SPATIALLY DISAGGREGATE PANEL MODELS OF CRASH AND INJURY COUNTS: THE EFFECT OF SPEED LIMITS AND DESIGN

Young-Jun Kweon
Graduate Student Researcher
Department of Civil Engineering
The University of Texas at Austin
ECJ 6.9, Austin, Texas 78712
Email: fire264@mail.utexas.edu

Kara M. Kockelman
(Corresponding Author)
Clare Boothe Luce Assistant Professor of Civil Engineering
The University of Texas at Austin
ECJ 6.9, Austin, Texas 78712,
Tel: (512) 471-4379
FAX: (512) 475-8744
Email: kkockelm@mail.utexas.edu

Presented at the 83rd Annual Meeting of the Transportation Research Board, January 2004

ABSTRACT

This work statistically examines the impacts of the 1996 speed limit changes based on over 6,000 Washington State highway segments. Fixed-effects and random-effects Poisson and negative binomial regression specifications are employed to estimate six crash count measures during the 1993-1996 period. The modeled crash measures are the numbers of fatalities, injuries, crashes, fatal crashes, injury crashes, and property-damage-only (PDO) crashes.

The average segment length is just 0.131 mile, permitting tight control of geometric characteristics, such as curvature and grade, as well as vehicle miles traveled (VMT). A 10 mph speed limit increase, typical of U.S. state policies pursued in the mid-1980s (on rural interstates) and in the mid-1990s, is estimated to increase fatalities and injuries by 78 and 24 percent, respectively, assuming other factors remain constant. Speed limit effects on total crashes (and property-damage-only crashes), however, are predicted to be slightly negative, suggesting that crashes become more severe, but not more common. As expected, tighter horizontal curves, fewer lanes, and lower traffic volumes (per lane) increase fatal and injurious crash rates.

Key Words: Crash models; Panel data estimation; Speed limits; Poisson; Negative binomial; Fixed-effects and Random-effects; Highway Safety Information System (HSIS)

1. INTRODUCTION

Changing speed limits on high-speed roads has been a contentious issue. The prevailing hypothesis is that speed limit increases compromise traffic safety, due to reduced response times and higher speeds upon impact. However, some suggest that raising speed limits reduces speed variations in traffic streams and diverts traffic from slower, low-design roads to faster, high-design roads, thus enhancing overall safety. Others emphasize the value of reduced travel times. Studies examining safety hypotheses generally have used spatially aggregate data. Of course, spatial aggregation obscures the impacts of roadway design features, so this work focuses on a series of short roadway sections, with average lengths under 700 feet.

Many studies on the safety effects of speed limit changes exist (e.g., Baum et al., 1991; Lave and Ellias 1994 and 1997; Houston, 1999; Ossiander and Cummings, 2002). These have used a variety of statistical methods, including before-after comparisons, time-series analysis, cross-sectional regressions, and panel data analysis. However, all have relied on data that are highly spatially aggregate, encompassing entire state networks by class of road (e.g., rural interstate versus urban interstate). Data aggregation invites an "ecological fallacy" in results, where individual-level relationships cannot be inferred (Robinson, 1950; Fridstrom and Ingebrigtsen, 1991; Davis, 2002).

Fortunately, relatively disaggregate road-section-specific data are available, such as those detailed by the U.S. Highway Safety Information System (HSIS) and provided by the Federal Highways Administration (FHWA). Such data have been used to investigate traffic safety factors other than speed limits, including horizontal curvature and truck vehicle miles-traveled (VMT) (e.g., Miaou et al., 1992, 1993, 1994; Mohamedshah et al., 1993). By using data that characterize traffic and road design, one can better control for all relevant factors, helping avoid bias in parameters associated with speed limits (since speed limits are tied to design, and often to traffic).

Using a four-year panel of the HSIS data, this study investigates the safety effects of the speed limit increases that took place across Washington State in March 1996. It employs panel regression methods for crash count data, using both fixed-effects and random-effects models to capture correlation in repeat observations of roadway sections. It examines six different crash count measures, permitting a more robust understanding of the effects of speed limit changes on crash frequency and severity.

2. LITERATURE REVIEW

This section discusses the results of past models of crash occurrence and investigations into the safety effects of speed limit changes. It is followed by sections that detail the data and models employed in the present study.

Models of Crash Occurrence

Researchers around the globe have modeled crash occurrence. The best methods employ count models, owing to the rarity of crashes and the non-negativity of their counts. For example, Fridstrom and Ingebrigtsen (1991) assembled monthly crash data over 17 years for 18 Norwegian counties, and modeled crash counts as a function of various factors (including weather, alcohol sales, network length, and annual maintenance expenditures). They employed negative binomial specifications, thus allowing overdispersion[1] in crash counts. Their model specifications suggested use of panel models; however, they did not mention how serial trends were captured, so it is unclear whether these were reasonably accommodated. Moreover, they did not control for speed limits, as done here.

Miaou et al. (1992) used 1985-1987 HSIS data for 1,644 rural interstate segments in Utah to explore the effects of geometric features on truck crashes, using Poisson regression models. Although the data tested positive for overdispersion (based on a test by Wedderburn [1974]), their conclusions regarding the nature and strength of different relationships was not predicted to change. Miaou and Lum (1993) later used the same data to compare the results of ordinary least squares models with Poisson count models, and found the former severely lacking. This is to be expected, given the count nature of the data. Neither Miaou et al. (1992) nor Miaou and Lum (1993) models recognized the panel nature of the data. And they did not control for speed limits. Shankar et al. (1995) used negative binomial models for crash counts, while controlling for roadway geometry and weather (i.e., snow and rainfall). Their data set focused on ten 3.8 mile (6.1 km) sections of I-90 in Washington State over the period of 1988-1993. Shankar et al. (1997) then investigated the applicability of zero-inflated Poisson (ZIP) and negative binomial (ZINB) models using a two–year data set. Zero-inflated models typically use a binary logit model for segmentation of roadways into zero and non-zero crash counts, and a count model (Poisson or negative binomial) for those exhibiting crashes. They found the zero-inflated models performed better, in a statistically significant way, than the non-inflated models. This suggests that some road sections may be characterized by a very low to no-crash state, distinguishing themselves from crash-prone roadways. The zero-inflated models did not control for panel effects in the two-year data set, nor for speed limits.

Research on the Safety Impacts of Speed Limit Changes

Since 1974, three national-level changes in speed limit policy have occurred, and each has generated at least a few safety impact studies. The National Maximum Speed Limit (NMSL) of 55 mph was introduced across the U.S. in 1974; it was relaxed in 1985, and completely repealed in 1995. The 1985 changes drew substantial research attention. Since the 1995 repeal, only a few significant studies have been completed (Farmer et al., 1999; Moore, 1999; Patterson 2002). Among the earlier studies, inconsistencies in findings are relatively common, due in large part to different methodologies and data settings (including temporal and spatial aggregation). In general, however, an increase in speed limits has been estimated to result in more fatalities, per mile driven. Of course, increased speeds can result in saved travel time. Thus, it is important to quantify both the costs and benefits, to permit conclusions regarding the overall impacts of speed limits.

As for the impact of the 1985 relaxation of the NMSL, some studies employed time series models. McKnight and Klein (1990) used ARIMA[2] intervention models for yearly counts of fatal and injury crashes between 1982 and 1988. They estimated a 22% increase in fatal crashes on rural interstates with the higher (65-mph) limit. Wagenaar et al. (1990) also used an ARIMA intervention model, using all limited-access rural highways (with 65 and 55 mph speed limits) in Michigan and monthly data from 1978 to 1988. They predicted count increases for all injury levels on rural 65 mph highways, including a 19% increase in fatalities. They also reported a considerable speed-spillover[3] effect onto 55 mph rural highways, resulting in a startling 38%

increase in predicted fatalities. If such results exist, even *reductions* in crash counts on roadways with high speed limits may not offset spillover effects.

Rock (1995) employed an ARIMA intervention model using 1982 to 1991 monthly crash data for rural Illinois highways and reported adverse effects caused by the new, 65 mph speed limits across all injury levels on rural highways with and without speed limit changes. Ledolter and Chan (1996) adopted a similar model; using 1983-1991 quarterly crash data, they attributed a 57% increase in fatal crashes on rural Iowa interstates to the 65 mph limit.

Before-after comparisons are very common methods of analyzing crash counts. Using "naïve comparisons"[4], without statistical measures, Upchurch's (1989) plots for Arizona's rural interstates indicate 12% and 17% increases in fatal and injurious crash rates, respectively, between April 1987 and April 1988. He also noted a slight decline in crash rates on urban segments. Lynn and Jernigan (1992) found a 54% increase in fatal crashes and fatalities on rural Virginia interstates during this time, but noted no statistically significant increases in crash rates on urban interstates. However, their before-after comparison of counts, without any accounting for exposure (i.e., VMT), may not be entirely valid.

Through a before-after comparison with odds ratios for monthly fatalities across 38 states in 1987 and yearly fatalities across 48 states between 1987 and 1988, Baum et al (1989, 1991) concluded that the increased speed limits on rural interstates was harmful, with 15 and 29 percent increases in fatalities in 1987 and 1989, respectively. Gallaher et al. (1989) used a linear trend regression and before-after comparisons of annual fatal crash rates for New Mexico between 1982 and 1988, and concluded that the increase in speed limits resulted in a 93 percent higher crash rate.[5] These results are quite different – but both, like others discussed here, suggest substantial increases in crashes, death, and injury.

A few studies have investigated the impact of the 1985 speed limit change using cross-sectional regression methods. Garber and Graham (1990) modeled monthly fatalities on rural roads in 40 states and estimated median increases in crash fatalities to be 15% and 5%, on rural interstates and rural non-interstates, respectively. They also estimated that some states experienced *reductions* in fatalities (on rural non-interstates in Alabama, Florida, Indiana, and Montana, and on non-rural interstates in California, Colorado, Louisiana, and Texas). Lave and Ellias' (1994, 1997) 40 state-by-state linear regression models for monthly fatality rates between 1976 and 1990 also indicated *average reductions* of 3.4 to 5.1% in statewide fatality rates (following the speed limit increase).

There also are at least two studies where crash effects were ambiguous or negligible. Based on ARIMA models and before-after comparisons for 32 states between 1975 and 1988, Chang and Paniati (1990) could not arrive at a definite conclusion regarding effects of the 1985 speed limit changes; this may have been due to the short "after" period of their data. Using a Poisson model for monthly crash rates across rural Ohio highways between 1982 and 1988, Pant et al. (1992) found no appreciable effect.

Using similar specifications to Lave and Ellias's (however, they used state-by-state regressions) with some additional specifications such as indicators for years and data for different road types below the state level, Greenstone (2002) examined yearly fatality rates in all 50 states for the 1982-1990 period and concluded that fatality rates increased by 30% on rural interstates but decreased by 17% on urban non-interstates. His state-level aggregate models found no evidence of police resource reallocation after the speed limit increase, in contrast to Lave and Ellias's (1997) suggestions.

Based on a 20-year data set for Washington State, Ossiander and Cummings (2000) examined the effects of the 1985 speed limit changes (on rural interstates) by using Poisson and negative binomial regression models for fatal and total crash counts controlling for a logarithmic VMT with its coefficient equaling to 1.0. They estimated a striking 110 percent increase in fatal crash rates on rural highways but no substantive change in urban highway crash rates after 1987.

The work described in this paper relies on panel models for analysis of count data, and it examines the effects of the 1995/1996 speed limit changes. Panel models have become popular in the safety literature only recently, but most relate to the 1985 speed limit changes and are continuous in nature, providing estimates of crash rates (rather than crash counts). For example, Houston (1999) estimated continuous fixed-effects panel models of fatality rates for all 50 U.S. states using 1981-1995 data on different road types. He found rural interstate fatality rates to rise in the wake of higher speed limits, but noted lowered rates on all other road types, and lower overall fatality rates. Hausman tests statistically supported for fixed-effects over random-effects models; however, Houston's models did not account for improvements in vehicle and roadway design (often incorporated via a simple time trend variable).

Relatively few studies exist that relate to the most recent speed limit change, the repeal of the National Maximum Speed Limit in 1995. Moore (1999) reported reductions in all crash rates (including injury and fatal crash rates) based on before-after rate comparisons for the entire U.S. using 1995-1997 data. Farmer et al. (1999) predicted fatality rates and counts using time series regression models of the 1990-1997 period. Based on the data from all roads in 31 states, they estimated a 15% increase in fatalities and a 17% increase in fatality rates on interstate highways but no changes on non-interstate highways in the 24 states that raised their limits.

Patterson et al. (2002) estimated cross-sectional regression models of fatality rates using 1992-1999 data for 34 states with and without speed limit increases. They reported 35 to 38% increases in fatality rates following the 1996 speed limit changes. Haselton et al. (2002) used Hauer's observational before-after approach (1997) and compared it to results from standard before-after and cross-sectional regression analyses. Using crash rates of different severity on California freeways, they found the cross-sectional regression results to be unreliable; based on their before-after approach, they reported increases in crash rates (and crash counts).

It is striking how distinct – and even contradictory – many of these works' empirical results appear. Even so, a few expectations seem reasonable. For example, it may be that speed limit increases make more sense on high-design urban freeways than on rural interstates and similarly designed facilities. It may be that route substitution offsets higher crash rates on corridors whose limits were raised. However, speed spillover can raise crash rates elsewhere. It also is possible that some states truly experienced overall safer conditions following speed limit changes, particularly the 1995/1996 changes.

Certainly, the methods and data sets used in the research described here differ, often dramatically. A key question arises: What are the best methods and the best data? Most studies have relied on highly aggregate data; and only a few have made use of panel approaches and/or recent data sets, particularly for count data. This study focuses on the effects of the 1995 repeal of the NMSL using geographically disaggregate data and rigorous panel techniques for count data. It is hoped that the meticulous approach and the detailed data permit a true relationship to emerge, at least as experienced by the State of Washington.

3. DATA

This study focuses on roadway section-specific crash data contained in the HSIS data set. Among several states voluntarily providing HSIS data, Washington State was chosen for this analysis because it offers curve and grade information[6] and offers the possibility of linkage to a long history of loop detector data, at sites throughout the state. Only interstate sections are examined here, in order to provide focus while moderating the number and type of unobserved effects[7]. This focus produces 6,023 interstate segments, along I-5, I-82, I-90, I-205, and I-405.[8] Total one-way length of the included interstates is 747.58 miles (WSDOT, 2003).

The Washington HSIS data are available from 1993 to1996[9]. Yearly crash counts were assembled at a segment level, along with roadway design and other variables; these are summarized in Table 1. The average segment is 0.131 miles long, and the longest segment is 1.69 miles. Such spatial resolution is unusual and valuable. The vehicle miles-traveled (VMT) variable was computed by multiplying segment length and segment estimated annualized average daily traffic (AADT). An average of 2.7 crashes occurred per segment per year, injuring almost two persons (per segment, per year). The yearly counts are modeled using models that recognize both the panel nature of the data, and the integer crash values. These methods are described here now.

<Table 1>

4. METHODOLOGY

Poisson and negative binomial (NB) fixed-effects and random-effects models were used to predict six different crash measures. Pooled Poisson and NB models are described first since these overlook the panel nature of the data and thus are the simplest. Descriptions of their extended, panel specifications, based on fixed-effects and random-effects, follow. For more details, readers may refer to Cameron and Trivedi (1998) and Hardin and Hilbe (2001).

Pooled Models for Count Data

The standard Poisson model specification is the following:

$$P(Y = y_i) = \frac{e^{-\lambda_i} \cdot \lambda_i^{y_i}}{y_i!}, \text{ where } \lambda_i = E[y_i | \mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta}), \text{ and } y_i = 0, 1, 2, \ldots \quad (1)$$

Here, *i* indexes individual observational units (in this study, individual roadway segments), $\mathbf{x}_i$ denotes a vector of explanatory variables for *i*, and $\boldsymbol{\beta}$ is a vector of coefficients to be estimated. In the context of our analysis, y_i is the number of crashes, injuries, or deaths occurring in a roadway segment *i* during a given time period. This standard model does *not* recognize the possibility of repeat observations of units over time.

A distribution that is truly Poisson exhibits equidispersion, where mean equals variance: $\lambda_i = \sigma_i^2 = \exp(\mathbf{x}_i'\boldsymbol{\beta})$. However, if the underlying process is not Poisson in nature, or, for example, all sources of systematic variation are not appropriately controlled for, overdispersion may exist, and a Poisson specification will not be most appropriate. In the case of overdispersion, a popular alternative is the negative binomial distribution, which can be specified as including a random variable u_i in the mean function, $\exp(\mathbf{x}_i'\boldsymbol{\beta} + u_i)$ (Fridstrom and Ingebrigtsen, 1991; Poch et al. 1996; Cameron and Trivedi, 1998; Mohamed et al., 2000). If this random term u_i follows a Gamma distribution with shape and scale parameters both equalling θ, one has the following negative binomial model:

$$P(Y_i = y_i) = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{\Gamma(y_i + 1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{1}{1 + \alpha\mu_i}\right), \text{ where}$$

$$\mu_i = E[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta} + u_i) = \upsilon_i \exp(\mathbf{x}_i'\boldsymbol{\beta}) \tag{2}$$

Here, $\alpha = \frac{1}{\theta}$ is the overdispersion parameter and $\Gamma(\;)$ is the gamma function. Given the mean or expected count, μ_i, the count variance is $V[Y_i] = \sigma_i^2 = \mu_i(1 + \alpha \cdot \mu_i)$. Cameron and Trivedi (1986) refer to this mixed gamma-Poisson model as the Type II negative binomial model.[10]

Panel Models of Count Data

For each count model, Poisson and negative binomial panel model specifications involving fixed-effects and random-effects also are estimable.

Poisson Panel

The conditional joint distribution for a Poisson-based random-effects model is the following:

$$\Pr\left(y_{i1},\ldots,y_{iT_i}|\upsilon_i,\mathbf{x}_i\right) = \left[\prod_{t=1}^{T_i}\frac{\lambda_{it}^{y_{it}}}{y_{it}!}\right]\exp\left(-\upsilon_i\sum_{t=1}^{T_i}\lambda_{it}\right)\upsilon_i^{\sum_{t=1}^{T_i}y_{it}}, \text{ where}$$

$$\upsilon_i \sim gamma(\theta,\theta) \text{ with } E[\upsilon_i] = 1 \text{ and } V[\upsilon_i] = 1/\theta \tag{3}$$

Here, υ_i represents the random-effects applicable to each group *i* (such as a specific roadway segment) and $\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$ ensures non-negativity.

Integrating out the gamma-distributed υ_i results in the following (unconditional) distribution:

$$\Pr\left(y_{i1},\ldots,y_{iT_i}|\mathbf{x}_i\right) = \left[\prod_{t=1}^{T_i}\frac{\lambda_{it}^{y_{it}}}{y_{it}!}\right]\cdot\frac{\Gamma\left(\theta + \sum_{t=1}^{T_i}y_{it}\right)}{\Gamma(\theta)}\left(\frac{\theta}{\theta + \sum_{t=1}^{T_i}\lambda_{it}}\right)\left(\frac{1}{\theta + \sum_{t=1}^{T_i}\lambda_{it}}\right)^{\sum_{t=1}^{T_i}y_{it}} \tag{4}$$

For the fixed-effects version of this model, the fixed-effects, υ_i, can take on any value. By conditioning on $\sum_{t=1}^{T_i} y_{it}$, the fixed-effects disappear from the log-likelihood function, and the parameters $\boldsymbol{\beta}$ become reasonably estimable in a situation with hundreds of observation sections (roadway units, in this case). However, each unit's probabilities are conditioned on total crash counts for that section:

$$\Pr\left(y_{i1},\ldots,y_{iT_i}\middle|\sum_{t=1}^{T_i} y_{it},\mathbf{x}_i\right)=\frac{\left(\sum_{t=1}^{T_i} y_{it}\right)!}{\prod_{t=1}^{T_i} y_{it}!}\cdot\prod_{t=1}^{T_i}\left[\frac{\lambda_{it}}{\sum_{s=1}^{T_i}\lambda_{is}}\right]^{y_{it}} \quad (5)$$

Negative Binomial Panel

In a panel analysis of negative binomial models, the individual effects do not apply to the mean rate, but to the distribution of the dispersion parameter. Random-effects models are most appropriate when dispersion varies randomly across groups due to unidentified factors specific to groups. This is similar to the linear, or continuous, panel case: individual effects are randomly distributed in a linear random-effects model, but can take any value in a linear fixed-effects model. In the standard random-effects case, the dispersion varies randomly across groups according to a *beta* distribution with parameters *r* and *s*. The gamma distribution is a special case of the more general beta distribution; it provides the following random-effects negative binomial example:

$$Y_{it}|\gamma_{it} \sim Poisson(\gamma_{it}), \quad (6)$$

where $\gamma_{it}|\delta_i \sim gamma(\lambda_{it}, 1/\delta_i)$.

Here, $\lambda_{it}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta})$ and δ_i is a dispersion parameter. (StataCorp, 2003)

This specification has not been utilized to a great extent in the past. Recently, however, Noland (2003) used fixed-effects and random-effects negative binomial models to examine the effects of U.S. roadway improvements on traffic safety. And McCarthy (1999) used fixed-effects negative binomial models to investigate fatal crash counts in a panel data set of 418 U.S. cities and 57 U.S. regions. Both studies adopted count panel models, but used highly spatially aggregate data (at the state and city levels). In contrast, this study uses spatially disaggregate data, based on one-way roadway segments averaging just 0.13 miles in length, enabling detailed levels of control on speed limit and design variables.

A more general version of random-effects negative binomial models allows the dispersion parameter to vary such that $1/(1+\delta_i) \sim beta(r,s)$. This approach, used for this study, yields the following joint probability for the *i*th group:

$$\Pr\left(y_{i1},\ldots,y_{iT_i}\right)=\frac{\Gamma(r+s)\Gamma\left(r+\sum_{t=1}^{T_i}\lambda_{it}\right)\Gamma\left(s+\sum_{t=1}^{T_i}y_{it}\right)}{\Gamma(r)\Gamma(s)\Gamma\left(r+s+\sum_{t=1}^{T_i}\lambda_{it}+\sum_{t=1}^{T_i}y_{it}\right)}\prod_{t=1}^{T_i}\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)} \quad (7)$$

For a fixed-effects negative binomial model, each group's dispersion parameter can take on any value. By adopting a conditional likelihood estimation process, like the Poisson fixed-effects model described earlier, the dispersion parameter drops and the $\boldsymbol{\beta}$ parameters can be identified/estimated without simultaneously trying to tackle all the fixed-effects parameter values (Hausman et al. 1994). When conditioned on the group's total count, a fixed-effects joint probability for group *i*, can be expressed as follows:

$$\Pr\left(y_{i1},\ldots,y_{iT_i}\middle|\sum_{t=1}^{T_i}y_{it}\right)=\frac{\Gamma\left(\sum_{t=1}^{T_i}\lambda_{it}\right)\Gamma\left(\sum_{t=1}^{T_i}y_{it}+1\right)}{\Gamma\left(\sum_{t=1}^{T_i}\lambda_{it}+\sum_{t=1}^{T_i}y_{it}\right)}\prod_{t=1}^{T_i}\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)} \quad (8)$$

After initial tests using pooled fixed-effects and random-effects count models (both the Poisson and negative binomial specifications), the more complex panel models were employed. For further details on count panel specifications, one may consult Cameron and Trivedi (1998) and/or StataCorp (2003).

5. RESULTS AND DISCUSSIONS

Cross-sectional (or pooled) count data models were first estimated for six different crash measures: the numbers of dead occupants, injured occupants, all crashes, fatal crashes, injury crashes, and property-damage-only (PDO) crashes (Table 2). Given the large number of possible specifications (i.e., Poisson and negative binomial, pooled or panel), the first set of model estimates essentially disregarded the data set's panel properties, and assumed independence of observations[11]. Based on these results, the choice between Poisson and negative binomial specifications was made using the likelihood ratio test for the overdispersion parameter. Then fixed-effects (Table 3) and random-effects (Table 4) panel specifications, based on the selected Poisson or negative binomial model, were estimated and compared.

Several rate specifications were explored, with explanatory variables included either linearly or logarithmically within the exponential function. Those included logarithmically imply a multiplicative effect, subject to an exponential term, as illustrated for *VMT* in the following equation: $\lambda(x, VMT) = VMT^{\alpha}\exp(\beta x) = \exp(\beta x + \alpha \ln(VMT))$.

Generally, VMT is expected to have a purely multiplicative effect, with an exponent of α=1.0. And, in the end, this specification provided the best predictions (highest goodness of fit statistics). Tables 2 through 5 include all explanatory variables in a linear form of the exponential function, except for VMT. VMT enters the rate function in a multiplicative fashion, to allow for the possibility of constant crash rates, with an exponent of 1.0. The following describes the estimated effects of the numerous explanatory variables on the various crash crates (and the resulting counts).

Interpretation of Coefficients

Due to the exponential transformation used (to ensure process rate non-negativity), the effects of the model coefficients are not as obvious as in simple, linear models (i.e., where $E(y|x) = \beta x + \varepsilon$). In an exponential case, $E[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$, and each variable's marginal effect is as follows:

$$\frac{\partial E[y_i|\mathbf{x}_i]}{\partial x_{ji}} = \exp(\mathbf{x}_i'\boldsymbol{\beta}) \times \beta_j = E[y_i|\mathbf{x}_i] \times \beta_j \text{ or } \frac{\partial E[y_i|\mathbf{x}_i] / E[y_i|\mathbf{x}_i]}{\partial x_{ji}} = \beta_j \quad (9)$$

This implies that a unit change in the *j*th variable leads to a multiplicative change in $E[y_i|\mathbf{x}_i]$ of β_j. In other words, if β_j=1.1, a unit change in $x_{j.}$ will increase the mean by 110%.
Another way to interpret such effects is via an incidence rate ratio (IRR), also called factor change (Long, 1997; Long and Freeese, 2001; StataCorp, 2003):

$$IRR(x_{j.}) = \frac{\exp[\beta_1 x_1 + \cdots + \beta_j (x_j + 1) + \cdots + \beta_J x_J]}{\exp[\beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_J x_J]} = \exp(\beta_j) \quad (10)$$

Thus, if β_j = 1.1, then the $IRR(x_j)$ = exp(1.1) = 3.0, so a unit increase in $x_{j\cdot}$ would be estimated to increase the mean by 300%, assuming other factors remain constant.

In cases of logarithmically transformed variables such as VMT, the coefficient can be interpreted as an elasticity. Thus, if β_{VMT} equals 1.1, a doubling (100% increase) of VMT is expected to result in a 110% increase in average crash rate.

Regression Results

Based on comparisons of pooled-data model log-likelihoods, a Poisson model appeared adequate for counts of fatal crashes, whereas negative binomial specifications were needed for the other five crash counts, complete with statistically significant estimates of their dispersion parameters, as shown in Table 2. Given that crashes are rare events, perhaps described by a Bernoulli with a very low probability of occurrence in any short time interval, one may hypothesize crash counts to be Poisson distributed (Fridstrom and Ingebrigtsen, 1991). However, that hypothesis is often rejected in favor of overdispersion (e.g., Shankar et al., 1995; Poch and Mannering, 1996; Mohamed et al., 2000). Miaou and Lum (1993) have suggested three sources of such overdispersion; these are omitted variables (such as human factors and vehicle information), uncertainties in exposure data (such as possible sampling errors for AADT estimates), and heterogeneous roadway conditions (such as differences in lighting and weather conditions). All these may apply here.

<Table 2>

Panel Regressions

Fixed-effects and random-effects count panel regression models for each dependent variable were estimated, and the results of these models are condensed in Tables 3 and 4. Recall that pooled negative binomial specifications permit overdispersion and thus, to great extent, are able to reflect heterogeneity (that remains after controlling for various, explanatory factors). However, negative binomial *panel* models can offer further control, for remaining heterogeneity, particularly omitted variable biases (McCarthy, 1999). Unfortunately, such models also may *over*-control for data overdisperion.

Based on a likelihood-ratio test for comparison of random-effects and pooled models (analogous to Breusch and Pagan's Lagrange multiplier test in linear panel models [Greene, 2000]), the pooled models for counts of fatal crashes and fatalities appear statistically appropriate. Interestingly, results of the fixed-effects and random-effects models differ, in terms of the effect of the factor of primary interest: roadway speed limits.

Speed-limit effects in the random-effects models are consistent with those in the pooled models: higher speed limits have a positive effect on fatalities, yet a negative effect on total and PDO crash counts. They have no statistically significant effects on counts of injuries, fatal crashes, and injury crashes. In the random-effects and pooled model, a 1 mph increase in speed limits is predicted to result in fatal crash increases of 2.4 and 5.9 percent, respectively. This same increase, however, is predicted to result in 0.8 and 0.6 percent reductions in PDO and total crashes, using the random-effects specifications. In contrast, the fixed-effects models suggest increases in both PDO and total crash counts due to speed limit increases, and 2.1 and 4.1

percent increases in injuries and injury crashes, respectively, assuming a 1 mph increase in the posted speed limit.

<Table 3>

<Table 4>

Hausman's specification test (1978, 1984) can compare fixed-effects and random-effects in both linear and count panel specifications. Random-effects models were clearly rejected for injury and total crash count measures; pooled models were already chosen for fatalities and fatal crash counts. Such rejection suggests either misspecification or correlation between random-effects and explanatory variables. Given how many valuable variables are unobserved yet likely to be correlated with control variables (such as sight distances, which generally are correlated with speed limits, weather, which may be correlated with mountainous terrain, and vehicle type, which may be correlated with the number of lanes), correlation between random-effects and explanatory variables is a likely culprit.

Final Models

Based on the results of three different tests in each count case (one test for overdispersion, one for random-effects versus pooled estimators, and one for fixed-effects versus random-effects), the final model specifications were determined. The ostensibly best models based on three tests were fixed-effects negative binomial models (FENB) for counts of injuries and total crashes, random-effect negative binomial models (RENB) for counts of injury crashes and PDO crashes, pooled negative binomial models for fatalities, and pooled Poisson models for fatal crash counts. However, comparisons of actual counts and predicted rates showed that statistically insignificant constant term values were effectively biasing the results. Ultimately, the best models were found to be a fixed-effects negative binomial model for injuries, a pooled Poisson model for fatal crashes, and pooled negative binomial models for fatalities, injury crashes, PDO crashes, and total crashes. These are presented in Table 5. In data sets longer than four years, and/or those with more observations per roadway section (e.g., monthly data), panel models may perform best for all count variables, since there is more reliable or consistent information available per section.

<Table 5>

Speed limit increases are predicted to result in more fatalities and injuries. The adverse effects are estimated to be quite large for fatalities: a 1 mph speed limit increase is associated with a 5.9 percent increase in fatalities and 2.1 percent rise in injuries, but reductions of 0.7 and 0.4 percent in PDO crashes and total crashes, respectively. No statistically significant changes in the numbers of fatal and injury crashes were detected due to speed limit increases; however, there were significant increases in these numbers in 1996 (by 32 and 11 percent, respectively) due to unobserved changes occurring in that year. Increases in the numbers of victims yet no changes in crash counts in fatal and injury crashes are probably due to the presence of more severe – though not fatal – crashes, where more involved travelers experience an injury[12].

The variable AADT per lane acts as a proxy for congestion (helping relieve some of the burden on the VMT variable). As expected, it is predicted to have a negative effect on fatalities and fatal crashes. Roads with 1000 more vehicles per lane per day are predicted to have 6.0 and 8.3 percent fewer fatalities and fatal crashes, respectively. However, this variable appears to increase injury crashes, PDO crashes, and total crashes.

In terms of geometric features, highways with more lanes are estimated to experience fewer crashes and crash victims in any severity level, everything else constant. The effects are greatest for fatalities (26 percent) and fatal crashes (24 percent). However, more lanes typically mean more traffic, an urban area, and regular congestion. Such congestion slows the crash-involved vehicles and results in fewer deaths and injury, no doubt.

A wider median is predicted to reduce non-fatal and total crashes slightly (less than 1 percent with 10ft wider medians). Vertical grade appears to have no effect, except for PDO crashes (where it adds to crash rates), while an additional 100 ft in vertical curve length, is predicted to reduce fatalities and fatal crashes slightly (by 2 and 1 percent, respectively).

Longer *horizontal* curves were estimated to decrease injuries, injury crashes, PDO crashes, and total crashes, but have no effect on fatalities and fatal crashes. These reducing effects range from 6 to 8 percent for every 1000 additional feet of horizontal curve length, perhaps because drivers must turn more tightly in order to stay in their lane. The degree of curvature, which is inversely proportional to curve radius (Δ [deg/100 ft] = 5729/R, where R = radius [feet]), has a positive effect on all crash measures, as expected (due to higher centrifugal forces and tighter turning requirements). A one-degree increase is associated with 21 and 13 percent more fatalities and fatal crashes, respectively.

Remarkably, 71 percent *fewer fatalities* and 57 percent fewer fatal crashes are expected in mountainous terrain, compared to level terrain. However, 81, 83, and 113 percent *more injuries*, injury crashes, and PDO crashes are predicted in mountainous terrain. Rolling terrain's effects lie somewhere between those of mountainous and level terrain. Travelers appear to be at somewhat higher risk (to injury) but drivers appear to exercise more caution while in mountainous or rolling terrain than level terrain (resulting in fewer predicted fatalities). Under wet or snowy road-surface conditions, expected fatalities and fatal crashes are 46 and 47 percent lower than on dry surface conditions, respectively. This also is probably due to more careful driving, under poor driving conditions.

Model Limitations

McFadden's likelihood-ratio index (LRI) (1974) suggests the estimated models have a likelihood not too much higher than that based on a constants-only approach. However, low LRIs are common with count models and rare events, such as crashes, particularly in disaggregate studies such as this. For example, Shankar et al.'s models (1997), which controlled for geometry and weather, produced LRIs of 0.07 to 0.11. Model enhancement is possible. For example, zero-inflated Poisson and negative binomial models may better capture the nature of crash counts, based on two underlying states: an inherently safe state and an potential-for-crash state. Zero-inflated models explicitly reflect high zero crash counts by changing the conditional mean's structure (Greene, 2000). And additional explanatory variables, such as speed and weather information, add explanatory power.

In order to examine how high of LRI the current model can reach, simulations were performed assuming the model estimates are true parameters. For illustration, the negative binomial model for total crash counts was chosen for simulations. The total crash counts were generated based on the negative binomial model specification: $y_i \sim Poisson(\mu_i)$ and $\mu_i \sim Gamma(\frac{1}{\alpha}, \alpha\lambda_i)$ where $\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{X})$. Estimates of parameters, α and $\boldsymbol{\beta}$ from the negative binomial model for total crash counts were plugged in for the simulation. Using the simulated total crash count, y_i, the log-likelihood value was computed using the following equation:

$$LL = \sum_{i=1}^{N}\left[\ln\Gamma\left(\frac{1}{\alpha}+y_i\right)-\ln\Gamma(y_i+1)-\ln\Gamma\left(\frac{1}{\alpha}\right)+\frac{1}{\alpha}\ln\left(\frac{1}{1+\alpha\lambda_i}\right)+y_i\ln\left(1-\frac{1}{1+\alpha\lambda_i}\right)\right]$$

10 sets of simulations were performed and the average simulated LRI is 0.079, which is still low. This indicates that the true regression model for total crash counts attains 0.079 of LRI.

Although the low LRIs are commonly found in disaggregate geometric studies, models for the four non-fatal crash measures yielded very low LRIs and do not carry much weight. In addition, due to the low crash counts per section, even over a period of a year, these data were aggregated temporally. Such aggregation may result in "ecological fallacies" in interpretation of results. (Fridstrom and Ingebrigtsen 1991; Davis, 2002)

6. CONCLUSIONS

This study examined the HSIS data in order to model crash counts along short freeway segments in Washington State between 1993 and 1996. Poisson and negative binomial fixed-effects and random-effects models were used to recognize the count and panel nature of the response variables (crash counts). The safety effects of changing speed limits were the focus of this work, and six different crash measures were used: counts of injured occupants and deaths, as well as the number of all, fatal, injury, and PDO crashes. Design variables such as horizontal and vertical curve lengths, degree of (horizontal) curve, grade, median width, number of lanes, and indicators for mountainous and rolling terrains were controlled for, along with AADT per lane (as a proxy for congestion), VMT, and pavement wetness.

For each crash measure, an appropriate model was developed starting with the same set of explanatory variables. This resulted in a fixed-effects negative binominal model for injury counts. Pooled negative binomial models performed best for fatalities, injuries, PDO crashes, and total crashes, whereas a pooled Poisson model was appropriate for fatal crash counts. Through the statistical test, negative binomial panel models were selected for injury crashes, PDO crashes, and total crashes. However, they resulted in unrealistic predictions. Instead, pooled negative binomial models were chosen even after estimations of Poisson panel models for those three dependent variables.

Raised speed limits appear to increase injuries and fatalities. Raising speed limits by just 10 mph is estimated to bring about 78 percent more fatalities and 24 percent more injuries (but reductions of 7 and 4 percent in PDO and total crashes, respectively). More traffic counteracts the adverse effects, to some extent. For example, 1,000 more AADT per lane is predicted to result in 6 and 8 percent fewer fatalities and fatal crashes, respectively. And added lanes (which provide more space for maneuvering but also are likely to be correlated with congestion) reduce crash expectations by 4 to 26%. As one might expect sharper horizontal curves increase crash rate predictions, while driving in hilly or mountainous terrain reduces fatalities and fatal crashes. Interestingly, vertical grade has no effect on crash measures except for PDO crashes, though longer vertical curves operate more safely. Median width is estimated to add to injury crashes, PDO crashes, and total crashes.

Based on the crash rate estimates, travel time savings (due to higher speed limits) were compared to crash cost estimates. Blincoe's (1994) crash cost estimates (with a four percent inflation rate) are $3,600, $195,206, and $899,803 (all in 1996 U.S. dollars) per PDO crash[13], injury, and fatality. For time saving estimation, average speeds were assumed to equal to speed limits (which is a bit generous[14]). And an $19/vehicle-hour value of travel time (in 1996$) was applied (based on McFarland and Chui's [1987] findings, and inflated at an annual rate of 4 percent. For a hypothetical speed limit increase from 60 to 70 mph on a highway with the average characteristics (in this data set), the total increase in crash cost increase was estimated to be $750 million, versus a paltry time savings valuation of $2.61 million. Among the crash costs, roughly 90 percent results from increased injury predictions.

The study findings related to speed limits, geometric design, environmental factors, and traffic conditions are useful for roadway designers, law enforcement agencies and policy makers in that these stakeholders are in a position to affect a variety of factors, thereby influencing traffic safety. For example, policy makers and enforcement agencies might want to predict the effects (including monetarized benefits and costs) of changing speed limits, given roadway geometry and traffic conditions. Certainly, traffic engineers and roadway designers also can better appreciate the safety impacts of their management and design choices.

Taken all together, these results suggest speed limits play a central role in driving safety and traffic fatalities. They dominate most other factors, and the estimated cost impacts of raised limits overwhelm even the most liberal time saving estimates. Though increasing speed limits may have some positive network or system effects (as suggested by work by Lave and Ellias (1994, 1997) and Houston (1999)), they appear to rather dramatically raise local crash injury counts and fatalities. Though crashes claim more human lives in the U.S. than any other disease or accident, they remain a rare event. And, in the long term, vehicle design and driver choices may counteract more of the risks that Americans seem to be facing due to the raised speed limits.

Acknowledgements
This study was sponsored by the American Association of State Highway and Transportation Officials (AASHTO), in cooperation with the Federal Highway Administration (FHWA), and was conducted in the National Cooperative Highway Research Program (NCHRP) 17-23. The results of this work are preliminary and have not yet been approved by the NCHRP 17-23 project panel for publication as an NCHRP report. Authors would like to thank Mohamedshah Yusuf at LENDIS Corporation for providing HSIS data, and Dr. Daniel Powers at the University of Texas at Austin for his helpful comments.

Endnotes

[1] Overdispersion exists when the variance of a distribution exceeds its mean.
[2] AutoRegressive Integrated Moving Average (ARIMA) models were developed by Box and Jenkins in 1976. The ARIMA(*p*, *d*, *q*), model relies on an autoregressive parameter *p*, nonseasonal differencing parameter *d*, and moving average parameter *q*.
[3] Speed spillover implies that drivers, accustomed to raised speed limits on some corridors, are more likely to speed on roads whose limits have not been altered.
[4] The origin of this term is Hauer's (1997) "naïve" before-after studies. However, such applications do not have to be in the context of after vs. before, so here the term naïve comparison is used.
[5] He observed 2.9 fatal crashes per 100 million VMT relative to an expected 1.5 such crashes, as predicted from a model of the previous 5 years.
[6] Only four HSIS-reporting states provide this information.
[7] Interstate road design is relatively standard, apart from the many variables controlled for here. If a variety of roadway types were included, correlation of unobserved attributes with control variables may obscure true relationships while introducing substantial variation.
[8] It excludes I-705, which measures just 1.5 centerline miles in the State of Washington.
[9] Years 1999 through 2000 should become available by early 2004. Unfortunately, data from years 1997 and 1998 are incomplete and thus not available from the HSIS.
[10] A Type I version results in $V[Y_i] = \mu_i(1+\alpha)$, requiring that dispersion be constant. Type II dispersion increases with mean value.
[11] No study was found that permits a direct choice between fixed- and random-effects Poisson and NB models. In a panel setting, with individual effects recognized via either a fixed- or random-effects specification, a Poisson approach may be sufficient, particularly if there are theoretical grounds for a Poisson specification. In such cases, permission for overdisperion may result in "overkill".
[12] This result also may be due to higher occupancy rates and/or crashes involving more vehicles. Vehicle occupancy data are not available for these data. Number of vehicles per crash is, however, and it did not vary much over the 4 year period of 1993-1996; the average number of vehicles per crash went from 1.89 in 1993, to 1.92 in 1994, to 1.94 in 1997, to 1.90 in 1998.
[13] Blincoe's (1994) costs were per PDO-crash-involved vehicle, rather than per PDO crash. There are (slightly over) two vehicles involved in every PDO crash in this data set, so the per-PDO-crash cost used is double Blincoe's PDO-vehicle cost.
[14] Actual average speed increases may be more on the order of 50%. Ossiander and Cummings (2002) reported 5.5 mph speed increase following 10 mph speed limit increases on rural interstates. However, there is also a cost of paying traffic fines when one exceeds the limit, which is no longer a factor for those who travel below the new, higher limit, and that is a cost savings/benefit as well.

References

Baum, H. M., A. K. Lund, and J. K. Wells. The mortality consequences of raising the speed limit to 65 mph on rural interstates. *American Journal of Public Health*, Vol. 79, No. 10, 1989, pp. 1392-1395.

Baum, H. M., J. K. Wells, and A. K. Lund. The fatality consequences of the 65 mph speed limits, 1989, *Journal of Safety Research*, Vol. 22, No. 4, 1991, pp. 171-177.

Blincoe, L. J. *The Economic Cost of Motor Vehicle Crashes, 1994*. National Highway Traffic Safety Administration. http://www.nhtsa.dot.gov/people/economic/ecomvc1994.html. Accessed July 30, 2003.

Cameron, A. C., and P. K. Trivedi. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, Vol. 1, No.1, 1986, pp. 29-53.

Cameron, A. C., and P. K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, New York, 1998.

Chang, G.-L., and J. F. Paniati. Effects of 65-mph speed limit on traffic safety. *Journal of Transportation Engineering*, Vol. 116, No. 2, 1990, pp. 213-226.

Davis, G. A. Is the claim that 'variance kills' an ecological fallacy? *Accident Analysis and Prevention*, Vol.. 34, No. 3, 2002, pp. 343-346.

Farmer, C. M., R. A. Retting, and A. K. Lund. Changes in motor vehicle occupant fatalities after repeal of national maximum speed limit. *Accident Analysis and Prevention*, Vol. 31, No. 5, 1999, pp. 537-543

Fridstrom, L, and S. Ingebrigtsen. An aggregate accident model based on pooled, regional time-series data. *Accident Analysis and Prevention*, Vol. 23, No. 5, 1991, pp. 363-378.

Gallaher, M. M., C. M. Sewell, S. Flint, J. L. Herndon, H. Graff, J. Fenner, and H. F. Hull. Effects of the 65-mph speed limit on rural interstate fatalities in New Mexico. *Journal of the American Medical Association*, Vol. 262, No. 16, 1989, pp. 2243-2245.

Garber, S., and J. D. Graham. The effects of the new 65 mile-per-hour speed limit on rural highway fatalities: A state-by-state analysis. *Accident Analysis and Prevention*, Vol. 22, No. 2, 1990, pp. 137-149.

Greene, W. *Econometric Analysis*. Prentice-Hall, Upper Saddle River, New Jersey, 2000.

Greenstone, M. A reexamination of resource allocation responses to the 65-mph speed limit. *Economic Inquiry*, Vol. 40, No. 2, 2002, pp. 271-278.

Hardin, J., and J. Hilbe. *Generalized Linear Models and Extensions*. Stata Press, College Station, Texas, 2001.

Haselton, C. B., A. R. Gibby, and T. C. Ferrara. Methodologies used to analyze collision experience associated with speed limit changes on selected California highways. *Transportation Research Record 1784*, 2002, pp. 65-72.

Hauer, E. *Observational before-after studies in road safety*. Pergamon, Oxford, U.K., 1997

Hausman, J. Specification tests in econometrics. *Econometrica*, Vol. 46, No. 6, 1978, pp. 1251-1271.

Hausman, J., B. Hall, and Z. Griliches. Econometric models for count data with an application to the patents–R&D relationship. *Econometrica*, Vol. 52, No. 4, 1984, pp. 909-938.

Houston, D. J. Implications of the 65 mph speed limits for traffic safety. *Evaluation Review*, Vol. 23, No. 3, 1999, pp. 304-315.

Lave, C., and P. Elias. Did the 65 mph speed limit save lives? *Accident Analysis and Prevention*, Vol. 26, No. 1, 1994, pp. 49-62.

Lave, C., and P. Elias. Resource allocation in public policy: The effects of the 65-mph speed limit. *Economic Inquiry*, Vol. 35, 1997, pp. 614-620.
Ledolter, J., and K. S. Chan. Evaluating the impact of the 65 mph maximum speed limit on Iowa rural interstates. *The American Statistician*, Vol. 50, No. 1, 1996, pp. 79-85.
Long, J. S. *Regression Models for Categorical and Limited Dependent Variables*. SAGE Publications, Thousand Oaks, California., 1997.
Long, J. S., and J. Freese. *Regression Models for Categorical Dependent Variables Using STATA*. Stata Corporation, College Station, Texas, 2001.
Lynn, C. W., and J. D. Jernigan *The Impact of the 65 mph Speed Limit on Virginia's Rural Interstate Highways Through 1990*. Report VTRC 92-R12, Virginia Transportation Research Council, 1992.
McCarthy, P. S. Public policy and highway safety: a city-wide perspective. *Regional Science and Urban Economics*, Vol. 29, 1999, pp. 231-244.
McFadden, D. The measurement of urban travel demand. *Journal of Public Economics*, Vol. 3, No. 4, 1974, pp. 303-328.
McFarland, W. F., and M. Chui. The value of travel time: New elements developed using a speed choice model. *Transportation Research Record 1116*, 1987, pp. 15-21.
McKnight, A. J., and T. M. Klein. Relationship of 65-mph limit to speeds and fatal accidents. *Transportation Research Record 1281*, 1990, pp. 71-77.
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.
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.
Miaou, S.-P., S. H. Patricia, T. Wright, A. K. Rathi, and S. C. Davis. Relationship between truck accidents and highway geometric design: A Poisson regression approach. *Transportation Research Record 1376*, 1992, pp. 10-18.
Mohamedshah, Y. M., J. F. Paniati, and A. G. Hobeika. Truck accident models for interestates and two-lane rural roads. *Transportation Research Record 1407*, 1993, pp. 35-41
Moore, S. Speed doesn't kill: The repeal of the 55-MPH speed limit. *Policy Analysis*, Vol. 346, 1999, pp. 1-23.
Noland, R. B. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. *Accident Analysis and Prevention*, Vol. 35, No. 4, 2003, pp. 599-611.
Ossiander, E. M., and P. Cummings. Freeway speed limits and traffic fatalities in Washington state. *Accident Analysis and Prevention*, Vol. 34, No. 1, 2002, pp. 13-18.
Pant, P. D., J. A. Adhami, J. C. Niehaus. Effects of the 65-mph speed limit on traffic accidents in Ohio. *Transportation Research Record 1375*, 1992, pp. 53-60.
Patterson, T. L., W. J. Frith, L. J. Povey, and M. D. Keall. The effect of increasing rural interstate speed limits in the United States. *Traffic Injury Prevention*, Vol. 3, 2002, pp. 316-320.
Poch, M., and F. Mannering. Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, Vol. 122, No. 2, 1996, pp. 105-113.
Robinson, W. S. Ecological correlations and the behavior of individuals. *American Sociological Review*, Vol. 15, No. 3, 1950, pp. 351-357.
Rock, S. M. Impact of the 65 mph speed limits on accidents, deaths, and injuries in Illinois. *Accident Analysis and Prevention*, Vol. 27, No. 2, 1995, pp. 207-214.

Shankar, V., F. Mannering, and W. Barfield. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention*, Vol. 27, No. 3, 1995, pp. 371-389.

Shankar, V., J. Milton, and F. Mannering. Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis and Prevention*, Vol. 29, No. 6, 1997, pp. 829-837.

StataCorp. *STATA References: Release 8*. Stata Press, College Station, Texas, 2003.

Upchurch, J. Arizona's experience with the 65-mph speed limit. *Transportation Research Record 1244*, 1989, pp. 1-6.

Wagenaar, A. C., F. M. Streff, and R. H. Schultz. Effects of the 65 mph speed limit on injury morbidity and mortality. *Accident Analysis and Prevention*, Vol. 22, No. 6, 1990, pp. 571-585.

Washington State Department of Transportation (WSDOT) Washington State National Highway System Designated Highways. 2002. http://www.wsdot.wa.gov/ ppsc/hsp/NHSlist.pdf. Accessed June 30, 2003.

Washington State Department of Transportation (WSDOT) *Washington State Highway System Plan 2003-2022*, Transportation Planning Office, Olympia, WA, 2003.

List of Tables

TABLE 1. Variable Definitions and Summary Statistics (Annual Washington State HSIS Data)

Variables	Description	Mean	Std. Dev.	Min	Max
Dependent Variables					
NUMDEAD	Number of fatalities	0.019	0.161	0	4
NUMINJUR	Number of injuries	1.968	3.413	0	76
NUMFATCR	Number of fatal crashes	0.017	0.130	0	2
NUMINJCR	Number of injury crashes	1.217	1.912	0	44
NUMPDOCR	Number of PDO crashes	1.514	2.063	0	37
NUMCRASH	Number of crashes	2.747	3.599	1	81
Independent Variables					
SEGLENTH	Segment length (mile)	0.131	0.146	0	1.69
CURVLENTH	Horizontal curve length (ft)	515.0	983.5	0	12683
DEGCURVE	Degree of curve (degrees/100')	0.565	1.152	0	38
VCURLENTH	Vertical curve length (ft)	721.3	678.8	0	6700
PCTGRADE	Grade percent (%)	0.029	2.123	-5.130	5.550
MEDWIDTH	Median width (ft)	72.97	119.1	0	999
NUMLANES	Number of lanes	5.458	1.500	3	9
NDRYSURF	Indicator for non-dry surface (at time of crash)	0.423	0.416	0	1
SPDLIMIT	Speed limit (mph)	62.16	5.494	55	70
MOUNTAIN	Indicator for mountainous terrain	0.069	0.253	0	1
ROLLING	Indicator for rolling terrain	0.863	0.344	0	1
AADTPL	AADT per lane =AADT / NUMLANES	13076	8511	814	48251
VMT	Average daily vehicle miles traveled = segment length (miles) × Annual Avg. Daily Traffic (veh/day)	7234	8422	77	81869
YEAR94	Indicator for year 1994	0.238	0.426	0	1
YEAR95	Indicator for year 1995	0.249	0.432	0	1
YEAR96	Indicator for year 1996	0.280	0.449	0	1

TABLE 2. Pooled Count Data Regression Results for Six Crash Measures

Independent Variables	Dependent Variable & Selected Model					
	NUMDEAD	NUMINJUR	NUMFATCR	NUMINJCR	NUMPDOCR	NUMCRASH
	NegBin	NegBin	Poisson	NegBin	NegBin	NegBin
Constant	-13.07 (1.575)	-7.634 (7.24E-02)	-9.540 (0.316)	-8.719 (6.66E-02)	-8.168 (0.242)	-7.588 (0.184)
CURVLENGTH		-7.00E-05 (1.64E-05)		-8.48E-05 (1.51E-05)	-6.10E-05 (1.29E-05)	-6.69E-05 (1.08E-05)
DEGCURVE	0.186 (7.39E-02)	7.35E-02 (1.41E-02)	1.19E-01 (2.95E-02)	8.12E-02 (1.19E-02)	7.27E-02 (1.05E-02)	7.77E-02 (8.87E-03)
VCURLENGTH	-1.91E-04 (1.20E-04)	4.34E-05 (2.06E-05)	-1.24E-04 (1.00E-04)	3.33E-05 (1.85E-05)		2.65E-05 (1.34E-05)
PCTGRADE					6.83E-03 (4.58E-03)	
MEDWIDTH		-1.84E-04 (1.18E-04)		-2.61E-04 (1.11E-04)	-1.40E-04 (9.16E-05)	-1.48E-04 (7.58E-05)
NUMLANES	-0.270 (6.64E-02)	-7.25E-02 (9.60E-03)	-0.301 (5.96E-02)	-4.62E-02 (7.86E-03)	-4.54E-02 (7.53E-03)	-5.46E-02 (6.18E-03)
NDRYSURF	-0.609 (0.203)		-0.633 (0.179)	5.03E-02 (3.15E-02)	0.303 (2.77E-02)	0.188 (2.27E-02)
SPDLIMIT	5.76E-02 (2.13E-02)				-7.45E-03 (3.57E-03)	-3.73E-03 (2.75E-03)
MOUNTAIN	-1.220 (0.423)	0.389 (6.00E-02)	-0.842 (0.365)	0.607 (7.24E-02)	0.758 (6.28E-02)	0.647 (5.13E-02)
ROLLING	-0.643 (0.244)		-0.491 (0.191)	0.112 (5.36E-02)	0.199 (4.80E-02)	0.134 (3.86E-02)
RURAL		-0.152 (4.39E-02)			0.181 (4.06E-02)	6.44E-02 (3.26E-02)
AADTPL	-6.20E-05 (1.53E-05)	5.79E-06 (2.22E-06)	-8.70E-05 (1.18E-05)	1.57E-05 (1.53E-06)	1.02E-05 (1.84E-06)	1.01E-05 (1.49E-06)
YEAR94					-5.60E-02 (2.69E-02)	
YEAR95	-0.342 (0.194)	4.36E-02 (3.17E-02)		2.13E-02 (2.76E-02)	0.144 (2.50E-02)	
YEAR96		0.118 (3.05E-02)	0.276 (0.146)	0.108 (2.67E-02)		0.134 (2.02E-02)
lnVMT*	1.0	1.0	1.0	1.0	1.0	1.0
α **	14.94 2.870	1.253 (2.62E-02)	NA	0.501 (1.72E-02)	0.417 (1.31E-02)	0.348 (8.05E-03)
Num. Obs.	13,494	13,494	13,494	13,494	13,494	13,494
LL(Model)	-1,223	-25,042	-1,142	-19,858	-22,238	-27,978
LRI***	0.070	0.003	0.088	0.007	0.013	0.010

*: coefficient was constrained to 1.0

**: overdispersion parameter

***: McFadden's Likelihood Ratio Index (LRI) = $1 - LL(Model)/LL(Const\,Only)$

Notes: Standard errors are provided in parentheses. Variables with a p-value of 0.2 or higher remained in the model.

TABLE 3. Fixed-Effects (FE) Count Data Model Results for Six Crash Measures

Independent Variables	Dependent Variable & Selected Model					
	NUMDEAD	NUMINJUR	NUMFATCR	NUMINJCR	NUMPDOCR	NUMCRASH
	FENB	FENB	FEPois	FENB	FENB	FENB
Constant	-13.86 (1.286)	-8.586 (0.407)		1.332 (29.21)	6.417 (100.2)	11.82 (30.91)
CURVLENGTH		-7.04E-05 (5.07E-05)		5.09E-04 (2.85E-04)		
DEGCURVE	1.826 (0.792)	6.58E-02 (4.39E-02)				
VCURLENGTH				8.42E-04 (4.22E-04)	6.72E-04 (3.83E-04)	7.01E-04 (2.75E-04)
MEDWIDTH						-8.81E-03 (5.74E-03)
NUMLANES		-0.191 (2.77E-02)		-0.186 (0.121)		-0.159 (8.30E-02)
NDRYSURF	-0.513 (0.247)	-4.65E-02 (3.23E-02)	-0.567 (0.246)		0.157 (2.90E-02)	7.24E-02 (2.17E-02)
SPDLIMIT		2.12E-02 (4.44E-03)		4.01E-02 (5.16E-03)	9.78E-03 (3.91E-03)	1.25E-02 (2.88E-03)
MOUNTAIN		0.595 (0.278)				
ROLLING		0.453 (0.217)				
AADTPL				-2.88E-05 (1.00E-05)	-2.80E-05 (8.93E-06)	-2.85E-05 (6.75E-06)
YEAR94		4.02E-02 (2.62E-02)		4.73E-02 (2.47E-02)		
YEAR95				-0.106 (2.27E-02)		
YEAR96					0.152 (2.06E-02)	0.127 (1.53E-02)
lnVMT *	1.0	1.0	1.0	1.0	1.0	1.0
Num. Obs	565	10,412	565	10,373	11,012	11,509
Num. Grps	180	3,557	180	3,541	3,809	4,038
LL(Model)	-206	-10,664	-200	-7,701	-8,645	-10,733
LRI **	0.026	0.005	NA	0.005	0.009	0.009

*: coefficient was constrained to 1.0

**: McFadden's Likelihood Ratio Index (LRI) = $1 - LL(Model)/LL(ConstOnly)$

Notes: Standard errors are provided in parentheses. Variables with a p-value of 0.2 or higher remained in the model.

TABLE 4. Random-Effects (RE) Count Data Model Results for Six Crash Measures

Independent Variables	Dependent Variable & Selected Model					
	NUMDEAD	NUMINJUR	NUMFATCR	NUMINJCR	NUMPDOCR	NUMCRASH
	RENB	RENB	REPois	RENB	RENB	RENB
Constant	-9.095 (1.473)	-7.245 (9.15E-02)	-9.581 (0.321)	5.912 (64.12)	6.398 (57.88)	6.982 (41.83)
CURVLENGTH		-5.10E-05 (1.76E-05)		-6.37E-05 (1.78E-05)	-3.99E-05 (1.56E-05)	-4.42E-05 (1.36E-05)
DEGCURVE	0.123 (3.49E-02)	6.99E-02 (1.33E-02)	0.121 (3.29E-02)	6.98E-02 (1.35E-02)	6.54E-02 (1.23E-02)	6.82E-02 (1.10E-02)
VCURLENGTH	-1.33E-04 (1.05E-04)	3.76E-05 (2.31E-05)		4.45E-05 (2.30E-05)		3.34E-05 (1.76E-05)
MEDWIDTH				-1.81E-04 (1.35E-04)		-1.48E-04 (1.02E-04)
NUMLANES	-0.301 (6.26E-02)	-8.33E-02 (1.10E-02)	-0.304 (6.06E-02)	-5.47E-02 (1.05E-02)	-5.00E-02 (9.85E-03)	-6.14E-02 (8.52E-03)
NDRYSURF	-0.612 (0.180)		-0.641 (0.181)		0.217 (2.40E-02)	0.111 (1.86E-02)
SPDLIMIT	2.42E-02 (1.75E-02)				-8.21E-03 (4.89E-03)	-6.34E-03 (3.60E-03)
MOUNTAIN	-0.903 (0.386)	0.343 (6.56E-02)	-0.885 (0.373)	0.453 (6.71E-02)	0.670 (7.65E-02)	0.546 (6.66E-02)
ROLLING	-0.531 (0.211)		-0.532 (0.199)		0.193 (5.88E-02)	0.104 (4.97E-02)
RURAL		-0.136 (4.73E-02)		-0.186 (3.71E-02)	0.132 (4.99E-02)	
AADTPL	-7.57E-05 (1.40E-05)	-3.99E-06 (2.51E-06)	-8.49E-05 (1.17E-05)			-3.84E-06 (1.91E-06)
YEAR94		3.68E-02 (2.48E-02)			-3.12E-02 (2.15E-02)	
YEAR95		6.64E-02 (2.49E-02)		4.60E-02 (2.03E-02)	4.67E-02 (3.18E-02)	6.63E-02 (2.27E-02)
YEAR96		0.150 (2.42E-02)	0.274 (0.147)	0.138 (1.94E-02)	0.183 (3.13E-02)	0.184 (2.28E-02)
lnVMT *	1.0	1.0	1.0	1.0	1.0	1.0
r	9.395 (5.441)	4.539 (0.173)	NA	2.83E+06 (1.81E+08)	3.65E+06 (2.11E+08)	3.30E+06 (1.38E+08)
s	1.352 (1.642)	2.722 (0.119)	NA	2.131 (7.72E-02)	2.436 (8.23E-02)	2.723 (7.46E-02)
α **	NA	NA	0.467 (0.561)	NA	NA	NA
Num. Obs	13,494	13,494	13,494	13,494	13,494	13,494
Num. Grps	6,023	6,023	6,023	6,023	6,023	6,023
LL(Model)	-1,233	-24,678	-1,143	-19,098	-21,240	-26,110
LRI ***	0.076	0.004	0.015	0.004	0.010	0.009

*: coefficient was constrained to 1.0

**: overdispersion parameter

***: McFadden's Likelihood Ratio Index (LRI) = $1-LL(Model)/LL(Const\,Only)$

Notes: Standard errors are provided in parentheses. Variables with a p-value of 0.2 or higher remained in the model.

TABLE 5. Final Model Results for Six Crash Measures

Independent Variables	Dependent Variable & Selected Model					
	NUMDEAD	NUMINJUR	NUMFATCR	NUMINJCR	NUMPDOCR	NUMCRASH
	NegBin	FENB	Poisson	NegBin	NegBin	NegBin
Constant	-13.07 (1.575)	-8.586 (0.407)	-9.540 (0.316)	-8.719 (0.067)	-8.168 (0.242)	-7.588 (0.184)
CURVLENGTH		-7.04E-05 (5.07E-05)		-8.48E-05 (1.51E-05)	-6.10E-05 (1.29E-05)	-6.69E-05 (1.08E-05)
DEGCURVE	0.186 (0.074)	0.066 (0.044)	0.119 (0.030)	0.081 (0.012)	0.073 (0.010)	0.078 (8.87E-03)
VCURLENGTH	-1.91E-04 (1.20E-04)		-1.24E-04 (1.00E-04)	3.33E-05 (1.85E-05)		2.65E-05 (1.34E-05)
PCTGRADE					6.83E-03 (4.58E-03)	
MEDWIDTH				-2.61E-04 (1.11E-04)	-1.40E-04 (9.16E-05)	-1.48E-04 (7.58E-05)
NUMLANES	-0.270 (0.066)	-0.191 (0.028)	-0.301 (0.060)	-0.046 (7.86E-03)	-0.045 (7.53E-03)	-0.055 (6.18E-03)
NDRYSURF	-0.609 (0.203)	-0.047 (0.032)	-0.633 (0.179)	0.050 (0.031)	0.303 (0.028)	0.188 (0.023)
SPDLIMIT	0.058 (0.021)	0.021 (0.004)			-7.45E-03 (3.57E-03)	-3.73E-03 (2.75E-03)
MOUNTAIN	-1.220 (0.423)	0.595 (0.278)	-0.842 (0.365)	0.607 (0.072)	0.758 (0.063)	0.647 (0.051)
ROLLING	-0.643 (0.244)	0.453 (0.217)	-0.491 (0.191)	0.112 (0.054)	0.199 (0.048)	0.134 (0.039)
RURAL					0.181 (0.041)	6.44E-02 (0.033)
AADTPL	-6.20E-05 (1.53E-05)		-8.70E-05 (1.18E-05)	1.57E-05 (1.53E-06)	1.02E-05 (1.84E-06)	1.01E-05 (1.49E-06)
YEAR94		0.040 (0.026)			-5.60E-02 (0.027)	
YEAR95	-0.342 (0.194)			0.021 (0.028)	0.144 (0.025)	
YEAR96			0.276 (0.146)	0.108 (0.027)		0.134 (0.020)
lnVMT*	1.000	1.000	1.000	1.000	1.000	1.000
α **	14.943 (2.870)	NA	NA	0.501 (0.017)	0.417 (0.013)	0.348 (0.008)
Num. Obs.	13,494	10,412	13,494	13,494	13,494	13,494
Num.Grps.	NA	3,557	NA	NA	NA	NA
LL(Model)	-1,223	-10,664	-1,142	-19,858	-22,238	-27,978
LRI***	0.070	0.005	0.088	0.007	0.013	0.010

*: coefficient was constrained to 1.0

**: overdispersion parameter

***: McFadden's Likelihood Ratio Index (LRI) = $1 - LL(Model)/LL(Const\,Only)$

Notes: Standard errors are provided in parentheses. Variables with a p-value of 0.2 or higher remained in the model.