with an increase in traffic intensity (exposure), shoulder width, sight distance, and level of service (LOS). On the other hand, MV crashes were positively associated with an increase in traffic intensity, shoulder width, truck percentage, and number of traffic signals.
In a subsequent study, Ivan et al. (2000) reported that the time-of-day differently influenced for both categories of crashes. SV crashes occurred mostly during the evening and at night, as expected, whereas MV crashes occur more frequently during daylight and evening peak periods. This was mainly attributed to the higher traffic intensity (or exposure). Driveway density had a mixed effect on SV crashes. Driveways at gas stations and minor road intersections were negatively associated with SV crashes, whereas driveways located adjacent to apartment complexes seemed to be associated with an increase in SV crashes. MV crashes increased for all types of driveways.

Ivan (2004) modeled crashes, also using data collected in Connecticut, according to the manner of collision (i.e., number of vehicles involved and their direction of travel). The study showed that the expected number of SV crashes decreases with the increase in traffic volume, whereas all types of MV crashes increased with the increase in traffic flow during the evening time period. During daytime, SV and MV opposite-direction crashes decreased with an increase in traffic flow, whereas MV same-direction and intersecting-direction crashes increased with an augmentation in traffic flow.

Qin et al. (2004) developed zero-inflated Poisson models for different crash types (see Lord et al. 2005, 2007 for a discussion about the application of such models in highway safety). These authors examined crashes occurring on highway segments in Michigan and concluded that crashes are differently associated with traffic flows for different crash types. They noted, for example, that aggregated crash prediction model ignore significant variation in highway crashes. For SV crashes, the marginal crash rate was found to be high for low traffic flow levels and small for high traffic flow volumes. For all MV crashes (except multi-vehicle intersecting-direction crashes), the marginal crash rate was small at low traffic volumes, but high for large traffic volumes, probably because this type of crash is more likely to occur under short headways.

Lord et al. (2005) evaluated several functional forms for SV and MV crashes, and one that combined both (referred to as All Model), as a function of traffic flow, vehicle density and volume over capacity (V/C) ratio on urban freeway segments in Montreal, Canada. Using regression methods, they recommended developing different predictive models for SV and MV crashes rather than developing a common model for both crash categories. The output of the SV model showed that crashes initially increase, peak and then decrease as density or V/C increases. On the other hand, the MV crash model and the model grouping both SV and MV crashes (All Model) showed that crashes increased with an increasing vehicle density or V/C ratio, as expected.

Recently, Jonsson et al. (2007) developed distinct models for SV and different types of MV crashes occurring at intersections on rural four-lane highways in California. These authors concluded that the SV crash model loses some of the observed covariate effects through the aggregation of collision types. Different crash type models exhibited dissimilar relationships with traffic flow and other covariates. Furthermore, Jonsson et al. (2007) noted that the distribution by crash severity varied for different crash types.
In summary, numerous studies have shown that MV and SV crashes have vastly different associative relationships with exposure and geometric design features. These differences were noted for various types of highway facilities. Given these observations, some transportation safety analysts have indicated that distinct models should be estimated for SV and MV crashes when the objective consists of estimated the safety performance of highway segments.

**METHODOLOGY**

This section describes the probabilistic structure of the Poisson-gamma (aka negative binomial NB) and bivariate models, the functional form used for linking SV and MV crashes to traffic flow and the procedure employed for estimating the confidence intervals.

**PROBABILISTIC STRUCTURE OF POISSON-GAMMA MODELS**

Poisson and Poisson-gamma models belong to the family of generalized linear models (GLMs). Poisson-gamma models in highway safety applications have been shown to have the following probabilistic structure: the number of crashes at the $i$-th entity (road section, intersections, etc.) and $t$-th time period, $Y_{it}$, when conditional on its mean $\mu_{it}$, is assumed to be Poisson distributed and independent over all entities and time periods as:

$$ Y_{it} \mid \mu_{it} \sim \text{Po}(\mu_{it}) \quad i = 1, 2, \ldots, I \text{ and } t = 1, 2, \ldots, T $$

(1)

The mean of the Poisson is structured as:

$$ \mu_{it} = f(X; \beta) \exp(e_{it}) $$

(2)

where,

- $f(.)$ is a function of the covariates ($X$);
- $\beta$ is a vector of unknown coefficients; and,
- $e_{it}$ is a the model error independent of all the covariates.

With this characteristic, it can be shown that $Y_{it}$, conditional on $\mu_{it}$ and $\alpha$, is distributed as a Poisson-gamma random variable with a mean $\mu_{it}$ and a variance $\mu_{it} + \alpha \mu_{it}^2$ respectively. (Note: other variance functions exist for the Poisson-gamma model, but they are not covered here since they are seldom used in highway safety studies. The reader is referred to Cameron and Trevidi (1998) and Maher and Summersgill (1996) for a description of alternative variance functions.) The probability density function (PDF) of the Poisson-gamma structure described above is given by the following equation:

$$ f\left(y_{it}; \alpha, \mu_{it}\right) = \frac{\Gamma(y_{it} + \alpha^{-1})}{\Gamma(\alpha^{-1})} \frac{\alpha^{-1}}{(\mu_{it} + \alpha^{-1})^{y_{it} + 1}} \left(\frac{\mu_{it}}{\mu_{it} + \alpha^{-1}}\right)^{y_{it}} $$

(3)
Where,

\[ y_{it} = \text{response variable for observation } i \text{ and time period } t; \]

\[ \mu_{it} = \text{mean response for observation } i \text{ and time period } t; \] and,

\[ \alpha = \text{dispersion parameter of the Poisson-gamma distribution.} \]

The variance of the Poisson-gamma random variable is given by

\[ \text{Var}(y_{it}) = \mu_{it} + \alpha \mu_{it}^2 \]  \hspace{1cm} (4)

Note that if \( \alpha \rightarrow 0 \), the crash variance equals the crash mean and this model converges to the standard Poisson regression model.

The choice of functional form for linking the number of crashes with the model’s covariates is an important characteristic associated with the development of statistical relationships. In this study, crash counts were assumed to follow a nonlinear relationship with traffic volume. Although this functional form is very popular among transportation safety analysts, it may not be the most appropriate form to properly capture exposure at the boundary conditions (Lord, 2002; Lord et al., 2005). Furthermore, the segment length was assumed to be directly proportional to the crash frequency, meaning that the segment length has linear relation with the crash occurrence. Thus, the segment length is considered to be an offset rather than as a covariate.

The statistical model considered in this study is similar to that used elsewhere in the safety literature (see Lord and Bonneson, 2007), except that only one independent variable, namely Average Daily Traffic (ADT), was used. These models are often referred to as flow-only models. They are the most popular type of models utilized by transportation safety analysts (Hauer, 1997; Persaud et al., 2001). Furthermore, they are frequently preferred over models that include several covariates because they can be easily re-calibrated when they are developed in one jurisdiction and applied to another (Persaud et al., 2002; Lord and Bonneson, 2005). Although such models will suffer from an omitted variables bias (because many non-flow related factors are known to affect the frequency of crashes), the empirical assessment carried out in this work still provides valuable information for the development of SV and MV predictive models.

The mean of the crashes per year for segment \( i \) can be calculated by

\[ \mu_i = \beta_0 L_i F_i^{\beta_i} \]  \hspace{1cm} (5)

where,

\[ L_i = \text{length (in miles) of segment } i, \]

\[ F_i = \text{Average Daily Traffic (ADT)}, \]

\[ \beta_0 = \text{intercept (to be estimated)}, \]
\( \beta_1 = \) coefficient (to be estimated) associated with ADT.

It is usually assumed that the number of crashes increases at a decreasing rate as the traffic volume increases. This relationship is characterized in predictive models with the coefficient for the traffic volume parameter (\( \beta_1 \)) to be below 1.

**PROBABILISTIC STRUCTURE OF BIVARIATE NEGATIVE BINOMIAL MODELS**

There has not been a lot of research conducted on jointly modeling crash counts in highway safety. As stated by Park and Lord (2006), a multivariate model treats the correlated crash counts as interdependent variables and leads to more precise estimates for the effects of factors on crash risk, than a univariate model. The few studies that documented the application of multivariate models for analyzing crash data include Tunaru (2002), Bijleveld (2005), Miaou and Song (2006), Song et al. (2006), and Ma and Kockelman (2006).

The probability function of a bivariate negative binomial (BNB) model is given by the following equation (Subrahmainiam and Subrahmainiam, 1973):

\[
f(y_{1it}, y_{2it}) = \frac{(\mu_{1it} - \delta)^{y_{1it}} (\mu_{2it} - \delta)^{y_{2it}} \Gamma(r + y_{1it} + y_{2it})}{y_{1it} \cdot y_{2it} \cdot \Gamma(r)(1 + \mu_{1it} + \mu_{2it} - \delta)^{y_{1it} + y_{2it}}} S(y_{1it}, y_{2it})
\]

(6)

Where,

- \( y_{1it}, y_{2it} \) = two response variables for observation \( i \) and time period \( t \);
- \( \mu_{1it} \) = mean of the first response variable for observation \( i \) and time period \( t \);
- \( \mu_{2it} \) = mean of the second response variable for observation \( i \) and time period \( t \);
- \( r \) = combined inverse dispersion parameter;
- \( S(y_{1it}, y_{2it}) = \sum_{j=0}^{\min(y_{1it}, y_{2it})} \binom{y_{1it}}{j} \binom{y_{2it}}{j} (r + y_{1it} + y_{2it} - 1) \) (7)

with \( \tau = \left\{ \delta(1 + \mu_{1it} + \mu_{2it} - \delta)/[(\mu_{1it} - \delta)(\mu_{2it} - \delta)] \right\}; \) (8)

\[
\delta = \left( \frac{m_{11}}{r^2} - \frac{\bar{y}_1 \cdot \bar{y}_2}{r^2} \right)
\]

with \( m_{11} = \) first mixed central moment.

**GOODNESS-OF-FIT STATISTICS**

Different methods were used for evaluating the goodness-of-fit (GOF) and predictive performance of the models. The methods used in this research include the following:
Mean Absolute Deviance (MAD)

MAD provides a measure of the average mis-prediction of the model (Oh et al., 2003). It is computed using the following equation:

\[
\text{Mean Absolute Deviance (MAD)} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|
\]  

(10)

Mean Squared Predictive Error (MSPE)

MSPE is typically used to assess the error associated with a validation or external data set (Oh et al., 2003). It can be computed using Equation (12):

\[
\text{Mean Squared Predictive Error (MSPE)} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2
\]  

(11)

ESTIMATION OF CONFIDENCE INTERVALS

Confidence intervals can be used for selecting highway design alternatives where the safety performance is used as a screening criterion and for identifying hazardous sites. Wood (2005) has proposed a method for estimating the confidence intervals for the mean response (μ), for the gamma mean (m), and the predicted response (y) at a new site having similar characteristics as the sites used in the original dataset from which the model was developed. The following table gives the equations for calculating the confidence intervals. In this table, η is the logarithm of the estimated mean response, while φ is the inverse dispersion parameter.

<table>
<thead>
<tr>
<th>Table 1: Confidence Intervals for μ, m, y. (Wood, 2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>μ</td>
</tr>
<tr>
<td>m</td>
</tr>
<tr>
<td>y</td>
</tr>
</tbody>
</table>

Note:
\[\text{Var}(\hat{\eta}) = X(I^{-1})X^T\] where \(I^{-1}\) is the variance-covariance matrix and X is a matrix containing observed values in logarithmic form.
\[\lfloor x \rfloor\] denotes the largest integer less or equal than x

The primary purpose of this study was to evaluate whether or not there will be any significant difference in the prediction of confidence intervals by modeling SV and MV
crashes separately. To examine this difference, confidence intervals for the Poisson mean, the gamma mean and the predicted response were calculated for the total number of crashes (ALL) and for the sum of SV and MV crashes predicted by the SV+MV UNB model and the SV+MV BNB model. For the sake of simplicity, a fixed dispersion parameter was considered during the calculation of confidence intervals, although a varying dispersion parameter can reduce the width of the confidence intervals (Geedipally and Lord, 2008). Furthermore, since the database contains more than 1,500 observations, it was assumed that the inverse dispersion parameter is properly estimated (see Lord, 2006).

The modeling procedure was accomplished using the following 4-stage process:

1. In the initial step, the models for total (Fatal or Killed, Injury type A, Injury type B, Injury type C, Property damage only or KABCO) crashes, and fatal and severe injury (KAB) crashes were estimated using a fixed dispersion parameter in SAS (SAS, 2002). Individual models for SV crashes, MV crashes and the total number of crashes (ALL) were estimated using PROC GENMOD, whereas a joint model for SV and MV crashes was estimated using PROC NLMIXED in SAS. It should be noted that the number of years and the length for each site were used as an offset.

2. The confidence interval for Poisson mean \( \mu \), gamma mean \( m \), and the prediction interval \( y \) at a new site with the same traffic flow characteristics was then calculated using the equations in Table 1 for the ALL and SV+MV BNB models. In the equation, \( \eta \) is the logarithm of the estimated mean (\( \mu \)) while \( \phi \) is the inverse dispersion parameter estimated during the fitting process. \( \text{Var}(\eta) \) is calculated using \( XI^{-1}X^T \) where \( I^{-1} \) represents the variance-covariance matrix and \( X \) is the matrix containing the observed values. The variance-covariance matrix was provided by SAS.

3. The above step was also calculated for the SV+MV UNB model. The Poisson mean is now normally distributed with mean \( \mu = \mu_{sv} + \mu_{mv} \) and variance \( \sigma_{\eta}^2 = \mu^2 \text{Var}(\eta) \) where \( \text{Var}(\eta) \) is calculated as \( \frac{\mu^2 \text{Var}_{sv}(\eta) + \mu^2 \text{Var}_{mv}(\eta)}{(\mu_{sv} + \mu_{mv})^2} \) (This equation is based on the fact that SV crashes and MV crashes are two independent random variables). Here \( \mu_{sv} \) and \( \text{Var}_{sv}(\eta) \) were estimated from the SV crash model and \( \mu_{mv} \) and \( \text{Var}_{mv}(\eta) \) were estimated from the MV crash model. The confidence intervals were then calculated with the calculated means and variances.

4. The difference in the width of confidence intervals calculated by ALL, SV+MV UNB and SV+MV BNB was then estimated.
DATA DESCRIPTION

The statistical models were estimated using crash data collected at 1,552 Texas undivided four-lane highway segments. The data were obtained from the Texas Department of Public Safety and the Texas Department of Transportation (TxDOT) for the years 1997 to 2001. The data included information on segment length, ADT, number of intersections, number of horizontal curves and other variables that influence crash risk. The crash data also provided details about the severity as well as the number of vehicles involved in the collision. Only head-on and rear-end collisions were considered for MV crashes. A total of 3,283 SV crashes and 4380 MV crashes were extracted. Table 2 provides relevant descriptive statistics for key explanatory variables and the crash data.
Table 2: Descriptive statistics of independent variables and crash data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean (Std. dev)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment length (Miles)</td>
<td>0.1</td>
<td>6.275</td>
<td>0.55 (0.66)</td>
<td>848.29</td>
</tr>
<tr>
<td>AADT (Vehicles/day)</td>
<td>42</td>
<td>24800</td>
<td>6684.3 (4104.9)</td>
<td>--</td>
</tr>
<tr>
<td>Incapacitating Injury (A)</td>
<td>0</td>
<td>5</td>
<td>0.21 (0.61)</td>
<td>331</td>
</tr>
<tr>
<td>Non-Incapacitating Injury (B)</td>
<td>0</td>
<td>18</td>
<td>0.51 (1.29)</td>
<td>789</td>
</tr>
<tr>
<td>Possible Injury (C)</td>
<td>0</td>
<td>10</td>
<td>0.39 (0.95)</td>
<td>619</td>
</tr>
<tr>
<td>Fatal (K)</td>
<td>0</td>
<td>3</td>
<td>0.06 (0.26)</td>
<td>88</td>
</tr>
<tr>
<td>Non-Injury (O)</td>
<td>0</td>
<td>34</td>
<td>0.94 (2.12)</td>
<td>1456</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>0</td>
<td>64</td>
<td>2.12 (4.31)</td>
<td>3283</td>
</tr>
<tr>
<td>MV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incapacitating Injury (A)</td>
<td>0</td>
<td>9</td>
<td>0.29 (0.79)</td>
<td>458</td>
</tr>
<tr>
<td>Non-Incapacitating Injury (B)</td>
<td>0</td>
<td>18</td>
<td>0.62 (1.44)</td>
<td>966</td>
</tr>
<tr>
<td>Possible Injury (C)</td>
<td>0</td>
<td>20</td>
<td>0.97 (2.04)</td>
<td>1512</td>
</tr>
<tr>
<td>Fatal (K)</td>
<td>0</td>
<td>3</td>
<td>0.08 (0.32)</td>
<td>124</td>
</tr>
<tr>
<td>Non-Injury (O)</td>
<td>0</td>
<td>17</td>
<td>0.85 (1.81)</td>
<td>1320</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>0</td>
<td>48</td>
<td>2.82 (5.26)</td>
<td>4380</td>
</tr>
<tr>
<td>ALL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incapacitating Injury (A)</td>
<td>0</td>
<td>9</td>
<td>0.51 (1.12)</td>
<td>789</td>
</tr>
<tr>
<td>Non-Incapacitating Injury (B)</td>
<td>0</td>
<td>31</td>
<td>1.13 (2.24)</td>
<td>1755</td>
</tr>
<tr>
<td>Possible Injury (C)</td>
<td>0</td>
<td>22</td>
<td>1.37 (2.43)</td>
<td>2131</td>
</tr>
<tr>
<td>Fatal (K)</td>
<td>0</td>
<td>6</td>
<td>0.14 (0.45)</td>
<td>212</td>
</tr>
<tr>
<td>Non-Injury (O)</td>
<td>0</td>
<td>42</td>
<td>1.79 (3.14)</td>
<td>2776</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>0</td>
<td>99</td>
<td>4.94 (7.95)</td>
<td>7663</td>
</tr>
</tbody>
</table>

Figure 1 shows the proportion by severity levels for SV, MV and ALL crashes. As seen in this figure, SV crashes have a much larger percentage of non injury than MV crashes. This is expected since SV crashes on average have a lesser transfer of energy than MV crashes.
Table 3 shows the correlation matrix between SV and MV crashes for total crashes and fatal and serious injury crashes. As seen in this table, the correlation is larger for KAB crashes than KABCO crashes. Although the crash types are not highly correlated, they cannot be assumed to be independent either. Because of this, a bivariate model needs to be used for analyzing SV and MV crashes simultaneously.

### Table 3: Correlation between SV and MV crashes

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>MV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KABCO</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>1</td>
<td>0.372</td>
</tr>
<tr>
<td>MV</td>
<td>0.372</td>
<td>1</td>
</tr>
<tr>
<td><strong>KAB</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>1</td>
<td>0.431</td>
</tr>
<tr>
<td>MV</td>
<td>0.431</td>
<td>1</td>
</tr>
</tbody>
</table>

### RESULTS

This section describes the results of the comparison analysis between SV, MV and ALL crash models. The confidence intervals were calculated for total crashes (KABCO) and fatal and serious injury (KAB) crashes.

Table 4 provides the parameter estimates of the models with their associated standard errors.
### Table 4: Estimates of model coefficients (and standard errors)

<table>
<thead>
<tr>
<th>Crash type</th>
<th>ln(β₀)</th>
<th>β₁</th>
<th>α (1/φ)</th>
<th>ln(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNB</td>
<td>-5.9403 (0.4708)</td>
<td>0.6535 (0.0539)</td>
<td>1.0616 (0.0734)</td>
<td>--</td>
</tr>
<tr>
<td>KAB</td>
<td>-5.5348 (0.6075)</td>
<td>0.4878 (0.0695)</td>
<td>1.1895 (0.1321)</td>
<td>--</td>
</tr>
<tr>
<td>BNB</td>
<td>-3.4303 (0.5919)</td>
<td>0.4156 (0.06389)</td>
<td>--</td>
<td>-0.1816 (0.0581)</td>
</tr>
<tr>
<td>KAB</td>
<td>-5.0467 (0.5918)</td>
<td>0.4525 (0.06619)</td>
<td>--</td>
<td>-0.0451 (0.0746)</td>
</tr>
<tr>
<td><strong>MV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNB</td>
<td>-8.9067 (0.5650)</td>
<td>1.0504 (0.0645)</td>
<td>1.6883 (0.0877)</td>
<td>--</td>
</tr>
<tr>
<td>KAB</td>
<td>-8.9794 (0.6646)</td>
<td>0.9233 (0.0753)</td>
<td>1.4173 (0.1175)</td>
<td>--</td>
</tr>
<tr>
<td>BNB</td>
<td>-7.7371 (0.6006)</td>
<td>0.9288 (0.06497)</td>
<td>--</td>
<td>-0.1816 (0.0581)</td>
</tr>
<tr>
<td>KAB</td>
<td>-9.1087 (0.6186)</td>
<td>0.9357 (0.06866)</td>
<td>--</td>
<td>-0.0451 (0.0746)</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KABCO</td>
<td>-6.9613 (0.4075)</td>
<td>0.8839 (0.0467)</td>
<td>0.9981 (0.049)</td>
<td>--</td>
</tr>
<tr>
<td>KAB</td>
<td>-6.7263 (0.4915)</td>
<td>0.7288 (0.0560)</td>
<td>1.0093 (0.0737)</td>
<td>--</td>
</tr>
</tbody>
</table>

Tables 5 and 6 provide the variance-covariance matrices for univariate and bivariate models respectively.

### Table 5: Variance-covariance matrices for univariate NB models

<table>
<thead>
<tr>
<th>Crash type</th>
<th>(lnβ₀)</th>
<th>(lnβ₁)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KABCO</td>
<td>0.22164</td>
<td>-0.02530</td>
</tr>
<tr>
<td>KAB</td>
<td>-0.002530</td>
<td>0.002905</td>
</tr>
<tr>
<td><strong>MV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KABCO</td>
<td>0.31925</td>
<td>-0.03637</td>
</tr>
<tr>
<td>KAB</td>
<td>-0.03637</td>
<td>0.004164</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KABCO</td>
<td>0.16609</td>
<td>-0.01897</td>
</tr>
<tr>
<td>KAB</td>
<td>-0.01897</td>
<td>0.002178</td>
</tr>
</tbody>
</table>
Table 6: Variance-covariance matrices for bivariate NB models

<table>
<thead>
<tr>
<th>Crash type</th>
<th>SV</th>
<th>MV</th>
<th>ln(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(lnβ₀)</td>
<td>β₁</td>
<td>(lnβ₀)</td>
</tr>
<tr>
<td>KABCO</td>
<td>0.3503</td>
<td>-0.03755</td>
<td>0.2774</td>
</tr>
<tr>
<td></td>
<td>-0.03755</td>
<td>0.004082</td>
<td>-0.02942</td>
</tr>
<tr>
<td>KAB</td>
<td>0.1745</td>
<td>-0.01880</td>
<td>0.3827</td>
</tr>
<tr>
<td></td>
<td>-0.01885</td>
<td>0.002117</td>
<td>-0.04204</td>
</tr>
</tbody>
</table>

Table 7 gives the goodness of fit statistics for each type of model. This table shows that there is a clear difference in the prediction of total crashes with each type of the model. The SV+MV bivariate NB model is found to provide a slightly better statistical fit than the other two types of model (based on the MSPE) for KABCO crashes. There is no significant difference in the prediction of KAB crashes among each type of the model.

Table 7: Goodness of fit statistics

<table>
<thead>
<tr>
<th>Crash type</th>
<th>MAD</th>
<th>MSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>KABCO</td>
<td>4.022</td>
<td>48.683</td>
</tr>
<tr>
<td></td>
<td>4.114</td>
<td>47.857</td>
</tr>
<tr>
<td></td>
<td>4.044</td>
<td>48.719</td>
</tr>
<tr>
<td>KAB</td>
<td>1.512</td>
<td>6.096</td>
</tr>
<tr>
<td></td>
<td>1.525</td>
<td>6.159</td>
</tr>
<tr>
<td></td>
<td>1.522</td>
<td>6.141</td>
</tr>
</tbody>
</table>

Figure 2 illustrates the relationship between total crashes and traffic flow for the ALL, SV+MV (UNB) and SV+MV (BNB) models, respectively. From this figure, the crash-flow relationships indicate that the ALL model and SV+MV (BNB) model predicts the total number of crashes at a decreasing rate as traffic flow increases. This relationship basically means that there are proportionally less crashes per passing vehicles as the traffic flow increases and thus the crash risk per vehicle diminishes when traffic flow increases. However, the SV+MV (UNB) model shows a linear relationship between total crashes and traffic flow. Figure 2 also shows that the SV+MV (BNB) model always predicts more crashes than the ALL and SV+MV (UNB) crash model for traffic flows less than 7,000 veh/day. At higher traffic flows (flow>15,000), the SV+MV (UNB) predicts more crashes than other two models.
Figure 2: Crash-Flow Relationship for Total Crashes

Figure 3 shows the relationship between severe crashes (KAB) and traffic flow for the ALL, SV+MV (UNB) and SV+MV (BNB) models. All model types predict crashes at a decreasing rate as traffic flow increases. Both the SV+MV (UNB) and SV+MV (BNB) models predict almost the same number of crashes for the entire range of traffic flow volumes. The ALL crash model predicts fewer crashes than other two models at larger traffic flows.

Figure 3: Crash-Flow Relationship for KAB Crashes

Figure 4 shows the confidence intervals for the Poisson mean ‘μ’ for KABCO crashes, and KAB (fatal and serious injuries) crashes. This figure illustrates that the width of the interval for Poisson mean ‘μ’ for total crashes (i.e., the distance between the mean...
response and the upper confidence interval boundary) predicted by the SV+MV UNB and SV+MV BNB models is always wider than the width predicted by the ALL crash model for KABCO crashes. For fatal and serious injury crashes, the width of the interval for Poisson mean predicted by SV+MV UNB is narrower than the width predicted by the ALL crash model for most of the time. The SV+MV BNB model always predicts wider interval width than the ALL crash model, indicating that the two crash types are correlated. The maximum relative difference in the mean of predicted crashes between SV+MV UNB model and ALL crash model is 21% for KABCO crashes and 25% for KAB crashes. Similarly, the maximum relative difference in the mean of predicted crashes between SV+MV BNB model and ALL crash model is 73% for KABCO crashes and 40% for KAB crashes.

Figure 4: 95-Percentile Confidence Intervals for the Poisson Mean (μ)
Figure 5 shows the confidence intervals for the gamma mean ‘m’ for KABCO crashes and KAB crashes. The figure illustrates that the confidence interval for the gamma mean ‘m’ for KABCO crashes predicted by SV+MV BNB model is narrower than the ALL and SV+MV UNB model for flows less than 15,000 veh/day. For KAB crashes, the interval predicted by SV+MV BNB is wider than that predicted by ALL crash model. The SV+MV UNB model predicted narrower confidence intervals than the other two type of modeling frameworks.

Figure 6 shows the confidence intervals for the predictive response ‘y’ for KABCO and KAB crashes. Since the predicted number of crashes must be non-negative integer, the
estimated values were rounded to the nearest integer. The figure illustrates that the SV+MV UNB model predicts narrower confidence intervals than the ALL crash model for at least half of the time for KABCO crashes and at least 77% of the time for KAB crashes. On the other hand, the SV+MV BNB model predicts wider confidence intervals than the ALL crash model for at least half of the time for both KABCO crashes and KAB crashes. As discussed by Lord (2008), the predicted response and its associated variance are the most important values to be computed, since most of the time, analysts will apply the model to other datasets (e.g., comparing different highway design alternatives).

In sum, the comparison analysis showed that using two distinct models to predict crashes for highway segments most often provides wider confidence intervals than using one
model for all categories of crashes, especially for the Poisson and gamma mean values. Although each subset contains fewer crashes or observations and the functional forms are vastly different, the summation of the variances is higher than the variance estimated from the full dataset by a single model. Overall, a bivariate NB model provides wider confidence intervals than a univariate NB model, as expected. This is attributed to fact that the joint model takes correlation of the dependent variables into effect which increase the variance of the mean values and predicted response. Although the results of this analysis show that a separate model increases the variance of mean and predicted response, it still support previous work, which recommended that SV and MV crashes should be modeled separately. Furthermore, a joint model (which considers the correlation between dependent variables) is recommended when modeling SV and MV crashes, even though the confidence intervals will be larger.

SUMMARY AND CONCLUSIONS

This paper documented a research study that examined the potential differences in the prediction of confidence intervals of Poisson-gamma models when SV and MV crashes are modeled separately and together. Recently, some transportation safety analysts have recommended modeling both categories of crashes independently for estimating the safety performance of highway segments. To accomplish the comparison analysis, crash data were collected on four-lane rural highways in Texas. Then, KABCO and KAB models were estimated for SV, MV and ALL crashes along with the confidence intervals for the Poisson mean (μ), gamma mean (m), and predictive response (y), respectively.

The following results were obtained from the analysis:

1. There is a clear difference in the prediction of confidence intervals for the Poisson mean, gamma mean, and predictive response between aggregated crash prediction (ALL) and the summation of distinct models (SV and MV). Overall, the bivariate NB model provides wider confidence intervals than the univariate model.
2. Combining SV and MV models predicts wider confidence intervals than developing a single model for both total (KABCO) crashes and fatal and serious injuries (KAB) crashes. Although wider confidence intervals were observed with SV+MV UNB and BNB models, this paper still supports the research done by others that recommended analyzing SV and MV crashes separately. However, given the correlation, a joint NB model should be utilized.

There are several avenues for further work. First, it is recommended to conduct similar analyses using data collected on other types of highways, such as two-lane and divided highways as well as controlled-access facilities. As part of these analyses, the effect of the sample size for each subset should be examined more carefully. This would give us the opportunity to determine whether sample size plays an important role in the computation of the confidence intervals when the ratio between SV and MV crashes is different. Second, since the analysis was carried with only one covariate, namely ADT, one should also evaluate the influence of models containing several covariates on the computation of confidence intervals for a single versus two distinct models. The approach proposed by Wood (2002) is still appropriate for computing confidence
intervals for models that includes several covariates (see Lord, 2008). Finally, given the role confidence intervals can play in the identification of hazardous sites (i.e., hot spot identification) and select highway design alternatives, it is suggested to compare how well two distinct models can influence both kinds of application versus the use of a single model.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Samiran Sinha from the Department of Statistics at Texas A&M University for his help in deriving the equation for computing the variance of the sum of SV and MV crashes. The authors would also like to thank two anonymous reviewers for providing useful comments to improve this paper.

REFERENCES

Bijleveld, F. D., 2005. The Covariance Between the Number of Accidents and the Number of Victims in Multivariate Analysis of Accident Related Outcomes. Accident Analysis and Prevention, Vol. 37, No. 4, pp. 591-600.


Ivan, J., 2004. New approach for including traffic volumes in crash rate analysis and forecasting. Transportation Research Record 1897, 134-141.


Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention 38 (4), 751-766.

Lord, D. 2008. Methodology for estimating the variance and confidence intervals of the estimate of the product of baseline models and AMFs. Accident Analysis & Prevention 40 (3), 1013-1017.


Geedipally and Lord


