Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

The choice of statistical models in road safety countermeasure effectiveness studies in Iowa

Wen Li[a,*,1], Alicia Carriquiry[b,a,1], Michael Pawlovich[c], Thomas Welch[c]

[a] Department of Statistics, Iowa State University, United States
[b] Department of Statistics, Pontificia Universidad Católica de Chile, Chile
[c] Iowa Department of Transportation, United States

ARTICLE INFO

Article history:
Received 19 October 2007
Received in revised form 25 March 2008
Accepted 31 March 2008

Keywords:
Road diet
Hierarchical models
Deviance
Deviance information criterion
Markov chain Monte Carlo
Posterior distribution

ABSTRACT

With few exceptions, model selection in traffic safety studies does not receive as much attention as do the methods implemented to estimate the parameters in those models. In this manuscript, we focus on the modeling step in an intervention study and discuss issues associated with formulation, interpretation, comparison and selection of models for intervention studies. All of the statistical models we consider rely on an over-dispersed Poisson assumption for the crash densities, and are fitted by Bayesian methods. The crash data we use arose from a study by the Iowa Department of Transportation to evaluate the effectiveness of converting roads from four lanes to three lanes. Deviance and the deviance information criterion (DIC) are used for model selection. In the Iowa road diet study, a subset of best models (which fit the data better than others) was then also used to carry out posterior predictive checks to assess model fit.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

There has been considerable discussion recently on the benefits and limitations of Bayesian methods for the analysis of traffic safety data and in particular, of crash data arising from intervention studies (e.g., Al-Masaeid, 1990; Huang et al., 2002; Miaou and Lord, 2003; Pawlovich et al., 2006; Qin et al., 2004; Schlütler et al., 1997; Tunaru, 1999). The emphasis has been on the relative merits of traditional before/after studies and Bayesian approaches in various flavors including empirical (EB) and full (FB) Bayesian estimation (e.g., Hauer, 1997; Miaou and Lord, 2003; Lord and Miranda-Moreno, 2007; Persaud and Lyon, 2007).

In this manuscript, we argue that the form of the statistical model used to describe the crash data also deserves attention and that model assessment and comparison are important steps in any statistical analysis. In particular, we discuss the implications – on the marginal distribution of crashes – of different choices for the function that is used to model the Poisson mean. We view our work as complementary to the discussion in Miaou and Lord (2003) and in Lord et al. (2005) in which the issue of model formulation is approached from a first principles viewpoint. Here, we consider models that are plausible representations of crash data (that is, that can be justified from an engineering viewpoint) and investigate their statistical properties. In describing model formulation and model parameters we attempt to justify our choices by referring back to the actual application and the characteristics of the data we use for analysis. Thus, while our focus is on model selection from a statistical viewpoint, we formulate the collection of candidate models using first principles information. Lord and Miranda-Moreno (in press) have initiated the discussion by comparing the performance on the Poisson-Gamma and the Poisson-LogNormal models (two models we discuss here) via simulation and with a focus on the estimation of the dispersion parameter. Here, we focus on estimation of mean number of crashes and compare models using formal statistical procedures.

For illustration, we analyze data collected at sites matched manually by researchers in the course of an intervention study that was conducted by the Iowa Department of Transportation (IA-DOT). Some of the sites (treatment sites) received an intervention some time during the study period and their paired sites (controls) did not receive it. The typical goal in this type of study is to assess the effect of the intervention on safety.

Throughout, we implement Bayesian methods to estimate model parameters and to carry out model diagnostics and comparison. Given a statistical model, however, similar point estimates

* Corresponding author. Present address: 112 Snedecor Hall, Iowa State University, Ames, IA 50011, United States. Tel.: +1 515 294 3440; fax: +1 515 294 4040.
E-mail address: shirley@iastate.edu (W. Li).

[1] Li's and Carriquiry's work is partially funded through IA DOT Contract #07970. Carriquiry's work is also partially funded through NSF Grant #DMS 0502347.

0001-4575/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2008.03.015

of model parameters could have been obtained by proceeding within a classical likelihood-based inference framework. A potentially important difference between the classical and Bayesian approaches to estimation might arise when attaching measures of uncertainty (standard errors in the classical framework, posterior uncertainty or credible sets in the Bayesian framework) to point parameter estimates. It is in this aspect of the statistical analysis of crash data that the Bayesian approach shows an advantage over classical methods. The relative merits of what is known as the EB approach (e.g., Hauer et al., 2002) and the FB approach have been discussed (e.g., Miaou and Lord, 2003; Pawlovich et al., 2006; Persaud and Lyon, 2007), yet there appears to be some disagreement regarding what constitutes an EB or an FB analysis. We elaborate on this issue in the Discussion section of the manuscript. Throughout, we assume that the reader is familiar with the Bayesian estimation framework and with Markov chain Monte Carlo methods for approximating posterior distributions. Otherwise, a good reference for both is Gelman et al. (2004).

Absent from this manuscript is a discussion of traditional before/after methods to evaluate the effect of an intervention. Hauer (1997) provides an extensive discussion of traditional before/after studies in road safety. Persaud et al. (2001) also refer to observational before/after studies but in the framework of empirical Bayes methods. In the more recent literature, however, it has been argued (e.g., Srinivasan and Kockelman, 2002; Miaou and Song, 2005) that multivariate modeling approaches can more effectively isolate the marginal effect of an intervention on safety from the effect of potential confounders.

This manuscript is organized as follows. In Section 2 we briefly describe the data that were used for illustration in this manuscript and define terms to be used in the remainder of the paper. Because results from the analysis of this particular set of data have been published elsewhere (Pawlovich et al., 2006) the focus here is on the methodology rather than on the intervention study itself. A set of plausible statistical models for representing the crash data in the Iowa study are presented and discussed in Section 3. In Section 4 we discuss issues associated with model diagnostics and comparison and introduce the statistics that will be implemented in our particular application to evaluate various plausible statistical models and select the "best" from among them. In Section 5 we present results and implement various approaches for model selection. Additional results that arise from fitting the selected models to the Iowa crash data, together with posterior predictive diagnostic checks are presented and interpreted in Section 5 as well. Finally, we offer some additional discussion and conclusions in Section 6. The dataset used in our analysis as well as the WinBUGS and R code used to carry out the calculations can be requested from the corresponding author.

2. Study data

The data used in this study have been described in detail elsewhere (Pawlovich et al., 2006). Briefly, the dataset used for analysis includes 28 road segments in the State of Iowa, 14 of which were converted from four through lanes to three lanes (two through lanes and a center turning lane) sometime within the period 1982–2004. The other 14 sites in the study were selected to act as comparison or control sites; these sites did not receive the intervention and were considered to be similar enough (in terms of geometry, location, traffic volumes and other relevant characteristics) to the treatment sites to serve for comparison.

Sites were distributed across the State of Iowa and were located in population centers of varying size (approximately 1000–200,000 inhabitants according to the 2000 population census, although most locations had 15,000 or fewer inhabitants). The number of crashes per month was recorded at the sites over segments of different lengths (0.2–2.5 miles) between January 1982 and December 2004. Monthly traffic volumes at each of the sites were estimated by the IA-DOT using average daily traffic (ADT), and for the year 2000, ADT ranged between 2700 and 16,500 (approximately). The ADT at most sites was between 4000 and 12,000 vehicles. There is abundant crash information for the period preceding the intervention at all sites (see Table 2). On the other hand, crash information during the period following the intervention is rather limited at several of the study sites.

We use the term *crash frequency* to denote the number of crashes per mile observed at a site during a given period (monthly, annual, or averaged over a given number of years), and *crash density* to denote crashes per 1000 ADT during a specified period. *Crash rate* denotes crash frequency per 100 million ADT-miles during a specified period (or crashes per hundred million vehicle-miles traveled (HMVMT) in a given period). Table 1 shows, for each site, the average and the standard deviation (S.D.) of annual crash frequency during the years preceding the intervention (the "before period") and during the years following the intervention (the "after period"), as well as the percent reduction in crash density at the site. The average and S.D. of crash rates during the before and after periods and the percent reduction in crash rates are also shown in the table. For comparison sites, the before and after periods were defined as if an intervention had occurred at the same time as it was implemented in the corresponding paired treatment site. Table 2 displays summary statistics for the variables shown in Table 1.

With the exception of sites 5, 21 and 25, both crash frequency and crash rate appear to have decreased at all sites. A rough calculation that consists of averaging crash rates across all treatment sites and across all comparison sites during the years preceding the intervention and during the years following the intervention (along the lines of a standard before/after analysis) indicates a reduction in crash rate of approximately 56% (877 crashes/HMVMT before intervention and 388 crashes/HMVMT after intervention) and a reduction across all comparison sites of approximately 31% (710 crashes/HMVMT before versus 489 after). The number of crashes (and the crash rate) at a site is highly variable from year-to-year (and even more so from month-to-month). When the within-site variance in frequency or rate is high an average crash rate based on a small number of years of observation is not a reliable estimate of the site's long-run average (or expected) crash frequency (or crash rate). Further, a related challenge is that those sites at which the average crash rate over a few periods is highest will tend to show a much lower rate when an additional year of crash data is collected and vice versa. This is what is sometimes referred to as the regression to the mean problem (e.g., Hauer, 1997; Lord et al., 2008). We note finally that in an observational study such as this one, sampling bias is likely to have occurred. From Table 2 it appears that treatment sites exhibited slightly more crashes than control sites during the "before" period, but the difference is not statistically significant. If the intervention is found to be effective, the results should be interpreted with some caution. The S.D. of frequency and rate in Table 2 suggest that variances are, for most sites, higher than the means. Thus, appropriate models for these data include those which accommodate extra Poisson dispersion.

3. Statistical models to quantify the impact of an intervention

We considered several different statistical models to evaluate the effect of the intervention. In all cases, y_{it} denotes the observed number of crashes at site i during month t, and $y_{it} \sim \text{Poisson}(\theta_{it})$.

Table 1
Observed average and S.D. of crash frequency (left panel) and of crash rate (right panel) at each site during the years preceding and following intervention and observed percent reduction in both quantities

SID	Before		After		Reduction (mean) (%)	Before		After		Reduction (mean) (%)
	Mean frequency	S.D.	Mean frequency	S.D.		Mean rate	S.D.	Mean rate	S.D.	
Treatment sites										
1	45.6	8.2	20.2	6.6	55.8	2139.8	374.3	751.3	253.3	64.9
2	22.4	5.3	14.8	1.6	34	659.8	175.5	338.4	33.9	48.7
4	24.1	6.4	10.1	1.8	58.1	859.7	230.9	341.4	61.1	60.3
5	25	12.9	31.4	2.3	−25.5	750.7	380	769.8	58.1	−2.5
6	23.1	6.7	9.9	2.1	57.4	761.9	216.4	250.7	46.6	67.1
7	10.2	6.1	4.2	2.6	58.7	591.4	349.4	232.3	147.7	60.7
8	25.5	6.1	14.9	8.8	41.4	1144.5	277.9	669.2	425.7	41.5
9	56.3	7.5	31	12	45	1333.2	201.1	618.7	239.3	53.6
10	31.1	14.3	10.6	3.5	65.9	956.9	436.6	266.6	87.4	72.1
11	18.5	3.8	3.5	4.4	81	751	152.1	145.3	122.2	80.7
12	38.9	10.9	17	6.1	56.2	1326	362.8	476.5	174.8	64.1
13	17.9	3.7	17.8	5	0.7	557.2	117.6	383.6	134.4	31.1
14	4.6	3	2.4	1.4	49	199.2	129.5	70.5	42.4	64.6
15	6.5	2.6	3.3	1.9	49.8	250.4	106.3	111.9	73.5	55.3
Control sites										
16	29.1	8.4	19.6	5	32.7	1535.9	442.7	726.2	225.2	52.7
17	26	5.1	15.2	3.4	41.8	700.4	139.8	313.5	67.4	55.2
19	7.7	2.1	7.7	1.1	1.1	462.7	130.4	371.7	54.4	19.7
20	15.4	5.8	11.4	1.4	26.4	761.1	295.3	451.1	58	40.7
21	5.3	2.5	5	5.2	5	243.6	113.3	178.6	185.1	26.7
22	6.9	2.2	5	3	28.4	551.8	174.3	351.6	202.4	36.3
23	15.2	4.9	16.7	11	−10	770.1	243.6	673.9	455.1	12.5
24	34.9	5.3	35.1	6.7	−0.5	876.2	143.9	697.3	152.2	20.4
25	5.3	2.7	7	3.6	−30.8	157.3	79.8	184.5	99.8	−17.2
26	9.7	3.8	9.2	5.2	4.8	1055.7	402.8	783.4	344.8	25.8
27	20.7	4.6	17.9	6.5	13.6	846.5	207.8	667.9	225.4	21.1
28	44.5	7.5	37.7	3.8	15.3	1057.6	166.2	895	112.4	15.4
29	7.2	4.6	5.3	4.5	26.3	266	180.5	143.7	108	46
30	19.9	6.7	18.1	5.9	9.4	651.1	234.9	409	185.6	37.2

Sites 1–15 are treatment sites.

The Poisson mean θ_{it} can be written as the product $\lambda_{it}\nu_{it}$ where for site i during month t λ_{it} is a (rescaled) crash density, ν_{it} is a measure of traffic volume and for convenience we divide ν_{it} into 1000 (and later on rescale appropriately).

3.1. Probability model for the Poisson mean

Typically, some function of θ_{it} is expressed as a combination of fixed and random effects possibly associated with crash density, crash frequency or crash rate. Different probability models for θ_{it} (or for λ_{it}) result in different marginal expectation and variance for number of crashes y_{it} and thus deserve some attention.

Consider the following probability model for θ_{it}:

$$\theta_{it} \sim \mathrm{Ga}(\rho, \delta_{it}), \qquad \delta_{it} = \frac{\rho}{\mu_{it}}, \tag{1}$$

so that $E(\theta_{it}) = \rho/\delta_{it} = \mu_{it}$ and $\mathrm{Var}(\theta_{it}) = \rho/\delta_{it}^2 = \rho^{-1}\mu_{it}^2$. Then, the marginal mean and variance of number of crashes are given by

$$E(y_{it}) = E[E(y_{it}|\theta_{it})] = \mu_{it},$$

Table 2
Summary statistics for crash frequency and crash rate at study sites

Sites	Statistics	Average annual crash frequency			Average annual crash rate			Number of months	
		Before	After	Reduction (%)	Before	After	Reduction (%)	Before	After
All	Minimum	4.6	2.4	−30.8	157.3	70.5	−17.2	–	–
	Median	20.3	13.1	30.55	756.1	377.7	43.8	–	–
	Maximum	56.3	37.7	81	2139.8	895.0	80.7	–	–
	Mean	21.3	14.4	28.3	793.5	438.3	41.2	–	–
	S.D.	13.6	9.8	28.5	438.5	242.3	23.6	–	–
Treatment	Minimum	4.6	2.4	−25.5	199.2	70.5	−2.5	137	19
	Median	23.6	12.7	52.8	756.5	339.9	60.5	215	60
	Maximum	56.3	31.4	81	2139.8	769.8	80.7	256	138
	Mean	25	13.7	44.8	877.3	387.6	54.4	214.6	60.4
	S.D.	14.5	9.4	27.2	497.0	234.7	20.6	28.4	28.4
Control	Minimum	5.3	5	−30.8	157.3	143.7	−17.2	–	–
	Median	15.3	13.3	11.5	730.8	430.1	26.2	–	–
	Maximum	44.5	37.7	41.8	1535.9	895.0	55.2	–	–
	Mean	17.7	15.1	11.7	709.7	489.1	28.0	–	–
	S.D.	12.2	10.4	19	370.5	247.6	18.8	–	–

Note: A negative value in columns 5 and 8 refers to an increase of crashes in the after period.

$$V(y_{it}) = E(V(y_{it}|\theta_{it})] + V[E(y_{it}|\theta_{it})] = \mu_{it} + \frac{1}{\rho}\mu_{it}^2 = \mu_{it}\left[1 + \frac{\mu_{it}}{\rho}\right].$$

In this case it is straightforward to show (by integrating θ out of the joint distribution $p(y|\theta)p(\theta)$) that the marginal probability distribution for the number of crashes y has the negative binomial form (e.g., Agresti, 2002, p. 560). Note that even though we have parameterized the Poisson mean in (1) in a slightly different form than in, e.g., Lord and Miranda-Moreno (2007), the expressions for the marginal mean and variance of y_{it} are the same as in their work.

Suppose now that the probability model for θ_{it} is

$$\log(\theta_{it}) \sim N(\phi_{it}, \tau^2), \tag{2}$$

so that $E(\theta_{it}) = \exp\{\phi_{it} + (1/2)\tau^2\}$ and $\mathrm{Var}(\theta_{it}) = \exp\{2\phi_{it} + \tau^2\}[\exp(\tau^2) - 1]$. If we let $\mu_{it} = \exp\{\phi_{it} + (1/2)\tau^2\}$ then the marginal mean and variance of y_{it} can be written as

$$E(y_{it}) = E[E(y_{it}|\theta_{it})] = \mu_{it},$$

$$V(y_{it}) = E(V(y_{it}|\theta_{it})] + V[E(y_{it}|\theta_{it})] = \mu_{it} + (\exp(\tau^2) - 1)\mu_{it}^2.$$

In both cases the marginal variance of y_{it} is larger than the mean and therefore both probability models accommodate extra-Poisson dispersion in number of crashes. In both models for θ, the variance of y is a linear function of the mean number of crashes and of the square of the mean number of crashes, but dependence on the squared mean differs between the two models. In the case of the log-normal model for the Poisson mean, the variance of number of crashes grows very fast (relative to the mean) if the variance τ^2 of $\log(\theta_{it})$ is large. On the other hand, for very small ρ the Gamma probability model can put significant mass on very small values of θ_{it}, which might be appropriate when studying locations with low crash frequencies. Lord and Miranda-Moreno (2007) have argued that the Poisson-LogNormal model formulation is a better alternative in safety studies where sample sizes are small and mean number of crashes is low. However, their work focuses only on the estimation of the dispersion parameters in the models. Here, our aim is to obtain good predictors of crash frequencies (and rates). The choice of one model over the other is not always clear, however and it then becomes important to refer back to characteristics of the data themselves and to carry out model assessment and diagnostics.

While over-dispersion is often present in crash data, under-dispersion might be present in some cases. Recently, Lord et al. (2008) proposed application of the Conway–Maxwell-Poisson (COM-Poisson) model (Conway and Maxwell, 1961; Shmueli et al., 2005; Kadane et al., 2006), which accommodates both over- and under-dispersion in crash data through a single parameter estimated from the sample. Within a Bayesian framework, Kadane et al. (2006) derived the conjugate family of prior distributions for the parameters of the COM-Poisson. Miaou and Song (2005) have suggested that careful modeling of the random effects in the model (e.g., via spatial or temporal effects) may reduce the unexplained over dispersion typically observed in crash data. If so, then the COM-Poisson model may become an attractive alternative.

3.2. Mixed effects versus hierarchical models

Whether the probability model for the Poisson mean is Gamma (as in (1)) or log-normal (as in (2)), it is often of interest to associate covariates x with the appropriate function of that mean. (We use the term "covariate" loosely, to refer to all variables, fixed or random, on the right-hand side of a regression model.) For a vector of p unknown parameters β, we often let $f(\theta_{it}) = x'_{it}\beta$, where $f(.)$ is some function of the Poisson mean θ. When repeated observations at each site have been obtained (e.g., over different time periods) it is possible to fit a different model to each site by for example including a site-level random effect among the covariates, so that

$$f(\theta_{it}) = x'_{it}\beta + \alpha_i, \qquad \alpha_i \sim N(0, \sigma_\alpha^2). \tag{3}$$

The mixed linear regression model in (3) establishes that the effect of the covariates on crashes at each site is the same, but that the intercept is different across sites. The site-level effect α induces a correlation among observations obtained at the same site.

An alternative to model (3) is one in which all parameters (not just the intercept) are allowed to vary from site to site, so that we fit a different curve to the appropriate function of the Poisson mean at each site. This formulation of the model permits both site and population-level inferences but requires estimation of a very large number of parameters and the inclusion of probability models for the population distributions of those parameters. These multi-level models are often known as hierarchical models and in the general case can be written as

$$f(\theta_{it}) = x'_{it}\beta_i, \qquad \beta_i \sim (\mu_\beta, \Sigma_\beta), \tag{4}$$

where μ_β is the $p \times 1$ mean vector and Σ_β is the $p \times p$ covariance matrix of the population distribution that generates the site-level regression coefficient vectors β_i. The population parameters μ_β, Σ_β are assigned their own hyper-prior distribution. Typically, the population distribution in model (4) is the p-variate normal model. If an intercept is not already included among the covariates x_{it}, we can add the site-level intercept α_i to model (4).

An intermediate approach is to proceed as in the general area known as *small area estimation* (e.g., Battese et al., 1988; Ghosh and Rao, 1994). In this case, sites are grouped into small *domains* or groups based on site-level attributes such as traffic volume, geometry, location or other. Suppose that γ_j is used to denote the jth small group of sites. Then, we can extend model (3) as follows:

$$f(\theta_{itj}) = x'_{it}\beta + \gamma_j + \alpha_{ij}, \quad \gamma_j \sim N(0, \sigma_\gamma^2), \;\; \alpha_{ij} \sim N(0, \sigma_\alpha^2), \tag{5}$$

where now θ_{itj} is the Poisson mean for site i in domain j during time t and $\sigma_\gamma^2, \sigma_\alpha^2$ represent variability between and within groups, respectively. Note that model (5) fits a different mean curve at the group level, and within group, fits curves with varying intercepts at the site level. Model (4) can be viewed as an extended small area model by noticing that in the hierarchical formulation, the domains are defined by the combinations of different levels of the p covariates.

3.3. Modeling the effect of covariates on the Poisson mean

In this work, we investigate the association between crash density λ and covariates x. We focus on piecewise linear functions of the covariates to accommodate a possible change in the slope of crash density on time at treatment sites that might be attributable to the intervention. We use t_{0i} to denote the month during which the intervention is completed for treated site i and (fictitiously) for the matched comparison site. Further, we let x_{1i} be an indicator of whether the site was an intervention site ($x_{1i} = 1$) or a comparison site ($x_{1i} = 0$) and define

$$\begin{aligned} S_t &= 1 \quad \text{if } t \text{ is a winter month (December, January, and February)},\\ &= 2 \quad \text{if } t \text{ is a spring month (March, April, and May)},\\ &= 3 \quad \text{if } t \text{ is a summer month (June, July and August)},\\ &= 4 \quad \text{if } t \text{ is a fall month (September, October and November)}, \end{aligned}$$

so that

$$x_{2t} = \cos\left(\frac{2\pi \times S_t}{4}\right), \qquad x_{3t} = \cos\left(\frac{4\pi \times S_t}{4}\right), \qquad x_{4t} = \sin\left(\frac{2\pi \times S_t}{4}\right)$$

are smoothly evolving and periodic variables used to represent seasonal effects on crashes. Other covariates in the model for crash density include time, a change point to accommodate a possible change in the slope of crash density on time, a sudden drop in crash density at the time of the intervention, a site-level random effect, a yoked pair-level random effect and various two-way interactions (see below). In addition to choosing a probability model for the Poisson mean, a main objective in this work is to select the subset of the covariates to include in that probability model.

We first consider the log-normal model (2) for crash density and then write the log crash density as a piecewise linear function of covariates, such that the function is continuous at the change point t_{0i}. We include a site-level random effect as in (3). The model is

$$\log(\lambda_{it}) = \beta_1 + \beta_2 x_{1i} + \beta_3 t + \beta_4 (t - t_{0i}) I_{(t>t_{0i})} + \beta_5 x_{1i} t + \beta_6 x_{1i}(t - t_{0i}) I_{(t>t_{0i})} + \beta_7 x_{2t} + \beta_8 x_{3t} + \beta_9 x_{4t} + \alpha_i \quad (6)$$

where $\alpha_i \sim N\left(0, \sigma_\alpha^2\right)$, $I_{(t>t_{0i})}$ is an indicator function that takes on the value 0 if $t < t_{0i}$ and takes on the value 1 if $t \geq t_{0i}$ and σ_α^2 is the between-site variance in monthly log crash density. The two-way interactions allow for different slopes of log crash rate on time before and after the intervention and also across the two groups (treatment and comparison). We provide an interpretation of the various model parameters in Section 5.

We choose proper, semi-conjugate but non-informative prior distributions for the regression coefficients and for σ_α^2, to guarantee that the joint posterior distribution integrable and to let the data “speak for themselves”. Because the number of monthly observations available for each site as well as the number of sites was large enough, we anticipate that the priors will have little if any influence on the posterior distribution. The priors and the value of the prior parameters were

$$\beta \sim N_9(0, \Sigma_\beta),$$
$$\Sigma_\beta = \mathrm{diag}\{\sigma_{\beta_i}^2 : \sigma_{\beta_i}^2 = 1000, i = 1, \ldots, 9\},$$
$$\sigma_\alpha^2 \sim \mathrm{Ga}(0.01, 0.01).$$

Model (6) accounts for potential differences across treatment and comparison groups; within a group, sites can differ but only with respect to the general level of the curve. While the model allows for inference at the site level, it forces all sites in a group to conform to the same relationship between crash density and covariates.

In practice, and given that changes at the treatment sites are not gradual, we might expect a sudden drop (or increase) in crash density at those sites which received the intervention. Such an effect can be accommodated by an additional parameter in model (6) in which case

$$\log(\lambda_{it}) = \beta_1 + \beta_2 x_{1i} + \beta_3 t + \beta_4 (t - t_{0i}) I_{(t>t_{0i})} + \beta_5 x_{1i} t + \beta_6 x_{1i}(t - t_{0i}) I_{(t>t_{0i})} + \sum_{j=2}^{4} \beta_{j+5} x_{jt} + \beta_{10} x_{1i} I_{(t>t_{0i})} + \alpha_i \quad (7)$$

The additional parameter β_{10} in the log link function of the Poisson regression model is also normally distributed a priori, independent of other parameters. Model (7) is also non-hierarchical in that fixed values are assigned to the parameters of the prior distributions.

The design of the study included yoked (or matched) pairs of sites. Since the matched comparison sites were chosen to be as similar to treatment sites as possible, this may induce a correlation in log crash density between sites within pairs. Thus, we now define a second random effect δ_j for the jth pair and let α_i denote the random effect associated with site i. Model (8) below is identical to model (6) except that we include a possible correlation across sites induced by the pairing in the design:

$$\log(\lambda_{it}) = \beta_1 + \beta_2 x_{1i} + \beta_3 t + \beta_4 (t - t_{0i}) I_{(t>t_{0i})} + \beta_5 x_{1i} t + \beta_6 x_{1i}(t - t_{0i}) I_{(t>t_{0i})} + \sum_{j=2}^{4} \beta_{j+5} x_{jt} + \delta_j + \alpha_i, \quad (8)$$

where $\delta_j \sim N(0, \sigma_\delta^2)$ is the between-pair variability in log crash density and $\sigma_\delta^{-2} \sim \mathrm{Ga}(0.01, 0.01)$ is its prior distribution.

Sites in the intervention and in the comparison groups may differ in more than just the intercept due to geometric, driver, weather and other characteristics, even within group. To capture the potential differences in crash density over time and in the effect of the four to three lanes conversion, we extend models (6) and (7) and make them fully hierarchical. By allowing different values of each regression coefficient for each site, we capture between site, within group variability and enable meaningful inference both at the site and at the group levels. The underlying assumption is that sites within groups are exchangeable, so that given treatment and period group, the curves that represent the evolution of log crash densities over time are random deviations from a population or mean curve for each group. In the expression below for $\log(\lambda_{it})$ we allow all regression coefficients (except those associated with season) to vary across sites and include third level population distributions for those parameters. We assume that the effect of season on log crash densities is constant across sites, however, since seasonalities are likely to affect all sites in the State of Iowa in a similar manner. The expression for log crash density is then

$$\log(\lambda_{it}) = \beta_{1i} + \beta_{2i} x_{1i} + \beta_{3i} t + \beta_{4i} (t - t_{0i}) I_{(t>t_{0i})} + \beta_{5i} x_{1i} t + \beta_{6i} x_{1i}(t - t_{0i}) I_{(t>t_{0i})} + \sum_{j=2}^{4} \beta_{j+5} x_{jt} + \beta_{10} x_{1i} I_{(t>t_{0i})}, \quad (9)$$

where now $\beta_1 - \beta_6, \beta_{10} \sim N_7(\mu_\beta, \Sigma_\beta)$, $\beta_7, \beta_8, \beta_9 \sim \mathrm{iid}\, N(0, 1000)$, $\mu_\beta \sim N_7(0, \Sigma_{\mu_\beta})$, $\Sigma_{\mu_\beta} = \mathrm{diag}\{1000\}$, $\Sigma_\beta = \mathrm{diag}\{\sigma_\beta^2\}$ and $\sigma_\beta^{-2} \sim \mathrm{Ga}(0.001, 0.001)$.

A hierarchical version of (6) is similar to (9) but does not include additional parameters for possible drops or jumps on log crash density at the treatment sites following the intervention. The expression for log crash density in the hierarchical model without jump parameters is not shown here but results from fitting the model are presented later.

The last model formulation we consider is based on the Poisson-Gamma probability model (1). Here, the log mean of crash density is regressed on the covariates as above. Using the same notation as before,

$$\log(\mu_{it}) = \beta_1 + \beta_2 x_{1i} + \beta_3 t + \beta_4 (t - t_{0i}) I_{(t>t_{0i})} + \beta_5 x_{1i} t + \beta_6 x_{1i}(t - t_{0i}) I_{(t>t_{0i})} + \beta_7 x_{2t} + \beta_8 x_{3t} + \beta_9 x_{4t} + \alpha_i \quad (10)$$

for $\mu_{it} = \rho/\delta_{it}$, $\lambda_{it} \sim \mathrm{Ga}(\rho, \delta_{it})$. Priors are similar to those proposed earlier for the Poisson-LogNormal case, so that $\beta_j \sim N(0, 1000), j = 1, \ldots, 9$, $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\sigma_\alpha^{-2} \sim \mathrm{Ga}(0.01, 0.01)$ and

$\rho \sim \text{Ga}(0.01, 0.01)$. Christiansen and Morris (1997) propose a hierarchical Poisson-Gamma model like model (10) but with a slightly different parameterization and a prior for ρ that is similar in form to the logistic distribution. The Poisson-Gamma model can be extended in the same ways as the Poisson-LogNormal model and can also be formulated in a hierarchical manner.

4. Model selection and model diagnostics

We have considered several plausible model formulations to summarize the crash data in our study. Practitioners typically wish to decide whether one model formulation is "better" than others and if so, whether the selected model fits the data well. Thus, the model evaluation step will typically consist of two separate tasks: select the best set of models and then carry out diagnostics to determine whether the selected models fit the data. In practice we often select the "best" model by examining the fit of the models to the data thus blurring the distinction between the comparison and the diagnostic tasks.

4.1. Model comparison

The standard Bayesian approach to selecting between competing models relies on posterior model probabilities or on Bayes factors (BF) (e.g. Gelman et al., 2004). Schwartz's Bayesian Information Criterion (BIC; Schwartz, 1978), while shown to provide an approximation to the BF, requires specification of the dimension of the model. This presents a challenge in the case of hierarchical models where the size of the model is not clearly defined. Furthermore, BIC converges asymptotically to BF only when the number of model parameters is fixed and this assumption is violated in models where the number of random effects grows with the number of experimental units (in our case, road segments).

A general measure of model fit is the *deviance* $D(y, \theta)$ defined as twice the negative of the log-likelihood (Gelman et al., 2004). We denote the estimated deviance by $\hat{D}_{\text{avg}}(y)$ as in Gelman et al. (2004). The deviance measures the fit of the (nested) model to the observed data. If we wish to select a model with good out-of-sample predictive ability, one approach is to compute the expected predictive deviance which can be approximated by the *d* eviance information criterion (DIC; Spiegelhalter et al., 2002, 2003)

$$\text{DIC} = 2\hat{D}_{\text{avg}}(y) - D(y, \hat{\theta}(y)),$$

where $\hat{\theta}(y)$ is a point estimate of θ. The DIC penalizes model complexity and can be used to compare non-nested models. In general, the lower the DIC, the more accurate the model will be in terms of prediction of future data. While the DIC is an appropriate measure of model predictive ability in the generalized linear modeling context, it fails in other contexts including mixture modeling and models with unobserved latent variables (Gelfand and Ghosh, 2006).

4.2. Model diagnostics

While the DIC criterion is useful for selecting a model, a related question is whether the selected model fits the data well. Posterior predictive checking complements traditional techniques (such as residual analysis) and consists in comparing the distribution of *discrepancy statistics* of interest computed using data generated from the fitted model to the value of the statistic obtained using the observed data. We use $T(y, \theta)$ to denote a discrepancy statistic that depends on the data y and perhaps also on the parameters θ. The choice of the statistics $T(y, \theta)$ depends on the application and on the attributes of the data which we hope our model will accommodate (Gelman et al., 2004).

Here we wish to evaluate the model relative to its ability to accommodate the variability in monthly crashes at any one site in addition to accurately estimating the impact of the intervention at treatment sites. We therefore consider the following statistics:

$$\begin{aligned} T_1(y_i, \theta) &= \max\{y_{it}\}, \\ T_2(y_i, \theta) &= \text{standard deviation of } y_{it}. \end{aligned}$$

We anticipate that the selected model will fit the difference in mean crash densities better than it will fit $T_1(y_i, \theta)$, $T_2(y_i, \theta)$ both of which are associated with the variance in monthly crash densities.

To carry out posterior predictive checks, we first generate replicated data sets y^{rep} from the fitted model. We then compute the value of the statistic $T_j(y^{\text{rep}}, \theta)$ using each replicated dataset and compare $T_j(y^{\text{obs}}, \theta)$ (where y^{obs} denotes the observed crash data) to the distribution of the $T_j(y^{\text{rep}}, \theta)$ over the replicates. This procedure mimics the classical approach in that it considers the behavior of the statistics under repeated sampling.

5. Results

5.1. Estimates of model parameters

We estimated the posterior distributions of the parameters in the models discussed in Section 3 using Markov chain Monte Carlo (MCMC) methods and the freeware WinBUGS (Lunn et al., 2000; Spiegelhalter et al., 2003; Cowles, 2004). We monitored convergence of the chains using the Gelman–Rubin statistic (Gelman and Rubin, 1992; Cowles and Carlin, 1996) and also checked the autocorrelation functions (Cowles, 2004). The posterior mean and standard deviation as well as the 2.5th and 97.5th percentiles of the posterior distributions for the parameters in some of the models are given in Tables 3 and 4. The parameter σ_α is the square root of the between-site variance in log crash density and σ_δ is the between-pair standard deviation in log crash density.

Parameters β_7, β_8, and β_9 in Table 3 are significantly different from zero because their posterior credible sets are fully below or above zero for all models. Therefore, there are seasonal differences in safety in the State of Iowa (not a surprising find). In order to interpret the meaning of the remaining parameters we focus on model (6) and letting C and T denote control and treatment, respectively, note that:

$$E(\log(\lambda_{it})|C,\ \text{before}) = \beta_1 + \beta_3 t + \sum_{j=2}^{4} \beta_{j+5} x_{jt},$$

$$\begin{aligned} E(\log(\lambda_{it})|C,\ \text{after}) &= (\beta_1 - \beta_4 t_{0i}) + (\beta_3 + \beta_4)t + \sum_{j=2}^{4} \beta_{j+5} x_{jt}, \\ &= \beta_1^* + \beta_3^* t + \sum_{j=2}^{4} \beta_{j+5} x_{jt}, \end{aligned}$$

$$\begin{aligned} E(\log(\lambda_{it})|T,\text{before}) &= (\beta_1 + \beta_2) + (\beta_3 + \beta_5)t + \sum_{j=2}^{4} \beta_{j+5} x_{jt}, \\ &= \beta_1^{**} + \beta_3^{**} t + \sum_{j=2}^{4} \beta_{j+5} x_{jt}, \end{aligned}$$

Table 3
Summary of the posterior distributions of parameters in non-hierarchical models

Parameter	$\hat{\theta}$% credible set			
	Model (6)	Model (7)	Model (8)	Model (10)
β_1: intercept	−4.892 (−5.461, −4.370)	−4.832 (−5.342, −4.249)	−4.913 (−5.403, −4.411)	−4.847 (−5.405, −4.228)
β_2: treatment group	0.014 (−0.697, 0.861)	−0.126 (−0.883, 0.582)	−0.033 (−0.539, 0.509)	−0.125 (−0.761, 0.804)
β_3: time	−0.001 (−0.002, −0.001)	−0.001 (−0.002, −0.001)	−0.001 (−0.002, −0.001)	−0.001 (−0.002, −0.001)
β_4: intervention date	−0.005 (−0.007, −0.003)	−0.005 (−0.007, −0.003)	−0.005 (−0.007, −0.003)	−0.005 (−0.007, −0.003)
β_5: treatment by time	−0.001 (−0.001, 0.000)	0.000 (0.000, 0.001)	0.000 (−0.001, 0.000)	0.000 (−0.001, 0.000)
β_6: treatment by date	−0.008 (−0.011, −0.006)	−0.003 (−0.005, 0.000)	−0.008 (−0.010, −0.006)	−0.008 (−0.011, −0.006)
β_7: seasonal 1	0.036 (0.012, 0.059)	0.037 (0.014, 0.061)	0.035 (0.012, 0.057)	0.035 (0.010, 0.060)
β_8: seasonal 2	−0.046 (−0.062, −0.030)	−0.046 (−0.062, −0.031)	−0.045 (−0.061, −0.030)	−0.045 (−0.063, −0.028)
β_9: seasonal 3	0.086 (0.064, 0.108)	0.085 (0.063, 0.107)	0.085 (0.066, 0.107)	0.089 (0.065, 0.114)
β_{10}: discontinuity		−0.437 (−0.541, −0.331)		
τ				15.070 (11.920, 19.310)
σ_α: between site standard deviation	0.974 (0.738, 1.304)	0.980 (0.752, 1.302)	0.907 (0.145, 1.714)	1.107 (0.600, 1.772)
σ_δ: between pair standard deviation			0.637 (0.421, 0.979)	

$$E(\log(\lambda_{it})|T,\text{after}) = [\beta_1+\beta_2-(\beta_4+\beta_6)t_{0i}]+(\beta_3+\beta_4+\beta_5+\beta_6)t+\sum_{j=2}^{4}\beta_{j+5}x_{jt}, \\ = \beta_1^{***}+\beta_3^{***}t+\sum_{j=2}^{4}\beta_{j+5}x_{jt}.$$

From Table 3 we note that β_2, β_5 are not significantly different from zero in model (6) and that β_3, β_4, β_6 are significantly negative. This implies that in a comparison of the before and after periods at control sites, $\beta_1^* > \beta_1$ and $\beta_3^* < \beta_3$, which means that the decrease in log crash density at control sites accelerated in later years. However, we also note that $\beta_3^{***} < \beta_3^*$, meaning that at treatment sites, the decrease in log crash density was even more pronounced than at control sites. These results were observed as well for models (8) and (10). In model (7) (see Table 3), β_{10} (associated with a possible "jump" or "drop" effect) is significantly negative while β_5 is not significantly different from zero to weakly positive. However, in absolute value the posterior mean of β_5 is smaller than the posterior mean of $\beta_4 + \beta_6$. Thus there appears to be a sudden drop in log crash density at intervention sites in addition to a steeper decline in log crash density over time at these sites relative to the comparison sites even though the slope of log crash density on time in the after period is not as negative as when a jump is not included in the model. The introduction of the jump parameter in the model resulted in a smaller change in the slope of log crash density on time after the intervention.

One interesting finding is that the between pairs standard deviation corresponding to the yoked pair random effect in model (8), while smaller than the between site (within pairs) standard deviation, is still considerable, suggesting that at least in the non-hierarchical version of the model, accounting for the yoked pairs design of the study might be important. Results arising from the model with both a discontinuity effect and a random effect for pair-induced correlations are similar to those obtained for models (7) and (8) and are not shown here.

Table 3 also shows the results obtained when fitting the non-hierarchical Poisson-Gamma model (10). Results do not differ noticeably from those obtained when fitting model (6), which differs from model (10) only in the form of the probability model for crash density. This is not surprising since in our study, the prevalence of zero crashes was very low. Thus, the advantage of the Poisson-Gamma formulation in situations where crashes are very sparse is not realized here.

Table 4 shows results obtained from fitting fully hierarchical versions of the Poisson-LogNormal regression model. In addition to the posterior distributions summarized in the tables, we could have also shown the site-level posterior distributions for

Table 4
Summary of the posterior distributions of population parameters in two hierarchical models without and with parameters to accommodate a sudden drop in crash density at treatment sites

Parameter	Estimate: posterior mean and 95% credible set	
	Model w/o jump	Model w/jump (model (9))
β_7: seasonal 1	0.036 (0.012, 0.057)	0.037 (0.016, 0.061)
β_8: seasonal 2	−0.046 (−0.060, −0.031)	−0.046 (−0.062, −0.031)
β_9: seasonal 3	0.085 (0.063, 0.107)	0.084 (0.061, 0.105)
μ_{β_1}: mean of intercept	−4.772 (−5.313, −4.201)	−4.812 (−5.433, −4.272)
μ_{β_2}: mean of treatment group	−0.120 (−1.027, 0.626)	−0.161 (−1.036, 0.843)
μ_{β_3}: mean of time	−0.002 (−0.008, 0.004)	−0.001 (−0.008, 0.005)
μ_{β_4}: mean of intervention date	−0.004 (−0.015, 0.006)	−0.003 (−0.013, 0.008)
μ_{β_5}: mean of treatment by time	0.001 (−0.011, 0.014)	0.000 (−0.013, 0.013)
μ_{β_6}: mean of treatment by date	−0.020 (−0.040, −0.002)	−0.010 (−0.029, 0.009)
μ_C: mean of discontinuity at C sites		−0.060 (−0.251, 0.127)
μ_T: mean of discontinuity at T sites		−0.415 (−0.645, −0.223)
σ_{β_1}: standard deviation of intercept	0.992 (0.704, 1.382)	1.003 (0.715, 1.389)
σ_{β_2}: standard deviation of treatment group	0.354 (0.031, 1.169)	0.350 (0.029, 1.232)
σ_{β_3}: standard deviation of time	0.011 (0.008, 0.016)	0.012 (0.008, 0.016)
σ_{β_4}: standard deviation of intervention date	0.018 (0.012, 0.026)	0.017 (0.012, 0.024)
σ_{β_5}: standard deviation of treatment by time	0.016 (0.011, 0.026)	0.017 (0.011, 0.026)
σ_{β_6}: standard deviation of treatment by date	0.023 (0.014, 0.038)	0.021 (0.013, 0.035)
σ_C: standard deviation of discontinuity at C sites		0.245 (0.061, 0.488)
σ_T: standard deviation of discontinuity at T sites		0.244 (0.072, 0.475)

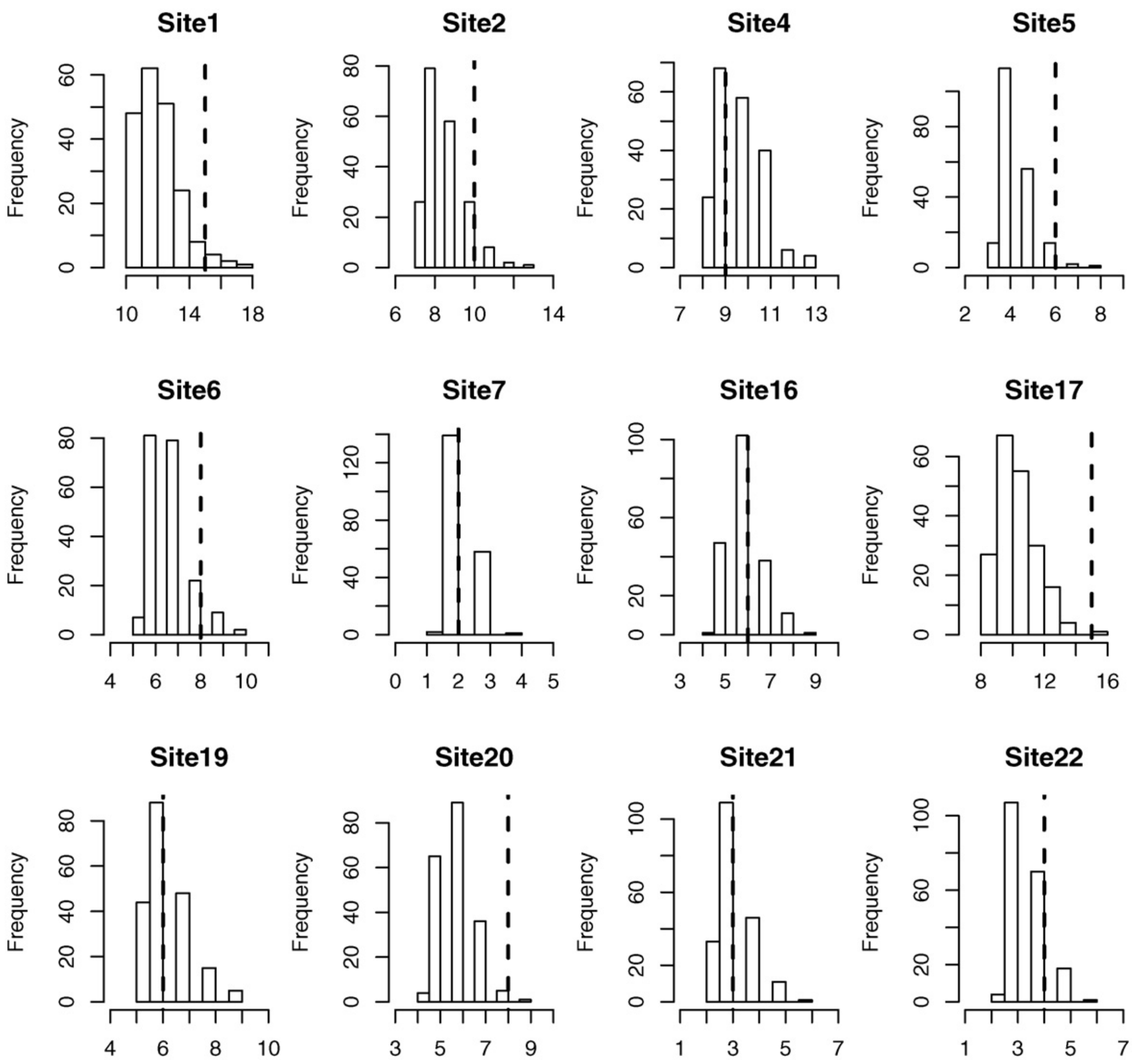

Fig. 1. Distribution of discrepancy statistic $T_1(y, \theta)$ in replicated datasets and observed value of the statistic for 12 sites. Replicated datasets were obtained from the hierarchical model with no jump parameters.

the regression coefficients associated with treatment group, time, intervention date and the interactions of treatment group with time and intervention dates (28×7 additional posterior distribution summaries). Instead, we show the posterior distributions for the population-level parameters of the seven regression coefficients and recall that under the hierarchical structure of the models, these population distributions generate the site-level regression coefficients.

While the fit of model (9) has not yet been evaluated, estimates obtained for the population-level model parameters are in line with our expectations. We note that:

1. The population-level mean μ_4 of β_4 is not significantly different from zero, so that the association between log crash density and time in comparison sites does not change over time.
2. The population-level mean μ_5 is not significant, so that there is no significant difference between comparison sites and treatment sites before the intervention regarding the association between log crash density and time.

Table 5
Deviance and DIC of six models

Index	Model	Deviance	DIC	Rank
Model (6)	Log-normal, non-hierarchical	8805.14	8828.04	5
Model (7)	Log-normal, non-hierarchical, with jump	8767.60	8788.21	3
Model (8)	Log-normal, non-hierarchical, pair effect	8799.05	8823.94	4
Not shown	Log-normal, hierarchical	8679.94	8741.02	2
Model (9)	Log-normal, hierarchical, jump	8647.62	8716.75	1
Model (10)	Gamma, non-hierarchical	8816.33	8845.10	6

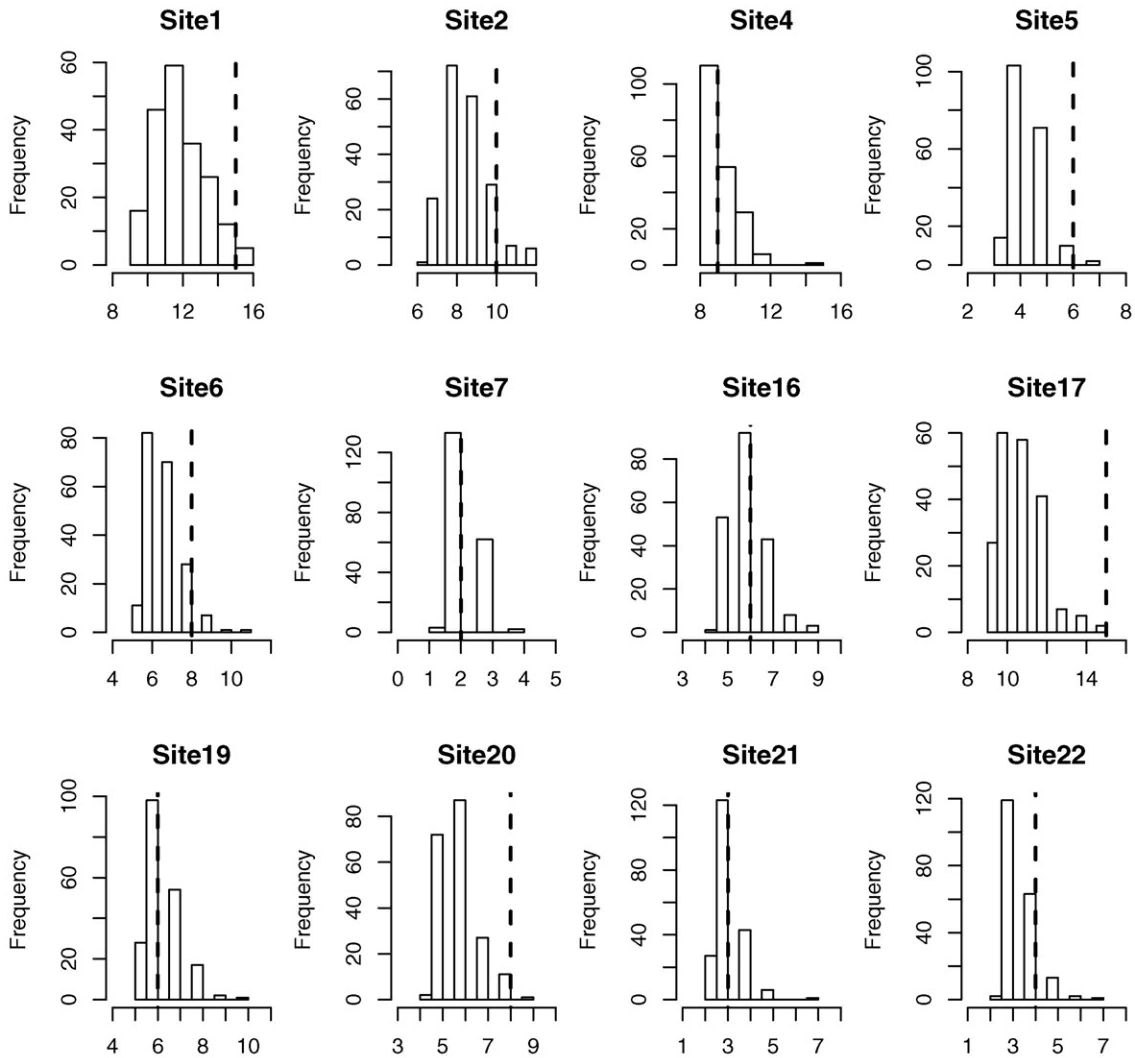

Fig. 2. Distribution of discrepancy statistic $T_1(y, \theta)$ in replicated datasets and observed value of the statistic for 12 sites. Replicated datasets were obtained from hierarchical model (9).

3. The population-level mean μ_6 is significantly negative, so that the slope of log crash density on time decreases significantly after the intervention at treatment sites.

These inferences at the population level may not coincide exactly with inferences at the level of each site. Because sites within group are allowed to differ regarding the effect of the various factors on log crash density, it may well be that for particular sites we find that the intervention did not significantly affect the association between log crash rate and time, for example. We can, however, argue that on the average, the effect of the intervention is statistically significant.

In comparing results obtained from the two hierarchical model formulations we find that while most of the parameter estimates remained approximately the same even after introduction of the discontinuity parameter, one change was noticeable. While at comparison sites we do not find that overall there is a sudden change in log crash density, the opposite is true at treatment sites. Note that μ_T is significantly smaller than zero, meaning that treatment sites overall exhibited a drop in log crash density immediately following the intervention . When the sudden drop parameter is included in the model, however, the change in the average slope of log crash density on time is no longer significant at intervention sites. This means that the change that seems to have occurred in log crash density at treatment sites after the intervention can be represented as a discontinuity (a sudden drop) or as a gradual drop. To decide which of the two representations is more consistent with the data requires that we implement model comparison and diagnostics.

5.2. Deciding on a "best" model

All models are defensible from a first principles viewpoint and produce believable results, but some may fit the data better and may produce more accurate predictions of the effect of implementing road diets on roadways like those considered in the study. We first select a small subset of the models using the deviance and the DIC criteria and then carry out posterior predictive assessment

Table 6
Posterior mean and 95% credible set for the difference and percent reduction in expected crash frequency (using monthly traffic volume estimated for each site) and in crash rate (100,000,000) AADT for model (9)

Site	Difference: after–before		Percent reduction	
	Crash frequency	Crash rate	Crash frequency	Crash rate
Treatment sites				
1	−25.3 (−28.7, −21.3)	−1391.5 (−1613.4, −1164.7)	55.8 (50.3, 60.9)	65.1 (59.8, 69.8)
2	−6.7 (−10.4, −1.6)	−291.9 (−386.2, −173.3)	30.1 (7.4, 46.3)	45.2 (27.1, 58.4)
4	−13.1 (−16.2, −9.8)	−487 (−645.0, −347.6)	55 (43.1, 65.7)	57.2 (44.6, 67.5)
5	6.4 (−2.3, 14.1)	10.5 (−228.4, 224.6)	−26.6 (−61.7, 9.0)	−2.4 (−36.1, 27.2)
6	−11.8 (−15.7, −8.1)	−470.3 (−578.3, −369.8)	52.2 (36.4, 67.0)	63.1 (50.7, 74.6)
7	−5.4 (−8.8, −1.9)	−324.8 (−515.8, −132.7)	54.4 (21.4, 76.0)	56.2 (25.3, 76.8)
8	−11.8 (−15.7, −8.6)	−535.5 (−706.8, −396.3)	46.5 (34.5, 58.1)	47 (35.0, 58.5)
9	−25.3 (−29.7, −20.9)	−708 (−811.2, −604.5)	46 (38.6, 53.2)	54.4 (48.0, 60.0)
10	−18.2 (−24.0, −12.0)	−619.9 (−780.7, −460.8)	60 (39.9, 76.9)	67.1 (50.5, 81.0)
11	−14 (−16.4, −11.4)	−555.4 (−657.6, −437.9)	77.7 (69.0, 85.6)	76.9 (66.3, 85.9)
12	−19.7 (−23.1, −16.4)	−794.6 (−889.0, −689.4)	52.1 (43.9, 59.5)	61 (54.1, 67.2)
13	−1.5 (−4.4, 2.1)	−205.8 (−269.6, −117.7)	8.4 (−12.7, 24.7)	37.4 (22.3, 48.7)
14	−2 (−3.3, −0.4)	−124.1 (−173.7, −64.3)	41.6 (10.0, 66.1)	59.7 (37.7, 76.6)
15	−2.7 (−4.6, −1.0)	−125.5 (−192.5, −62.0)	42.4 (15.7, 67.7)	50.9 (27.2, 73.7)
Control sites				
16	−8.6 (−12.8, −3.8)	−790.9 (−1030.6, −574.0)	30 (15.4, 43.3)	51.7 (41.8, 61.0)
17	−7.3 (−11.7, −1.3)	−311.8 (−405.9, −187.3)	28.6 (5.2, 44.3)	45.2 (27.9, 56.9)
19	−0.2 (−1.9, 1.5)	−91.1 (−248.1, 54.4)	1.9 (−20.9, 24.2)	17.5 (−15.5, 44.0)
20	−2.8 (−6.2, 0.1)	−247.5 (−506.1, −26.7)	18.2 (−0.9, 39.1)	31.7 (4.5, 55.3)
21	2.1 (−0.3, 5.2)	22.7 (−66.0, 128.7)	−42.1 (−103.8, 4.8)	−10.1 (−58.0, 26.2)
22	−2.4 (−4.0, −0.6)	−230.9 (−355.3, −91.5)	34.7 (9.2, 53.8)	41.7 (18.8, 59.0)
23	−0.5 (−3.6, 2.9)	−173.9 (−317.3, −17.2)	2.8 (−22.2, 22.5)	22.7 (2.7, 38.2)
24	−0.9 (−4.8, 3.4)	−197.3 (−285.9, −112.8)	2.4 (−10.2, 13.8)	22.8 (13.7, 31.8)
25	1.1 (−1.3, 3.8)	9.9 (−55.9, 82.6)	−20.2 (−70.3, 25.6)	−7 (−52.3, 34.1)
26	−2.6 (−4.8, 0.1)	−435.8 (−665.2, −171.1)	26.2 (−1.0, 45.9)	40.8 (17.5, 57.5)
27	−3.7 (−6.7, −1.2)	−206.6 (−319.7, −109.5)	18.1 (6.2, 31.5)	24.7 (14.0, 36.7)
28	−5.8 (−10.1, −1.3)	−138 (−242.3, −32.7)	13.3 (3.1, 22.8)	13.4 (3.3, 22.7)
29	−0.7 (−3.5, 2.1)	−66.7 (−155.8, 16.4)	8.9 (−34.2, 46.7)	26.1 (−7.6, 55.7)
30	−0.9 (−4.7, 2.8)	−230.2 (−320.0, −147.3)	4.5 (−14.7, 23.9)	36.2 (23.5, 49.5)
All sites				
–	−6.6 (−25.3,3.5)	−346.9 (−988.6,14.4)	25.8 (−31.6,65.8)	39.1 (−8.0,70.3)

of the models with lowest DIC. The deviance and the DIC for each model are given in Table 5. When calculating the deviance and DIC for the Poisson-Gamma model (10), the mean of λ_{it} (μ_{it}) was used instead of λ_{it} itself.

The model with the highest DIC (worst fit to the data) is (10). Among non-hierarchical models, the model that includes a discontinuity at the time of the intervention has the lowest DIC. The hierarchical versions of models (6) and (7) have the lowest DIC overall, indicating that the added complexity of the model pays off in terms of fit. We focus only on these two models in the remainder of this manuscript.

5.3. *Posterior predictive assessment of the hierarchical models*

We carried out posterior predictive assessment of the two hierarchical models by investigating how well the observed values of the discrepancy statistics $T_1(y^{\text{obs}}, \theta)$, $T_2(y^{\text{obs}}, \theta)$ described in Section 4.2 compare to the distributions of the statistics computed using replicated datasets generated from the models. For model (9), we generated 50 replicated datasets identical in size and structure to the observed set of crash data as follows:

1. For a draw of the population-level parameters μ_β, Σ_β we generated 28 six-dimensional vectors of regression coefficients β_i (one for each site).
2. Given those regression coefficient vectors, a draw of the parameters associated with seasonal effects and the values of the covariates for each site, we computed the log crash density for each site and each time point.
3. We then obtained the Poisson mean for a site and a month by re-scaling the crash density and multiplying it by the monthly average daily traffic (MADT) for the site.
4. We finally drew a crash from a Poisson distribution for each month of observation at each site.

The process followed to generate replicate datasets from the hierarchical model with a jump was identical except that it also required drawing values of the parameters associated with the sudden log crash drop using the appropriate draws from the population distributions for those parameters.

For reasons of space, we only show results corresponding to T_1 in Figs. 1 and 2 and only for the first six treatment sites and the first six comparison sites. Results obtained for the other sites and other discrepancy measures are comparable. In Figs. 1 and 2 we display the distributions of $T_1(y^{\text{rep}}, \theta)$ over replicated datasets generated from the two models, respectively. Each histogram represents the distribution of the discrepancy statistic across replicate datasets for each of the 12 sites included in the figures. The vertical lines in each histogram shows the value of the statistic computed from the observed data at the site.

Both models seem to do a good job of accommodating the variation in crash density across months within site in terms of maximum frequency of monthly crashes. For most of the sites (including those for which figures are not included here), the models appear to adequately reproduce functions of the mean (not shown), the maximum and the variability (not shown) in crash frequency. Thus, deciding in favor of one of the two hierarchical models in this study is more a matter of suitability to the application itself than to differences between models in term of their fit or predictive ability. From a statistical point of view, either one of the two models appears to be equally appropriate.

The Iowa DOT was interested in an estimate, at the site level, of the expected change in crash frequency at sites in which an intervention is implemented. We therefore computed the expected difference in mean crash frequency at each site before and after the intervention and the percent reduction in mean crash frequency after the intervention. We repeated the calculation, but this time focusing on crash rate. Results are shown in Table 6 for model (9). All values displayed in the table are posterior means with the corresponding 95% credible sets shown in parentheses.

As expected, intervention sites show a larger decrease in crash frequency and in crash rate than do comparison sites. In fact, with the exception of site 5 (which exhibited a non-significant *increase* in crash frequency) and sites 13 and 14 (at which the decrease in crash frequency was not statistically different from zero) all other intervention sites showed a significant decrease in crash frequency after the four lane to three lane conversion was implemented. With the exception of sites 16, 17, 22, 26 and 27, at which a small but statistically significant decrease in crash frequency was estimated, no change in frequency was observed at comparison sites. When considering crash rates we observe similar results. Even though now several of the comparison sites show statistically significant reductions in crash rates, the magnitude of the decrease is, for most sites, significantly smaller than the decrease estimated at the corresponding paired treatment site.

Site-level estimates corresponding to the formulation of the hierarchical Poisson-LogNormal model that does not include a term associated to a possible sudden change in log crash rate at intervention sites are not shown here. However, results were similar to those resulting from model (9) except that the increase in crash frequency at site 5 after the intervention was weakly significant.

6. Conclusions and discussion

We have discussed in some detail issues associated with model formulation, interpretation, comparison and selection in the context of a study to evaluate the effectiveness of an intervention. While our approach has been mostly statistical (given a set of plausible models, we have asked ourselves which of them exhibits better performance), the underlying thrust of this work is consistent with the discussion in Lord et al. (2005) and in Mitra and Washington (2007).

Lord et al. (2005) convincingly argue that it is important to let substantive knowledge guide the choice of models for crash data. As an example, they contrast the standard Poisson probability model for crashes with the mixture Poisson model in which one component restricts crashes to zero. Since the existence of perfectly safe sites is difficult to justify from a substantive viewpoint, they claim that the model itself is implausible. We take their discussion one step further by focusing here on the interpretation and comparison of models which while plausible from a substantive point of view might still exhibit poor performance from a statistical point of view. Since in many applications practitioners adopt multi-level models for crash data, it can be easy to lose track of the implications, at the crash data level, of model choices made down stream. Thus, we comment on the implied models for crash frequency when different probability models are selected for the mean of those frequencies.

We used data collected by the IA DOT to evaluate the effectiveness of a four- to three-lane conversions in certain types of roadways in Iowa. Results from this study were reported elsewhere (Pawlovich et al., 2006), so the focus of this work is on the methodology rather than on the conclusions from the study. The statistical models under consideration included the log-normal and the Gamma models for the Poisson mean and were either hierarchical or not. Further, we attempted to use knowledge about the nature of crashes themselves when proposing different sets of covariates to represent crash density at a site. For this particular set of data, findings were as we might have been anticipated given what we know about four-lane to three-lane conversions in undivided roadways in Iowa. We also found that the design of the study in yoked pairs could be ignored in the analysis, since the correlation in crashes between sites within pairs while not negligible, did not affect inferences drawn from the model. This again is not surprising given that the paired comparison sites were selected *ex post* rather than as part of a designed experiment.

In recent years, discussion of the relative merits of what is known as the EB and FB methodologies has appeared in the traffic safety literature (e.g., Miaou and Lord, 2003; Pawlovich et al., 2006; Persaud and Lyon, 2007; Lord and Miranda-Moreno, 2007). In this manuscript, we have purposely avoided using EB or FB to refer any of our models and analysis, because as currently used in the safety literature (in particular as it refers to FB), the terminology can be confusing and is not always consistent with that used in the statistics literature as we explain below.

In the safety literature (e.g., Hauer et al., 2002; Miaou and Lord, 2003), the EB label appears to be used to denote a point estimate of a parameter that combines information from various sources. Strictly speaking, if a model has a p-dimensional vector of parameters λ (which may include location and dispersion parameters) and if the parameters θ of their prior distribution $p(\lambda|\theta)$ are estimated by maximizing the marginal distribution $p(y|\theta)$ with respect to θ, then we talk about an EB approach (see, e.g., Carlin and Louis, 2000). In practice, the term has been used (less than correctly) for analysis in which the estimate of θ, denoted $\hat{\theta}$ is based on data and once estimated, is considered "known" and fixed. The important point here is that inference about the first level parameter vector λ is then based on the posterior distribution $p(\lambda|y, \hat{\theta})$ and thus is conditional on the fixed estimate $\hat{\theta}$. This results in over-confident estimation for λ because the uncertainty in θ is not incorporated in the inference. In this light, estimation arising from our models (6)–(8) is similar to an EB analysis since the prior distributions for model parameters are indexed by parameters that are assumed to be known. To really account for all parameter uncertainties, one assumes that the second level parameters θ are not known a priori (even as point estimates) but are instead generated by a hyper-prior distribution (or population distribution) $p(\theta|\phi)$ where typically ϕ is assigned a non-informative prior (e.g., Gelman et al., 2004, Chapter 5). Inference about λ is then based on $p(\lambda|y)$, obtained by integrating $p(\lambda|y, \theta)$ with respect to the distribution of θ. We argue that the FB label would be more appropriate when inference about first level parameters is *not* conditional on a point estimate of the prior parameters. A natural approach to carry out a FB analysis is then to formulate the model in a hierarchical way.

We propose that a more accurate nomenclature would be along the lines of non-hierarchical (with informative or with non-informative priors) or hierarchical Bayesian approaches. If the model is not hierarchical and the prior parameters are obtained by maximizing the appropriate marginal likelihood, then this would correspond to the EB label. Hierarchical models would correspond to the current FB label. A full accounting of all uncertainties associated with model parameters can only be realized when the model is formulated in a hierarchical manner. Thus, we propose that FB be used (if at all) to refer to estimation based on multi-level or hierarchical models.

Acknowledgments

The authors wish to thank the two anonymous reviewers as well as Associate Editor Abdel-Aty for their constructive comments and suggestions. We have greatly benefited from discussions with our colleagues at the Center for Transportation Research and Education (CTRE) at Iowa State University.

References

Agresti, A., 2002. Categorical Data Analysis, 2nd edition. Wiley, New York.

Al-Masaeid, H.R., 1990. Bayesian approach to the estimation of expected number of accidents. Ph.D. Dissertation, Purdue University.

Battese, G.E., Harter, R.M., Fuller, W.A., 1988. An error component model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association 83, 28–36.

Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis, 2nd edition. Chapman & Hall/CRC, London, England.

Christiansen, C.L., Morris, C.M., 1997. Hierarchical Poisson regression modeling. Journal of the American Statistical Association 92, 618–632.

Conway, R.W., Maxwell, W.L., 1961. A queuing model with state-dependent service rates. Journal of Industrial Engineering 12, 132–136.

Cowles, M.K., 2004. Review of WinBUGS 1.4. Statistical computing software reviews. Journal of the American Statistical Association 89, 330.

Cowles, M.K., Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association 91, 883–904.

Gelfand, A.E., Ghosh, M., 2006. Generalized linear models: a Bayesian view. In: Dey, D.K., Ghosh, S.K., Mallik, B.K. (Eds.), Generalized Linear Models: A Bayesian Perspective. CRC Press/Taylor & Francis, Boca Raton, FL.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd edition. Chapman & Hall, London, England.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–472.

Ghosh, M., Rao, J.N.K., 1994. Small area estimation: an appraisal. Statistical Science 9, 55–76.

Hauer, E., 1997. Observational Before–After Studies in Road Safety. Pergamon, Oxford, UK.

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S., 2002. The empirical Bayes method for estimating safety: a tutorial. Transportation Research Record 1784, 126–131.

Huang, H.F., Stewart, J.R., Zegeer, C.V., 2002. Evaluation of lane reduction "road diet" measures on crashes and injuries. Transportation Research Record 1784, 80–90.

Kadane, J.B., Shmueli, G., Minka, T.P., Borle, S., Boatwright, P., 2006. Conjugate analysis of the Conway–Maxwell-Poisson distribution. Bayesian Analysis 1, 363–374.

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46, 751–770.

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 & Prevention 37, 35–46.

Lord, D., Guikema, S.D., Geedipally, S.R., 2008. Application of the Conway–Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. Accident Analysis & Prevention 40, 1123–1134.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, 325–337.

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., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion and spatial dependence. Accident Analysis and Prevention 37, 699–720.

Mitra, S., Washington, S., 2007. On the nature of over—in motor vehicle crash prediction models. Accident Analysis & Prevention 39, 459–468.

Pawlovich, M., Li, W., Carriquiry, A.L., Welch, T., 2006. Iowa's experience with "road diet" measures: impacts on crash frequencies and crash rates assessed following a Bayesian approach. Transportation Research Record 1953, 163–171.

Persaud, B., Retting, R., Garder, P., Lord, D., 2001. Observational before–after study of U.S. roundabout conversions using the empirical Bayes method. Transportation Research Record 1751, 1–8.

Persaud, B., Lyon, C., 2007. Empirical Bayes before–after studies: lessons learned from two decades of experience and future directions. Accident Analysis & Prevention 39, 546–555.

Qin, X., Ivan, J.N., Ravishankar, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis & Prevention 36, 183–191.

Schlütler, P.J., Deely, J.J., Nicholson, A.J., 1997. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. The Statistician 46, 293–316.

Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, P., 2005. A useful distribution for fitting discrete data: revival of the Conway–Maxwell-Poisson distribution. Journal of the Royal Statistical Society, Part C 54, 127–142.

Schwartz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6, 461–464.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, Part B 64, 583–640.

Spiegelhalter, D.J., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS Users Manual, Version 1.4.

Srinivasan, S., Kockelman, K.M., 2002. The impacts of bypasses in small and medium-sized communities: an econometric analysis. Journal of Transportation and Statistics 5, 57–69.

Tunaru, R., 1999. Hierarchical Bayesian models for road accident data. Traffic Engineering & Control 40, 318–324.