# Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking

Jonathan Aguero-Valverde and Paul P. Jovanis

**Traditionally, highway safety analyses have used univariate Poisson or negative binomial distributions to model crash counts for different levels of crash severity. Because unobservables or omitted variables are shared across severity levels, however, crash counts are multivariate in nature. This research uses full Bayes multivariate Poisson lognormal models to estimate the expected crash frequency for different levels of crash severity and then compares those estimates to independent or univariate Poisson lognormal estimates. The multivariate Poisson lognormal model fits better than the univariate model and improves the precision in crash-frequency estimates. The covariances and correlations among crash severities are high (correlations range from 0.47 to 0.97), with the highest values found between contiguous severity levels. Considering this correlation between severity levels improves the precision of the expected number of crashes. The multivariate estimates are used with cost data from the Pennsylvania Department of Transportation to develop the expected crash cost (and excess expected cost) per segment, which is then used to rank sites for safety improvements. The multivariate-based top-ranked segments are found to have consistently higher costs and excess costs than the univariate estimates, which is due to higher multivariate estimates of fatalities and major injuries (due to the random effects parameter). These higher estimated frequencies, in turn, produce different rankings for the multivariate and independent models. The finding of a high correlation between contiguous severity levels is consistent with some of the literature, but additional tests of multivariate models are recommended. The improved precision has important implications for the identification of sites with promise (SWiPs), because one formulation includes the standard deviation of crash frequencies for similar sites as part of the assessment of SWiPs.**

Traditionally, highway safety analyses have used univariate Poisson or negative binomial distributions to model crash counts for different levels of crash severity. The counts are often treated as a whole, combining crashes of different severities of outcome. Alternatively, safety analysts may arbitrarily choose outcome levels for inclusion in a study (e.g., fatal and injury crashes only).

It can be argued that crash counts are fundamentally multivariate in nature because unobservables or omitted variables are shared across levels of crash severity. Crash investigations have commonly found that a difference of mere microseconds may differentiate a crash that causes severe injury from one that causes property damage only (PDO). These small differences in the evolution of crash outcomes lead to the notion that there are potentially strong correlations across outcome levels: correlations that are not recognized when counts are aggregated or when separate models are estimated for each outcome level. Hierarchical Bayesian models have the flexibility to assess this correlation. The purpose of this research is to use properly formulated hierarchical Bayesian models to explore the nature of the correlations across levels of crash-outcome severity and then to assess the implications of these new model formulations in common safety analyses, specifically the identification of sites with promise (SWiPs) (*1*).

Concerning model formulation, separate univariate models of interrelated outcomes (such as crash severity counts) may ignore information shared in those unobservables common to the outcomes. Neglecting to include correlation structures for these multivariate outcomes may distort estimates of variance components and regression coefficients (*2*) and result in reduced efficiency and possibly biased parameter estimates. In addition, crash counts are frequently characterized by low sample mean values (*3*) due to the relatively low frequency of crashes. Therefore, crash data often exhibit excess zeroes compared to those expected under a Poisson process (*4*, *5*). Even after adjusting for extra-Poisson variation in the data by including a gamma- or lognormal-distributed random effect term in the model, analysis of data sets with low sample mean values often results in crash frequency estimates with low precision (i.e., high standard deviation). Estimation precision in data sets presenting low sample mean values may benefit by pooling strength in terms of the underlying multivariate distributions (*2*). This is particularly critical in cases that involve the most severe (but also most infrequent) crashes: those that involve fatalities, major injuries, or incapacitating injuries. Because the final objective of any highway safety analysis is to reduce the frequency and severity of crashes, it follows that obtaining the most precise estimates of crash frequency at the highest severity levels is highly desirable.

One of the most important tasks in highway safety analysis is the identification of locations or sites that might be in need of engineering improvements, to reduce the number of crashes. The extensive literature on the subject of identifying SWiPs includes a series of papers by Hauer (*1*) and Hauer et al. (*6*, *7*). The latest paper in the

J. Aguero-Valverde, Programa de Investigación en Desarrollo Urbano Sostenible, Universidad de Costa Rica, Los Profesores, Calle B, No. 11, Mercedes, San Pedro, San José, 11503, Costa Rica. P. P. Jovanis, Department of Civil and Environmental Engineering and Pennsylvania Transportation Institute, Pennsylvania State University, 212 Sackett Building, University Park, PA 16802-1408. Corresponding author: J. Aguero-Valverde, jaguero@produs.ucr.ac.cr.

*Transportation Research Record: Journal of the Transportation Research Board, No. 2136,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 82–91.
DOI: 10.3141/2136-10

series (*7*) suggests the use of severity-weighted models of crash frequency as the most cost-effective way to select sites for engineering improvements. In addition, Tarko and Kanodia (*8*) have suggested the use of an "index of crash cost," which is the sum of the expected cost of the excess crash frequency at each severity level divided by the expected variance of the estimate. In line with these suggestions, this paper discusses the application of multivariate models in the estimation of total expected crash cost and the excess crash cost, while ranking candidate sites for engineering improvements. Note that the approach suggested by Tarko and Kanodia (*8*) uses an estimate of the variance, highlighting the importance of the variable in safety analyses.

In summary, this research seeks to estimate and better understand the implications of correlations that are present across severity of crash outcomes. The specific modeling tool used to conduct this assessment is a full Bayes hierarchical model specifically formulated for this purpose. The approach estimates and compares two sets of models with identical specifications: one set estimated with the assumption of independence across outcomes, the second set with correlations estimated across outcomes as an explicit part of the model formulation (a nonindependence assumption). Both model parameter estimates and their variances are discussed. Finally, these two models are used in the development of SWiPs, highlighting the advantages of using the Bayesian formulation.

This paper is organized as follows. First, there is a review of the literature concerning application of hierarchical models to similar safety problems. Next, the statistical methodology is reviewed and the data set is described. The paper concludes with a discussion of results, conclusions, and recommendations for future research.

## LITERATURE REVIEW

Until recently, crash-frequency models for different severities were estimated using independent univariate Poisson, negative binomial (or Poisson–gamma), and zero-inflated Poisson distributions. Tanaru (*9*) proposed a multivariate Poisson lognormal model for modeling crash severities at the link level (road sections between two major junctions). Full Bayes hierarchical models were estimated with four levels of severity: fatality or serious injury from crashes involving one vehicle, fatality or serious injury from crashes involving more than one vehicle, slight injury from crashes involving one vehicle, and slight injury from crashes involving more than one vehicle. The expected crash frequency for each severity level was used to produce rankings of sites; therefore, four different rankings were produced. No covariates were included in this analysis. All the entries in the covariance matrix were found to be significantly different from zero, but the author did not report the correlation.

Simultaneous negative binomial models of outcomes from crashes involving fatalities or injuries were proposed by Ladrón de Guevara et al. (*10*) to account for shared omitted variables and measurement errors. The authors were interested in forecasting crashes at the planning level; the unit of study was the traffic analysis zone. The authors found a high and significant correlation ($\rho = 0.85$) between fatal-crash and injury-crash outcomes.

Miaou and Song (*11*) used Bayesian generalized linear mixed models with multivariate spatial random effects to rank sites by crash cost rate. Three crash severities were analyzed at the county level: fatal, incapacitating injury, and nonincapacitating injury. The authors were mainly interested in ranking criteria and did not discuss correlation structures.

Bijleveld (*12*) proposed a method to estimate the covariances between crash and injury counts from within the accident data, using individual accident information on the relevant outcomes such as the number of victims and fatalities. This method, however, estimates only covariances within an observation and ignores covariances between observations.

Song et al. (*13*) further explored the issue of multivariate spatial models. Instead of analyzing correlated severities, the authors studied correlated crash types: intersection crash, intersection-related crash, driveway-access crash, and nonintersection crash. The authors found that the model with correlated spatial random effects fit the data better than did the model with independent random effects. Although summary statistics of the correlation coefficients between the responses were not provided, plots of the posterior distribution of the coefficients indicated significant correlations ranging from 0.3 to 0.6.

Ma and Kockelman (*14*) used multivariate Poisson regression within a Bayesian framework to estimate the number of victims per road segment in five different severity classes: fatal, disabling injury, nondisabling injury, possible injury, and noninjury. The model did not allow for extra-Poisson variation in the data and restricted the covariances for the five different severity levels to be identical and positive.

Park and Lord (*15*) extended the Bayesian multivariate Poisson specification to include extra-Poisson variation and relaxed the covariance specification to allow for different covariances among severity levels by including a lognormal random effect. The authors analyzed crash counts for three-legged unsignalized intersections and five severity levels: fatal, incapacitating injury, nonincapacitating injury, minor injury, and PDO. The posterior means of the correlation matrix indicated high correlations between different crash severities (from 0.69 to 0.85). Ma et al. (*16*) also used a Bayesian multivariate Poisson lognormal specification to model crash counts at the road-segment level. However, the correlations found by the authors—ranging from 0.01 to 0.43—were significantly lower than the correlations found in Park and Lord (*15*).

The literature review reveals that several authors have explored the analysis of crash outcomes and crash types in an integrated modeling structure, most often using hierarchical Bayes approaches. However, these analyses stopped short of quantifying the effect of those multivariate structures on the precision of the estimates of crash frequency and the implications of those estimation differences on the safety analyses used as examples in those analyses. This research conducts those studies, building on this aforementioned foundational literature.

## METHODOLOGY

Full Bayesian methods have been gaining popularity as the approach of choice when modeling multiple levels and incorporating random effects or complicated dependence structures, such as correlated errors (*17*). Around 1990, the "Markov chain Monte Carlo revolution" took place, when methods like the Gibbs sampler and the Metropolis algorithm were coupled with faster computing to enable evaluation of the complicated integrals that are usually found in Bayesian methods. Markov chain Monte Carlo, also known as Markov chain simulation, is a general method based on drawing values of the parameters of interest from approximate distributions and then correcting those draws to better approximate the target posterior distribution (*18*). Details about Markov chain simulations are beyond the scope of this paper; interested readers can refer to Gelman et al. (*18*) and Congdon (*19*) for more on this topic.

For the full Bayes multivariate Poisson lognormal models assessed in this research, the response variable is the number of persons injured by severity level. In this analysis, the count of crashes is used for PDO crashes since that type of crash does not involve injuries. The number of persons injured is Poisson distributed:

$$y_{ijt} \sim \text{Poisson}(\theta_{ijt}) \tag{1}$$

where $y_{ijt}$ is the observed number of persons injured in segment $i$ of the severity type $j$ at time $t$ (in years), and $\theta_{ijt}$ is the expected Poisson rate for segment $i$ of the severity type $j$ at time $t$. The Poisson rate is modeled as a function of the covariates following a lognormal distribution, as shown is Equation 2:

$$\log(\theta_{ijt}) = \beta_{0j} + \sum_{k=1}^{K} \beta_{jk} X_{ijtk} + v_{ij} \tag{2}$$

where

$\beta_{0j}$ = intercept for severity $j$,
$\beta_{jk}$ = coefficient for $k$ covariate and severity $j$,
$X_{ijtk}$ = value of the $k$ covariate for segment $i$ of the severity type $j$ at time $t$, and
$v_{ij}$ = random effects for each severity type.

These random effects capture the extra-Poisson heterogeneity among segments.

At the second stage, the coefficients ($\beta_{jk}$), including the intercepts, are modeled by using noninformative normal priors:

$$\beta_{jk} \sim \text{N}(0, 1000) \tag{3}$$

Now it can be assumed that the random effects for each severity type are independent and therefore have the following prior distributions:

$$v_{ij} \sim \text{N}(0, \tau_j^{-1}) \qquad j = 1, \ldots, 5 \tag{4}$$

where $\tau_j$ is the inverse of the variance (also known as precision). The precision has a gamma prior:

$$\tau_j \sim \text{gamma}(0.01, 0.001) \qquad j = 1, \ldots, 5 \tag{5}$$

with a mean of 10 and a variance of 10,000.

This specification is equivalent to the univariate Poisson lognormal specification for each severity level, but it offers the advantage of being directly comparable with the multivariate specification.

For the multivariate model, correlated priors in the random effects vector are estimated using multivariate normal priors (*15, 16*):

$$\mathbf{v}_i \sim \text{MN}(\boldsymbol{\mu}_i, \Sigma) \tag{6}$$

where $\boldsymbol{\mu}_i$ is a vector of zeroes $\boldsymbol{\mu}_i = (0, 0, 0, 0, 0)$ and $\Sigma$ is the variance–covariance matrix with a hyperprior defined by

$$\Sigma^{-1} \sim \text{Wishart}(\boldsymbol{R}, n) \tag{7}$$

where

$\Sigma^{-1}$ = symmetric positive definite matrix (also known as the precision matrix),
$\boldsymbol{R}$ = scale matrix =

$$\begin{bmatrix} 0.1 & 0.005 & 0.005 & 0.005 & 0.005 \\ 0.005 & 0.1 & 0.005 & 0.005 & 0.005 \\ 0.005 & 0.005 & 0.1 & 0.005 & 0.005 \\ 0.005 & 0.005 & 0.005 & 0.1 & 0.005 \\ 0.005 & 0.005 & 0.005 & 0.005 & 0.1 \end{bmatrix}$$

and
$n$ = degrees of freedom = 5.

The values of $\boldsymbol{R}$ and $n$ were recommended by Gelman et al. (*18*) and Carlin and Louis (*20*) to produce a noninformative prior for the precision matrix, analogous to the noninformative gamma distribution for the prior of the univariate precision shown in Equation 5.

When sites are ranked, the cost of the crash can be used. Two indexes are used: total crash cost and excess crash cost.

Total crash cost is defined by

$$\text{cost}_{it} = \sum_{j=1}^{5} \text{cost}_j \theta_{ijt} \tag{8}$$

where $\text{cost}_j$ is the cost associated with an injury of severity type $j$.

Excess crash cost is defined by

$$\delta - \text{cost}_{it} = \sum_{j=1}^{5} \text{cost}_j \delta_{ijt} \tag{9}$$

where $\delta_{ijt}$ is the expected excess injury frequency for segment $i$ of the severity type $j$ at time $t$, defined as [Aguero-Valverde and Jovanis (*21*)]:

$$\delta_{ijt} = e^{\beta_{0j} + \sum_{k=1}^{K} \beta_{jk} X_{ijtk}} \left(e^{v_i} - 1\right) \tag{10}$$

## DATA DESCRIPTION

The crash data for the models were aggregated by year at the road segment level and for state-maintained rural two-lane roads within District 2-0 of the Pennsylvania Department of Transportation (PennDOT). Figure 1 is a map of the District 2-0, including all state-maintained roads and rural two-lane roads in that region. The analysis included a total of 6,353 rural two-lane segments. Average crash cost data for the five levels of outcome severity used by PennDOT (for all of Pennsylvania) were obtained from the *2006 Pennsylvania Crash Facts and Statistics Book* (*22*).

A relational database was assembled with information from two different data sources: crash data and road inventory. All data were collected for calendar years 2003 to 2006.

### Crash Data

Crash data were obtained from the PennDOT Crash Reporting System. The data includes reportable crashes for road-segment locations only (i.e., those segments that do not include an intersection or ramp junction). Five severity levels are coded into the data: deaths, major injuries, moderate injuries, minor injuries, and PDO.

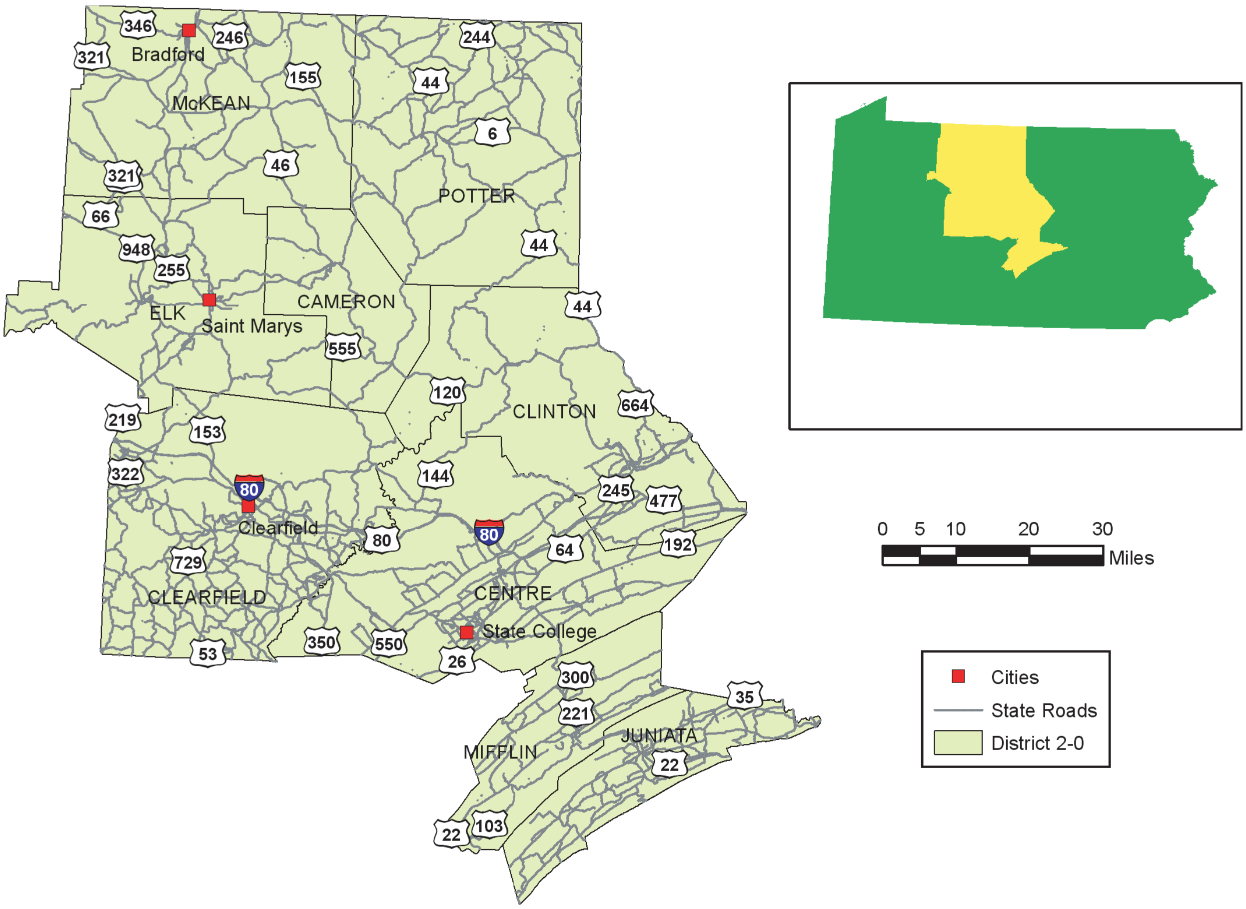


FIGURE 1 PennDOT District 2-0 roads.

A special location code was created for each crash by concatenating the county, route, and segment numbers in a single variable, creating a unique location identification for each road segment. Then crashes were summarized by location code and year.

## Road Inventory

Road data were obtained from the Pennsylvania Road Management System (RMS) for the study period. The Pennsylvania RMS includes data for each road segment such as county code (each county in Pennsylvania has a unique numeric identification code), State route number, road-segment number and length, average daily traffic (ADT), lane width, travel lane count, posted speed limit, and functional class. These data were joined with a digital map of roads in the state, obtained from Pennsylvania State University (*23*), which allowed researchers to map crash locations. Summary statistics for the data are shown in Table 1.

## RESULTS

Models were estimated using the open source software OpenBUGS (*24*). For the models, 1,000 iterations were discarded as burn-in, and the 20,000 iterations that followed were used to obtain summary statistics of the posterior distribution of parameters. Convergence was assessed by visual inspection of the Markov chains for the parameters. Furthermore, the number of iterations was selected so that the Monte Carlo error for each parameter in the model would be less than 10% of the value of the standard deviation of that parameter.

Table 2 presents the estimates for the independent and multivariate Poisson lognormal random effects models. Because the issue of primary interest is the differences and similarities between the two different modeling approaches, the following discussion of model fit and coefficient significance is focused in this direction.

The goodness-of-fit measures commonly used in full Bayesian statistics are presented in the table: the posterior mean of the deviance and the deviance information criterion (DIC) (*25*). The deviance is estimated in the same way for frequentist and Bayesian statistics while the DIC is the Bayesian equivalent of the Akaike information criterion. As in the case of their frequentist counterparts, the deviance and the DIC quantify the relative goodness-of-fit of the models; therefore, they are useful for comparing models.

In comparing the overall fit of the two models, it is clear from the summary statistics (which appear just after the PDO model table at the bottom of the right-hand column in Table 2) that the multivariate model better fits the data, as measured by the DIC. Although the multivariate model has a DIC more than 500 points lower than the independent model's DIC, the multivariate model's posterior deviance (Dbar) is only 57 points higher than that with the independent model. Furthermore, once the correlated random effects are included in the model, the effective number of parameters ($p_D$) drops significantly,

**TABLE 1 Summary Statistics of the Data by Segment and Year**

| Variable | Mean | SE | Min. | Max. |
|---|---|---|---|---|
| Deaths | 0.007 | 0.088 | 0 | 3 |
| Major injuries | 0.015 | 0.143 | 0 | 4 |
| Moderate injuries | 0.040 | 0.235 | 0 | 5 |
| Minor injuries | 0.111 | 0.452 | 0 | 11 |
| PDO crashes | 0.089 | 0.320 | 0 | 4 |
| Average daily traffic (ADT) | 1,683 | 2,251 | 23 | 18,749 |
| Length (mi) | 0.467 | 0.117 | 0.026 | 0.756 |
| Lane width <10 | 0.332 | 0.471 | | |
| Lane width ≥10 and <11 ft | 0.331 | 0.471 | | |
| Lane width ≥11 and <12 ft | 0.218 | 0.413 | | |
| Lane width = 12 ft | 0.087 | 0.282 | | |
| Lane width >12 ft | 0.032 | 0.176 | | |
| Shoulder width <3 ft | 0.447 | 0.497 | | |
| Shoulder width ≥3 and <4 ft | 0.158 | 0.365 | | |
| Shoulder width ≥4 and <6 ft | 0.256 | 0.436 | | |
| Shoulder width = 6 ft | 0.049 | 0.215 | | |
| Shoulder width >6 ft | 0.090 | 0.287 | | |
| Speed limit <35 mph | 0.052 | 0.223 | | |
| Speed limit ≥35 and <45 mph | 0.298 | 0.457 | | |
| Speed limit ≥45 and <55 mph | 0.256 | 0.436 | | |
| Speed limit ≥55 mph | 0.394 | 0.489 | | |

which explains the improvement in DIC. Indeed, even after controlling for the covariates, there is additional correlation between different severities that is not explained by the independence assumption.

The coefficients change little after including the multivariate random effects with the exception of the speed limit coefficients for fatalities, which decreased more than 10%. With a sample size this large (6,353 segments over 4 years), this result is not unexpected; the correlation among severities has little affect on coefficient estimates when large sample sizes are used. Consistently, only the smallest sample size outcome (fatalities) has any change at all and that change is only for one set of coefficients representing speed limit.

Notice, however, the pattern in the estimated standard deviation for the significant parameters. The multivariate model consistently provides far more precise estimates (i.e., smaller standard deviations) than does the independent model (e.g., 0.10 verses 0.11 for ADT in the fatalities model; 1.14 verses 1.52 for speed limit ≥35 and <45 mph in the fatalities model). This is the first indication of generally improved precision with the multivariate model.

Concerning the pattern of variables significance, the intercept and ADT are significant in each model, with ADT ranging from 0.48 to 0.69. Speed limit is significant in the fatalities model, with all segments with speeds above 35 mph having a positive parameter greater than 2.0 illustrating a highly nonlinear effect of speed on fatal-crash frequency.

Segment length is significant for all crash severity levels except for fatalities, ranging from 1.08 for major injuries to 0.82 for moderate injuries. Lane widths less than 10 ft show a protective effect, which seems counterintuitive; comparable results, however, have been found in an analysis of similar data using case-control methods by Gross and Jovanis (*26*). The use of narrow lanes may be a reflection of the terrain conditions in this part of Pennsylvania, which are included in the model.

Covariance estimates and correlations for the models are presented in Table 3. The table on the left summarizes the estimates of the mean and standard deviation (standard error) of the standard deviation of the crash estimates (for the independent model the covariance matrix is all zeroes in the off-diagonal cells, hence they are omitted for brevity). The table on the right summarizes both the correlation matrix and covariance matrix. The correlation matrix is highlighted by the use of gray cells. The correlations are significant and high, ranging from 0.47 to 0.97. Not surprisingly, adjacent severities are consistently highly correlated (e.g., 0.97 for correlation of fatal and major injury). The covariance matrix shows that covariances are all strongly significant (standard deviations in parenthesis). There is clearly a significant correlation between different crash severities, a finding consistent with Park and Lord (*15*), who used data from intersections in California.

The next step is to apply the model results to the identification of SWiPs. Using full Bayes methods, it is possible to estimate the precision of the expected crash frequency (and excess expected crash frequency) by calculating the variance and the standard deviation of the posterior distribution of these parameters. Table 4 presents the average standard deviations for the expected crash frequency in each road segment by severity level and correlation structure. Clearly, the estimates are more precise for the multivariate model for each class of severity. The improvements in precision are even more impressive with the higher severity levels; the average standard deviation decreases by 40.7% and 48% for, respectively, fatalities and major injuries.

The importance of these findings lies in the use of the standard deviation to assess SWiPs. As discussed in Hauer's seminal work (*1*) the standard deviation of the crash frequency (or excess crash frequency) may be used to consider the variability in crashes within a set of comparable sites. This variability can be used to give lower priority to a site that is a member of a set of comparable sites with high variability. In any assessment of fatal crashes in Pennsylvania, for example, it can now be ascertained whether a high frequency of crashes involving fatalities at a particular site is likely or unlikely given the standard deviation of frequencies (or excess frequencies) for the group of comparable sites as a whole.

Table 5 lists the top 40 road segments as ranked by total expected crash cost by the multivariate model. Table 5 also lists the total expected crash costs of these segments as estimated by the independent model, and also lists for these 40 segments the expected crash frequencies for both the multivariate and univariate (independent) models. According to the *2006 Pennsylvania Crash Facts and Statistics Book* (*22*), the costs per casualty associated with each level of crash severity are as follows: $3,043,560 for fatalities, $1,114,764 for major injury, $74,550 for moderate injury, $5,853 for minor injury, and $2,341 for PDO crashes. Given these costs, fatalities and major injuries represent an average of about 90% of the total cost of crashes per segment. For the top-ranked segments, the multivariate expected total crash cost is consistently higher than the independent estimate. In fact, out of the top 40 segments listed in Table 5, only three of them have higher expected cost from the independent model than from the multivariate model. The two severities that most affect the total cost, fatalities and major injuries, are also presented in Table 5. The expected number of fatalities and major injuries are also consistently higher for the multivariate model in comparison to the independent model. The road-segment rankings themselves differ significantly between the two models. Of the top 40 road segments as ranked by the multivariate model, 37.5% of them are not among the top 40 segments with the independent model, 30% of them are not among the top 100 from the independent model, and 10% of

**TABLE 2 Estimates for Independent and Multivariate Random Effects Models**

| | Independent | | | Multivariate | | |
|---|---|---|---|---|---|---|
| | Mean | SD | MC Error | Mean | SD | MC Error |
| **Fatalities** | | | | | | |
| Intercept | –12.91 | 1.93 | 0.1572 | –12.70 | 1.38 | 0.1070 |
| ADT | 0.65 | 0.11 | 0.0077 | 0.65 | 0.10 | 0.0071 |
| Length | 0.51 | 0.30 | 0.0094 | 0.50 | 0.30 | 0.0094 |
| Lane width <10 | –0.22 | 0.39 | 0.0169 | –0.27 | 0.39 | 0.0153 |
| Lane width ≥10 and <11 ft | 0.03 | 0.30 | 0.0106 | –0.01 | 0.30 | 0.0102 |
| Lane width ≥11 and <12 ft | 0.05 | 0.25 | 0.0063 | 0.02 | 0.25 | 0.0067 |
| Lane width >12 | –0.20 | 0.45 | 0.0074 | –0.21 | 0.45 | 0.0084 |
| Shoulder width <3 ft | 0.52 | 0.37 | 0.0154 | 0.53 | 0.36 | 0.0136 |
| Shoulder width ≥3 and <4 ft | –0.05 | 0.40 | 0.0135 | –0.01 | 0.39 | 0.0121 |
| Shoulder width ≥4 and <6 ft | 0.29 | 0.34 | 0.0123 | 0.29 | 0.33 | 0.0116 |
| Shoulder width >6 ft | 0.32 | 0.35 | 0.0110 | 0.38 | 0.35 | 0.0116 |
| Speed limit ≥35 and <45 mph | 2.34 | 1.52 | 0.1213 | 2.03 | 1.14 | 0.0851 |
| Speed limit ≥45 and <55 mph | 2.72 | 1.52 | 0.1216 | 2.41 | 1.15 | 0.0857 |
| Speed limit ≥55 mph | 2.76 | 1.52 | 0.1222 | 2.44 | 1.14 | 0.0859 |
| **Major Injuries** | | | | | | |
| Intercept | –9.60 | 0.92 | 0.0722 | –9.64 | 0.86 | 0.0664 |
| ADT | 0.55 | 0.08 | 0.0057 | 0.55 | 0.08 | 0.0054 |
| Length | 1.08 | 0.25 | 0.0100 | 1.05 | 0.23 | 0.0089 |
| Lane width <10 | –0.67 | 0.31 | 0.0140 | –0.69 | 0.30 | 0.0130 |
| Lane width ≥10 and <11 ft | 0.08 | 0.24 | 0.0096 | 0.10 | 0.22 | 0.0093 |
| Lane width ≥11 and <12 ft | 0.40 | 0.20 | 0.0068 | 0.37 | 0.20 | 0.0069 |
| Lane width >12 | 0.35 | 0.32 | 0.0075 | 0.54 | 0.30 | 0.0079 |
| Shoulder width <3 ft | 0.79 | 0.30 | 0.0140 | 0.80 | 0.28 | 0.0132 |
| Shoulder width ≥3 and <4 ft | 0.61 | 0.30 | 0.0126 | 0.66 | 0.29 | 0.0120 |
| Shoulder width ≥4 and <6 ft | 0.44 | 0.28 | 0.0117 | 0.47 | 0.26 | 0.0110 |
| Shoulder width >6 ft | 0.46 | 0.29 | 0.0114 | 0.50 | 0.28 | 0.0109 |
| Speed limit ≥35 and <45 mph | 0.49 | 0.45 | 0.0286 | 0.57 | 0.43 | 0.0257 |
| Speed limit ≥45 and <55 mph | 0.68 | 0.45 | 0.0290 | 0.75 | 0.43 | 0.0263 |
| Speed limit ≥55 mph | 0.70 | 0.45 | 0.0300 | 0.78 | 0.43 | 0.0271 |
| **Moderate Injuries** | | | | | | |
| Intercept | –6.81 | 0.50 | 0.0384 | –7.06 | 0.53 | 0.0412 |
| ADT | 0.48 | 0.05 | 0.0033 | 0.50 | 0.05 | 0.0033 |
| Length | 0.82 | 0.15 | 0.0052 | 0.81 | 0.15 | 0.0057 |
| Lane width <10 | –0.68 | 0.18 | 0.0075 | –0.72 | 0.18 | 0.0079 |
| Lane width ≥10 and <11 ft | –0.24 | 0.14 | 0.0049 | –0.26 | 0.14 | 0.0058 |
| Lane width ≥11 and <12 ft | –0.07 | 0.12 | 0.0033 | –0.10 | 0.12 | 0.0041 |
| Lane width >12 | 0.04 | 0.20 | 0.0041 | 0.04 | 0.20 | 0.0049 |
| Shoulder width <3 ft | –0.20 | 0.16 | 0.0065 | –0.10 | 0.17 | 0.0075 |
| Shoulder width ≥3 and <4 ft | 0.09 | 0.16 | 0.0059 | 0.19 | 0.17 | 0.0068 |
| Shoulder width ≥4 and <6 ft | –0.15 | 0.15 | 0.0052 | –0.08 | 0.15 | 0.0064 |
| Shoulder width >6 ft | –0.18 | 0.16 | 0.0049 | –0.06 | 0.17 | 0.0061 |
| Speed limit ≥35 and <45 mph | 0.54 | 0.26 | 0.0149 | 0.50 | 0.25 | 0.0145 |
| Speed limit ≥45 and <55 mph | 0.57 | 0.25 | 0.0150 | 0.54 | 0.25 | 0.0151 |
| Speed limit ≥55 mph | 0.43 | 0.25 | 0.0155 | 0.39 | 0.25 | 0.0154 |

| | Independent | | | Multivariate | | |
|---|---|---|---|---|---|---|
| | Mean | SD | MC Error | Mean | SD | MC Error |
| **Minor Injuries** | | | | | | |
| Intercept | –6.92 | 0.38 | 0.0301 | –6.96 | 0.34 | 0.0265 |
| ADT | 0.69 | 0.04 | 0.0029 | 0.70 | 0.03 | 0.0025 |
| Length | 1.00 | 0.11 | 0.0042 | 1.00 | 0.10 | 0.0039 |
| Lane width <10 | –0.45 | 0.12 | 0.0061 | –0.46 | 0.12 | 0.0061 |
| Lane width ≥10 and <11 ft | –0.18 | 0.10 | 0.0040 | –0.18 | 0.10 | 0.0043 |
| Lane width ≥11 and <12 ft | –0.04 | 0.08 | 0.0029 | –0.06 | 0.08 | 0.0032 |
| Lane width >12 | 0.03 | 0.13 | 0.0037 | –0.01 | 0.13 | 0.0042 |
| Shoulder width <3 ft | 0.18 | 0.12 | 0.0056 | 0.19 | 0.11 | 0.0052 |
| Shoulder width ≥3 and <4 ft | 0.03 | 0.12 | 0.0050 | 0.04 | 0.11 | 0.0047 |
| Shoulder width ≥4 and <6 ft | 0.02 | 0.11 | 0.0047 | 0.03 | 0.10 | 0.0044 |
| Shoulder width >6 ft | 0.12 | 0.12 | 0.0047 | 0.14 | 0.11 | 0.0043 |
| Speed limit ≥35 and <45 mph | 0.10 | 0.16 | 0.0089 | 0.08 | 0.15 | 0.0085 |
| Speed limit ≥45 and <55 mph | 0.08 | 0.16 | 0.0093 | 0.05 | 0.15 | 0.0090 |
| Speed limit ≥55 mph | –0.11 | 0.16 | 0.0096 | –0.14 | 0.15 | 0.0092 |
| **PDO** | | | | | | |
| Intercept | –6.40 | 0.32 | 0.0237 | –6.40 | 0.30 | 0.0223 |
| ADT | 0.66 | 0.03 | 0.0022 | 0.65 | 0.03 | 0.0019 |
| Length | 0.97 | 0.09 | 0.0033 | 0.96 | 0.09 | 0.0032 |
| Lane width <10 | –0.49 | 0.11 | 0.0049 | –0.50 | 0.11 | 0.0043 |
| Lane width ≥10 and <11 ft | 0.03 | 0.08 | 0.0031 | 0.02 | 0.08 | 0.0028 |
| Lane width ≥11 and <12 ft | 0.04 | 0.07 | 0.0018 | 0.03 | 0.07 | 0.0018 |
| Lane width >12 | 0.08 | 0.11 | 0.0022 | 0.07 | 0.11 | 0.0024 |
| Shoulder width <3 ft | 0.04 | 0.09 | 0.0036 | 0.07 | 0.10 | 0.0038 |
| Shoulder width ≥3 and <4 ft | 0.01 | 0.10 | 0.0030 | 0.04 | 0.10 | 0.0032 |
| Shoulder width ≥4 and <6 ft | –0.07 | 0.09 | 0.0027 | –0.04 | 0.09 | 0.0029 |
| Shoulder width >6 ft | –0.14 | 0.09 | 0.0026 | –0.11 | 0.09 | 0.0027 |
| Speed limit ≥35 and <45 mph | 0.07 | 0.13 | 0.0065 | 0.06 | 0.13 | 0.0071 |
| Speed limit ≥45 and <55 mph | –0.04 | 0.13 | 0.0067 | –0.05 | 0.14 | 0.0074 |
| Speed limit ≥55 mph | –0.15 | 0.13 | 0.0071 | –0.15 | 0.14 | 0.0077 |
| **Summary Statistics** | | | | | | |
| | Dbar | DIC | $p_D$ | Dbar | DIC | $p_D$ |
| | 39,896 | 43,026 | 3,130 | 39,953 | 42,439 | 2,487 |

NOTE: Shaded cells indicate significant coefficients at $\alpha = 5\%$ level. Dbar = posterior deviance, DIC = deviance information criterion, and $p_D$ = effective number of parameters.

**TABLE 3 Covariance Estimates for Random Effects Independent and Multivariate Models**

| Variance Independent Model | | | | Covariance/Correlation Matrix Multivariate Model[a] | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Severity | Mean | SD | MC Error | | Fatal | Major | Moderate | Minor | PDO |
| $\sigma_{FATAL}$ | 1.50 | 0.44 | 0.0334 | Fatal | 1.68 (0.41) | 1.73 (0.31) | 1.48 (0.15) | 1.03 (0.12) | 0.31 (0.07) |
| $\sigma_{MAJ}$ | 2.09 | 0.29 | 0.0179 | Maj | 0.97 (0.01) | 1.91 (0.28) | 1.64 (0.14) | 1.14 (0.10) | 0.31 (0.07) |
| $\sigma_{MOD}$ | 1.29 | 0.13 | 0.0068 | Mod | 0.93 (0.03) | 0.97 (0.02) | 1.51 (0.16) | 1.10 (0.07) | 0.31 (0.05) |
| $\sigma_{MIN}$ | 1.13 | 0.07 | 0.0031 | Min | 0.74 (0.07) | 0.77 (0.07) | 0.83 (0.04) | 1.16 (0.08) | 0.41 (0.04) |
| $\sigma_{PDO}$ | 0.26 | 0.05 | 0.0032 | PDO | 0.49 (0.11) | 0.47 (0.10) | 0.52 (0.09) | 0.79 (0.07) | 0.24 (0.04) |

NOTE: MC = Monte Carlo.
[a]Shaded cells are correlation coefficients; standard deviations in parentheses.

them are not among the top 200. Clearly, the multivariate model generates different rankings for priority safety investment than does the univariate approach (which assumes independence across severity levels).

Table 6 shows the top 40 road segments ranked by excess crash cost as estimated by the multivariate model. Table 6 also lists for these 40 segments the excess crash cost as estimated by the independent model. As was seen in the rankings by total cost, the rankings by excess cost indicate significant differences between the multivariate and independent models. Among the top 40 road segments ranked, only two have higher expected excess cost from the independent than from the multivariate model. Moreover, of the top 40 road segments as ranked by the multivariate model, 47.5% of them are not among the top 40 segments with the independent model; 22.5% of them are not among the top 100 from the independent model, and 5% of them are not among the top 200.

The differences in the rankings are due to the differences in the random effects in the multivariate model and independent models. The positive correlations mean that sites are "borrowing strength" from adjacent severity levels including the fatal and major injury categories, which have the highest average cost. As a result, random effects terms increase, which subsequently increases both expected and excess crash costs.

## CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Traditionally, highway safety analyses have used univariate Poisson or negative binomial distributions to model crash counts for different levels of crash severity. The counts are often treated as a whole, combining crashes of different severities of outcome. Alternatively, safety analysts may arbitrarily choose outcome levels for inclusion in a study (e.g., fatal and injury crashes only). The implications of making a decision to aggregate in whole or in part across severity levels are not well understood. There has been limited research into the nature of the correlation across outcome severity levels and even more limited examination of how these correlations may influence safety analyses.

This research sought to estimate and better understand the implications of the correlations present across severities of crash outcomes. The specific modeling tool used to conduct this assessment was a full Bayes hierarchical model specifically formulated for this purpose. The approach estimated and compared two sets of models with identical specifications: one set estimated with the assumption of independence across outcomes, the second set with correlations estimated across outcomes as an explicit part of the model formulation (a nonindependence assumption). Both model parameter estimates and their variances were discussed. Finally, these two models were used in the development of SWiPs, highlighting the advantages from using a formulation that explicitly includes the correlation structure.

A series of statistical models were developed, using crash roadway and traffic data for 4 years (2003 to 2006) from central Pennsylvania. The inclusion of multivariate random effects for different crash severities significantly improved the precision of the crash frequency estimates for each level of severity; the multivariate model also fit the data significantly better as measured by the DIC. The average standard deviation of the crash frequency estimates was reduced 20%, but, more importantly, the standard deviation of the most severe outcomes were reduced even further (almost 41% for fatalities and 48% for major injuries). The effect of pooling strength across severity classes by accounting for multivariate responses was very significant.

Considering estimated parameter values in models of expected and excess crash frequency, the mean values did not change significantly with the inclusion of multivariate random effects (likely because of the large sample size in this study). The exception is the speed limit coefficients for fatalities, which decreased more than 10%. The precision of the parameter estimates for significant variables in the multivariate models were much smaller than in the independent models. This is another effect of utilizing the multivariate formulation.

In terms of rankings, it was found that for the top-ranked segments, the multivariate estimates of total crash cost and excess crash cost are consistently higher than the independent estimates. The two severities that most affect crash cost, fatalities and major injuries, also present consistently higher frequency estimates for the multivariate model

**TABLE 4 Average Standard Deviations for the Expected Crash Frequency in Each Road Segment by Severity Level and Correlation Structure**

| Severity Level | Independent | Multivariate | % Decrease |
|---|---|---|---|
| Fatalities | 0.0107 | 0.0076 | 40.7 |
| Major injuries | 0.0243 | 0.0164 | 48.1 |
| Moderate injuries | 0.0431 | 0.0355 | 21.5 |
| Minor injuries | 0.0870 | 0.0787 | 10.5 |
| PDO | 0.0442 | 0.0367 | 20.5 |
| Total | 0.0419 | 0.0350 | 19.7 |

**TABLE 5 Ranking of Segments by Total Crash Cost**

| | Expected Total Crash Cost | | | | | | Expected Crash Frequency | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Multivariate | | | Independent | | | Fatal Multivariate | | Fatal Independent | | Major Injury Multivariate | | Major Injury Independent | |
| Segment | Rank | Mean | SD | Rank | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 575 | 1 | 1,463,900 | 624,447 | 1 | 1,031,360 | 696,124 | 0.293 | 0.154 | 0.296 | 0.228 | 0.414 | 0.183 | 0.032 | 0.048 |
| 2,499 | 2 | 1,130,960 | 496,363 | 5 | 659,386 | 376,976 | 0.181 | 0.100 | 0.057 | 0.069 | 0.445 | 0.202 | 0.396 | 0.279 |
| 4,395 | 3 | 1,013,830 | 472,112 | 12 | 461,039 | 291,719 | 0.170 | 0.097 | 0.013 | 0.019 | 0.388 | 0.190 | 0.340 | 0.256 |
| 1,290 | 4 | 956,511 | 501,013 | 2 | 784,511 | 446,272 | 0.150 | 0.094 | 0.016 | 0.024 | 0.404 | 0.220 | 0.642 | 0.395 |
| 299 | 5 | 935,428 | 434,679 | 3 | 691,820 | 389,023 | 0.118 | 0.075 | 0.051 | 0.065 | 0.446 | 0.215 | 0.441 | 0.301 |
| 3,097 | 6 | 915,089 | 447,281 | 15 | 447,148 | 297,856 | 0.156 | 0.091 | 0.054 | 0.067 | 0.347 | 0.176 | 0.223 | 0.196 |
| 543 | 7 | 855,290 | 410,288 | 23 | 412,754 | 262,171 | 0.143 | 0.080 | 0.026 | 0.036 | 0.330 | 0.170 | 0.254 | 0.211 |
| 5,621 | 8 | 786,423 | 384,643 | 18 | 437,118 | 281,466 | 0.084 | 0.056 | 0.006 | 0.010 | 0.423 | 0.217 | 0.331 | 0.250 |
| 5,231 | 9 | 726,874 | 470,666 | 6 | 655,988 | 478,153 | 0.115 | 0.086 | 0.013 | 0.019 | 0.303 | 0.208 | 0.543 | 0.426 |
| 1,420 | 10 | 725,354 | 405,004 | 7 | 646,187 | 427,428 | 0.124 | 0.081 | 0.127 | 0.123 | 0.276 | 0.155 | 0.222 | 0.188 |
| 1,491 | 11 | 705,876 | 366,960 | 10 | 537,698 | 340,845 | 0.118 | 0.072 | 0.074 | 0.081 | 0.272 | 0.148 | 0.259 | 0.210 |
| 2,555 | 12 | 634,121 | 336,862 | 17 | 439,775 | 296,403 | 0.095 | 0.061 | 0.011 | 0.018 | 0.269 | 0.147 | 0.352 | 0.261 |
| 5,460 | 13 | 623,345 | 334,139 | 117 | 181,982 | 124,709 | 0.108 | 0.069 | 0.013 | 0.020 | 0.218 | 0.125 | 0.075 | 0.094 |
| 1,438 | 14 | 589,249 | 301,954 | 75 | 220,195 | 146,619 | 0.072 | 0.047 | 0.013 | 0.018 | 0.287 | 0.157 | 0.107 | 0.119 |
| 3,108 | 15 | 568,936 | 331,974 | 13 | 449,295 | 380,585 | 0.102 | 0.073 | 0.112 | 0.119 | 0.183 | 0.114 | 0.082 | 0.104 |
| 248 | 16 | 537,807 | 308,085 | 111 | 190,391 | 178,921 | 0.098 | 0.066 | 0.045 | 0.057 | 0.175 | 0.105 | 0.018 | 0.033 |
| 3,727 | 17 | 525,154 | 296,584 | 63 | 239,184 | 188,631 | 0.083 | 0.055 | 0.038 | 0.051 | 0.212 | 0.125 | 0.082 | 0.097 |
| 1,362 | 18 | 519,977 | 303,460 | 35 | 309,998 | 237,867 | 0.075 | 0.053 | 0.013 | 0.021 | 0.230 | 0.139 | 0.228 | 0.205 |
| 3,642 | 19 | 512,164 | 309,731 | 74 | 220,579 | 203,506 | 0.083 | 0.057 | 0.051 | 0.064 | 0.211 | 0.133 | 0.031 | 0.048 |
| 316 | 20 | 498,521 | 329,751 | 24 | 403,896 | 297,372 | 0.089 | 0.066 | 0.051 | 0.066 | 0.182 | 0.125 | 0.214 | 0.194 |
| 3,203 | 21 | 496,431 | 296,435 | 272 | 120,637 | 90,081 | 0.073 | 0.051 | 0.015 | 0.023 | 0.217 | 0.137 | 0.029 | 0.045 |
| 2,692 | 22 | 487,661 | 319,249 | 36 | 307,178 | 255,942 | 0.098 | 0.072 | 0.060 | 0.074 | 0.142 | 0.097 | 0.090 | 0.107 |
| 3,148 | 23 | 483,899 | 342,137 | 8 | 614,018 | 364,048 | 0.082 | 0.066 | 0.011 | 0.018 | 0.195 | 0.145 | 0.517 | 0.324 |
| 5,048 | 24 | 481,550 | 282,088 | 120 | 180,686 | 147,406 | 0.096 | 0.065 | 0.020 | 0.030 | 0.132 | 0.083 | 0.082 | 0.103 |
| 3,045 | 25 | 474,246 | 339,219 | 4 | 676,385 | 556,765 | 0.095 | 0.081 | 0.197 | 0.181 | 0.137 | 0.100 | 0.060 | 0.073 |
| 4,840 | 26 | 470,119 | 273,827 | 343 | 108,073 | 83,047 | 0.068 | 0.046 | 0.014 | 0.021 | 0.197 | 0.125 | 0.027 | 0.043 |
| 1,430 | 27 | 469,893 | 293,454 | 38 | 305,625 | 242,845 | 0.081 | 0.057 | 0.057 | 0.069 | 0.176 | 0.116 | 0.101 | 0.109 |
| 21 | 28 | 466,355 | 282,426 | 104 | 194,405 | 175,456 | 0.071 | 0.049 | 0.044 | 0.055 | 0.196 | 0.124 | 0.027 | 0.043 |
| 4,399 | 29 | 450,997 | 295,586 | 72 | 223,216 | 197,116 | 0.084 | 0.062 | 0.042 | 0.055 | 0.154 | 0.105 | 0.072 | 0.092 |
| 1,073 | 30 | 450,811 | 265,916 | 134 | 174,020 | 125,372 | 0.076 | 0.053 | 0.014 | 0.021 | 0.161 | 0.104 | 0.079 | 0.095 |
| 356 | 31 | 448,026 | 309,474 | 21 | 421,627 | 269,530 | 0.090 | 0.074 | 0.041 | 0.052 | 0.133 | 0.099 | 0.257 | 0.196 |
| 3,552 | 32 | 446,649 | 304,110 | 9 | 591,583 | 391,711 | 0.071 | 0.060 | 0.084 | 0.096 | 0.185 | 0.130 | 0.291 | 0.233 |
| 1,206 | 33 | 445,245 | 269,330 | 26 | 359,304 | 255,456 | 0.049 | 0.036 | 0.013 | 0.019 | 0.229 | 0.150 | 0.263 | 0.223 |
| 2,674 | 34 | 441,738 | 266,688 | 27 | 347,122 | 253,092 | 0.065 | 0.049 | 0.016 | 0.024 | 0.181 | 0.117 | 0.243 | 0.215 |
| 3,023 | 35 | 433,324 | 254,072 | 413 | 100,276 | 68,869 | 0.066 | 0.044 | 0.012 | 0.017 | 0.178 | 0.110 | 0.021 | 0.035 |
| 3,079 | 36 | 425,854 | 258,394 | 114 | 187,208 | 143,978 | 0.071 | 0.048 | 0.018 | 0.027 | 0.164 | 0.105 | 0.093 | 0.106 |
| 5,027 | 37 | 425,204 | 261,600 | 118 | 181,856 | 141,073 | 0.078 | 0.055 | 0.018 | 0.028 | 0.138 | 0.091 | 0.085 | 0.100 |
| 2,497 | 38 | 424,659 | 264,163 | 41 | 292,267 | 239,289 | 0.074 | 0.052 | 0.055 | 0.068 | 0.153 | 0.099 | 0.097 | 0.109 |
| 3,684 | 39 | 422,109 | 264,282 | 258 | 123,824 | 132,815 | 0.066 | 0.048 | 0.029 | 0.042 | 0.176 | 0.114 | 0.017 | 0.031 |
| 1,374 | 40 | 420,934 | 255,905 | 150 | 165,040 | 145,262 | 0.069 | 0.053 | 0.026 | 0.038 | 0.144 | 0.093 | 0.057 | 0.078 |

**TABLE 6 Ranking of Segments by Excess Crash Cost**

| | Multivariate | | | Independent | | |
|---|---|---|---|---|---|---|
| Segment | Rank | Mean | SD | Rank | Mean | SD |
| 575 | 1 | 1,381,400 | 623,209 | 1 | 925,998 | 699,163 |
| 2,499 | 2 | 1,231,000 | 540,879 | 6 | 601,361 | 374,500 |
| 4,395 | 3 | 1,095,810 | 515,308 | 13 | 427,013 | 288,649 |
| 1,290 | 4 | 992,647 | 511,624 | 2 | 741,458 | 446,328 |
| 299 | 5 | 915,394 | 436,201 | 5 | 623,317 | 387,846 |
| 3,097 | 6 | 910,194 | 461,987 | 16 | 393,131 | 302,125 |
| 5,621 | 7 | 854,429 | 410,229 | 12 | 427,851 | 286,236 |
| 543 | 8 | 829,056 | 427,465 | 23 | 326,320 | 259,349 |
| 5,231 | 9 | 773,622 | 494,341 | 4 | 626,249 | 484,309 |
| 1,420 | 10 | 764,266 | 427,776 | 7 | 592,645 | 452,763 |
| 1,491 | 11 | 692,627 | 404,399 | 11 | 440,102 | 344,248 |
| 2,555 | 12 | 675,869 | 363,010 | 14 | 414,690 | 296,487 |
| 5,460 | 13 | 621,519 | 347,861 | 103 | 145,124 | 126,905 |
| 3,108 | 14 | 579,860 | 356,155 | 15 | 397,529 | 386,824 |
| 1,362 | 15 | 557,514 | 338,080 | 31 | 269,661 | 238,103 |
| 1,438 | 16 | 543,349 | 313,527 | 81 | 169,760 | 149,338 |
| 248 | 17 | 527,173 | 321,883 | 95 | 147,698 | 178,387 |
| 3,148 | 18 | 514,689 | 355,561 | 8 | 591,843 | 369,936 |
| 3,727 | 19 | 496,039 | 301,461 | 60 | 202,668 | 197,478 |
| 316 | 20 | 493,089 | 336,880 | 22 | 349,589 | 287,986 |
| 1,634 | 21 | 459,372 | 312,547 | 38 | 255,188 | 250,432 |
| 5,048 | 22 | 458,180 | 295,973 | 132 | 123,234 | 140,077 |
| 1,206 | 23 | 449,461 | 292,492 | 24 | 312,597 | 269,249 |
| 1,073 | 24 | 445,121 | 282,450 | 115 | 131,435 | 123,199 |
| 3,642 | 25 | 444,099 | 306,317 | 87 | 162,252 | 204,164 |
| 2,692 | 26 | 442,257 | 312,746 | 44 | 248,478 | 257,264 |
| 3,203 | 27 | 439,753 | 289,737 | 308 | 71,254 | 90,260 |
| 4,840 | 28 | 437,523 | 283,425 | 329 | 62,845 | 82,440 |
| 1,374 | 29 | 432,002 | 272,646 | 106 | 143,601 | 150,785 |
| 356 | 30 | 430,080 | 284,964 | 18 | 381,518 | 273,252 |
| 3,045 | 31 | 428,647 | 298,846 | 3 | 633,512 | 568,306 |
| 2,695 | 32 | 428,444 | 292,071 | 76 | 178,250 | 190,523 |
| 4,399 | 33 | 421,215 | 293,546 | 67 | 187,186 | 194,607 |
| 4,357 | 34 | 418,641 | 286,661 | 17 | 389,961 | 291,422 |
| 21 | 35 | 417,030 | 276,093 | 98 | 147,041 | 178,820 |
| 4,456 | 36 | 412,911 | 264,204 | 144 | 118,481 | 130,660 |
| 4,302 | 37 | 410,727 | 269,200 | 199 | 101,597 | 108,408 |
| 1,127 | 38 | 408,167 | 282,649 | 47 | 241,840 | 213,928 |
| 2,017 | 39 | 405,971 | 281,972 | 48 | 239,240 | 211,598 |
| 4,638 | 40 | 405,704 | 266,430 | 183 | 106,236 | 101,755 |

as compared with the independent model. This, in turn, produces different rankings for the multivariate and independent models. Therefore, the implication of the differences in precision is that the identification of SWiPs results in different rankings: an important practical consequence of the treatment of severity level.

There are several fruitful areas for future research that evolve from this research. Investigating and quantifying the effect of multivariate modeling when small samples are used is an interesting future research area. It is possible that differences in parameter estimates as well as precision may result in studies with smaller data sets.

Future replications with different data sets and infrastructures (e.g., intersections, ramps) are needed to gather better estimates of these correlations and their implications. Improvements in precision on the order of 40% and more were unexpected; the researchers expected some improvement, but the magnitude was a surprise. More extensive tests with additional data sets will allow researchers to reach more general expectations concerning precision.

## ACKNOWLEDGMENT

The authors acknowledge the contribution of the Pennsylvania Department of Transportation, which provided the data for this analysis.

## REFERENCES


1. Hauer, E. Identification of Sites with Promise. In *Transportation Research Record 1542*, TRB, National Research Council, Washington, D.C., 1996, pp. 54–60.
2. Congdon, P. *Bayesian Statistical Modeling.* John Wiley & Sons, West Sussex, United Kingdom, 2001.
3. Lord, D. 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 and Prevention*, Vol. 38, 2006, pp. 751–766.
4. Lord, D., S. P. Washington, and J. N. Ivan. Poisson, Poisson–Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention,* Vol. 37, No. 1, 2005, pp. 35–46.
5. Lord, D., S. P. Washington, and J. N. Ivan. Further Notes on the Application of Zero-Inflated Models in Highway Safety. *Accident Analysis and Prevention,* Vol. 39, No. 1, 2007, pp. 53–57.
6. Hauer, E., J. Kononov, B. Allery, and M. S. Griffith. Screening the Road Network for Sites with Promise. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 27–32.
7. Hauer, E., B. K. Allery, J. Kononov, and M. S. Griffith. How Best to Rank Sites with Promise. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 48–54.
8. Tarko, A. P., and M. Kanodia. Effective and Fair Identification of Hazardous Locations. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 64–70.
9. Tanaru, R. Hierarchical Bayesian Models for Multiple Count Data. *Austrian Journal of Statistics,* Vol. 31, No. 3, 2002, pp. 221–229.
10. Ladrón de Guevara, F., S. P. Washington, and J. Oh. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 191–199.
11. Miaou, S.-P., and J. J. Song. Bayesian Ranking of Sites for Engineering Safety Improvements: Decision Parameter, Treatability Concept, Statistical Criterion, and Spatial Dependence. *Accident Analysis and Prevention,* Vol. 37, No. 4, 2005, pp. 699–720.
12. Bijleveld, F. D. 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, 2005, pp. 591–600.
13. Song, J. J., M. Ghosh, S. Miaou, and B. Mallick. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis,* Vol. 97, 2006, pp. 246–273.

14. Ma, J., and K. M. Kockelman. Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1950,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 24–34.
15. Park, E. S., and D. Lord. Multivariate Poisson–Lognormal Models for Jointly Modeling Crash Frequency by Severity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2019,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–6.
16. Ma, J., K. M. Kockelman, and P. Damien. A Multivariate Poisson Lognormal Regression Model for Prediction of Crash Counts by Severity, Using Bayesian Methods. *Accident Analysis and Prevention,* Vol. 40, No. 3, 2008, pp. 964–975.
17. Banerjee, S., B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall/CRC, Boca Raton, Fla., 2004.
18. Gelman, A., J. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis,* 2nd ed. Chapman & Hall/CRC, Boca Raton, Fla., 2003.
19. Congdon, P. *Applied Bayesian Modeling.* John Wiley & Sons, West Sussex, United Kingdom, 2003.
20. Carlin, B. P., and T. A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall, London, 1996.
21. Aguero-Valverde, J., and P. P. Jovanis. Identifying Road Segments with High Risk of Weather-Related Crashes Using Full Bayesian Hierarchical Models. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
22. Bureau of Highway Safety and Traffic Engineering, Pennsylvania Department of Transportation. *2006 Pennsylvania Crash Facts and Statistics Book.* www.dot.state.pa.us/Internet/Bureaus/pdBHSTE.nsf/BHSTEHomepage?OpenFrameset. Accessed March 20, 2008.
23. Pennsylvania State University. *Pennsylvania Spatial Data Access.* www.pasda.psu.edu/. Accessed May 2007.
24. Thomas, A., B. O'Hara, U. Ligges, and S. Sturtz. Making BUGS Open. *R News,* Vol. 6, No. 1, 2006, pp. 12–17.
25. Spiegelhalter, D., N. Best, B. P. Carlin, and A. Linde. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B,* Vol. 64, Part 4, 2002, pp. 583–639.
26. Gross, F., and P. P. Jovanis. Estimation of the Safety Effectiveness of Lane and Shoulder Width: Case-Control Approach. *Journal of Transportation Engineering,* Vol. 133, No. 6, 2007, pp. 362–369.

*The Statistical Methodology and Statistical Computer Software in Transportation Research Committee sponsored publication of this paper.*