# Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity

Jianming Ma and Kara M. Kockelman

**In practice, crash and injury counts are modeled by using a single equation or a series of independently specified equations, which may neglect shared information in unobserved error terms, reduce efficiency in parameter estimates, and lead to biases in sample databases. This paper offers a multivariate Poisson specification that simultaneously models injuries by severity. Parameter estimation is performed within the Bayesian paradigm with a Gibbs sampler for crashes on Washington State highways. Parameter estimates and goodness-of-fit measures are compared with a series of independent Poisson equations, and a cost–benefit analysis of a 10-mph speed limit change is provided as an example application.**

In the United States, traffic crashes bring about more loss of human life (as measured in person-years) than any other cause except cancer and heart disease (*1*). The annual cost of such crashes is estimated to be \$231 billion, or \$820 per capita in 2000 (*2*). The costs do not include the cost of delays imposed on other travelers, which are also significant, particularly when crashes occur on busy roadways. For example, Schrank and Lomax (*3*) estimate that more than one-half of all traffic delays are due to nonrecurring events, such as crashes, and cost on the order of \$1,000 per peak-period driver per year, particularly in urban areas. Thus, while vehicle and roadway design are improving and growing congestion may be reducing impact speeds, crashes are becoming more critical in many ways, particularly in societies that continue to motorize.

There has been considerable crash prediction research and analysis (e.g., *4–11*). Crash frequencies are commonly collected by severity on relatively homogeneous roadway segments. In virtually all cases, frequency is modeled separately from severity.

There are several drawbacks to separate analyses. First, such approaches may result in a substantial decrease in estimator efficiency, since any relationship between crash severity and frequency is ignored. (For example, sites that are more prone to crashes may exhibit higher proportions of less severe injuries.) Second, severity analysis can only be conducted once a crash has occurred—and thus only on sites where crashes have transpired, which results in a biased site sample. Finally, joint probabilities (of crash occurrence and severity) better characterize overall risk than marginal or conditional probabilities.

By using a multivariate Poisson specification, as well as Bayesian techniques, this paper presents a joint model of crash frequency and severity (as measured in terms of crash-involved occupants). A Gibbs sampler was constructed to create distributions of all parameter estimates. The data come from the Highway Safety Information System (HSIS) database and cover all Washington State highways in 1996. The results lend themselves to recommendations for highway safety treatments and design policies.

University of Texas at Austin, 6.9 E. Cockrell Jr. Hall, Austin, TX 78712-1076.

*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.

## LITERATURE REVIEW

Models of crash (or injury) counts can be classified into two major streams: (*a*) the conventional univariate Poisson and related models, such as the negative binomial, and (*b*) potentially more realistic specifications, such as the multivariate Poisson (MVP). The first stream of models has provided a means of investigating associations between crash frequency and many crucial factors, such as traffic volume, access density, posted speed limit, and number of lanes (e.g., *12–23*). There also has been considerable interest in models that allow for excessive zeros, such as zero-inflated Poisson and zero-inflated negative binomial regression approaches (e.g., *11, 24–33*).

Thanks to computational and statistical advances, panel data, in which a cross section (of segments, intersections, etc.) is observed over time, have become more amenable to rigorous analysis. In traffic crash analyses, many unobserved explanatory variables affect frequencies and severities. Panel data can be used to deal with heterogeneity in the individuals. To address the heterogeneity issue across individuals, many recent studies have used (univariate) panel count data models, such as fixed- and random-effects negative binomial specifications (*9, 34*).

Such past research endeavors, however, have neglected the role of unobserved factors across different types of counts (e.g., the number of fatalities and the number of debilitating injuries). Recognizing the need for such considerations, Ladron de Guevara et al. (*35*) investigated the simultaneity of fatality and injury crash outcomes. Bijleveld (*36*) also examined the correlation structure between crash and injury counts. As expected, he found significant correlations. However, he did not control for any covariates. Multivariate models (of count data) can correct for this. A particular MVP application of such models is the focus of this paper.

Ideally, the frequency of traffic crashes by severity is simultaneously modeled by using multivariate count data models, such as an MVP or multivariate zero-inflated Poisson (MVZIP) regression model. [See, e.g., Li et al. (*37*) for their MVZIP model of manufacture defects.]

Unfortunately, parameters in most MVP model specifications are difficult to estimate. Karlis (*38*) developed an expectation maximization algorithm for estimating the class of such models that is described in the following section. Christiansen et al. (*39*) developed a univariate hierarchical Bayesian Poisson model for investigating crash counts. MacNab (*40*) proposed and applied a Bayesian hierarchical model in his investigations of crashes using surveillance data. Miaou and Song (*29*) used Bayesian methodologies in ranking roadway sites for safety improvements; they adopted a multivariate spatial generalized linear mixed model to predict crash counts by severity.

However, to the authors' knowledge, no study has applied Bayesian methods to estimate MVP models of injury frequencies by severity. Of course, Bayesian methods generate a multivariate posterior distribution across all parameters of interest, as opposed to the traditional maximum likelihood estimation, which only offers the mode of parameters (and relies on asymptotic properties to ascertain covariance).

This paper introduces an MVP approach to model injury counts by severity simultaneously. A Gibbs sampler and Metropolis–Hastings (M-H) algorithms are established to estimate the parameters of interest for the Bayesian statistical inference. For comparison purposes, independent (univariate) Poisson models for injury counts are also estimated.

## MODEL STRUCTURE AND ESTIMATION

### Mathematical Formulation

For ease of presentation, a trivariate MVP mathematical formulation for analyzing counts of crash-involved persons across three levels of injury severity is described. Extending the specification to accommodate additional levels of severity (e.g., five levels) is conceptually and mathematically straightforward. Consider a sample $\{\mathbf{y}_i; i = 1, 2, \ldots, n\}$ from a trivariate Poisson distribution, where $\mathbf{y}_i = [y_{i1}, y_{i2}, y_{i3}]'$ denotes the number of crash-involved persons on the *i*th roadway segment in the sample experiencing no injury ($y_{i1}$), injury ($y_{i2}$), or fatal injury ($y_{i3}$), over a given time period (such as a year). According to Karlis (*38*), the general trivariate Poisson model is specified as follows:

$$\mathbf{y}_i = A\mathbf{z}_i \tag{1}$$

where

$$A = \begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \text{ and}$$

$$A_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Substituting matrix $A$ into Equation 1 gives the following:

$$\begin{aligned} y_{i1} &= z_{i1} + z_{i12} + z_{i13} + z_{i123} \\ y_{i2} &= z_{i2} + z_{i12} + z_{i23} + z_{i123} \\ y_{i3} &= z_{i3} + z_{i13} + z_{i23} + z_{i123} \end{aligned} \tag{2}$$

where all $z_{ik}$'s are independently Poisson distributed random variables with parameters $\theta_{ik}$, $k \in \{1, 2, 312, 13, 23, 213\}$. Parameters $\theta_{ikj}$ are actually covariance parameters between $Y_{ik}$ and $Y_{ij}$, and $\theta_{ijkl}$ is a common three-way covariance parameter among $Y_{ik}$, $Y_{ij}$, and $Y_{il}$.

For ease of implementation, the following assumption is made for the trivariate Poisson distribution, as used by Tsionas (*41*) for his models of forest damage:

$$\begin{aligned} y_{i1} &= z_{i1} + \delta_i \\ y_{i2} &= z_{i2} + \delta_i \\ y_{i3} &= z_{i3} + \delta_i \end{aligned} \tag{3}$$

where $z_{i1}$, $z_{i2}$, $z_{i3}$, $\delta_i$ have independent Poisson distributions with parameters $\theta_{i1}$, $\theta_{i2}$, $\theta_{i3}$, $\lambda$, respectively, for each $i = 1, 2, \ldots, n$.

Like the univariate Poisson regression, the MVP regression model is constructed so that the parameters depend on explanatory variables $\mathbf{x}_{is}$ ($s = 1, 2, 3$).

$$\begin{aligned} \theta_{i1} &= E^{\alpha_1} \exp(\mathbf{x}_{i1}'\gamma_1) \\ \theta_{i2} &= E^{\alpha_2} \exp(\mathbf{x}_{i2}'\gamma_2) \\ \theta_{i3} &= E^{\alpha_3} \exp(\mathbf{x}_{i3}'\gamma_3) \end{aligned} \tag{4}$$

where $\mathbf{x}_{is}$ and $\boldsymbol{\gamma}_s$ are $p_s \times 1$ column vectors. $E^{\alpha_s}$ denotes an exposure measure [such as vehicle miles traveled (VMT)], and the exponential transformation ensures nonnegativity of crash rates. Equation 4 can be further expressed as follows:

$$\begin{aligned} \theta_{i1} &= \exp[\mathbf{x}_{i1}'\gamma_1 + \alpha_1 \ln(E)] & & \theta_{i1} = \exp[\mathbf{x}_{i1}'\boldsymbol{\beta}_1] \\ \theta_{i2} &= \exp[\mathbf{x}_{i2}'\gamma_2 + \alpha_2 \ln(E)] & \Rightarrow \quad & \theta_{i2} = \exp[\mathbf{x}_{i2}'\boldsymbol{\beta}_2] \\ \theta_{i3} &= \exp[\mathbf{x}_{i3}'\gamma_3 + \alpha_3 \ln(E)] & & \theta_{i3} = \exp[\mathbf{x}_{i3}'\boldsymbol{\beta}_3] \end{aligned}$$

where

$\mathbf{x}_{i1}'\boldsymbol{\beta}_1 = \mathbf{x}_{i1}'\boldsymbol{\gamma}_1 + \alpha_1 \ln(E)$,
$\mathbf{x}_{i2}'\boldsymbol{\beta}_2 = \mathbf{x}_{i2}'\boldsymbol{\gamma}_2 + \alpha_2 \ln(E)$, and
$\mathbf{x}_{i3}'\boldsymbol{\beta}_3 = \mathbf{x}_{i3}'\boldsymbol{\gamma}_3 + \alpha_3 \ln(E)$.

In this way the set of regressors (and their number) may differ across $\theta_{is}$'s. It is also assumed that $\delta_i$ is independent of the $\mathbf{x}_{is}$'s.

For application of computational Bayesian models, the MVP regression model requires a distributional assumption for $\delta_i$, as well as knowledge of each observational unit's contribution to the likelihood, $\mathbf{y}_i | \delta_i, \boldsymbol{\beta}, \mathbf{x}$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)'$ and $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3')'$. Here, the $\delta_i$ is assumed to come from a univariate Poisson distribution with parameter $\lambda$. According to Equation 3, the likelihood contribution by the *i*th segment is a product of univariate Poisson distributions with rate parameters $\theta_{i1} + \lambda$, $\theta_{i2} + \lambda$, $\theta_{i3} + \lambda$. Thus, the joint probability function of $\mathbf{y}_i | \delta_i, \boldsymbol{\beta}, \mathbf{x}$ can be expressed as follows:

$$\begin{aligned} p(\mathbf{y}_i | \delta_i, \boldsymbol{\beta}, \mathbf{x}) = & \frac{\exp(\mathbf{x}_{i1}'\boldsymbol{\beta}_1)^{y_{i1}-\delta_i}}{\exp[\exp(\mathbf{x}_{i1}'\boldsymbol{\beta}_1)](y_{i1}-\delta_i)!} \\ & \times \frac{\exp(\mathbf{x}_{i2}'\boldsymbol{\beta}_2)^{y_{i2}-\delta_i}}{\exp[\exp(\mathbf{x}_{i2}'\boldsymbol{\beta}_2)](y_{i2}-\delta_i)!} \\ & \times \frac{\exp(\mathbf{x}_{i3}'\boldsymbol{\beta}_3)^{y_{i3}-\delta_i}}{\exp[\exp(\mathbf{x}_{i3}'\boldsymbol{\beta}_3)](y_{i3}-\delta_i)!} \end{aligned} \tag{5}$$

which is simply the product of the individual univariate probability mass functions for each of $y_{i1}, y_{i2}, y_{i3}$. Let $L(\boldsymbol{\beta}, \{\delta_i, i = 1, 2, \ldots, n\} | \mathbf{x}, \mathbf{y}) = \Pi_{i=1}^{n} p(\mathbf{y}_i | \delta_i, \boldsymbol{\beta}, \mathbf{x})$ be the likelihood function. According to Bayes' theorem, the posterior distribution is proportional to the product of the likelihood function and the joint prior of all parameters, so it must be given by $\Pi_{i=1}^{n}\left[p(\mathbf{y}_i | \delta_i, \boldsymbol{\beta}, \mathbf{x})\, p(\delta_i | \lambda)\right] p(\boldsymbol{\beta}, \lambda)$. Therefore, the kernel posterior distribution of the model is obtained as follows:

$$p(\boldsymbol{\beta},\lambda,\delta|\mathbf{x},\mathbf{y}) \propto \prod_{i=1}^{n}\left\{\frac{\exp(\mathbf{x}'_{i1}\boldsymbol{\beta}_1)^{y_{i1}-\delta_i}}{\exp[\exp(\mathbf{x}'_{i1}\boldsymbol{\beta}_1)](y_{i1}-\delta_i)!}\right.$$
$$\left.\times\frac{\exp(\mathbf{x}'_{i2}\boldsymbol{\beta}_2)^{y_{i2}-\delta_i}}{\exp[\exp(\mathbf{x}'_{i2}\boldsymbol{\beta}_2)](y_{i2}-\delta_i)!}\times\frac{\exp(\mathbf{x}'_{i3}\boldsymbol{\beta}_3)^{y_{i3}-\delta_i}}{\exp[\exp(\mathbf{x}'_{i3}\boldsymbol{\beta}_3)](y_{i3}-\delta_i)!}\right\}$$
$$\times\exp(-n\lambda)\prod_{i=1}^{n}\frac{\lambda^{\delta_i}}{\delta_i!}p(\boldsymbol{\beta},\lambda) \tag{6}$$

where $\delta_i \leq \min(y_{i1}, y_{i2}, y_{i3})$, $i = 1, 2, \ldots, n$. This constraint is caused by the fact that the variables following Poisson distributions take on only nonnegative integers. Simply put, it is assumed that $\boldsymbol{\beta}$ and $\lambda$ are independent of $\mathbf{x}$. The parameters $(\boldsymbol{\beta}, \lambda)$ can be assumed to have the following flat (uninformative) prior:

$$p(\boldsymbol{\beta},\lambda) \propto \lambda^{-1} \tag{7}$$

The nature of computational techniques for Bayesian analysis allows one to handle any arbitrary priors for the regression coefficients. Both flat and conjugate priors are assumed in the following series of Markov chain Monte Carlo (MCMC) simulation techniques for parameter estimation.

## Estimating Parameters via MCMC

Bayesian inference is primarily based on the MCMC simulation techniques, such as the Gibbs sampler and the M-H algorithm (e.g., *42–48*). The Gibbs sampler and the M-H algorithm set up a Markov chain in the parameter space. The Gibbs sampler is logically simpler but requires knowledge of the conditional distributions. It generates random draws from a joint density $\pi(\boldsymbol{\theta}) = \pi(\theta_1, \theta_2, \ldots, \theta_K)$, where $\boldsymbol{\theta}$ is the parameter vector. Let $\pi(\theta_k | \boldsymbol{\theta}_{-k})$ denote the full conditional density of $\theta_k$ given values of other components $\boldsymbol{\theta}_{-k} = (\theta_j, j \neq k)$, $k = 1, 2, \ldots, K$, where $K$ is the number of blocks of parameters. Given a starting point $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_K^{(0)})$, successive random draws are made from each of the conditional distributions $\pi(\theta_k | \boldsymbol{\theta}_{-k})$, where $k = 1, 2, \ldots, K$, by using the following subroutine:

Draw a value $\theta_1^{(m+1)}$ from $\pi(\theta_1 | \boldsymbol{\theta}_{-1}^{(m)})$;
Draw a value $\theta_2^{(m+1)}$ from $\pi(\theta_2 | \theta_1^{(m+1)}, \theta_3^{(m)}, \ldots, \theta_K^{(m)})$;
⋮
Draw a value $\theta_K^{(m+1)}$ from $\pi(\theta_K | \boldsymbol{\theta}_{-K}^{(m+1)})$

where $m = 1, 2, \ldots, M$. Iterating the subroutine $M$ times produces $M$ draws from the joint density $\pi(\boldsymbol{\theta})$. Thus the problem of sampling a multivariate distribution is reduced to the much easier problem of sampling from a series of univariate distributions. Under mild regularity conditions (*49*), the sample $\{\boldsymbol{\theta}^{(m)}; m = 1, 2, \ldots, M\}$ converges in distribution to $\pi(\boldsymbol{\theta})$. In practice, one often wants to investigate the marginal distributions of parameters of interest. The Gibbs sampler and M-H algorithms are the best devices for exploring such distributions.

To make draws from the posterior distribution in Equation 6, one has to provide the conditional distributions of parameters and determine how to obtain random draws from these distributions. Such posterior conditional distributions can be easily extracted from the joint posterior distribution in Equation 6. For example, the posterior conditional distribution of parameter $\lambda$ is given by

$$p(\lambda|\boldsymbol{\beta},\boldsymbol{\delta},\mathbf{x},\mathbf{y}) \propto \lambda^{\sum_{i=1}^{n}\delta_i - 1}\exp(-n\lambda) \tag{8}$$

which is a two-parameter gamma distribution with shape parameter $\sum_{i=1}^{n} \delta_i$ and scale parameter $1/n$.

The posterior conditional distribution of each $\delta_i$ is

$$p(\delta_i|\boldsymbol{\beta},\lambda,\mathbf{x},\mathbf{y}) \propto \frac{\exp(\mathbf{x}'_{i1}\boldsymbol{\beta}_1)^{y_{i1}-\delta_i}}{(y_{i1}-\delta_i)!}\times\frac{\exp(\mathbf{x}'_{i2}\boldsymbol{\beta}_2)^{y_{i2}-\delta_i}}{(y_{i2}-\delta_i)!}$$
$$\times\frac{\exp(\mathbf{x}'_{i3}\boldsymbol{\beta}_3)^{y_{i3}-\delta_i}}{(y_{i3}-\delta_i)!}\times\frac{\lambda^{\delta_i}}{\delta_i!} \tag{9}$$

for $\delta_i = 0, 1, \ldots, \min(y_{i1}, y_{i2}, y_{i3})$.

The conditional distribution of $\delta_i$ given the values of $(\boldsymbol{\beta}, \lambda, \mathbf{x}, \mathbf{y})$ is discrete, so it is easy to make random draws.

The posterior conditional distribution of $\boldsymbol{\beta}_s$ ($s = 1, 2, 3$) can be simplified as a posterior of regression coefficients in the following univariate Poisson regression model:

$$p(\boldsymbol{\beta}_s|\lambda,\boldsymbol{\delta},\mathbf{x},\mathbf{y}) \propto \prod_{i=1}^{n}\frac{\exp(\mathbf{x}'_{is}\boldsymbol{\beta}_s)^{y_{is}-\delta_i}}{\exp[\exp(\mathbf{x}'_{is}\boldsymbol{\beta}_s)]} \tag{10}$$

However, the conditional distribution of $\boldsymbol{\beta}_s$ is nonstandard, and thus it is difficult to generate random draws by using the Gibbs sampler. The M-H algorithm allows random draws from such nonstandard distributions.

The M-H algorithm generates a sequence of samples from the probability distribution of variables of interest. The key to this algorithm is creating a sampling strategy that satisfies a "detailed balance" requirement: the probability of being in state $\theta_a$ and moving to state $\theta_b$ must be the same as moving from $\theta_b$ to $\theta_a$. Notationally, this means the following:

$$p[\theta^{(m-1)} = \theta_a, \theta^{(m)} = \theta_b] = p[\theta^{(m-1)} = \theta_b, \theta^{(m)} = \theta_a]$$

The sequence of draws is accomplished by proposal and acceptance or rejection of candidate values $\theta^*$. A candidate point $\theta^*$ is sampled through a proposal function $q[\theta^* | \theta^{(m-1)}]$, the form of which is arbitrary. To satisfy this balance requirement, a probability

$$\alpha[\theta^* | \theta^{(m-1)}] = \min\left\{\frac{p(\theta^*)q[\theta^{(m-1)} | \theta^*]}{p[\theta^{(m-1)}]q[\theta^* | \theta^{(m-1)}]}, 1\right\}$$

is used here. If $\alpha[\theta^* | \theta^{(m-1)}]$ is greater than $U$ [where $U$ is uniformly distributed on (0, 1)], $\theta^{(m)} = \theta^*$; otherwise, $\theta^{(m)} = \theta^{(m-1)}$. There are three commonly used options for the proposal function $q[\theta^* | \theta^{(m-1)}]$: random walk chains, independence chains, and autoregressive chains. Further details about the M-H algorithm can be found elsewhere (*46–48, 50*).

## DATA DESCRIPTION

The crash data sets used here were collected from Washington State through the HSIS. After filtering off unreasonable observations (such as segments with zero speed limits), a total of 40,718 Washington State highway segments remained. Vehicular accidents caused 299 fatal injuries, 1,637 disabling injuries, 6,570 nondisabling injuries, and 11,858 possible injuries along these segments in 1996; there were 20,100 crash-involved persons experiencing no injury. These segments serve as distinct observational units and contain information on crash-involved vehicle and person characteristics, roadway design features (including speed limits), environmental conditions at the time of crash, and basic crash information (such as injury severity and time and type of crash). Table 1 contains summary statistics of all variables expected to be of interest.

## MODEL ESTIMATION AND DISCUSSIONS

### Model Estimation

The MVP regression model described in Equations 3 through 6 was estimated by using a Bayesian approach. Starting values came from distinct univariate Poisson models [using the method of maximum likelihood estimation (MLE)]. A Gibbs sampler (with nested M-H algorithms) was coded in R-language (an open-source statistical computing environment described at www.r-project.org). The Gibbs sampler was implemented to obtain $M = 25{,}000$ draws for each of the 96 parameters. The initial 5,000 draws were discarded as burn-ins. To help ensure chain convergence, the Gibbs sampler was implemented by using two sets of initial values, and both converge at the same posterior distribution of parameters. Estimation results are presented in Tables 2 through 6, along with MLE results for the univariate Poisson models.

**TABLE 1 Summary Statistics of Variables for Washington State Highway Segments in 1996**

| Variable Name | Variable Description | Mean | Std. Err. | Min. | Max. |
|---|---|---|---|---|---|
| Dependent variables | | | | | |
| FATAL | Number of fatal injuries in a segment per year | 0.007343 | 0.1659 | 0 | 10 |
| DISABLING | Number of disabling injuries in a segment per year | 0.04020 | 0.4222 | 0 | 13 |
| NONDISAB | Number of nondisabling injuries in a segment per year | 0.1614 | 0.9290 | 0 | 30 |
| POSSIBLE | Number of possible injuries in a segment per year | 0.2912 | 1.663 | 0 | 54 |
| NOINJURY | Number of no injuries in a segment per year | 0.4936 | 2.250 | 0 | 84 |
| Independent variables | | | | | |
| CURV_LGT | Horizontal curve length (ft) | 317.8 | 695.1 | 0 | 12,683 |
| DEG_CURV | Degree of curvature (°/100 ft) | 1.522 | 3.269 | 0 | 23.97 |
| VCUR_LGT | Vertical curve length (ft) | 393.3 | 509.5 | 0 | 6,000 |
| PCT_GRAD | Vertical grade (%) | 1.804 | 1.833 | 0 | 11.22 |
| RSHDWIDT | Total right shoulder width (ft) | 6.506 | 6.271 | 0 | 50 |
| NUMLANES | Total number of lanes | 2.618 | 1.196 | 1 | 9 |
| MEDIAN | Indicator for presence of median (1: presence of median, 0: no median) | 0.1787 | 0.3831 | 0 | 1 |
| SPD_LIMT | Posted speed limit (mph) | 51.54 | 10.30 | 25 | 70 |
| SPDLMTSQ | Posted speed limit squared | 2,763 | 997.5 | 625 | 4,900 |
| MOUNTAIN | Indicator for mountainous terrain (1: presence of mountainous terrain, 0: otherwise) | 0.08338 | 0.2765 | 0 | 1 |
| ROLLING | Indicator for rolling terrain (1: presence of rolling terrain, 0: otherwise) | 0.7182 | 0.4499 | 0 | 1 |
| RURALCOL | Indicator for rural collector (1: rural collector, 0: otherwise) | 0.2187 | 0.4134 | 0 | 1 |
| RURALINT | Indicator for rural Interstate (1: rural Interstate, 0: otherwise) | 0.05022 | 0.2184 | 0 | 1 |
| URBANART | Indicator for urban arterial (1: urban arterial, 0: otherwise) | 0.1734 | 0.3786 | 0 | 1 |
| URBANCOL | Indicator for urban collector (1: urban collector, 0: otherwise) | 0.007441 | 0.08594 | 0 | 1 |
| URBANINT | Indicator for urban Interstate (1: urban Interstate, 0: otherwise) | 0.04924 | 0.2164 | 0 | 1 |
| ACCCNTRL | Indicator for access control (1: presence of access control, 0: otherwise) | 0.2588 | 0.4380 | 0 | 1 |
| VMT | Annual vehicle miles traveled on a segment | 319,971 | 1,040,530 | 376.0 | 93,420,800 |
| LNVMT | Logarithm of annual vehicle miles traveled on a segment | 11.23 | 1.687 | 5.929 | 18.35 |
| No. of observations | | | | | 40,718 |

**TABLE 2 Fatal Injury Frequency Models for Washington State Crash Data, 1996**

| | Univariate Poisson Regression (MLE) | | | Multivariate Poisson Regression (Gibbs Sampler) | | | |
|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | *P*-Value | Mean | Std. Err. | 95% (2.5%–97.5%) HDR | |
| Constant | −13.14 | 0.7778 | 0.000 | −12.92* | 1.433 | −15.71 | −10.10 |
| CURV_LGT | 1.894E-04 | 4.997E-05 | 0.000 | −6.639E-05 | 9.423E-05 | −2.522E-04 | 1.160E-04 |
| DEG_CURV | | | | 0.01212* | 0.006019 | 0.0003532 | 0.02395 |
| VCUR_LGT | −1.909E-04 | 1.105E-04 | 0.084 | 5.526E-05 | 1.246E-04 | −1.875E-04 | 3.005E-04 |
| PCT_GRAD | | | | 0.01286 | 0.01927 | −0.02470 | 0.05098 |
| RSHDWIDT | | | | −0.01992* | 0.005541 | −0.03088 | −0.0091049 |
| NUMLANES | −0.2130 | 0.07369 | 0.004 | −0.02792 | 0.07470 | −0.1728 | 0.1195 |
| MEDIAN | −0.4290 | 0.2475 | 0.083 | 0.08228 | 0.3733 | −0.6455 | 0.8162 |
| SPD_LIMT | 0.03435 | 0.009882 | 0.001 | 0.01214* | 0.005055 | 0.002259 | 0.02202 |
| SPDLMTSQ | | | | −9.432E-05 | 1.860E-04 | −4.599E-04 | 2.702E-04 |
| MOUNTAIN | −1.782 | 0.5943 | 0.003 | 1.943 | 2.853 | −3.657 | 7.524 |
| ROLLING | −0.3199 | 0.1335 | 0.017 | 0.2211 | 0.3013 | −0.3655 | 0.8197 |
| RURALCOL | −0.7587 | 0.3087 | 0.014 | 0.08142 | 0.2868 | −0.4803 | 0.6472 |
| RURALINT | 1.157 | 0.2793 | 0.000 | −0.03326 | 0.3041 | −0.6300 | 0.5658 |
| URBANART | 0.6766 | 0.1911 | 0.000 | 0.9335 | 1.285 | −1.572 | 3.439 |
| URBANCOL | | | | −29.37 | 32.26 | −92.84 | 33.40 |
| URBANINT | 0.6593 | 0.3343 | 0.049 | 0.8876 | 1.168 | −1.402 | 3.155 |
| ACCCNTRL | −0.4500 | 0.2025 | 0.026 | −0.2981 | 0.3508 | −0.9797 | 0.3906 |
| LNVMT | 0.6035 | 0.05141 | 0.000 | 0.5964* | 0.1053 | 0.3887 | 0.8037 |

NOTE: An asterisk (*) signifies that parameters differ from zero in a statistically significant way, based on the 95% (2.5%–97.5%) high density region (HDR).

**TABLE 3 Disabling Injury Frequency Models for Washington State Crash Data, 1996**

| | Univariate Poisson Regression (MLE) | | | Multivariate Poisson Regression (Gibbs Sampler) | | | |
|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | *P*-Value | Mean | Std. Err. | 95% (2.5%–97.5%) HDR | |
| Constant | −13.46 | 0.5977 | 0.000 | −13.80* | 0.9266 | −15.62 | −11.97 |
| CURV_LGT | | | | −8.342E-06 | 1.266E-05 | −3.330E-05 | 1.673E-05 |
| DEG_CURV | −0.029889 | 0.01342 | 0.026 | 0.01656 | 0.01875 | −0.02038 | 0.05278 |
| VCUR_LGT | | | | −1.680E-05 | 4.585E-05 | −1.075E-04 | 7.296E-05 |
| PCT_GRAD | | | | −0.0007990 | 0.0007548 | −0.002276 | 0.0006793 |
| RSHDWIDT | −0.010369 | 0.004750 | 0.029 | −0.02583* | 0.0037537 | −0.03311 | −0.01848 |
| NUMLANES | | | | −0.07253* | 0.01834 | −0.10842 | −0.03691 |
| MEDIAN | | | | −0.09199* | 0.01729 | −0.1258 | −0.05860 |
| SPD_LIMT | 0.07685 | 0.02420 | 0.001 | 0.1103* | 0.005038 | 0.1004 | 0.1202 |
| SPDLMTSQ | −8.429E-04 | 2.585E-04 | 0.001 | −7.478E-04* | 1.262E-04 | −9.944E-04 | −5.026E-04 |
| MOUNTAIN | | | | −0.1128 | 0.1061 | −0.3216 | 0.09505 |
| ROLLING | 0.2266 | 0.06124 | 0.000 | −0.1176* | 0.04095 | −0.1973 | −0.03646 |
| RURALCOL | −0.3861 | 0.1252 | 0.002 | 0.02386 | 0.2950 | −0.5587 | 0.5963 |
| RURALINT | 0.7683 | 0.1515 | 0.000 | 0.9377 | 1.168 | −1.368 | 3.235 |
| URBANART | 0.4916 | 0.07447 | 0.000 | 0.7872 | 1.117 | −1.388 | 2.956 |
| URBANCOL | | | | 0.4243 | 0.4852 | −0.5311 | 1.375 |
| URBANINT | 0.3399 | 0.1141 | 0.003 | 0.8374 | 1.143 | −1.409 | 3.085 |
| ACCCNTRL | −0.4546 | 0.08604 | 0.000 | −0.5668* | 0.1488 | −0.8576 | −0.2726 |
| LNVMT | 0.6966 | 0.02237 | 0.000 | 0.6018* | 0.0857263 | 0.4337 | 0.7693 |

NOTE: An asterisk (*) signifies that parameters differ from zero in a statistically significant way, based on the 95% (2.5%–97.5%) high density region (HDR).

**TABLE 4 Nondisabling Injury Frequency Models for Washington State Crash Data, 1996**

| | Univariate Poisson Regression (MLE) | | | Multivariate Poisson Regression (Gibbs Sampler) | | | |
|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | *P*-Value | Mean | Std. Err. | 95% (2.5%–97.5%) HDR | |
| Constant | −10.51 | 0.2777 | 0.000 | −10.68* | 0.6411 | −11.94 | −9.414 |
| CURV_LGT | −7.410E-05 | 1.912E-05 | 0.000 | −3.011E-05 | 3.638E-05 | −1.007E-04 | 4.071E-05 |
| DEG_CURV | −0.01315 | 0.006691 | 0.049 | 0.09084* | 0.008010 | 0.07527 | 0.1065 |
| VCUR_LGT | −8.778E-05 | 2.532E-05 | 0.001 | 9.737E-05* | 3.377E-05 | 3.005E-05 | 1.630E-04 |
| PCT_GRAD | | | | −0.007937 | 0.008190 | −0.02397 | 0.008120 |
| RSHDWIDT | −0.022034 | 2.48E-03 | 0.000 | −0.01438* | 0.0050373 | −0.02424 | −0.0044529 |
| NUMLANES | 0.1402 | 0.01271 | 0.000 | −0.1204* | 0.02506 | −0.1692 | −0.07163 |
| MEDIAN | −0.3593 | 0.05444 | 0.000 | −0.1547* | 0.06110 | −0.2758 | −0.03424 |
| SPD_LIMT | 0.02260 | 0.01157 | 0.051 | 0.01581* | 0.0066693 | 0.002672 | 0.02888 |
| SPDLMTSQ | −3.528E-04 | 1.267E-04 | 0.005 | −1.891E-04* | 5.132E-05 | −2.897E-04 | −8.917E-05 |
| MOUNTAIN | 0.1759 | 0.08498 | 0.038 | 0.9582 | 1.481 | −1.948 | 3.891 |
| ROLLING | 0.1946 | 0.03344 | 0.000 | 0.09585* | 0.04854 | 0.00044875 | 0.1909 |
| RURALCOL | −0.6237 | 0.07331 | 0.000 | 0.1386* | 0.03894 | 0.06282 | 0.2160 |
| RURALINT | 0.5760 | 0.08743 | 0.000 | 0.6055 | 0.9070 | −1.134 | 2.379 |
| URBANART | 0.5305 | 0.04265 | 0.000 | 0.9056 | 2.071 | −3.145 | 4.997 |
| URBANCOL | 0.4142 | 0.1401 | 0.003 | 1.925 | 3.143 | −4.329 | 8.075 |
| URBANINT | 0.6587 | 0.07187 | 0.000 | 1.305 | 2.171 | −2.949 | 5.589 |
| ACCCNTRL | −0.1219 | 0.04608 | 0.008 | −0.1300 | 0.1747 | −0.4716 | 0.2095 |
| LNVMT | 0.6583 | 0.01181 | 0.000 | 0.6859* | 0.06625 | 0.5573 | 0.8168 |

NOTE: An asterisk (*) signifies that parameters differ from zero in a statistically significant way, based on the 95% (2.5%–97.5%) high density region (HDR).

**TABLE 5 Possible Injury Frequency Models for Washington State Crash Data, 1996**

| | Univariate Poisson Regression (MLE) | | | Multivariate Poisson Regression (Gibbs Sampler) | | | |
|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | *P*-Value | Mean | Std. Err. | 95% (2.5%–97.5%) HDR | |
| Constant | −12.21 | 0.2159 | 0.000 | −12.74* | 0.6148 | −13.95 | −11.54 |
| CURV_LGT | −5.735E-05 | 1.505E-05 | 0.000 | −5.186E-05 | 6.551E-05 | −1.792E-04 | 7.439E-05 |
| DEG_CURV | −0.04531 | 0.005896 | 0.000 | 0.05432* | 0.01280 | 0.02934 | 0.07935 |
| VCUR_LGT | −1.148E-04 | 2.000E-05 | 0.000 | −5.111E-05 | 6.231E-05 | −1.732E-04 | 7.072E-05 |
| PCT_GRAD | | | | 1.919E-05 | 5.309E-05 | −8.368E-05 | 1.236E-04 |
| RSHDWIDT | −0.02654 | 0.001673 | 0.000 | −0.02326* | 0.00347 | −0.02996 | −0.01638 |
| NUMLANES | 0.1340 | 0.008782 | 0.000 | −0.1147* | 0.01599 | −0.1458 | −0.08345 |
| MEDIAN | −0.1094 | 0.03516 | 0.002 | −0.1051* | 0.02948 | −0.1627 | −0.04696 |
| SPD_LIMT | 0.07957 | 0.009240 | 0.000 | 0.08179* | 0.001190 | 0.07944 | 0.08410 |
| SPDLMTSQ | −1.300E-03 | 1.030E-04 | 0.000 | −8.133E-04* | 6.558E-05 | −9.417E-04 | −6.841E-04 |
| MOUNTAIN | 0.1940 | 0.08537 | 0.023 | 0.4045* | 0.1129 | 0.1817 | 0.6250 |
| ROLLING | 0.2380 | 0.02560 | 0.000 | 0.1598* | 0.04869 | 0.06274 | 0.2554 |
| RURALCOL | −1.025 | 0.08422 | 0.000 | −0.01044* | 0.004692 | −0.01955 | −0.001110 |
| RURALINT | 0.8093 | 0.07637 | 0.000 | 1.107 | 2.035 | −2.865 | 5.118 |
| URBANART | 0.7882 | 0.03455 | 0.000 | 1.092 | 3.023 | −4.879 | 7.029 |
| URBANCOL | 0.4641 | 0.1101 | 0.000 | 1.298 | 3.148 | −4.881 | 7.465 |
| URBANINT | 1.248 | 0.05267 | 0.000 | 1.713 | 4.060 | −6.252 | 9.596 |
| ACCCNTRL | | | | −0.009948* | 0.003844 | −0.01748 | −0.002380 |
| LNVMT | 0.7758 | 0.009092 | 0.000 | 0.7520* | 0.04727 | 0.6595 | 0.8436 |

NOTE: An asterisk (*) signifies that parameters differ from zero in a statistically significant way, based on the 95% (2.5%–97.5%) high density region (HDR).

**TABLE 6 No Injury Frequency Models for Washington State Crash Data, 1996**

| | Univariate Poisson Regression (MLE) | | | Multivariate Poisson Regression (Gibbs Sampler) | | | |
|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | *P*-Value | Mean | Std. Err. | 95% (2.5%–97.5%) HDR | |
| Constant | −7.967 | 0.09185 | 0.000 | −8.875* | 0.3890 | −9.637 | −8.108 |
| CURV_LGT | −9.409E-05 | 1.142E-05 | 0.000 | 1.987E-06 | 2.377E-05 | −4.457E-05 | 4.846E-05 |
| DEG_CURV | −0.01209 | 0.003744 | 0.001 | 0.01880* | 0.004550 | 0.009799 | 0.02760 |
| VCUR_LGT | | | | −2.293E-05* | 1.536E-06 | −2.595E-05 | −1.990E-05 |
| PCT_GRAD | 0.009696 | 0.004586 | 0.034 | 0.01110* | 0.003297 | 0.004656 | 0.01756 |
| RSHDWIDT | −0.02215 | 0.001369 | 0.000 | −0.02500* | 0.002516 | −0.02995 | −0.02006 |
| NUMLANES | 0.1835 | 6.92E-03 | 0.000 | −0.1563* | 0.01467 | −0.1852 | −0.1274 |
| MEDIAN | −0.3139 | 0.03036 | 0.000 | −0.3152* | 0.04708 | −0.4086 | −0.2235 |
| SPD_LIMT | −0.03771 | 0.001096 | 0.000 | 0.01261* | 0.004838 | 0.003166 | 0.02207 |
| SPDLMTSQ | | | | −2.031E-04* | 6.725E-05 | −3.337E-04 | −7.265E-05 |
| MOUNTAIN | 0.4520 | 0.04845 | 0.000 | 0.4736* | 0.07205 | 0.3327 | 0.6142 |
| ROLLING | 0.1621 | 0.01941 | 0.000 | 0.1480 | 0.3594 | −0.5581 | 0.8513 |
| RURALCOL | −0.7315 | 0.05075 | 0.000 | −0.5923 | 1.050 | −2.652 | 1.472 |
| RURALINT | 0.8061 | 0.04084 | 0.000 | 0.8565 | 1.873 | −2.818 | 4.492 |
| URBANART | 0.6673 | 0.02553 | 0.000 | 0.8327 | 1.686 | −2.404 | 4.144 |
| URBANCOL | 0.7253 | 0.06918 | 0.000 | 0.8854 | 1.421 | −1.907 | 3.675 |
| URBANINT | 0.8895 | 0.04065 | 0.000 | 0.9667 | 2.042 | −2.966 | 4.964 |
| ACCCNTRL | 0.08235 | 0.02710 | 0.002 | 0.1025 | 0.05967 | −0.01394 | 0.2192 |
| LNVMT | 0.6752 | 0.006900 | 0.000 | 0.6817* | 0.05125 | 0.5826 | 0.7820 |

NOTE: An asterisk (*) signifies that parameters differ from zero in a statistically significant way, based on the 95% (2.5%–97.5%) high density region (HDR).

Figure 1 shows the estimates of posterior distributions for these regression coefficients. On the basis of the posterior density of $\lambda$ (shown in Figure 1*h*), positive correlations between crash counts at different levels of severity within the segment appear to exist in a statistically significant way among counts of different injury levels. The univariate models are a special case of the MVP, with $\lambda = 0$, so the MVP predictions should prove better. Calculation of average likelihood values for the estimated models versus constant-only cases provides likelihood ratio indices (LRIs) as a measure of goodness of fit. These are 0.323 for the suite of univariate models and 0.766 for the MVP approach, which suggests that the latter is superior. Both approaches predict total counts (by severity) across all roadway segments with almost no error.

## Interpretation of Results

In addition to producing a substantially higher LRI and better estimates of total crash-involved persons (or total injuries), the MVP model's estimation results offer more intuitive interpretations. For example, fatal injury rates (per VMT) rise with speed limit in the MVP models. This potentially key variable was not found to be statistically significant in the univariate model for fatal crash counts. However, the MVP model's Bayesian results suggest far fewer statistically significant control variables.

The following discussion of results emphasizes fatal and disabling injuries (Tables 2 and 3), since these arguably are of greatest concern to agencies and policy makers. Moreover, the data on such outcomes are more likely to be reported and more reliably recorded than are data for other crash outcomes. Tables 4 through 6 provide person-count model estimates for the other three severity levels. The signs of most coefficients are consistent throughout the models, indicating robust directions of effect for almost all control variables, at least in the case of severe injury (fatal and debilitating).

Parameter estimates shown in Tables 2 and 3 suggest that roadway design plays an important role in injury counts. For example, with all other factors fixed, more fatal injuries are expected on sharper horizontal curves, while wider shoulders tend to reduce rates of both fatal and disabling injuries. On the basis of an average road segment's attributes and the MVP model's average parameter estimates, Table 7 provides estimates of percentage changes in crash frequencies as a function of various design details. For example, a 10-ft increase in shoulder width (from 10 to 20 ft) is predicted to result in 18% and 23% fewer fatal and disabling injury cases, respectively, per 100 million VMT. Added lanes are predicted to reduce disabling injuries by 11%; an added median by 8.8%. Removal of access control is predicted to increase the number of disabling injuries by 36%. Oddly, none of these three key variables was predicted to have a statistically significant impact on fatal injury counts (in the MVP model). Perhaps fatal crash counts are so few on short homogeneous roadway segments that they cannot be clearly linked to many design attributes. Nevertheless, disabling injuries may serve as a valuable proxy for fatal crash relationships. And the MVP model offers several statistically (and practically) significant insights into these injury counts' dependence on roadway design attributes.

## Example Application: A Cost–Benefit Analysis of Raised Speed Limits

Results in Tables 2 through 7 offer several suggestions for design changes that transportation agencies might consider. As indicated in

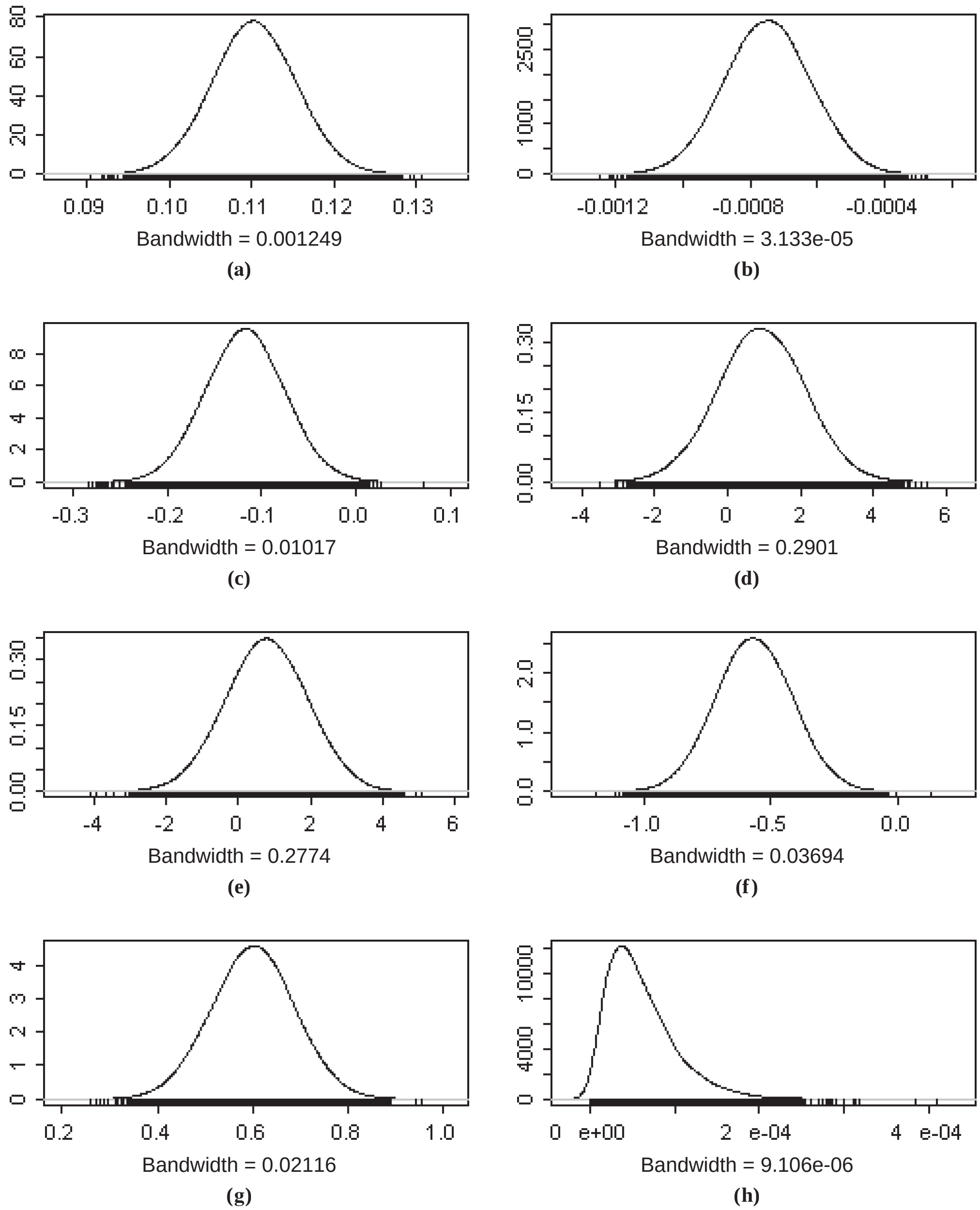


**FIGURE 1 Posterior densities for a subset of parameter estimates (from the disabling injury frequency model); $N$ = 20,000 in all cases: (*a*) SPD_LIMT, (*b*) SPDLMTSQ, (*c*) ROLLING, (*d*) RURALINT, (*e*) URBANART, (*f*) ACCCNTRL, (*g*) LNVMT, and (*h*) $\lambda$.**

Table 7, a speed limit increase of 10 mph (from 55 to 65 mph on the "average" roadway section in the database) is predicted to increase fatal and disabling injury rates by 0.95% and 11.13%, respectively (according to the MVP model's average parameter values). One might argue that travel time savings due to an increase in limits can offset the costs of increases in these and other crash outcomes. This section considers this question as an example application of the model results.

Table 8 presents estimates of injury costs. Its first two rows summarize a NHTSA study by Blincoe et al. (*2*). The first row presents the "market costs" of injuries (based on medical treatment, emergency services, losses in market and household productivity, insurance administration, workplace cost, and legal costs). The second row gives comprehensive costs based on quality-adjusted life years, as well as pain and suffering by family members. Since the HSIS database recognizes five injury levels (rather than six), injury costs were calculated by using a weighted average of the six Maximum Abbreviated Injury Scale (MAIS) costs. [MAIS denotes the highest (maximum) abbreviated injury severity score that corresponds to a crash victim's incurred injuries. It can take on values from 0 (minor injuries) to 5 (fatal injury).]

Table 9 presents driving speed increases that have been observed in a variety of published studies following speed limit increases. [Most of the studies listed here (except that in NCHRP Project 17-23) examined speeds on rural Interstate highways following a change from 55 to 65 mph. The NCHRP study examined an urban and a rural site, both with a 5-mph increase (the resulting average speed change was there-

**TABLE 7 Expected Percentage Changes in Injury Rates Corresponding to Changes in Variables**

| | | | Percentage Change in Injury Rates (per 100 million VMT) | | | | |
|---|---|---|---|---|---|---|---|
| Variable | Averages | Change in Variable | Killed | Disabling Injury | Nondisabling Injury | Possible Injury | No Injury |
| DEG_CURV | 2 (°/100 ft) | +2 | 2.45 | — | 19.92 | 11.48 | 3.83 |
| VCUR_LGT | 400 (ft) | +100 | — | — | 0.97 | — | –0.23 |
| RSHDWIDT | 10 (ft) | +10 | –18.03 | –22.75 | –13.40 | –20.73 | –22.14 |
| NUMLANES | 3 | +1 | — | –6.99 | –11.35 | –10.84 | –14.47 |
| MEDIAN | No | Yes | — | –8.79 | –14.36 | –9.96 | –27.00 |
| SPD_LIMT | 55 (mph) | +10 | 0.95 | 11.13 | –6.55 | –14.56 | –11.16 |
| MOUNTAIN | No (rolling) | Yes | — | — | — | 27.84 | 38.50 |
| ROLLING | Yes | No (level) | — | 13.09 | –21.31 | –32.45 | — |
| ACCCNTRL | Yes | No | — | 36.22 | — | 19.94 | — |

NOTE: The data set's average VMT value (78,358 mi) was used in these calculations.

fore doubled in that case, to estimate the change that would have occurred had the speed limit change been 10 mph).] On the basis of Table 9, there is approximately a 3.1-mph increase in average observed traffic speeds if speed limits are raised 10 mph. Thus, the time savings per 100 million VMT due to a 10-mph increase in speed limits is estimated to be 106,879 h. This time savings is equivalent to $1,450,687, assuming a $15.04 per vehicle-hour value of travel time savings (*60, 61*). A 10-mph increase in speed limits is predicted to result in 0.029 and 1.9 more fatal and disabling injuries, respectively, and in 4.87, 13.96, and 17.16 fewer nondisabling, possible, and no injury outcomes (per 100 million VMT), respectively. The equivalent average cost estimate for such shifts in injury types is estimated to be $3.34 million (in 2000 dollars, on the basis of the values of crash costs in the last row of Table 8). [Mrozek and Taylor (*62*) investigated the value of a statistical life (VOSL) by using a meta-analysis. On the basis of 33 previous studies, they recommended a VOSL of $1.5 million to $2.5 million, which is considerably lower than NHTSA's $3.37 million recommendation. However, the average VOSL of the 33 studies is about $5.59 million. If this $5.59 million value (per life) were used here and other injury costs were inflated by a ratio of 1.66 (5.59 million/3.37 million), the cost–benefit ratio would become 1:2.21, which suggests that speed limits could offer some valuable time savings benefits.] Therefore, the estimated cost–benefit ratio is 2.3:1. These results suggest that raising speed limits does not offer adequate time savings benefits. However, if actual travel speeds were to increase one-to-one with speed limits (i.e., by 10 mph rather than 3.1 mph), this ratio would change to 0.71:1. Thus, the result depends on how much speeds change after a speed limit change.

## CONCLUSIONS

This study developed a model that allows researchers to model crash outcomes simultaneously by severity on the basis of a type of MVP specification that can be estimated within a Bayesian framework with Gibbs sampling. Crash counts for more than 40,000 homogeneous segments of Washington State highways in 1996 were used to estimate the model. As expected, positive correlation in unobserved factors affecting count outcomes was found across severity levels, which resulted in a statistically significant additive latent term.

Thanks to MCMC simulation techniques, the marginal posterior distributions of all parameters of interest were obtained, and estimation results from the MVP approach offered more intuitive interpretations and better predictions than those from the univariate Poisson models. As anticipated, the results lend themselves to several recommendations for highway safety treatments and design policies. For example, access control and wide shoulders are important for reducing severe injury, as are medians and added lanes. Moreover, on the basis of a cost–benefit approach and assumptions about travel

**TABLE 8 NHTSA Estimate of Injury Costs (in 2000 Dollars) (*2*)**

| | PDO | MAIS 0 | MAIS 1 | MAIS 2 | MAIS 3 | MAIS 4 | MAIS 5 | Fatal |
|---|---|---|---|---|---|---|---|---|
| Market cost ($) | 2,532 | 1,962 | 10,562 | 66,820 | 186,097 | 348,133 | 1,096,161 | 977,208 |
| Comprehensive ($) | 2,532 | 1,962 | 15,017 | 157,958 | 314,204 | 731,580 | 2,402,997 | 3,366,388 |
| % crashes unreported (by type) | | 21.42 | 22.74 | 15.83 | 6.52 | 0.67 | 0.00 | |
| Persons involved in reported crashes | | 2,002,667 | 3,599,995 | 366,987 | 117,694 | 36,264 | 9,463 | |
| Persons involved in all crashes | | 2,548,571 | 4,659,585 | 436,007 | 125,903 | 36,509 | 9,463 | |
| Weight (% of persons involved) | | 25.62 | 46.06 | 4.70 | 1.51 | 0.46 | 0.12 | |
| Cost per injury ($) | 2,532 | | 10,351 | | | 232,890 | 2,402,997 | 3,366,388 |

NOTE: PDO = property damage only.

**TABLE 9 Speed Increases Following a 10-mph Speed Limit Increase (from 55 to 65 mph)**

| Study | Change in Observed Speeds (mph) |
|---|---|
| Brown et al. (*51*) | 2.4 |
| Freedman and Esterlitz (*52*) | 2.8 |
| Mace and Heckard (*53*) | 3.5 |
| NHTSA (*54*) | 1.9 |
| NHTSA (*55*) | 3.4 |
| Parker (*56*) | 0.2–2.3 |
| Pfefer, Stenzel, and Lee (*57*) | 4–5 |
| NCHRP (*58*) (Speed Choice in NW Washington State) | 3.4–4.8 |
| TRB (*59*) | 4 |
| Average | 3.1 |

speed changes, model results suggest that time savings from raising speed limits 10 mph (from 55 to 65 mph) may not be worth the added crash cost.

Several enhancements can be made in this work. The model specification relied on a one-way covariance structure and assumed the presence of an added constant across all count types. This implies that the covariances are nonnegative and identical within the segment and that within-segment covariances are the same across segments. A more general covariance structure would allow for different correlations across all pairs of count outcomes, and a multiplicative approach may better reflect the distinctions in count magnitudes (across severities). Other forms of overdispersion and correlation also should be explored, including the mixed multinomial-Poisson model (*63*) and the multivariate negative binomial model [as employed by Kockelman (*64*) and others and currently under investigation by the authors]. The use of panel data would allow sources of heterogeneity to be distinguished. Acquisition of other potentially valuable variables (such as distances to the nearest hospital and clear zone width) would also be helpful. Nevertheless, a Bayesian approach appears to offer great potential for new and different model specifications leading to richer sets of results and better predictive power. Such approaches may be critical in an area as important to human health and welfare as highway safety, even in the presence of large data sets (where classical approaches also tend to perform reasonably).

## ACKNOWLEDGMENTS

The authors thank the Texas Department of Transportation for funding this research. The authors also are grateful to FHWA's Yusuf Mohamedshah for provision of the crash data sets, Paul Damien and Shaw-Pin Miaou for useful discussions related to methods of analysis, and Annette Perrone for editorial assistance.

## REFERENCES

1. *Traffic Safety Facts: Research Note.* Report DOT HS 809 831. NHTSA, Washington, D.C., Jan. 2005.
2. Blincoe, L., A. Seay, E. Zaloshnja, T. Miller, E. Romano, S. Luchter, and R. Spicer. *The Economic Impact of Motor Vehicle Crashes, 2000.* U.S. Department of Transportation, Washington, D.C., 2002.
3. Schrank, D., and T. Lomax. *The 2002 Urban Mobility Report.* Texas Transportation Institute, Texas A&M University System, 2002.
4. Hauer, E. On the Estimation of the Expected Number of Accidents. *Accident Analysis and Prevention,* Vol. 18, No. 1, 1986, pp. 1–12.
5. Hauer, E. *Observational Before–After Studies in Road Safety.* Pergamon, Oxford, United Kingdom, 1997.
6. Hauer, E. Overdispersion in Modeling Accidents on Road Sections and in Empirical Bayes Estimation. *Accident Analysis and Prevention,* Vol. 33, No. 6, 2001, pp. 799–808.
7. Abdel-Aty, M. A., and A. E. Radwan. Modeling Traffic Accident Occurrence and Involvement. *Accident Analysis and Prevention,* Vol. 32, No. 5, 2000, pp. 633–642.
8. Ulfarsson, G. F., and V. N. Shankar. Accident Count Model Based on Multiyear Cross-Sectional Roadway Data with Serial Correlation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 193–197.
9. Kweon, Y.-J., and K. M. Kockelman. Safety Effects of Speed Limit Changes: Use of Panel Models, Including Speed, Use, and Design Variables. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 148–158.
10. Lord, D., and B. N. Persaud. Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1717,* TRB, National Research Council, Washington, D.C., 2000, pp. 102–108.
11. 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.
12. Miaou, S.-P., P. S. Hu, T. Wright, S. C. Davis, and A. K. Rathi. *Development of Relationship Between Truck Accidents and Geometric Design: Phase I.* Publication FHWA-RD-91-124. FHWA, U.S. Department of Transportation, 1993.
13. Miaou, S.-P., and H. Lum. Modeling Vehicle Accidents and Highway Geometric Design Relationships. *Accident Analysis and Prevention,* Vol. 25, No. 6, 1993, pp. 689–709.
14. Miaou, S.-P. The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions. *Accident Analysis and Prevention,* Vol. 26, No. 4, 1994, pp. 471–482.
15. Miaou, S.-P. *Measuring the Goodness-of-Fit of Accident Prediction Models.* Publication FHWA-RD-96-040. FHWA, U.S. Department of Transportation, 1996.
16. Miaou, S.-P. *Estimating Roadside Encroachment Rates with the Combined Strengths of Accident- and Encroachment-Based Approaches.* Publication FHWA-RD-01-124. FHWA, U.S. Department of Transportation, 2001.
17. Fridstrøm, L., J. Ifver, S. Ingebrigtsen, R. Kulmala, and L. K. Thomsen. Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention,* Vol. 27, No. 1, 1995, pp. 1–20.
18. Johansson, P. Speed Limitation and Motorway Casualties: A Time-Series Count Data Regression Approach. *Accident Analysis and Prevention,* Vol. 28, No. 1, 1996, pp. 73–87.
19. Vogt, A., and J. G. Bared. *Accident Models for Two-Lane Rural Roads: Segments and Intersections.* Publication FHWA-RD-98-133. FHWA, U.S. Department of Transportation, 1998.
20. Vogt, A. *Crash Models for Rural Intersections: Four-Lane by Two-Lane Stop-Controlled and Two-Lane by Two-Lane Signalized.* Publication FHWA-RD-99-128. FHWA, U.S. Department of Transportation, 1999.
21. Balkin, S., and J. K. Ord. Assessing the Impact of Speed-Limit Increases on Fatal Interstate Crashes. *Journal of Transportation and Statistics,* Vol. 4, No. 1, 2001, pp. 1–26.
22. Zegeer, C. V., J. R. Stewart, H. H. Huang, and P. A. Lagerwey. *Safety Effects of Marked Vs. Unmarked Crosswalks at Uncontrolled Locations: Executive Summary and Recommended Guidelines.* Publication FHWA-RD-01-075. FHWA, U.S. Department of Transportation, 2002.
23. Pernia, J., J. J. Lu, and H. Peng. *Safety Issues Related to Two-Way Left-Turn Lanes.* Department of Civil and Environmental Engineering, University of South Florida, Tampa, 2004.
24. Shankar, V., J. Milton, and F. L. Mannering. Modeling Accident Frequency as Zero-Altered Probability Processes: An Empirical Inquiry. *Accident Analysis and Prevention,* Vol. 29, No. 6, 1997, pp. 829–837.

25. Garber, N. J., and L. Wu. *Stochastic Models Relating Crash Probabilities with Geometric and Corresponding Traffic Characteristics Data.* Publication UVACTS-5-15-74. Center for Transportation Studies, University of Virginia, Charlottesville, 2001.
26. Lee, J., and F. L. Mannering. Impact of Roadside Features on the Frequency and Severity of Run-off-Roadway Accidents: An Empirical Analysis. *Accident Analysis and Prevention,* Vol. 34, No. 2, 2002, pp. 149–161.
27. Kumara, S. P., and H. C. Chin. Modeling Accident Occurrence at Signalized T Intersections with Special Emphasis on Excess Zeros. *Traffic Injury Prevention,* Vol. 4, No. 1, 2003, pp. 53–57.
28. Miaou, S.-P., and D. Lord. Modeling Traffic Crash–Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 31–40.
29. 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.
30. Rodriguez, D. A., M. Rocha, A. J. Khattak, and M. H. Belzer. Effects of Truck Driver Wages and Working Conditions on Highway Safety: Case Study. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1833,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 95–102.
31. Shankar, V. N., G. F. Ulfarsson, R. M. Pendyala, and M. B. Nebergall. Modeling Crashes Involving Pedestrians and Motorized Traffic. *Safety Science,* Vol. 41, No. 7, 2003, pp. 557–640.
32. Noland, R. B., and M. A. Quddus. A Spatially Disaggregate Analysis of Road Casualties in England. *Accident Analysis and Prevention,* Vol. 36, No. 6, 2004, pp. 973–984.
33. Qin, X., J. N. Ivan, and N. Ravishanker. Selecting Exposure Measures in Crash Rate Prediction for Two-Lane Highway Segments. *Accident Analysis and Prevention,* Vol. 36, No. 2, 2004, pp. 183–191.
34. Chin, H. C. C., and M. A. Quddus. Applying the Random Effect Negative Binomial Model to Examine Traffic Accident Occurrence at Signalized Intersections. *Accident Analysis and Prevention,* Vol. 35, No. 2, 2003, pp. 253–259.
35. Ladron 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.
36. 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, No. 4, 2005, pp. 591–600.
37. Li, C. C., J. C. Lu, J. Park, K. Kim, P. A. Brinkley, and J. P. Peterson. Multivariate Zero-Inflated Poisson Models and Their Applications. *Technometrics,* Vol. 41, No. 1, 1999, pp. 29–38.
38. Karlis, D. An EM Algorithm for Multivariate Poisson Distribution and Related Models. *Journal of Applied Statistics,* Vol. 30, No. 1, 2003, pp. 63–77.
39. Christiansen, C. L., C. N. Morris, and O. J. Pendleton. *A Hierarchical Poisson Model with Beta Adjustments for Traffic Accident Analysis.* Technical Report 103. Center for Statistical Sciences, University of Texas at Austin, 1992.
40. MacNab, Y. C. A Bayesian Hierarchical Model for Accident and Injury Surveillance. *Accident Analysis and Prevention,* Vol. 35, No. 1, 2003, pp. 91–102.
41. Tsionas, E. G. Bayesian Multivariate Poisson Regression. *Communications in Statistics—Theory and Methods,* Vol. 30, No. 2, 2001, pp. 243–255.
42. Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics,* Vol. 21, No. 6, 1953, pp. 1087–1092.
43. Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika,* Vol. 57, No. 1, 1970, pp. 97–109.
44. Tanner, M. A., and W. H. Wong. The Calculation of Posterior Distribution by Data Augmentation. *Journal of the American Statistical Association,* Vol. 82, 1987, pp. 520–540.
45. Gelfand, A. E., and F. M. Smith. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association,* Vol. 85, 1990, pp. 398–409.
46. Smith, A. F. M., and G. O. Roberts. Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B,* Vol. 55, No. 1, 1993, pp. 3–23.
47. Tierney, L. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics,* Vol. 22, No. 4, 1994, pp. 1701–1728.
48. Lee, P. M. *Bayesian Statistics: An Introduction.* Oxford University Press, New York, 2004.
49. Roberts, G. O., and F. M. Smith. Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis–Hastings Algorithms. *Stochastic Processes and Their Applications,* Vol. 49, No. 2, 1994, pp. 207–216.
50. Chib, S., and E. Greenberg. Understanding the Metropolis–Hastings Algorithm. *American Statistician,* Vol. 49, No. 4, 1995, pp. 327–335.
51. Brown, D. B., S. Maghsoodloo, and M. E. Ardle. The Safety Impact of 65 MPH Speed Limit: A Case Study Using Alabama Accident Records. *Journal of Safety Research,* Vol. 21, No. 4, 1990, pp. 125–139.
52. Freedman, M., and J. R. Esterlitz. Effect of the 65-mph Speed Limit on Speeds in Three States. In *Transportation Research Record 1281,* TRB, National Research Council, Washington, D.C., 1990, pp. 52–61.
53. Mace, D. J., and R. Heckard. *Effect of the 65 mph Speed Limit on Travel Speeds and Related Crashes.* Publication DOT HS 807 764. NHTSA, U.S. Department of Transportation, 1991.
54. *Report to Congress on the Effects of the 65 mph Speed Limit During 1987.* NHTSA, U.S. Department of Transportation, 1989.
55. *Report to Congress on the Effects of the 65 mph Speed Limit Through 1990.* NHTSA, U.S. Department of Transportation, 1992.
56. Parker, M. R. *Effects of Raising and Lowering Speed Limits on Selected Roadway Sections.* Publication FHWA-RD-92-084. FHWA, U.S. Department of Transportation, 1997.
57. Pfefer, R. C., W. W. Stenzel, and B. D. Lee. Safety Impact of the 65-mph Speed Limit: A Time Series Analysis. In *Transportation Research Record 1318,* TRB, National Research Council, Washington, D.C., 1991, pp. 22–33.
58. Kockelman, K., J. Bottom, Y.-J. Kweon, J. Ma, and X. Wang. *NCHRP Web-Only Document 90: Safety Impacts and Other Implications of Raised Speed Limits on High-Speed Roads.* Final Report, NCHRP Project 17-23. Transportation Research Board of the National Academies, Washington, D.C., 2006. www.trb.org/publications/nchrp/nchrp_w90.pdf.
59. *Special Report 254: Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits.* TRB, National Research Council, Washington, D.C., 1998.
60. *Departmental Guidance for the Valuation of Travel Time in Economic Analysis.* U.S. Department of Transportation, 1997. ostpxweb.dot.gov/policy/Data/VOT97guid.pdf. Accessed June 6, 2005.
61. *Revised Departmental Guidance: Valuation of Travel Time in Economic Analysis.* U.S. Department of Transportation, 2003. ostpxweb.dot.gov/policy/Data/VOTrevision1_2-11-03.pdf. Accessed July 4, 2005.
62. Mrozek, J. R., and L. O. Taylor. What Determines the Value of Life? A Meta-Analysis. *Journal of Policy Analysis and Management,* Vol. 21, No. 2, 2002, pp. 253–270.
63. Terza, J. V., and P. Wilson. Analyzing Frequencies of Several Types of Events: A Mixed Multinomial-Poisson Approach. *Review of Economics and Statistics,* Vol. 72, No. 10, 1990, pp. 108–115.
64. Kockelman, K. A Model for Time- and Budget-Constrained Activity Demand Analysis. *Transportation Research B,* Vol. 35, No. 3, 2001, pp. 255–269.


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