Accident Analysis and Prevention 50 (2013) 289–297

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates

Jonathan Aguero-Valverde*

Programa de Investigación en Desarrollo Urbano Sostenible, Universidad de Costa Rica, Barrio Los Profesores, Calle B, No 11, Mercedes, San Pedro, San José 11503, Costa Rica

ARTICLE INFO

Article history:
Received 1 February 2012
Received in revised form 13 April 2012
Accepted 30 April 2012

Keywords:
Full Bayes
Zero inflated models
Random effects
Ranking of sites

ABSTRACT

In recent years, complex statistical modeling approaches have being proposed to handle the unobserved heterogeneity and the excess of zeros frequently found in crash data, including random effects and zero inflated models. This research compares random effects, zero inflated, and zero inflated random effects models using a full Bayes hierarchical approach. The models are compared not just in terms of goodness-of-fit measures but also in terms of precision of posterior crash frequency estimates since the precision of these estimates is vital for ranking of sites for engineering improvement. Fixed-over-time random effects models are also compared to independent-over-time random effects models.

For the crash dataset being analyzed, it was found that once the random effects are included in the zero inflated models, the probability of being in the zero state is drastically reduced, and the zero inflated models degenerate to their non zero inflated counterparts. Also by fixing the random effects over time the fit of the models and the precision of the crash frequency estimates are significantly increased.

It was found that the rankings of the fixed-over-time random effects models are very consistent among them. In addition, the results show that by fixing the random effects over time, the standard errors of the crash frequency estimates are significantly reduced for the majority of the segments on the top of the ranking.

1. Introduction

One of the most important tasks of highway safety practitioners is the identification of locations in need of engineering improvements to reduce the number of crashes. The literature on the subject of identifying those sites includes a series of papers by Hauer (1996) and Hauer et al. (2002, 2004). Most ranking techniques rely on the estimation of the crash frequency for each segment or a function of it, such as the difference between the expected crash frequency in the site and the expected crash frequency in similar sites (Hauer, 1996; Tarko and Kanodia, 2004; Aguero-Valverde and Jovanis, 2007, 2009).

For modeling the expected crash frequency as well as better understanding the factors that affect the risk of a crash, researchers have used several count models of varying complexity, from Poisson to zero state Markov switching models. Jovanis and Chang (1986) presented one of the early uses of the Poisson regression for modeling crash frequencies. Other early adopters of Poisson regression include Jones et al. (1991) and Miaou and Lum (1993). Later, researchers adopted Poisson gamma (also known as negative binomial) models to account for the overdispersion frequently found in crash data (Shankar et al., 1995; Persaud and Mucsi, 1995; Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000; Donnell and Mason, 2006).

In recent years, more complex statistical modeling approaches have being proposed to handle the unobserved heterogeneity and the excess of zeros frequently found in crash data. Among them, random effects and random parameter models have gained popularity in the field. Random effects models were proposed to account for correlation among observations in the data (Lord and Mannering, 2010). Some applications in highway safety can be found in Johansson (1996), Shankar et al. (1998), MacNab (2004), Miaou and Lord (2003), Aguero-Valverde and Jovanis (2006) and Quddus (2008). More recently, random parameters models were proposed to control for unexplained heterogeneity in the data. Some examples of this approach are Anastasopoulos and Mannering (2009) and El-Basyounny and Sayed (2009).

Zero-inflated (ZI) models have being proposed to account for the excess of zeros, compared to the number of zeros expected under a Poisson or Poisson-gamma model, exhibited by many crash datasets. One of the first reported works to incorporate zero inflated models in road safety was the paper by Shankar et al. (1997). Other examples of zero inflated models in highway safety are Carson and Mannering (2001), Shankar et al. (2004), and Qin et al. (2005). These

* Tel.: +506 22834927; fax: +506 22834815.
E-mail address: jaguero@produs.ucr.ac.cr

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2012.04.019

models operate under the assumption of two states existing for the data: the "zero" state and the normal count state. This fact has led some researchers to criticize the application of these models in highway safety (Lord et al., 2005, 2007), since the segments in the zero state have a long term mean equal to zero implying that those segments are totally safe.

To overcome the issue of the long term mean equal to zero Malyshkina and Mannering (2010) proposed a Markov switching model that allows individual segments to change states over time. This model was implemented in a full Bayes approach since the likelihood function does not have a closed form; therefore, maximum likelihood estimation was not feasible. Furthermore, full Bayes estimation allows for direct statistical estimation of the road segment state.

Using also a full Bayes approach, this work proposes models that do not allow segments change states over time but concentrates on including random effects in the specification. Random effects, zero inflated, and zero inflated models with random effects are estimated. The models are compared not just in terms of goodness-of-fit measures but also in terms of precision of posterior crash frequency estimates since the precision of these estimates is vital for ranking of sites for engineering improvement. Other objective is to compare models with and without random effects fixed over time.

Even though, previous studies have analyzed and compared zero-inflated and Poisson or Poisson-gamma models for crash frequency in terms of goodness-of-fit, the effect of zero inflated models in the precision of crash frequency estimates is unknown. Ideally, one would try to assess the accuracy rather than precision of those estimates, but the "real" value of crash frequency is unknown. As an alternative, one can use simulated data for comparing the models so that the "real" value of crash frequency is known but this would defeat the purpose of comparing the models; the statistical model whose distribution more closely resembles the probability distribution used to simulate the dataset will always have a better fit. To circumvent this issue one can use the precision of the estimates. Estimates whose standard deviation is lower (higher precision) are statistically better.

In the present formulation, the segment state for each year is independent of the other years; therefore, it allows changes in state between years. Furthermore, all the parameters in Bayesian statistics are regarded as random in nature; hence, the expectation for the number of crashes cannot be equal to zero unless the probability of being in the zero state concentrates all its mass at one.

The fit of the models and the precision of the frequency estimates are analyzed for models where the random effects are fixed over time. Analog to the spatial correlation term (Aguero-Valverde and Jovanis, 2006, 2008, 2010), by using a fixed-over-time random effects, the segments estimates 'pool strength' from neighboring years improving model estimation. This is especially true in circumstances with high random variability in the data, such as is the case of most crash data, particularly when a high number of zeros is present.

Finally, the ranking of sites based on the posterior crash frequency estimates is compared for all proposed models in order to explore the effect that different modeling approaches have in the rankings. This paper is organized as follows: the next section describes the statistical methodology used; then, the dataset is described, followed by the discussion of results, conclusions and recommendations for future research.

2. Methodology

Full Bayes hierarchical approach is used for all the models estimated in this work. Markov chain Monte Carlo, also known as Markov chain simulation is employed to draw samples from the target posterior distribution of the parameters (Carlin and Louis, 1996). Details about Markov chain simulation are beyond the scope of this paper, interested readers can refer to Carlin and Louis (1996) and Gelman et al. (2003).

The formulation of the models is presented from the traditional Poisson-gamma to the more complex zero inflated models with random effects. Five count distributions are tested: Poisson gamma or negative binomial, Poisson lognormal, Poisson zero inflated, Poisson zero inflated lognormal and Poisson zero inflated gamma.

2.1. Poisson gamma

Poisson gamma or negative binomial (NB) is the most popular count distribution used in highway safety for crash frequency models. The model is specified as follows: at the first stage the crash counts are modeled as a Poisson process:

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

where y_{it} are the observed number of crashes in segment i at time t (in years), and θ_{it} are the expected Poisson rate (i.e. the expected crash frequency) for segment i at time t.

The Poisson rate is modeled as a function of the covariates following the log-link shown in Eq. (2):

$$\log \quad (\theta_{it}) = \beta_0 + \sum_k \beta_k x_{itk} + \varepsilon_{it} \tag{2}$$

where β_0 is the intercept, β_k is the coefficient for covariate k, x_{itk} is the value for the kth covariate (or any suitable transformation of the covariate), for segment i, at time t and ε_{it} is the error term for segment i at time t.

At the second stage, the coefficients are modeled using non-informative Normal priors (i.e. $\beta_k \sim N(0, 1000)$) while the exponent of the error term or random effect is modeled as a gamma distribution:

$$e^{\varepsilon_{it}} \left| \phi \overset{iid}{\sim} \text{Gamma} \quad (\phi, \phi) \right. \tag{3}$$

where ϕ controls the amount of extra-Poisson variation due to heterogeneity among segments. The gamma distribution of the error term has a mean of one and a variance of $1/\phi$. The parameter ϕ is also known as the dispersion parameter. A gamma (0.01, 0.01) was used as hyper-prior for ϕ. This hyper-prior was selected because it introduces little prior information into the model, reduces convergency times, and improves model identifiability.

The error term was also modeled as a random effect fixed over time. As in the case of the prior model, the exponent of the error term has a gamma prior distribution:

$$e^{\varepsilon_i} \left| \phi \overset{iid}{\sim} \text{Gamma} \quad (\phi, \phi) \right. \tag{4}$$

2.2. Poisson lognormal

Poisson lognormal models have been used only recently in highway safety since the marginal distribution of this model does not have a closed form like the Poisson gamma model; therefore, they are typically implemented using the Bayesian approach. Some recent applications of these models in highway safety are: Miranda-Moreno et al. (2005), Aguero-Valverde and Jovanis (2006), Ma et al. (2007), and Park and Lord (2007). For this model the Poisson rate is modeled using a log-normal distribution:

$$\log \quad (\theta_{it}) = \beta_0 + \sum_k \beta_k x_{itk} + \nu_{it} \tag{5}$$

where the random effect v_{it} has a normal prior:

$$v_{it} \overset{iid}{\sim} N\left(0, \frac{1}{\tau_v}\right) \tag{6}$$

and τ_v controls the extra-Poisson variation and has a hyper-prior gamma (0.01, 0.01).

A fixed-over-time random effects model was also fitted to the data with a normal prior for the error term as shown in Eq. (7):

$$v_i \overset{iid}{\sim} N\left(0, \frac{1}{\tau_v}\right) \tag{7}$$

2.3. Zero inflated Poisson

Mixture models such as zero inflated Poisson (ZIP) and zero inflated Poisson-gamma (i.e. zero inflated negative binomial, ZINB) have being used in highway safety to account for overdispersion in the data. Let the number of crashes y_{it} be distributed according to a zero-inflated Poisson distribution that can be represented as:

$$y_{it} = V_{it}(1 - B_{it}) \tag{8}$$

where B_{it} is a Bernoulli (p_{it}) random variable and V_{it} independently of B_{it} has a Poisson (θ_{it}) distribution. Then, the probability mass function of the ZIP distribution, hereafter represented as ZIP (p_{it}, θ_{it}) is given by:

$$P(y_{it} = 0) = p_{it} + (1 - p_{it})e^{-\theta_{it}}$$

$$P(y_{it} = k) = (1 - p_{it})\frac{\theta_{it}^k e^{-\theta_{it}}}{k!} \quad k = 1, 2, \ldots \tag{9}$$

where p_{it} is the probability of the outcome belonging to a degenerated distribution at zero ($0 \le p_{it} \le 1$) and θ_{it} is the Poisson rate.

The expectation and the variance of the random variable are defined as:

$$E(y_{it}) = (1 - p_{it})\theta_{it} \tag{10}$$

$$\mathrm{VAR}(y_{it}) = (1 - p_{it})\theta_{it} + p_{it}(1 - p_{it})\theta_{it}^2 \tag{11}$$

Note that, as mentioned previously, in Bayesian inference the only way for the mean (or expected) number of crashes to be zero is that $p_{it} = 1$, or in other words, the whole mass of the p_{it} distribution is concentrated at one (the Poisson rate is by definition greater than one).

Now, assuming independently distributed responses y_{it}'s sampled from a ZIP (p_{it}, θ_{it}), the link functions are given by Eqs. (12) and (13):

$$\log(\theta_{it}) = \beta_0 + \sum_k \beta_k x_{itk} \tag{12}$$

$$\mathrm{logit}(p_{it}) = \log\left(\frac{p_{it}}{1 - p_{it}}\right) = \gamma_0 + \sum_k \gamma_k x_{itk} \tag{13}$$

where γ_k's are the coefficients for the zero state covariates x_{itk}.

2.4. Zero inflated Poisson gamma

The previous ZIP specification has no random effects; however, in order to compare this model with the Poisson-gamma specification, a random effect is included in the Poisson rate as given by Eq. (2). The random effects are defined by Eq. (3). A random effect is not included in the logit part of the ZIP-gamma since this will result in unstable model fitting (Agarwal et al., 2002). A fixed-over-time version of this model was also fitted to the data as in the case of the previous models as defined by Eq. (4).

Table 1
Summary statistics of the data by segment and year.

Variable	Mean	Std. dev.	Min.	Max.
Crashes	0.310	0.691	0	7
Volume (AADT)	2636.4	3197.7	45	18,749
Length (miles)	0.464	0.107	0.039	0.751
Indicators				
Functional Class Expressway and Arterial	0.295	0.456		
Functional Class Collector and Local	0.705	0.456		
Speed Limit ≤35 MPH	0.331	0.471	20	55
Speed Limit >35 MPH	0.669	0.471		
Lane width <10 ft.	0.616	0.486	6	23.5
Lane width >10 ft. and <12 ft.	0.098	0.298		
Lane width 12 ft.	0.017	0.128		
Lane width >12 ft. and <14 ft.	0.011	0.106		
Lane width ≥14 ft.	0.095	0.294		
Shoulder width <4 ft.	0.608	0.488	0	14
Shoulder width >4 ft. and <6 ft.	0.161	0.367		
Shoulder width 6 ft.	0.096	0.295		
Shoulder width >6 ft. and <10 ft.	0.099	0.299		
Shoulder width ≥10 ft.	0.036	0.187		

2.5. Zero inflated Poisson lognormal

A ZIP model with lognormal random effects for the Poisson rate was estimated. As in the case of previous models, two models, an independent-over-time and fixed-over-time random effects, were fitted to the dataset, as defined by Eqs. (6) and (7) respectively.

2.6. Model comparison

Two different goodness-of-fit measures are used for model comparison: posterior mean deviance and the Deviance Information Criterion (DIC). The posterior mean deviance ($\bar{D}$) can be taken as a Bayesian measure of fit or 'adequacy' (Spiegelhalter et al., 2002). Another criterion for model comparison is the Bayes factor (Gelman et al., 2003) which is defined as the ratio of the marginal likelihood of the models. As in the case of the posterior deviance, the Bayes factor does not consider model complexity.

To account for model complexity the Deviance Information Criterion was proposed. The DIC is considered the Bayesian equivalent of the Akaike Information Criterion (AIC). DIC is defined as an estimate of fit plus twice the effective number of parameters as in Eq. (14):

$$\mathrm{DIC} = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \tag{14}$$

where $D(\bar{\theta})$ is the deviance evaluated at $\bar{\theta}$, the posterior means of the parameters of interest, p_D is the effective number of parameters in the model, and $\bar{D}$ is the posterior mean of the deviance statistic $D(\theta)$. As with AIC, models with lower DIC values are preferred. The zero-inflated models are mixture models where the data can take the zero state or the count state. Each state has a different set of parameters; hence, the DIC is undefined since the deviance evaluated at the posterior means of the parameters chances according to the state the segment is in.

Other penalized goodness-of-fit measures such as the Bayesian Information Criterion (BIC) (Carlin and Louis, 1996) also rely on a measure of the number of parameters. The problem with the penalized goodness-of-fit measures in a hierarchical setting, such as the one used here, is that the number of parameters is unknown due to the multilevel nature of the models.

Table 2
Poisson gamma and Poisson lognormal models for center country data.

	Poisson gamma					Poisson gamma fixed-over-time				
	Mean	Std. dev.	MC error	2.50%	97.50%	Mean	Std. dev.	MC error	2.50%	97.50%
Intercept	−5.9090	0.46397	0.0062	−6.8275	−5.0186	−5.8047	0.51117	0.0054	−6.8183	−4.8074
Volume (AADT)	0.7222	0.05643	0.0009	0.6135	0.8329	0.7073	0.06229	0.0009	0.5871	0.8308
Functional Class Expressway and Arterial	−0.0337	0.12522	0.0027	−0.2764	0.2139	−0.0214	0.14484	0.0026	−0.3033	0.2614
Speed Limit >35 MPH	−0.2166	0.07901	0.0006	−0.3714	−0.0623	−0.2162	0.09159	0.0006	−0.3969	−0.0376
Lane width <10 ft.	−0.4850	0.17232	0.0020	−0.8259	−0.1492	−0.5059	0.18893	0.0019	−0.8822	−0.1405
Lane width >10 ft. and <12 ft.	−0.0232	0.09359	0.0015	−0.2021	0.1639	−0.0394	0.11439	0.0017	−0.2647	0.1850
Lane width >12 ft. and <14 ft.	−0.1212	0.23035	0.0027	−0.5856	0.3181	−0.1673	0.26494	0.0028	−0.6922	0.3453
Lane width ≥14 ft.	0.3730	0.22253	0.0020	−0.0663	0.8084	0.4363	0.27783	0.0023	−0.1060	0.9867
Shoulder width <4 ft.	0.2151	0.12274	0.0015	−0.0242	0.4578	0.2333	0.14451	0.0017	−0.0499	0.5183
Shoulder width >4 ft. and <6 ft.	0.0890	0.11952	0.0013	−0.1466	0.3252	0.1232	0.14190	0.0015	−0.1541	0.4005
Shoulder width >6 in. and <10 in.	0.0411	0.11660	0.0013	−0.1872	0.2691	0.0699	0.13823	0.0013	−0.2009	0.3407
Shoulder width ≥10 ft.	0.0824	0.16320	0.0016	−0.2400	0.3975	0.0513	0.19727	0.0016	−0.3381	0.4374
Theta	2.7023	0.62458	0.0316	1.7883	4.2569	3.1401	0.61244	0.0125	2.1840	4.5766
Mean zero probability	0.7779	0.00591	0.0001	0.7662	0.7893	0.7756	0.00567	0.0000	0.7644	0.7866
Deviance	3925.8	65.0	2.8	3798.2	4053.5	3994.3	36.6	0.5	3924.0	4067.4
DIC	4202.33					4179.47				

	Poisson lognormal					Poisson lognormal fixed-over-time				
	Mean	Std. dev.	MC error	2.50%	97.50%	Mean	Std. dev.	MC error	2.50%	97.50%
Intercept	−6.1090	0.46925	0.0048	−7.0294	−5.1960	−6.0029	0.51251	0.0053	−7.0125	−5.0005
Volume (AADT)	0.7238	0.05663	0.0007	0.6133	0.8348	0.7141	0.06232	0.0008	0.5915	0.8363
Functional Class Expressway and Arterial	−0.0316	0.12714	0.0019	−0.2802	0.2195	−0.0343	0.14584	0.0026	−0.3177	0.2524
Speed Limit >35 MPH	−0.2211	0.07947	0.0004	−0.3773	−0.0648	−0.2250	0.09330	0.0006	−0.4097	0.0423
Lane width <10 ft.	−0.4860	0.17510	0.0015	−0.8327	−0.1469	−0.5190	0.19106	0.0020	−0.8941	−0.1440
Lane width >10 ft. and <12 ft.	−0.0222	0.09666	0.0011	−0.2123	0.1668	−0.0419	0.11544	0.0017	−0.2661	0.1869
Lane width >12 ft. and <14 ft.	−0.1271	0.23478	0.0019	−0.5999	0.3242	−0.1567	0.26661	0.0027	−0.6903	0.3583
Lane width ≥14 ft.	0.3696	0.22238	0.0014	−0.0722	0.7975	0.4205	0.27365	0.0020	−0.1210	0.9542
Shoulder width <4 ft.	0.2218	0.12418	0.0011	−0.0194	0.4666	0.2389	0.14640	0.0016	−0.0466	0.5269
Shoulder width >4 ft. and <6 ft.	0.0923	0.12200	0.0010	−0.1467	0.3340	0.1106	0.14443	0.0020	−0.1747	0.3947
Shoulder width >6 ft. and <10 ft.	0.0451	0.11917	0.0009	−0.1869	0.2816	0.0823	0.14180	0.0013	−0.1935	0.3625
Shoulder width ≥10 ft.	0.0916	0.16752	0.0011	−0.2399	0.4158	0.0822	0.20045	0.0016	−0.3144	0.4733
Sigma square	0.3695	0.06839	0.0021	0.2434	0.5112	0.3221	0.05625	0.0008	0.2223	0.4432
Mean zero probability	0.7783	0.00585	0.0001	0.7667	0.7897	0.7758	0.00565	0.000	0.7646	0.7868
Deviance	3932.0	57.1	1.5	3819.4	4044.1	4003.0	33.5	0.4	3937.4	4068.8
DIC	4238.41					4202.48				

Gray cells indicate significance at 97.5% level.

2.7. Mean zero probability

One of the reasons frequently cited to use zero-inflated models is the fact that regular Poisson and Poisson-gamma models often predict less zeros than those observed in crash datasets, even after accounting for overdispersion, as in the case of Poisson-gamma models. In these instances it is useful to compare the models in terms of the expected number of zeros given the coefficient estimates. The conventional way to accomplish this task is to estimate the probability of observing zero crashes by evaluating the Poisson rate at the mean of the covariates. This has obvious shortcomings, like the fact that mean of the covariates might not have a physical meaning (i.e. indicator variables). This problem can be easily overcome using full Bayes hierarchical approach through Markov Chain Monte Carlo simulation. In this framework, any quantity of interest can be sampled from its posterior distribution. Furthermore, the random effects can be specifically incorporated in the estimation. Eq. (15) shows the mean zero probability for the Poisson models:

$$\text{Mean} \quad (P(Y_{it}=0)) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} e^{-\theta_{it}}}{N \times T} \tag{15}$$

where N is the total number of segments and T is the total number of years of data.

For the zero-inflated models the mean zero probability is defined by:

$$\text{Mean} \quad (P(Y_{it}=0)) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} p_{it} + (1-p_{it})e^{-\theta_{it}}}{N \times T} \tag{16}$$

3. Data description

This dataset was introduced previously (Aguero-Valverde and Jovanis, 2008). The data for the models correspond to the state-maintained rural two-lane network of Centre County, located in Central Pennsylvania and part of the District 2-0 of the Pennsylvania Department of Transportation. The dataset is defined by segment and year, from 2003 to 2006. A total of 865 rural two-lane segments were included in the analysis. A relational database was assembled with information from the crash databases and road inventory.

3.1. Crash data

Crash data were obtained from the PennDOT Crash Reporting System. The data includes reportable crashes for road segment locations only (i.e. those that do not occur at an intersection or ramp junction). A total of 77.98% of observations by segment and year were zeros.

Table 3
Zero inflated models for center country data.

	ZIP gamma					ZIP gamma pooled-over-time random effects					ZIP				
	Mean	Std. dev.	MC error	2.50%	97.50%	Mean	Std. dev.	MC error	2.50%	97.50%	Mean	Std. dev	MC error	2.50%	97.50%
Intercept	−6.0217	0.4921	0.0318	−7.1507	−5.1361	−5.7516	0.4993	0.0268	−6.6898	−4.8287	−4.9116	0.6315	0.0414	−6.0933	−3.5465
Volume (AADT)	0.7350	0.0598	0.0039	0.6289	0.8708	0.7012	0.0594	0.0032	0.5932	0.8181	0.6368	0.0727	0.0048	0.4867	0.7720
Functional Class Expressway and Arterial	−0.0481	0.1270	0.0058	−0.3012	0.1937	0.0117	0.1406	0.0051	−0.2813	0.2657	−0.0273	0.1273	0.0058	−0.2836	0.2208
Speed Limit >35 MPH	−0.2154	0.0789	0.0016	−0.3687	−0.0586	−0.2213	0.0923	0.0018	−0.4068	−0.0429	−0.2176	0.0766	0.0018	−0.36875	−0.0691
Lane width <10 ft.	−0.4611	0.1749	0.0060	−0.8127	−0.122	−0.5131	0.1864	0.0051	−0.8787	−0.1486	−0.4883	0.1744	0.0055	−0.8301	−0.1483
Lane width >10 ft. and <12 ft.	−0.0098	0.0953	0.0034	−0.1967	0.1795	−0.0392	0.1138	0.0033	−0.2639	0.1807	−0.0294	0.0919	0.0031	−0.2062	0.1490
Lane width >12 ft. and <14 ft.	−0.1178	0.2335	0.0048	−0.5853	0.3316	0.15480	0.2624	0.0045	−0.6824	0.3488	−0.1370	0.2193	0.0039	−0.5792	0.2787
Lane width ≥14 ft.	0.3824	0.2285	0.0044	−0.0704	0.8298	0.4353	0.2768	0.0043	−0.1094	0.9810	0.3045	0.2035	0.0034	−0.1034	0.6867
Shoulder width <4 ft.	0.2220	0.1223	0.0044	−0.0223	0.4595	0.2242	0.1463	0.0050	−0.0679	0.5111	0.1708	0.1236	0.0047	−0.0718	0.4096
Shoulder width >4 ft. and <6 ft.	0.0949	0.1219	0.0036	−0.1429	0.3258	0.1141	0.1435	0.0041	−0.1695	0.3994	0.0465	0.1182	0.0040	−0.1806	0.2756
Shoulder width >6 ft. and <10 ft.	0.0461	0.1141	0.0028	−0.1756	0.2701	0.0648	0.1399	0.0032	−0.2097	0.3424	0.0354	0.1123	0.0030	−0.1854	0.2578
Shoulder width ≥10 ft.	0.0892	0.1656	0.0035	−0.2349	0.4059	0.0453	0.1968	0.0037	−0.3421	0.4306	0.0520	0.1563	0.0037	−0.2582	0.3515
Intercept (zero state)	−2.699	32.420	0.764	−65.351	68.174	−4.251	30.357	0.375	−64.488	54.463	0.532	1.274	0.082	−2.180	2.641
Volume (zero state)	−27.742	19.109	0.177	−72.299	0.541	−26.766	18.922	0.227	−71.916	0.364	0.262	0.135	0.009	−0.502	0.022
Length (zero state)	4.364	30.402	0.671	−52.429	64.403	1.679	31.444	0.797	−58.051	63.141	0.488	0.497	0.025	−1.440	0.528
Theta	2.7192	0.6824	0.0358	1.7631	4.4925	3.1566	0.6308	0.0124	2.1788	4.6441					
Mean probability, *p*	0.0003	0.0018	0.0001	0.00000	0.0055	0.0007	0.0025	0.0001	0.0000	0.0104	0.2953	0.0655	0.0036	0.1592	0.4126
Mean zero probability	0.7781	0.0061	0.0002	0.7658	0.7898	0.7755	0.0057	0.0001	0.7642	0.7865	0.7774	0.0066	0.0001	0.7644	0.7902
Deviance	3926.7	67.2	3.0	3795.7	4062.9	3994.6	36.9	0.5	3923.6	4068.4	3746.9	118.8	5.6	3523.5	3995.6

	ZIP log-normal					IP log-normal pooled-over-time random effect				
	Mean	Std. dev.	MC error	2.50%	97.50%	Mean	Std. dev.	MC error	2.50%	97.50%
Intercept	−5.9255	0.4491	0.0347	−6.7207	−5.0355	−6.1336	0.4617	0.0245	−7.0697	−5.2912
Volume (AADT)	0.7035	0.0543	0.0042	0.5992	0.7972	0.7302	0.0559	0.0030	0.6269	0.8439
Functional Class Expressway and Arterial	−0.0079	0.1250	0.0074	−0.2480	0.2411	−0.0536	0.1398	0.0051	−0.3261	0.2170
Speed Limit >35 MPH	−0.2235	0.0795	0.0025	0.3790	−0.0707	−0.2215	0.0932	0.0017	−0.4029	−0.0381
Lane width <10 ft.	−0.5004	0.1734	0.0072	−0.8376	−0.1659	−0.4910	0.1838	0.0049	−0.8519	−0.1286
Lane width >10 ft. and <12 ft.	−0.0240	0.0962	0.0044	−0.2088	0.1696	−0.0261	0.1127	0.0033	−0.2459	0.2021
Lane width >12 ft. and <14 ft.	−0.1036	0.2322	0.0067	−0.5721	0.3511	−0.1548	0.2618	0.0045	−0.6729	0.3549
Lane width ≥14 ft.	0.3612	0.2159	0.0055	−0.0623	0.7797	0.4280	0.2672	0.0045	−0.1058	0.9472
Shoulder width <4 ft.	0.2014	0.1204	0.0062	−0.0365	0.4383	0.2491	0.1450	0.0046	−0.0311	0.5346
Shoulder width >4 ft. and <6 ft.	0.0789	0.1162	0.0050	−0.1407	0.3150	0.1168	0.1426	0.0039	−0.1598	0.3990
Shoulder width >6 ft. and <10 ft.	0.0354	0.1134	0.0038	−0.1840	0.2563	0.0859	0.1377	0.0031	−0.1797	0.3575
Shoulder width ≥10 ft.	0.0882	0.1635	0.0045	−0.2399	0.4032	0.0886	0.1955	0.0037	−0.3006	0.4698
Intercept (zero state)	−3.376	30.764	0.563	−64.294	55.617	−4.951	29.816	0.394	−63.332	53.908
Volume (zero state)	−27.572	19.151	0.271	−72.530	0.803	−25.911	18.901	0.314	−70.642	0.205
Length (zero state)	5.464	30.842	0.948	−58.993	67.876	0.973	31.778	0.897	−61.698	63.472
Sigma square	0.3334	0.0730	0.0054	0.2038	0.4882	0.2980	0.0607	0.0020	0.1877	0.4254
Mean probability, *p*	0.0003	0.0015	000.01	0.0000	0.0034	0.0010	0.0027	0.0001	0.0000	0.0106
Mean zero probability	0.7766	0.0061	0.0002	0.7647	0.7886	0.7754	0.0057	0.0001	0.7640	0.7864
Deviance	3958.3	62.0	4.0	3834.8	4076.5	4013.4	36.2	0.9	3944.0	4086.4

Gray cells indicate significance at 97.5% level.

3.2. Road inventory

Road data were obtained from the Pennsylvania Road Management System (RMS) for the study period. RMS includes data for each road segment such as County Number, State Route Number, Segment Number, Segment Length, Average Daily Traffic, Lane Width, Travel Lane Count, Posted Speed Limit, Divisor Type, Functional Class, and Urban/Rural Code. These data were complemented with the State Roads Digital Map from Pennsylvania Spatial Data Access (Pennsylvania State University, 2007) to be able to "map" crash locations. Summary statistics of the inventory data for the area of study are shown in Table 1.

Table 4
Mean standard error of estimated crash frequency (Poisson rate) for the models.

Model	Standard error of Poisson rate
Poisson gamma	0.1764
Poisson gamma fixed-over-time	0.1378
Poisson lognormal	0.1823
Poisson lognormal fixed-over-time	0.1393
ZIP gamma	0.1766
ZIP gamma fixed-over-time	0.1377
ZIP lognormal	0.1747
ZIP lognormal fixed-over-time	0.1361
ZIP	0.1428

4. Results

Models were estimated using the open source software OpenBUGS (Thomas et al., 2006). For the models, 5000 iterations were discarded as burn-in. The following 50,000 iterations were used to obtain summary statistics of the posterior distribution of parameters. Convergence was assessed by visual inspection of the Markov chains for the parameters. Furthermore, the number of iterations was selected such that the Monte Carlo error for each parameter in the model would be less than 10% of the value of the standard deviation of that parameter.

Tables 2 and 3 present the estimates for the Poisson random effects and zero-inflated models. The main interest is in the differences and similarities between the different modeling approaches, so the discussion of model results is focused in this direction. Table 2 presents the results for the Poisson-gamma and Poisson log-normal models. From the results, it is evident the overdispersion on the data since both ϕ of the Poisson-gamma model and the variance of the Poisson log-normal model are significantly different from zero.

Fixed-over-time random effects also improve significantly the fit of the model as measured by the DIC. Evidently, by fixing the random effects over time, the degrees of freedom on the models are reduced which increases the posterior deviance ($\bar{D}$). On the other hand, the reduction on the effective number of parameters compensates the increase on the deviance which results in a lower DIC, favoring the more parsimonious fixed-over-time random effects models. Furthermore, the reductions on DIC are of approximately 23 and 36 for the Poisson gamma and Poisson lognormal models respectively.

Overall, the fixed-over-time Poisson gamma model has the lowest DIC of the models in Table 2, with a DIC of approximately 20 points less than the equivalent Poisson lognormal model. Also note that the mean zero probability for all the models is very close to the observed frequency of zeros of 77.98%.

Table 3 shows the results for the zero inflated models. The first result from the table is the fact that the ZIP model (i.e. without random effects) has by far the lowest deviance of all the models. This was expected since the model has more degrees of freedom and the data can be fitted to two probability distributions: the degenerated zero distribution and the Poisson distribution. Furthermore, the differences between observed and predicted counts are reduced since many segments that presented zero crashes fell into the 'zero state'; hence, the difference between observed and predicted is zero. On the other hand, crash count models have always positive crash rates which means that there is always a difference between observed and predicted counts; therefore, increasing the deviance and decreasing the goodness-of-fit of the model. In addition, consistent with the non zero inflated models, the ZIP gamma models have lower deviance than the ZIP lognormal models and the fixed-over-time models have higher deviance than the independent-over-time random effects models. In addition, the ZIP models have lower posterior deviances than their counterparts in Table 2.

Table 3 also presents the probability p of the segment being in the 'zero state'. All the ZIP models with random effects present mean probabilities p of less than 0.001 whereas the ZIP model shows a mean probability p of 0.29. Moreover, the maximum and minimum expected value of p are 0.12 and 1.87E−8 for the ZIP gamma model with fixed-over-time random effects and 0.56 and 0.16 for the ZIP model. This range of p is smaller than the range of 0.2 to 0.8 found by Malyshkina and Mannering (2010). For the dataset, this result shows that once the random effects are included, the probability of being in the 'zero state' decreases to virtually zero for most segments. Furthermore, only 64 observations out of the total of 3460 have expected p greater than 0.001 (approximately 1.8% of observations). Clearly, by including fixed-over-time random effects, the models account for the unobserved heterogeneity of specific segments by considering the shared effects among those segments, particularly shared effects of the same segment over time. In the absence of fixed-over-time random effects, the ZIP model assigns those presumed independent observations into the zero state. This is particularly important for segments that present a non-zero observation for at least 1 year: the fixed over time random effect will pull the crash estimate away from zero; hence, decreasing the probability of being in the zero state, as shown in the model results.

The mean standard error for the expected crash frequency or Poisson rate estimates is presented in Table 4. The expected crash frequency can be used to rank the sites for engineering improvements; hence, the importance of the precision of the estimates. As shown in the table, the standard error decreases around 30% when the random effects are fixed-over-time. In addition, all the fixed-over-time random effects models present lower mean standard errors than the ZIP model even though the difference is rather modest in comparison with the independent time random effects models.

Table 5 presents the rankings by expected crash frequency for the last year of data. The rankings are ordered according to the Poisson-gamma fixed-over-time random effects model that presents the lower DIC among the non zero-inflated models. The table shows that the rankings are very similar across models where the random effects are fixed over time. It is also evident from the table that the estimates of crash frequency for the top sites from the zero-inflated models are also very similar to their non zero-inflated counterparts. This also suggests that once the random effects are included in the models the effect of the 'zero state' is significantly reduced and the zero-inflated models degenerate to their non zero-inflated counterparts. In addition, the results show that by fixing the random effects over time, the standard errors of the crash frequency estimates are significantly reduced for the majority of the segments on the top of the ranking.

A very interesting result from the table is that although the mean standard errors for the fixed-over-time models are lower than that of the ZIP model, the standard errors for the top segments are lower

Table 5
Ranking of sites by crash frequency for Poisson and zero-inflated models.

Segment	Poisson gamma			Poisson gamma fixed-over-time			Poisson lognormal			Poisson lognormal fixed-over-time			ZIP gamma		
	Mean	s.e.	Rank	Mean	s.e.	Rank	Mean	s.e.	Rank	Mean	s.e.	Rank	Mean	s.e.	Rank
303	3.066	1.172	1	2.920	0.711	1	3.148	1.338	1	2.984	0.758	1	3.072	1.180	1
327	1.316	0.620	18	2.467	0.623	2	1.280	0.664	18	2.688	0.735	2	1.313	0.619	18
306	1.879	0.880	2	2.332	0.629	3	1.802	0.868	3	2.285	0.638	3	1.884	0.883	2
54	1.410	0.666	41	2.130	0.583	4	1.372	0.700	13	2.194	0.648	4	1.424	0.676	11
292	0.937	0.495	36	1.859	0.514	5	0.899	0.490	44	1.974	0.598	5	0.937	0.495	37
342	1.737	0.678	3	1.792	0.484	6	2.076	0.995	2	1.951	0.575	6	1.749	0.682	3
305	1.352	0.712	13	1.762	0.536	7	1.319	0.680	15	1.705	0.529	8	1.358	0.720	13
291	1.552	0.672	9	1.650	0.488	8	1.573	0.791	8	1.689	0.540	9	1.552	0.674	8
4	0.810	0.429	61	1.641	0.471	9	0.784	0.438	64	1.772	0.568	7	0.812	0.433	62
307	1.734	0.751	4	1.590	0.484	10	1.731	0.852	4	1.572	0.507	10	1.732	0.759	4
302	0.700	0.436	82	1.493	1.460	11	0.712	0.403	80	1.504	0.502	11	0.708	0.445	82
328	1.140	0.544	23	1.470	0.457	12	1.49	0.597	25	1.499	0.504	12	1.134	0.542	24
309	1.561	0.680	8	1.417	0.456	13	1.581	0.789	7	1.404	0.483	13	1.550	0.678	9
48	1.350	0.591	15	1.306	0.422	14	1.396	0.718	11	1.313	0.464	15	1.356	0.598	14
296	0.883	0.473	48	1.300	0.416	15	0.855	0.473	51	1.305	0.457	16	0.889	0.474	48
304	1.680	0.788	5	1.272	0.454	16	1.617	0.804	5	1.233	0.431	19	1.679	0.792	5
294	0.872	0.461	52	1.271	0.405	17	0.843	0.466	53	1.278	0.447	17	0.878	0.469	51
293	0.657	0.355	96	1.260	0.381	18	0.637	0.367	99	1.403	0.480	14	0.658	0.352	96
298	1.258	0.592	19	1.238	0.417	19	1.229	0.644	19	1.207	0.434	22	1.263	0.598	19
334	0.974	0.525	31	1.232	0.944	20	0.944	1.520	33	1.233	0.483	20	0.975	0.523	31
78	0.567	0.353	128	1.228	0.400	21	0.577	0.339	122	1.256	0.449	18	0.567	0.354	128
338	1.207	0.583	20	1.205	0.448	22	1.20	0.631	22	1.208	0.473	21	1.201	0.588	20
147	1.382	0.756	12	1.201	0.464	23	1.350	0.713	14	1.194	0.443	24	1.386	0.762	12
323	0.877	0.467	49	1.200	0.406	24	0.843	0.467	54	1.185	0.436	25	0.879	0.470	50
79	0.701	0.448	81	1.172	0.439	25	0.718	0.425	79	1.124	0.429	28	0.698	0.456	84
320	0.537	0.335	138	1.170	0.382	26	0.535	0.315	138	1.206	0.437	23	0.532	0.335	141
142	1.610	0.758	7	1.152	0.438	27	1.547	0.771	9	1.110	0.415	31	1.618	0.762	7
145	1.320	0.624	17	1.148	0.409	28	1.289	0.665	17	1.111	0.407	30	1.327	0.628	17
321	1.150	0.581	232	1.123	0.423	29	1.125	0.618	24	1.092	0.427	33	1.150	0.577	22
335	1.007	0.544	30	1.122	0.436	30	0.980	0.543	31	1.109	0.448	32	1.009	0.543	30

Segment	ZIP gamma fixed-over-time			ZIP lognormal			ZIP lognormal fixed-over-time			ZIP		
	Mean	s.e.	Rank	Mean	s.e.	Rank	Mean	s.e.	Rank	Mean	s.e.	Rank
303	2.917	0.713	1	0.713	1.302	1	2.943	0.745	1	2.060	0.313	1
327	2.459	0.619	2	1.248	0.623	18	2.651	0.721	2	1.274	0.133	16
306	2.329	0.629	3	1.812	0.851	3	2.278	0.640	3	2.056	0.224	2
54	2.137	0.582	4	1.357	0.676	11	2.167	0.632	4	1.405	0.140	9
292	1.866	0.514	5	0.892	0.478	39	1.944	0.594	5	1.136	0.112	24
342	1.791	0.480	6	1.977	0.919	2	1.901	0.570	6	0.969	0.106	43
305	1.759	0.536	7	1.317	0.633	15	1.703	0.520	8	1.801	0.196	4
291	1.656	0.494	8	1.547	0.738	8	1.661	0.536	9	1.232	0.139	17
4	1.643	0.470	9	0.792	0.427	61	1.741	0.559	7	0.963	0.100	45
307	1.588	0.488	10	1.681	0.794	4	1.560	0.498	10	1.388	0.167	10
302	1.497	0.464	11	0.727	0.403	76	1.470	0.494	12	0.686	0.590	89
328	1.465	0.457	12	1.091	0.49	25	1.472	0.494	11	1.044	0.109	31
309	1.415	0.460	13	1.549	0.775	7	1.385	0.484	13	1.283	0.164	14
48	1.313	0.420	14	1.328	0.670	14	1.308	0.455	15	1.000	0.112	40
296	1.303	0.415	15	0.844	0.441	52	1.297	0.451	16	1.041	0.104	32
304	1.269	0.450	17	1.596	0.761	5	1.245	0.427	18	1.720	0.187	5
294	1.273	0.412	16	0.831	0.443	57	1.267	0.440	17	1.031	0.102	34
293	1.264	0.379	18	0.629	0.346	102	1.378	0.472	14	0.716	0.071	81
298	1.240	0.420	19	1.209	0.613	19	1.195	0.410	23	1.214	0.120	18
334	1.229	0.456	20	0.936	0.504	33	1.220	0.464	20	1.148	0.172	21
78	1.228	0.395	21	0.580	0.334	120	1.244	0.437	19	0.543	0.445	150
338	1.207	0.447	22	1.157	0.611	22	1.210	0.466	21	1.095	0.164	29
147	1.205	0.466	24	1.346	0.695	12	1.197	0.435	22	1.876	0.378	3
323	1.206	0.414	23	0.836	0.42	55	1.170	0.425	25	1.008	0.098	38
79	1.172	0.436	25	0.720	0.402	77	1.113	0.421	29	0.622	0.593	111
320	1.164	0.381	26	0.550	0.300	138	1.173	0.420	24	0.548	0.416	147
142	1.159	0.445	27	1.524	0.728	9	1.125	0.414	27	1.615	0.175	7
145	1.154	0.410	28	1.277	0.644	17	1.115	0.402	28	1.286	0.128	12
321	1.123	0.425	29	1.091	0.582	26	1.082	0.417	35	1.034	0.215	33
335	1.122	0.438	31	0.986	0.43	31	1.107	0.435	31	1.205	0.181	19

than those from all the other models for almost all the top thirty segments. That might be explained by the fact that by 'dividing' the dataset in two distributions the overdispersion within the count distribution decreases significantly; hence, reducing the standard errors of the segments in the tails of the distribution including the top segments.

5. Conclusions and recommendations for future research

A series of full Bayes statistical models have been developed using crash roadway and traffic data for 4 years (2003–2006) from Centre County, Pennsylvania. The models range from simple Poisson gamma to more complex zero inflated with random effects approaches. Although zero inflated models present better fit to the data, as measured by the posterior deviance, once the random effects are included, the probability of being in the 'zero state' decreases to virtually zero for the vast majority of road segments.

By fixing the random effects over time the fit of the models is significantly increased. Furthermore, the reductions on DIC are of approximately 23 and 36 for the Poisson gamma and Poisson lognormal models respectively compared to their independent-over-time random effects counterparts. Fixed-over-time random effects models also produce significantly lower standard errors of the crash frequency estimates with reductions of about 30%.

The mean zero probability for all the models is very close to the observed frequency of zeros of 77.98%. There is no indication of underprediction of zeros for the non-zero inflated models; however, the observed frequency of zeros is relatively low compared to other crash datasets. A selection of a subset of crash types or severities with significantly higher percentage of zeros is a natural next step in order to compare the performance of random effects and zero inflated models under more extreme conditions.

For this dataset, the Poisson gamma models perform better than the Poisson lognormal models not only in terms of posterior deviance but also in terms of DIC for the non mixed models.

Full Bayes models offer the possibility of collecting summary statistics of any quantity of interest; hence, the posterior probability of being in the zero state for each segment and year (p) was estimated. The mean p for the ZIP model was 0.29 while the mean probability for the ZI random effects models was lower than 0.001 for all the cases. Furthermore, the maximum and minimum of p are 0.12 and 1.87E−8 for the ZIP gamma model with fixed-over-time random effects and 0.56 and 0.16 for the ZIP model.

It was found that the rankings of the fixed-over-time random effects models are very consistent among them, at least for the top sites. In addition, the results show that by fixing the random effects over time, the standard errors of the crash frequency estimates are significantly reduced for the majority of the segments on the top of the ranking. Nevertheless, the model that presented the lower standard errors among the top sites was the ZIP model without random effects. This might be explained by the fact that by 'dividing' the dataset in two distributions the overdispersion within the count distribution decreases significantly.

Future replications with other datasets could help corroborate the finding of significant reductions on the standard errors of the crash frequency estimates when fixing the random effects over time. More extensive tests with additional data sets will allow researchers to reach more general conclusions concerning precision of estimates.

In summary, for the crash dataset being analyzed, it was found that once the random effects are included in the zero inflated models, the probability of being in the zero state is drastically reduced, and the zero inflated models degenerate to their non-zero inflated counterparts. Clearly, by including fixed-over-time random effects, the models account for the unobserved heterogeneity of specific segments by considering the shared effects among those segments, particularly shared effects of the same segment over time. Also by fixing the random effects over time the fit of the models and the precision of the crash frequency estimates are significantly increased.

Acknowledgments

The author gratefully acknowledges the contribution of the Pennsylvania Department of Transportation which provided the data for this analysis. It is also important to acknowledge the contribution of anonymous reviewers that help improve the quality of this article.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32 (5), 633–642.
Agarwal, D.K., Gelfand, A.E., Citron-Pousty, S., 2002. Zero-inflated models with application to spatial count data. Environmental and Ecological Statistics 9, 341–355.
Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38 (3), 618–625.
Aguero-Valverde, J., Jovanis, P.P., 2007. Identifying road segments with high risk of weather-related crashes using full Bayesian hierarchical models. In: Presented at 86th Annual Meeting of the Transportation Research Board (on Conference CD).
Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transportation Research Record 2061, 55–63.
Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. Transportation Research Record 2136, 82–91.
Aguero-Valverde, J., Jovanis, P.P., 2010. Spatial correlation in multilevel crash frequency models: effects of different neighboring structures. Transportation Research Record 2165, 21–31.
Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameter count models. Accident Analysis and Prevention 41 (1), 153–159.
Carlin, B.P., Louis, T.A., 1996. Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall, London.
Carson, J., Mannering, F., 2001. The effect of ice warning signs on accident frequencies and severities. Accident Analysis and Prevention 33 (1), 99–109.
Donnell, E.T., Mason, J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accident Analysis and Prevention. 38 (3), 590–599.
El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accident Analysis and Prevention 41 (5), 1118–1123.
Gelman, A., Carlin, J., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis, 2nd ed. Chapman & Hall/CRC.
Hauer, E., Kononov, J., Allery, B., Griffith, M., 2002. Screening the road network for sites with promise. Transportation Research Record 1784, 27–32.
Hauer, E., 1996. Identification of sites with promise. Transportation Research Record 1542, 54–60.
Hauer, E., Allery, B., Kononov, J., Griffith, M., 2004. How best to rank sites with promise. Transportation Research Record 1897, 48–54.
Johansson, P., 1996. Speed limitation and motorway casualties: a time series count data regression approach. Accident Analysis and Prevention 28 (1), 73–87.
Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention 23 (3), 239–255.
Jovanis, P.P., Chang, H.L., 1986. Modeling the relationship of accidents to miles traveled. Transportation Research Record 1068, 42–51.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44, 291–305.
Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37 (1), 35–46.
Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero-inflated models in highway safety. Accident Analysis and Prevention 39 (1), 53–57.
Ma, J., Kockelman, K.M., Damien, P., 2007. Bayesian multivariate Poisson-lognormal regression for crash prediction on rural two-lane highways. In: Presented at 86th Annual Meeting of the Transportation Research Board.
MacNab, Y.C., 2004. Bayesian spatial and ecological models for small-area crash and injury analysis. Accident Analysis and Prevention 36 (6), 1019–1028.
Malyshkina, N., Mannering, F., 2010. Zero-state Markov switching count data models: an empirical assessment. Accident Analysis and Prevention 42 (1), 122–130.
Miaou, S.-P., Lord, D., 2003. Modeling traffic crash–flow relationships for intersections: dispersion parameter, functional form and Bayes versus empirical Bayes. Transportation Research Record 1840, 31–40.
Miaou, S.-P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention 25 (6), 689–709.
Miranda-Moreno, L.F., Fu, L., Saccomanno, F.F., Labbe, A., 2005. Alternative risk models for ranking locations for safety improvement. Transportation Research Record 1908, 1–8.
Park, E.S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019, 1–6.
Pennsylvania Spatial Data Access, 2007. Pennsylvania State University. http://www.pasda.psu.edu/ (accessed May 2007).
Persaud, B.N., Mucsi, K., 1995. Microscopic accident potential models for two-lane rural roads. Transportation Research Record 1485, 134–139.
Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering 122 (2), 105–113.
Qin, X., Ivan, J.N., Ravishanker, N., Liu, J., 2005. Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov chain Monte Carlo modeling. Journal of Transportation Engineering 131 (5), 345–351.
Quddus, M.A., 2008. Time series count data models: empirical application to traffic accidents. Accident Analysis and Prevention 40 (5), 1732–1741.

Shankar, V.N., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis and Prevention 27 (3), 371–389.

Shankar, V.N., Albin, R.B., Milton, J.C., Mannering, F.L., 1998. Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using random effects negative binomial model. Transportation Research Record 1635, 44–48.

Shankar, V.N., Chayanan, S., Sittikariya, S., Shyu, M.-B., Juvva, N.K., Milton, J.C., 2004. Marginal impacts of design, traffic, weather, and related interactions on roadside crashes. Transportation Research Record 1897, 156–163.

Shankar, V.N., Milton, J.C., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accident Analysis and Prevention 29 (6), 829–837.

Spiegelhalter, D., Best, N., Carlin, B.P., Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B 64 (4), 583–639.

Tarko, A.P., Kanodia, M., 2004. Effective and fair identification of hazardous locations. Transportation Research Record 1897, 64–70.

Thomas, A., O'Hara, B., Ligges, U., Sturtz, S., 2006. Making BUGS open. R News 6 (1), 12–17.