Accident Analysis and Prevention 42 (2010) 1118–1127

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A fully Bayesian multivariate approach to before–after safety evaluation

Eun Sug Park[a,*], Jaebeom Park[b,1], Timothy J. Lomax[a,2]

[a] Texas Transportation Institute, SPPE, Texas A&M University System, College Station, TX 77843-3135, United States
[b] Highway & Transportation Technology Institute, Korea Expressway Corporation, 50-5 Sancheok-Ri, Dongtan-Myeon, Hwaseong-Si, Gyonggi-Do 445-812, South Korea

ARTICLE INFO

Article history:
Received 13 May 2009
Received in revised form
23 December 2009
Accepted 27 December 2009

Keywords:
Fully Bayesian approach
Before–after evaluation with a comparison group/comparison groups
Multivariate Poisson-lognormal model
Change-point model
Markov chain Monte Carlo
Safety effectiveness

ABSTRACT

This paper presents a fully Bayesian multivariate approach to before–after safety evaluation. Although empirical Bayes (EB) methods have been widely accepted as statistically defensible safety evaluation tools in observational before–after studies for more than a decade, EB has some limitations such that it requires a development and calibration of reliable safety performance functions (SPFs) and the uncertainty in the EB safety effectiveness estimates may be underestimated when a fairly large reference group is not available. This is because uncertainty (standard errors) of the estimated regression coefficients and dispersion parameter in SPFs is not reflected in the final safety effectiveness estimate of EB.

Fully Bayesian (FB) methodologies in safety evaluation are emerging as the state-of-the-art methods that have a potential to overcome the limitations of EB in that uncertainty in regression parameters in the FB approach is propagated throughout the model and carries through to the final safety effectiveness estimate. Nonetheless, there have not yet been many applications of fully Bayesian methods in before–after studies. Part of reasons is the lack of documentation for a step-by-step FB implementation procedure for practitioners as well as an increased complexity in computation. As opposed to the EB methods of which steps are well-documented in the literature for practitioners, the steps for implementing before–after FB evaluations have not yet been clearly established, especially in more general settings such as a before–after study with a comparison group/comparison groups. The objectives of this paper are two-fold: (1) to develop a fully Bayesian multivariate approach jointly modeling crash counts of different types or severity levels for a before–after evaluation with a comparison group/comparison groups and (2) to establish a step-by-step procedure for implementing the FB methods for a before–after evaluation with a comparison group/comparison groups.

The fully Bayesian multivariate approach introduced in this paper has additional advantages over the corresponding univariate approaches (whether classical or Bayesian) in that the multivariate approach can recover the underlying correlation structure of the multivariate crash counts and can also lead to a more precise safety effectiveness estimate by taking into account correlations among different crash severities or types for estimation of the expected number of crashes. The new method is illustrated with the multivariate crash count data obtained from expressways in Korea for 13 years to assess the safety effectiveness of decreasing the posted speed limit.

1. Introduction

The use of Bayesian methods in traffic safety studies has become very popular over the last two decades (Aul and Davis, 2006; Carriquiry and Pawlovich, 2004; Davis and Yang, 2001; Fitzpatrick and Park, 2009; Harwood et al., 2002; Hauer, 1986, 1997; Hauer et al., 2002; Li et al., 2008; Miaou and Lord, 2003; Park and Lord, 2007; Patel et al., 2007; Pawlovich et al., 2006; Persaud and Hauer, 1984; Persaud, 1988; Persaud et al., 1997, 2004, 2009; Persaud and Lyon, 2007; Schluter et al., 1997; Lan et al., 2009; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009). Especially in before–after evaluation, empirical Bayes (EB) methods have been widely used and regarded as statistically defensible methods that can cope with several threats to validity of observational before–after studies such as the regression-to-the-mean bias. Extensive documentations of the EB methods are available (Hauer, 1986, 1997; Persaud, 1988; Persaud et al., 1997, 2004, 2009; Hauer et al., 2002; Harwood et al., 2002; Patel et al., 2007; Persaud and Lyon, 2007; Fitzpatrick and Park, 2009). Persaud and Lyon (2007) provided a comprehensive review on the Empirical Bayes before–after safety studies including the basics of EB evaluation and the need for and validity of the EB approach, and addressed the issues that are

* Corresponding author. Tel.: +1 979 845 9942; fax: +1 979 845 6008.
E-mail addresses: e-park@tamu.edu (E.S. Park), jbpark@ex.co.kr (J. Park), t-lomax@tamu.edu (T.J. Lomax).
[1] Tel.: +82 31 371 3313; fax: +82 31 371 3319.
[2] Tel.: +1 979 845 9960; fax: +1 979 845 6008.

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.12.026

critical to the proper conduct and interpretation of EB evaluations.

Although not as widely applied as EB yet, fully Bayesian (FB) approaches in highway safety have been explored more recently by several researchers (Aul and Davis, 2006; Carriquiry and Pawlovich, 2004; Davis and Yang, 2001; Li et al., 2008; Miaou and Lord, 2003; Park and Lord, 2007; Pawlovich et al., 2006; Schluter et al., 1997; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009; Lan et al., 2009; Persaud et al., 2009). Carriquiry and Pawlovich (2004) provided an excellent introduction and overview on the fully Bayesian approaches (though not in a before–after study setting) and highlighted the differences between the empirical Bayes and the fully Bayesian approaches. Whereas the EB methods have been widely applied in the before–after safety evaluations, there have not been many applications of the FB methods in before–after studies and so documentation on how to implement fully Bayesian before–after studies is rare. Notable exceptions are Pawlovich et al. (2006), Li et al. (2008), Aul and Davis (2006), Lan et al. (2009), and Persaud et al. (2009). Pawlovich et al. (2006) introduced a hierarchical Poisson regression model with a change point to before–after evaluation, and Li et al. (2008) explored a range of variations of the model introduced in Pawlovich et al. (2006) and addressed an issue of model selection in safety evaluation within a fully Bayesian framework. Aul and Davis (2006) estimated crash modification factors (CMFs) using a hybrid Bayesian method. Lan et al. (2009) conducted a validation study of a full Bayes methodology for before–after safety evaluation through simulation and applied the method to the data on rural signal conversions. Lan et al. (2009), however, adopted the framework that is close to the EB in that they compared the predicted crash count without treatment ($\pi_{i,t}$) in the after period to the *observed* crash count ($y_{i,t}$), not to the *expected* crash count, in the after period. The difference between the traditional EB approach and the FB approach by Lan et al. (2009) is that the predicted crash count without treatment in Lan et al. (2009) was obtained by the regression model estimated using data from both reference sites and the before period of treated sites whereas the predicted crash count without treatment in EB is obtained as a weighted sum of the before period observed crash count at the treated site and the expected crash count from the regression model estimated using data from reference sites only. Persaud et al. (2009) compared the EB and FB through two empirical applications, and reported that two approaches lead to comparable results. However, those comparisons were made only for the cases where a fairly large reference group was available for developing the SPFs (of EB), and the FB approach of Persaud et al. (2009) was the same approach in Lan et al. (2009) of which framework is close to that of EB.

In addition to the short history, two other main reasons for non-prevalent use of the FB methods in before–after safety evaluations appear to be (1) additional complexity in modeling and estimation and (2) the lack of clear guidelines for step-by-step implementation of the before–after FB evaluations. As discussed in Li et al. (2008), there could be several different ways of implementing the FB methods especially in terms of model specifications, and as a matter of fact, this flexibility is one of the advantages of FB methods. Nonetheless, it is still desirable to establish clear step-by-step procedures for implementing FB methods for practitioners.

Recently, Park and Lord (2007) introduced a new multivariate fully Bayesian approach for jointly modeling crash counts by severity data based on Multivariate Poisson-lognormal models (MVPLN). The MVPLN models are extended in this paper for before–after studies by adapting a change-point modeling framework introduced by Pawlovich et al. (2006). The MVPLN models with a change-point (MVPLNC) approach introduced in this paper has an additional advantage over the corresponding univariate before–after evaluation approaches (whether classical or Bayesian) in that the multivariate approach can recover the underlying correlation structure of the multivariate crash counts as well as lead to a more precise safety effectiveness estimate by incorporating correlations across different crash severities or types into estimation of the expected number of crashes.

We will also present in the paper a step-by-step multivariate FB procedure tailored to the before–after study with a comparison group/comparison groups. Although the data from such evaluation design have been analyzed by FB methods elsewhere (Pawlovich et al., 2006; Li et al., 2008), the comparison group was used merely for comparing the results from it to those from the treatment group and was not used for formally accounting for the difference in before–after periods resulting from extraneous factors in the safety reduction estimates for the treatment group. In this paper, we will show how the FB method can incorporate the crash information from the comparison group(s) into the crash reduction estimates for the treatment group(s) to account for general trends between the before and after periods. The new method is illustrated with the multivariate crash count data obtained from expressways in Korea for 13 years to assess the safety effectiveness of decreasing the posted speed limit.

2. Methodology

The focus of this paper is to present a general methodological framework for implementing fully Bayesian multivariate methods for before–after evaluations with comparison groups. The safety evaluation methodologies for the effectiveness of the countermeasures can be divided into two types: before–after evaluation methods and cross-sectional data analysis methods depending on the study design and the nature of the data, especially the availability of the information on the installation of the countermeasures. The before–after evaluation is usually considered superior to cross-sectional data analysis in that before–after evaluation can cope with site-to-site variability more effectively.

2.1. Study design for rigorous before–after evaluation

Note that in the safety evaluation study the outcome variable is usually the number of crashes per road segment per unit of time (often a year or a month). The main factor of interest (treatment) is the countermeasure. It is well-known that for before–after field observational studies the naïve before–after study will not be able to distinguish the effect of treatment from the effects of other factors that have also changed from the before to the after period (Hauer, 1997). In addition to the main treatment factor, countermeasure, the study design should incorporate any other important variables that can possibly affect safety such as roadway characteristics and traffic volumes. These variables need to be controlled either by being included as factors or blocks or being held at a constant level for all road segments (each road segment corresponds to a site) so as not to be confounded with the effect of the countermeasure when assessing the safety performance. To take into account the effects of other extraneous factors (factors not included in the model) that might change with the passage of time, it is strongly recommended that a comparison group is selected. A comparison group is a group of sites selected as being similar enough to the treatment sites with respect to roadway and traffic characteristics to adequately reflect what would have happened to the treatment sites, had they not been treated. Often a comparison group provides only control for some of the threats to validity of before–after observational studies (Campbell and Ross, 1968; Campbell and Stanley, 1963) such as changes in weather, vehicle fleet, driver characteristics, and the study design with a comparison group is strongly preferred to those without it.

In this paper, we will adapt a before–after evaluation design with a comparison group/comparison groups as a study design, and will show how the fully Bayesian multivariate approach can be implemented to analyze the data obtained from such design. Although the crash data obtained from such design have been analyzed elsewhere in a univariate way (Pawlovich et al., 2006; Li et al., 2008), there is still a room for further development. For example, in both of Pawlovich et al. (2006) and Li et al. (2008), the expected change in crash frequency at comparison sites were compared in an ad-hoc way with the expected change in crash frequency at treatment sites (computed as the expected difference in mean crash frequency at each site before and after the treatment or the percent reduction in mean crash frequency after the treatment), drawing the conclusions such that the latter was statistically significant while the former was insignificant. The expected changes in the comparison group were not used to adjust the expected changes in crash frequency at treatment sites. Recall that the purpose of having a comparison group in the study design is to account for the effects of unmeasured factors that might change from the before to the after period. To fulfill this objective, the estimated before–after change in the comparison group should actually be utilized to adjust the estimated change in the treatment group. We will show how this can be done as well as provide a way to analyze the multivariate crash data simultaneously, and introduce a clear step-by-step procedure for implementing a fully Bayesian multivariate before–after evaluation.

2.2. Fully Bayesian (FB) multivariate approach

The key idea of a Bayesian approach is that any prior knowledge or extra information on the crash prediction models or potential safety effectiveness of the countermeasures can be incorporated into the evaluation in the form of the prior distribution for the model parameters. In the fully Bayesian (FB) approach, prior information and all available data are incorporated into a single coherent statistical model and integrated into posterior distributions on which inferences on the parameter of interest can be based. By setting hyper-prior (second-level prior) distributions on the parameters of prior distributions, uncertainty in the parameters (e.g., regression parameters for the crash prediction) of prior distribution for the expected number of crashes is propagated throughout the model and carries through to the final safety effectiveness estimates. On the other hand, in the EB approach the parameters of prior distribution for the expected number of crashes are estimated externally (based on data from reference sites) using SPFs and are treated as if they were the true parameters once they are estimated. Thus, the associated uncertainties (standard errors) in the regression model parameters or dispersion parameters of SPFs are not incorporated into the final safety effectiveness estimate and may lead to underestimation of true uncertainty in the final safety effectiveness estimate. The interested reader may refer to Carriquiry and Pawlovich (2004) for more discussion on the difference between FB and EB.

The FB approach needs the specification of models for crash frequency up front. Model specification in FB is very flexible. Any reasonable model for crash frequencies can be employed (see, e.g., Carriquiry and Pawlovich, 2004; Li et al., 2008; Park and Lord, 2007; Pawlovich et al., 2006). Park and Lord (2007) presented the multivariate Poisson-lognormal (MVPLN) models that can simultaneously analyze the crash frequencies of different severities or crash types obtained from multiple intersections (cross-sectional data). We generalize the MVPLN model in Park and Lord (2007) to encompass a change-point model that can analyze before–after data with a comparison group/comparison groups.

Analyzing the multivariate crash data jointly has advantages over a univariate approach analyzing each crash type separately in that the underlying correlation structure of the multivariate crash counts can be recovered as well as the expected number of crashes can be estimated more precisely by incorporating correlations among different crash types (see, e.g., Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009), which in turn will result in more precise estimation of the safety effectiveness of a countermeasure. In previous before–after approaches such as naïve analysis, comparison group analysis, or EB approach, there have been no attempts to model multivariate crash counts simultaneously. The generalized modeling framework of MVPLN models that can analyze the before–after crash counts of different severities or crash types simultaneously is presented in the following section.

2.3. Modeling framework for multivariate before–after crash counts

Let $\boldsymbol{y_{it}} = (y_{it}^1, y_{it}^2, \ldots, y_{it}^J)$ denote a multivariate observation, consisting of counts of J different types of crashes at site i ($i = 1, \ldots, I$) during time (year) t ($t = 1, \ldots, m$). That is, y_{it}^j is the number of crashes of jth type occurred in year t at site i. Let K be the number of covariates and $X_{it} = (1, X_{1it}, \ldots, X_{6it}, X_{7i}, \ldots, X_{Ki})$ be a $(K+1)$-dimensional vector of covariates. Let $\boldsymbol{\beta^j} = (\beta_0^j, \beta_1^j, \ldots, \beta_K^j)'$ denote the $(K+1)$-dimensional column vector of the regression coefficients for the crash count of jth crash type. Let $\boldsymbol{b_{it}} = (b_{it}^1, b_{it}^2, \ldots, b_{it}^J)$ denote a vector of yearly random effects corresponding to site i and year t for J different types of crashes, explaining extra-Poisson variability. Suppose that, conditional on b_{it}^j and $\boldsymbol{\beta^j} \in R^{K+1}$, the crash count of the jth crash type at site i in year t, y_{it}^j, follows a Poisson distribution with mean μ_{it}^j, i.e.,

$$y_{it}^j \left| b_{it}^j, \boldsymbol{\beta^j} \right. \sim Poisson(\mu_{it}^j) \tag{1}$$

where

$$\mu_{it}^j = \exp(X_{it}\boldsymbol{\beta^j} + b_{it}^j). \tag{2}$$

The y_{it}^js are independent given the μ_{it}^js. As in Park and Lord (2007), the correlations among the crash counts of J different crash types at site i can be accounted for by assuming the vector of random effects corresponding to J crash types, $\boldsymbol{b_{it}} = (b_{it}^1, b_{it}^2, \ldots, b_{it}^J)$, have the following distribution:

$$\boldsymbol{b_{it}} | \Sigma \sim N_J(0, \Sigma) \tag{3}$$

where Σ is an unrestricted covariance matrix and $N_J(\cdot,\cdot)$ denotes J-dimensional multivariate normal distribution.

Model in Eqs. (1)–(3) can be regarded as a generalization of the Multivariate Poisson-lognormal (MVPLN) models given in Park and Lord (2007) in that the model has an additional time component as well as the components for multiple sites and crash types. Let the elements of the covariate vector $X_{it} = (X_{1it}, \ldots, X_{6it}, X_{7i}, \ldots, X_{Ki})$ be

$$\begin{aligned}
X_{1it} &= \log(AADT_{it}) = \ln AADT_{it},\\
X_{2it} &= T_i,\\
X_{3it} &= t,\\
X_{4it} &= (t - t_{0i})\boldsymbol{I}[t > t_{0i}],\\
X_{5it} &= T_i t,\\
X_{6it} &= T_i(t - t_{0i})\boldsymbol{I}[t > t_{0i}],
\end{aligned}$$

$X_{7i}, \ldots, X_{Ki}$: roadway characteristic variables such as lane width, shoulder width, number of lanes, etc. for the ith site, where

$T_i = 1$ if the ith site is a treatment site and is zero otherwise,
$t = t$th year in the study period ($t = 1, 2, \ldots, m$),
t_{0i} = year in which the countermeasure was installed at site i (for a site in the comparison group, it is defined to be the same year as that for the corresponding treatment group),
$\mathbf{I}[t > t_{0i}] = 1$ if t belongs to the after period and zero otherwise.

Then, Eq. (2) can be re-written as follows:

$$\begin{aligned}\mathrm{Log}(\mu_{it}^{j}) = \beta_0^j + \beta_1^j \ \ln AADT_{it} + \beta_2^j T_i + \beta_3^j t + \beta_4^j (t - t_{0i})\boldsymbol{I}[t > t_{0i}] \\ + \beta_5^j T_i t + \beta_6^j T_i (t - t_{0i})\boldsymbol{I}[t > t_{0i}] + \beta_7^j X_{7i} + \cdots + \beta_K^j X_{Ki} + b_{it}^j\end{aligned} \quad (4)$$

This model can be viewed as a change-point model which assumes that at the time of implementation, there is a possible change in the intercept and slope with respect to time at treatment sites that might be attributable to the implementation of the countermeasure. Note that the comparison group also has the imaginary before and after periods defined the same as those for the matching treatment group although no treatment is applied to sites in the comparison group (see Hauer, 1997). Note that for each group (Comp: Comparison, Trt: Treatment) and period (B: Before, A: After), Eq. (4) can be represented in terms of *log*(*mean crash count*) versus *time* (t) as follows:

$$\begin{aligned}
\mathrm{Log}(\mu_{it}^{j})_{Comp,B} &= \beta_0^j + \beta_1^j \ln AADT_{it} + \beta_3^j t + \beta_7^j X_{7i} + \cdots + \beta_K^j X_{Ki} + b_{it}^j, \\
\mathrm{Log}(\mu_{it}^{j})_{Comp,A} &= (\beta_0^j - \beta_4^j t_{0i}) + \beta_1^j \ln AADT_{it} + (\beta_3^j + \beta_4^j)t + \beta_7^j X_{7i} + \cdots + \beta_K^j X_{Ki} + b_{it}^j, \\
\mathrm{Log}(\mu_{it}^{j})_{Trt,B} &= (\beta_0^j + \beta_2^j) + \beta_1^j \ln AADT_{it} + (\beta_3^j + \beta_5^j)t + \beta_7^j X_{7i} + \cdots + \beta_K^j X_{Ki} + b_{it}^j, \\
\mathrm{Log}(\mu_{it}^{j})_{Trt,A} &= \{\beta_0^j + \beta_2^j - (\beta_4^j + \beta_6^j)t_{0i}\} + \beta_1^j \ln AADT_{it} + (\beta_3^j + \beta_4^j + \beta_5^j + \beta_6^j)t + \beta_7^j X_{7i} + \cdots + \beta_K^j X_{Ki} + b_{it}^j
\end{aligned}$$

As can be seen from the above equations, the intercept and slope of *log*(*mean crash count*) on time can be different depending on whether it is a comparison site or a treatment site and also depending on whether it is a before period or an after period. As explained in Pawlovich et al. (2006) and Li et al. (2008), such model allows estimation and accounting for time effects, countermeasure implementation, as well as the effects of other covariates in the model. Note that unlike the models in Pawlovich et al. (2006) or Li et al. (2008) which were developed for monthly crash data and included seasonal dummy variables, the multivariate model given in Eqs. (1)–(4) describes yearly crash data and includes AADT and roadway characteristic variables as covariates but not seasonal dummy variables. More importantly, the above multivariate model can incorporate the correlations in J different types of crashes into estimation, which was not possible in previous univariate change-point models.

A fully Bayesian analysis of model given in Eqs. (1)–(4) requires the (second-level) prior distributions for the parameters, $\beta_0^j, \beta_1^j, \beta_2^j, \ldots, \beta_K^j$ as well as Σ, to be chosen. Implementation of such model is not straightforward and it is necessary to adapt simulation-based methods such as a Markov chain Monte Carlo (MCMC) method (see, e.g., Chib and Greenberg, 1995; Gilks et al., 1996; Liu, 2001; Tierney, 1994).

2.4. Estimation of parameters

The MCMC simulation is employed for parameter estimation, which can be implemented in MATLAB, following the algorithm presented in Park and Lord (2007). For the prior on the parameters, we assume that $(\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^J, \Sigma)$ independently follow the distributions $\boldsymbol{\beta}^{\boldsymbol{j}} \sim N_{K+1}(c_0, C_0^{-1}), j = 1, \ldots, J$, and $\Sigma^{-1} \sim Wishart(R_0, r_0)$, where $N_{K+1}(c_0, C_0^{-1})$ is the $(K+1)$-variate normal distribution with mean vector c_0 and precision matrix C_0, *Wishart* (R_0, r_0) is the Wishart distribution (see, e.g., Anderson, 1984) with scale matrix R_0 and degrees of freedom parameter r_0, and (c_0, C_0, r_0, R_0) are pre-specified hyperparameters. When there is not much prior information available on $\boldsymbol{\beta}$ or Σ, the hyperparameter values can be chosen so that they lead to proper but diffuse priors. When there exists good prior information on the parameters, however, it can be incorporated by the use of more informative prior distribution. For example, any prior knowledge or expert opinion on a range of plausible values for the regression parameters β can be incorporated in selecting c_0 and C_0. Schluter et al. (1997) provides a good discussion on elicitation of priors in crash data analysis.

The posterior distribution of the parameters of interest, μ_{it}^j, $\beta_k^j (k = 0, 1, \ldots, K)$, and Σ, can be estimated by using the posterior samples generated from MCMC. See Park and Lord (2007) for more details on the implementation of MCMC. Among those parameters, the coefficients controlling slopes of log mean crash frequency on time for treatment and comparison sites before and after installation of the countermeasure, β_3^j, β_4^j, β_5^j, and β_6^j are of particular interest as well as the expected crash frequencies μ_{it}^j. If there are significant effects of a countermeasure on crash reduction, the estimate of the coefficient β_6^j controlling the difference in the slope of log mean crash frequency on time between treatment and comparison sites and also between before and after periods will be negative, which will subsequently suggest that the decrease in crash frequency accelerates after installing the countermeasure at treatment sites. The expected crash frequencies μ_{it}^j can also be used as a basis for quantifying the effect of the countermeasure in crash reduction of the jth crash type. Posterior distributions of the mean expected crash frequencies of crash type j before and after implementation of the countermeasure for each of the treatment group and the comparison group (μ_{TB}^j, μ_{TA}^j, μ_{CB}^j, and μ_{CA}^j), can be obtained as an average of the expected crash frequencies over the appropriate years and the sites. One of the main advantages of using the MCMC methods is that posterior distributions of any functions of model parameters can be easily obtained from the MCMC samples. If there were positive safety effects of countermeasures on the jth crash type, a more pronounced reduction in crashes for the treatment group ($\mu_{TB}^j - \mu_{TA}^j$ or μ_{TA}^j/μ_{TB}^j) is expected to be seen compared to that of the comparison group ($\mu_{CB}^j - \mu_{CA}^j$ or μ_{CA}^j/μ_{CB}^j). As in Pawlovich et al. (2006) and Li et al. (2008), the FB approach here addresses the regression-to-the-mean problem by focusing on estimation of the *expected* number of crashes at the site (given observations at the site) for both before and after periods without directly using the *observed* crash count in the comparison, recognizing that the observed crash count at a site from any 1 year is a noisy measurement of the true long-run mean crash frequency. Note that Lan et al. (2009) and Persaud et al. (2009) compared the predicted crash count obtained by the regression model estimated based on both reference sites and the before period of treated sites (to address the regression-to-the-mean bias) to the *observed* crash count ($y_{i,t}$) for the after period, which is rather similar to the EB approach than to the approach suggested in this paper or those by Pawlovich et al. (2006) and Li et al. (2008). Although the observed after period crash count is an unbiased estimate of the expected after period crash count, it is still an estimate that is subject to uncertainty (not the true value). In our fully Bayesian approach, this uncertainty as well as uncertainty in other model parameters is incorporated into the final safety effectiveness estimate.

2.5. Steps for implementing fully Bayesian before–after evaluations with a comparison group

The steps below can be used for implementing fully Bayesian before–after evaluations with a comparison group in assessing the safety effectiveness of a countermeasure. The ratio μ_{CA}^j/μ_{CB}^j can be explicitly incorporated into evaluation to adjust the prediction of what would have happened at the treatment sites had the countermeasure not been applied for general trends between the before and after periods.

Step 1. Specify the hyperparameter values (c_0, C_0, r_0, R_0) for prior distribution of model parameters.
Step 2. Obtain the draws of model parameters ($\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^J, \Sigma$), and the expected annual crash frequency for each site (i), year (t), and crash type (j), μ^j_{it}, by MCMC.
Step 3. Obtain posterior distributions of crash frequencies of crash type j during the before period for the treatment group (μ^j_{TA}), during the after period for the treatment group (μ^j_{TA}), during the before period for the comparison group (μ^j_{CB}), and during the after period for the comparison group (μ^j_{CA}) by taking an average of the expected crash frequencies over the appropriate years and the sites.
Step 4. Obtain a posterior distribution of the ratios of the expected crash frequencies of crash type j for before and after periods for the comparison group (comparison ratio) by:

$$R^j_C = \frac{\mu^j_{CA}}{\mu^j_{CB}}.$$

Step 5. Obtain a posterior distribution of the predicted crash frequencies of crash type j in the after period for the treatment group, had the countermeasure not been implemented, as:

$$\pi^j = \mu^j_{TB} R^j_C.$$

Step 6. Obtain a posterior distribution of the index of effectiveness (of the countermeasure) for the crashes of type j as

$$\theta^j = \frac{\mu^j_{TA}}{\pi^j} = \frac{\mu^j_{TA}}{(\mu^j_{TB} R^j_C)},$$

and/or the reduction of crashes (of type j) as

$$\delta^j = \pi^j - \mu^j_{TA} = \mu^j_{TB} R^j_C - \mu^j_{TA}.$$

Step 7. Obtain the point estimates for β^j_k ($k=0,\ldots,K$), θ^j, and/or δ^j as the sample means of the corresponding posterior distributions.
Step 8. Obtain the uncertainty estimates for β^j_k ($k=0,\ldots,K$), θ^j, and/or δ^j as the sample standard deviations of the corresponding posterior distributions.
Step 9. Construct the 95% (or 90%) credible intervals of β^j_k ($k=0,\ldots,K$), θ^j, and/or δ^j using the 2.5th (or 5th) percentiles and the 97.5th (or 95th) percentiles of the corresponding posterior distributions. If the credible interval contains the value 1, then no significant effect has been observed. The credible interval placed below 1 (i.e., the upper limit of the interval is less than 1) implies that the countermeasure has a significant positive effect (i.e., a reduction in crashes) on safety. The credible interval placed above 1 (i.e., the lower limit of the interval is greater than 1) implies that the countermeasure has a significant negative effect (i.e., an increase in crashes) on safety.

2.6. *Steps for implementing fully Bayesian before–after evaluations with multiple comparison groups*

It is possible that a countermeasure may be installed to several sites in different years. In such a case, different treatment sites (with different installation years) will have different before and after periods and as a result will need different comparison groups. Hauer (1997) discussed the cases where different entities/treatment sites have different comparison ratios. In the same vein, Step 4–Step 6 in Section 2.5 can be modified when there are multiple (say G) comparison groups as follows:

Step 4. Obtain a posterior distribution of the comparison ratio for crash type j for the gth comparison group by:

$$R^j_{C(g)} = \frac{\mu^j_{CA(g)}}{\mu^j_{CB(g)}}, \quad g = 1, \ldots, G.$$

Step 5. Obtain a posterior distribution of the predicted frequencies of type j crashes that would have occurred without treatment in the after period for the gth treatment group as:

$$\pi^j_{(g)} = \mu^j_{TB(g)} R^j_{C(g)}.$$

Step 6. Obtain a posterior distribution of the index of effectiveness for the crashes of type j as

$$\theta^j = \frac{\sum_{g=1}^{G} \mu^j_{TA(g)}}{\sum_{g=1}^{G} \pi^j_{(g)}} = \frac{\sum_{g=1}^{G} \mu^j_{TA(g)}}{\sum_{g=1}^{G} \{\mu^j_{TB(g)} R^j_{C(g)}\}}$$

and/or the reduction of crashes (of type j) as

$$\delta^j = \sum_{g=1}^{G} \pi^j_{(g)} - \sum_{g=1}^{G} \mu^j_{TA(g)} = \sum_{g=1}^{G} \{\mu^j_{TB(g)} R^j_{C(g)}\} - \sum_{g=1}^{G} \mu^j_{TA(g)}.$$

The remaining steps are the same as those in Section 2.5.

3. Application to Korean expressway crash data: safety evaluation of decreasing the speed limit

The proposed FB method was applied to the crash data obtained from expressways in Korea to assess safety benefits of decreasing the posted speed limit. The crash data consisting of detailed information about all crashes occurred between 1994 and 2006 along with the roadway characteristics, AADT, and the speed limit change information (implementation date and location) were obtained from Korea Expressway Corporation. The standard speed limit for most expressways in Korea is 100 km/h with one exception of 80 km/h (for one expressway, 88-Sun). There have been changes in the speed limit, decreasing to 80 km/h (or to 70 km/h for 88-Sun) in years 1995, 2000, 2001, and 2003 at some locations selectively due to high crash history. The before study period extended from the beginning of the first year for which crash data at the site were available (1994) to the end of the last calendar year before the change in the speed limit. The after study period extended from the beginning of the year after the speed limit was changed to the end of the last year for which crash data at the site were available (2006). It is likely that the regression-to-the-mean bias may be present in the crash data because the high crash frequency was the main reason for implementation of decreasing the posted speed limit for most treated sites. It is well known that this type of bias cannot be accounted for by traditional methods such as naïve before–after evaluations or traditional comparison group method (Hauer, 1997; Persaud and Lyon, 2007). Only the Bayesian methods, either EB or FB, can cope with this bias in the data. Crashes of the following types were analyzed for safety evaluation of decreasing the posted speed limit:

Table 1
Summary of data for 33 treatment sites.

Variable	Mean	Minimum	Maximum
Years before	4.55	1	9
Years after	7.45	3	11
Crashes/km-year			
Total before	1.88	0	6.90
Total after	0.78	0	1.83
Speed before	0.50	0	2.76
Speed after	0.22	0	1.17
A before	0.04	0	0.34
A After	0.02	0	0.10
B before	0.50	0	2.07
B after	0.18	0	0.52
C before	1.35	0	4.48
C after	0.59	0	1.75
AADT before	19,205	3655	29,900
AADT after	22,031	3114	50,379
Length (km)	2.53	0.7	11.7
Number of lanes	2	1	4

Note: AADT represents the one-way flow.

Table 2
Summary of data for 44 comparison sites.

Variable	Mean	Minimum	Maximum
Years before	4.36	1	9
Years after	7.64	3	11
Crashes/km-year			
Total before	1.73	0	5.88
Total after	0.80	0.06	2.57
Speed before	0.49	0	2.94
Speed after	0.22	0	1.28
A before	0.07	0	0.91
A after	0.02	0	0.09
B before	0.56	0	2.78
B after	0.21	0	1.02
C before	1.10	0	4.48
C after	0.57	0	1.75
AADT before	17,855	3655	44,231
AADT after	23,703	3114	58,128
Length (km)	7.98	0.7	28.2
Number of lanes	2.20	1	4

Note: AADT represents the one-way flow.

1. Total: Crashes of all types.
2. Speed: Crashes with “Cause of accident” reported as speed or speed violation.
3. A: Crashes satisfying one of the following conditions: involving at least 3 fatalities, involving at least 20 injuries, involving at least 10 cars, or resulting in at least $10,000 damage in road facilities.
4. B: Crashes dissatisfying all conditions of type A but satisfying one of the following conditions: involving 1–2 fatalities, involving 5–19 injuries, involving 5–19 cars, or resulting in $2500–$10,000 damage in road facilities.
5. C: Crashes dissatisfying all conditions of type A and of type B but satisfying one of the following conditions: involving 1–4 injuries, involving 3–4 cars, or resulting in $300–$2500 damage in road facilities.

For example, if a crash involved 3 fatalities, 6 injuries and 4 cars, it will be classified as Type A crash because one of the conditions for Type A crashes is satisfied.

The before–after evaluation design with four comparison groups (so that each of four treatment groups corresponding to four different implementation years has a separate comparison group) was employed as a study design. The treatment sites for this study are 83.4 km of expressways where the posted speed limit was decreased during 1995 through 2003. The whole contiguous segments of expressway in each direction with the speed limit decreased constitute a treatment site as long as there is no other change in the main roadway characteristics or AADT (for the same year). There were 33 treatment sites for this study. A comparison site was also defined as the untreated contiguous road segments (i.e., continuous segments without any speed limit change) within which the roadway characteristics and AADT (for the same year) do not change. For each treatment site, at least one comparison site was chosen. Those comparison sites were selected from the same expressway and direction (whenever possible) and also by ensuring that the AADTs are in the similar range as that for the corresponding treatment site. The comparison group consists of 44 untreated sites that correspond to 351.1 km of expressways. The data for 33 treatment sites and 44 comparison sites are summarized in Tables 1 and 2, respectively.

3.1. *Fully Bayesian multivariate approach for evaluating the safety effectiveness of speed limit decrease*

Because there were only handful of Type A crashes during the study period, Crashes of Type A and Type B were combined and denoted as A + B. The crashes of four different types, Total crashes, Speed crashes, Type A + B crashes, and Type C crashes, were analyzed simultaneously using the multivariate FB approach to before–after evaluation presented in this paper. The expected number of those crashes per site were fitted by the MVPLN model with a change point in Eqs. (1)–(4) with predictors, Log of Annual Average Daily Traffic (AADT), an indicator function specifying whether a site is a treatment site or a comparison site, time trend (year), an indicator functions specifying whether it belongs to the before or the after period, treatment by time, treatment by implementation date, Number of Lanes by direction (NL), and Log of Site Length (Length). Because the expressways in Korea are standardized in terms of major roadway characteristic variables such as shoulder width and lane width (i.e., the shoulder width is 3 m and the lane width is 3.6 m for almost all segments of expressways), those variables were not included in the model. The median was not included in the model either because all but one expressway (88-Sun) in Korea have the median, and the 88-Sun was the only two-lane (considering both directions) expressway.

The steps for implementing fully Bayesian before–after evaluations with comparison groups presented in the previous section were followed. For the prior distributions of the model parameters, proper but diffuse priors were used to reflect the lack of precise knowledge on the parameters a priori. The inferences on the parameters of interest were made based on the samples from the posterior distribution obtained by the MCMC algorithm implemented in MATLAB as noted in Section 2.4.

Table 3 summarizes the results from the FB analysis based on 1000 posterior samples. Note that the regression coefficient for time, β_3, is negative and significant for all four crash types suggesting that crashes decreased over time in general. The regression coefficient β_6 (controlling the difference in the slope of log crash frequency on time between treatment and comparison sites and between before and after periods) is negative and significant for Total, Speed and Type C crashes, suggesting that the decrease in crash frequency for those crash types accelerated after the speed limit was decreased at treatment sites. The estimated index of effectiveness ($\hat{\theta}$) was obtained by accounting for the changes in unmeasured factors between the before and the after period using the comparison ratio following Steps 4–6 of Section 2.6. The uncertainty estimates for the estimated index of effectiveness, the posterior standard deviation (Std Dev) and 95% credible interval, play the same role as the standard error and the 95% confidence interval in traditional or EB approaches. It can be concluded that there have been statistically significant decreases in Type C crashes

Table 3
Results of FB safety evaluation for decreasing the posted speed limit.

Regression coefficients	Predictors	Crash type			
		Total	Speed	A + B	C
β_0	Intercept	**−4.1727**	**−7.6237**	**−4.1631**	**−5.5150**
β_1	Log(AADT)	**0.4380**	**0.6267**	**0.3151**	**0.5326**
β_2	T	0.1076	−0.0663	−0.0026	0.1506
β_3	t	**−0.1655**	**−0.2231**	**−0.2164**	**−0.1459**
β_4	$(t-t_0)\mathbf{I}[t>t_0]$	0.0423	**0.1115**	0.0090	**0.0504**
β_5	$T\times t$	0.0562	**0.1662**	0.0729	0.0482
β_6	$T(t-t_0)\mathbf{I}[t>t_0]$	**−0.0885**	**−0.2367**	−0.0627	**−0.0918**
β_7	NL	0.1125	0.0455	**0.2080**	0.0566
β_8	Log(Length)	**1.1522**	**1.1887**	**1.2137**	**1.1373**

Index of effectiveness	Total	Speed	A + B	C
$\hat{\theta}$	*0.8553*	0.9123	1.0358	**0.7915**
Std dev	0.0792	0.1610	0.1717	0.0860
95% credible interval	(0.7145, 1.0195)	(0.6298, 1.2556)	(0.7296, 1.4144)	(0.6406, 0.9742)
90% credible interval	(0.7387, 0.9941)	(0.6644, 1.1864)	(0.7770, 1.3366)	(0.6579, 0.9444)
$100(1-\hat{\theta})$: percent reduction	*14.4*	8.8	−3.6	**20.8**

Notes: (1) $\hat{\theta}$ is the estimated index of effectiveness; (2) Std Dev represents the posterior standard deviation for θ; (3) $100(1-\hat{\theta})$ denotes the estimated percent crash reduction; (4) statistically significant results with 95% (90%) probability are shown in bold (in italic).

(20.8%) with 95% probability and in Total crashes (14.4%) with 90% probability.

Table 4 contains the MCMC estimates of the correlation matrix of the latent effects, ***b***, generating the correlation structure in the multivariate crash counts. It can be observed from the table that there is a positive correlation between each of the latent effects in the counts of four crash types.

3.2. Comparison with other methods

For comparison purposes, the naïve before–after evaluation method and the empirical Bayes before–after evaluation method were also applied to the crash data from the same treatment sites. The index of effectiveness estimates by the naïve method was obtained by comparing the total number of crashes during the after period to the total number of crashes during the before period that is adjusted only for the difference in duration between the before and after periods.

For the EB method, it is well-known that developing and calibrating a reliable safety performance function (SPF) based on the valid reference group (that is similar to the treatment group in terms of geometric design, traffic volumes, vehicle fleet, weather, etc.) is crucial in implementation of the method because the EB results could be sensitive to the SPFs (Persaud and Lyon, 2007). Because a reference group consisting of the 44 comparison sites used in the FB analysis was not large enough to develop reliable SPFs, an expanded reference group consisting of 126 untreated sites corresponding to 1185.7 km of expressways was used for the EB analysis. Table 5 contains the summary for 126 reference sites.

The SPFs for different crash types were calibrated by using the negative binomial model given below:

$$E(\mu_{it}) = \exp\{\alpha_t + \beta_1 \log(AADT_{it}) + \beta_2 NL_i + \beta_3 \log(Length_i)\} \quad (5)$$

where μ_{it} is the expected crash frequency at site i in year t, $E(\mu_{it})$ is the mean of the expected crash frequencies in year t for all sites in the imagined reference population of site i, α_t is the parameter for year t accounting for the effect of general trends in safety not related to the AADT, the number of lanes, or the site length, and NL_i and $Length_i$ are the number of lanes (by direction) and the length for the ith site, respectively. The regression parameters α_t, β_1, β_2, β_3, and the negative binomial dispersion parameter were estimated by PROC GENMOD in SAS although those model coefficients are not reported here for space. The steps in Hauer (1997) were followed to obtain the SPF predictions for each year in the before and after periods at each treatment site and the predicted number of crashes during the after period had the speed limit not been lowered, and to obtain the index of effectiveness estimates along with standard errors.

Table 6 summarizes the evaluation results for decreasing the posted speed limit obtained by the naïve analysis and the empirical Bayes analysis, in terms of the percent crash reduction estimates, $100(1-\hat{\theta})$, and the corresponding uncertainty estimates (standard errors). For comparison purposes, the EB evaluation was also conducted by using the same 44 untreated sites used in the FB analysis while recognizing that the accuracy of the SPF coefficients estimated based on such a small reference group may not be high. The standard errors of the estimated negative binomial regression coefficients and dispersion parameters based on the 44 reference sites were almost twice as large compared to those based on the 126 reference sites, which indicates that the resulting SPFs may not be very reliable. For fully Bayesian approaches, the crash reduction estimates from both univariate Poisson-lognormal (Univariate) and multivariate Poisson-lognormal (Multivariate) approaches are included in the table along with the corresponding

Table 4
Posterior means of the correlation matrix of the latent effects.

	Total	Speed	A + B	C
Total	1.0000			
Speed	0.6109	1.0000		
A + B	0.4893	0.4428	1.0000	
C	0.6457	0.6032	0.3705	1.0000

Table 5
Summary of data for 126 reference sites.

Variable	Mean	Minimum	Maximum
Years	13	13	13
Crashes/km-year			
Total	1.33	0.08	3.53
Speed	0.32	0	1.81
A	0.03	0	0.13
B	0.38	0	1.12
C	0.92	0.07	2.58
AADT	29,660	3316	91,739
Length (km)	9.41	0.7	28.2
Number of lanes	2.58	1	4

Table 6
Comparison of safety effectiveness estimates for decreasing the posted speed limit obtained by different before–after evaluation approaches.

Approach	Percent crash reduction (uncertainty estimate)			
	Total crash	Speed related crash	Type A or B crash (A + B)	Type C crash
Naïve	**61% (3%)**	**61% (6%)**	**60% (6%)**	**61% (4%)**
EB				
With 126 reference sites	**14% (6%)**	**19% (11%)**	−4% (11%)	**20% (7%)**
With 44 reference sites	4% (7%)	5% (13%)	−12% (12%)	8% (8%)
FB				
Univariate (DIC = 4875.9)	*14% (9%)*	5% (17%)	−6% (18%)	**20% (9%)**
Multivariate (DIC = 4280.2)	*14% (8%)*	9% (16%)	−4% (16%)	**21% (8%)**

Note: Statistically significant results with 95% (90%) confidence/probability are shown in bold (in italic).

posterior standard deviations. The deviance information criterion (DIC) (see, e.g., Spiegelhalter et al., 2002) is also provided for each of FB Univariate and FB Multivariate in Table 6. For FB Univariate, DIC was computed as the sum of four univariate DICs obtained under the Poisson-lognormal model for each crash type, following El-Basyouny and Sayed (2009). Note that DIC for FB Multivariate is substantially smaller than that for FB Univariate, which suggests that the MVPLN model provides a superior fit over the univariate Poisson-lognormal model as in El-Basyouny and Sayed (2009).

The following points can be made from Table 6. The percent crash reduction estimated by naïve before–after evaluation is the largest among all the estimates compared and the corresponding standard errors are smallest. It should be noted, however, that the estimates from the naïve method are subject to unknown biases and their accuracy is questionable because none of the differences between before and after periods in traffic volumes, weather, crash reporting practices, and vehicle fleet as well as the regression-to-the-mean bias are accounted for by the naïve before–after evaluation method.

For Type C crashes, there seems to be a statistically significant crash reduction at 95% confidence level (or with 95% probability) based on both the EB with 126 reference sites and FB approaches. The reduction amounts from both approaches are comparable. For Total crashes, the crash reduction was significant at 95% confidence level for the EB with 126 reference sites approach and with 90% probability for the FB approaches. The reduction amounts for Total crashes from both approaches are again comparable. The percent change in Type A + B crashes is not statistically significant for any of EB or FB approaches. The biggest difference between EB with 126 reference sites and FB crash reduction estimates is observed for Speed related crashes. Speed related crashes had the smallest number of crashes among the four crash types considered in this study, and the result seems to be more sensitive to the form of the SPFs and underlying distributional assumptions. In general, the uncertainty estimates for FB estimates are larger than those for EB estimates, which could be the result of incorporating uncertainty in estimated parameters of the crash prediction models into the final safety effectiveness estimates of FB.

The results from EB with 44 reference sites are noticeably different from those of all other approaches, which is deemed to have been caused by inaccurate estimation of the SPF coefficients. Note that EB does not have a mechanism to reflect the number of reference sites or standard errors of the SPF coefficients in the final safety effectiveness estimate because EB uses only the point estimates of SPF coefficients and dispersion parameter (not their standard errors) obtained from reference sites in annual SPF predictions at treatment sites (and the crash reduction estimate). In other words, regardless of number of reference sites, the SPF coefficient estimates obtained from those sites will be treated as if they were true SPF coefficients (i.e., as if they were equally accurate). If the SPF coefficients are accurately estimated based on a fairly large reference group, ignoring the uncertainty associated with those coefficient estimates may not practically matter and EB and FB may lead to similar crash reduction estimates and uncertainty. However, if SPFs are estimated based on a small number of reference sites (like 44 reference sites here), the chance to obtain inaccurate SPF coefficient estimates will increase and ignoring uncertainty associated with those coefficient estimates would make the final safety effectiveness estimate (crash reduction estimate) look much more precise than it actually is. Note that the uncertainty estimates for EB with 44 reference sites are given to be smaller (which seem to have been underestimated significantly in this case) than those of FB. Although Persaud et al. (2009) reported that two approaches (EB and FB) lead to comparable results (or in some cases the FB method results in the smaller uncertainty estimates in contrast to the indication by Carriquiry and Pawlovich, 2004), their comparisons were made only for the cases where a fairly large reference group was available for estimating a safety performance function to apply the EB approach. Also, as noted earlier, the FB safety effectiveness estimate of Persaud et al. (2009) is based on the framework that is close to EB in that the observed crashes are used in the numerator of the expected crash reduction rate (CRR). The impact of ignoring uncertainty in the estimated SPFs on the EB crash reduction estimate will show up when the SPF is poorly estimated based on a small reference group (as shown in Table 6). Persaud et al. (2009) also pointed out that even with the relatively small comparison group, the FB still can provide fairly good results.

The uncertainty estimates for the FB multivariate approach are slightly smaller than the FB univariate approach. This seems to be a natural consequence of accounting for correlations in the crashes of different types by a multivariate approach, which leads to more precise estimation of the expected crash frequencies and percent crash reduction. Other researchers including Aguero-Valverde and Jovanis (2009) and El-Basyouny and Sayed (2009) demonstrated that the multivariate Poisson-lognormal approach incorporating the correlations across crashes of different severity levels leads to significantly higher precision of expected crash frequencies compared to the univariate Poisson-lognormal models. Although for this data the correlation in Table 4 was moderate rather than being high and the difference between the univariate FB and multivariate FB was not big, for the multivariate data with higher correlation the difference in precision between univariate FB and multivariate FB estimates is expected to be bigger.

4. Summary and conclusions

Although the fully Bayesian (FB) methods in before–after safety evaluation have not been explored by many transportation researchers yet, they are emerging as the state-of-the-art methods that can overcome the limitations of the previous empirical Bayes (EB) methods such as the heavy dependence on the SPFs and potential underestimation of true uncertainties in the EB safety effectiveness estimates. In this paper, a new fully Bayesian multivariate approach tailored to the before–after studies with a comparison group/comparison groups was presented along with a step-by-step implementation procedure.

Similarities and differences between fully Bayesian and empirical Bayes methods discussed in Carriquiry and Pawlovich (2004) are still applicable here. The FB approaches implemented via MCMC can provide entire distributions for model parameters, expected crashes, index of effectiveness, or any other quantity of interest without additional computational cost. In the FB approach, uncertainty in regression parameters is propagated throughout the model and carries through to the final safety effectiveness estimate. On the other hand, in the EB approach the estimated regression model parameters of the SPF are treated as if they were the true parameters once they are estimated, and associated uncertainties (standard errors) are not incorporated into the final safety effectiveness estimate. If the regression model parameters of the SPF are estimated accurately based on a fairly large reference group, the uncertainty associated with estimation of regression parameters may not be big and it may not make a noticeable difference whether or not to include parameter uncertainty into final safety effectiveness estimate. In that case, the EB results may be similar to the FB results. However, if the size of the reference group is small and regression parameters of SPFs are subject to a large uncertainty, ignoring this uncertainty can make a big difference in the final safety effectiveness estimate and true uncertainty in the final safety effectiveness estimate by EB will be underestimated.

The fully Bayesian multivariate before–after evaluation approach presented in this paper has an additional advantage that it leads to more precise crash predictions and safety effectiveness estimates than those from the univariate approaches. This implies that the safety effectiveness estimates of countermeasures from the multivariate FB will genuinely be more precise than those from the traditional methods, EB approach, or FB univariate approach (i.e., additional precision is gained from accounting for correlation in the multivariate crash counts not from underestimating the true uncertainties in the estimates).

The FB (both multivariate and univariate) methods for before–after evaluation as well as the EB methods (with 126 reference sites and with 44 reference sites) and the traditional naïve before–after evaluation method were applied to the multivariate crash counts obtained from expressways in Korea for 13 years to assess the safety effectiveness of decreasing the posted speed limit. The analysis suggested that the naïve before–after evaluation method significantly overestimates the safety effectiveness of the speed limit decrease, which is a natural consequence of accounting for none of the regression-to-the-mean bias, traffic volume differences, and the general time trends between the before and after periods. The results from EB with 126 reference sites and FB methods were in general comparable although the uncertainty estimates from FB were larger than those from the EB approaches, which could be the result of incorporating parameter uncertainty of the crash prediction models into the final safety effectiveness estimates of FB. The results from EB with 44 reference sites were noticeably different from those of all other approaches (including EB with 126 reference sites), which seems to have been caused by poorly estimated SPFs. Although the estimated negative binomial regression coefficients and dispersion parameter based on 44 reference sites have considerably larger uncertainty (standard errors are almost twice as large) than those based on 126 reference sites, this increased uncertainty is not accounted for in the final safety effectiveness estimate and leads to a noticeable difference from other approaches. The FB multivariate approach was also compared to the corresponding FB univariate approach to make a more fair comparison between multivariate and univariate approaches. The uncertainty estimates from the FB multivariate approach were slightly smaller than those from the FB univariate approach, which suggests that the multivariate approach leads to (genuinely) more precise safety effectiveness estimates compared to the corresponding univariate approach. As the correlation across crashes of different types or severities increases, the gain in precision of the multivariate approach is expected to be larger.

As mentioned in Persaud and Lyon (2007), the FB approach has also some obstacles such that the methodology is quite complex and the implementation of MCMC is not straightforward. It is hoped that this paper will contribute to resolving these issues by providing a clear step-by-step FB implementation procedure as well as offering an additional tool for implementing FB methods on multivariate crash count data in before–after safety evaluation studies.

Acknowledgments

The authors thank two anonymous referees for their constructive comments, which led to substantial improvements in the manuscript.

References

Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson log-normal models for crash severity modeling and site ranking. In: Transportation Research Board 88th Annual Meeting.

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis, 2nd ed. Wiley, New York.

Aul, N., Davis, G., 2006. Use of propensity score matching method and hybrid Bayesian method to estimate crash modification factors of signal installation. Transportation Research Record 1950, 17–23.

Campbell, D.T., Stanley, J.C., 1963. Experimental and Quasi-experimental Designs for Research on Teaching. In: Gage, N.L. (Ed.), Handbook of Research on Teaching. Rand McNally, Chicago.

Campbell, D.T., Ross, H.L., 1968. The Connecticut crackdown on speeding: time-series data in quasi-experimental analysis. Law and Society Review 3, 33–54.

Carriquiry, A., Pawlovich, M.D., 2004. From empirical Bayes to full Bayes: methods for analyzing traffic safety data. http://www.iowadot.gov/crashanalysis/pdfs/eb_fb_comparison_whitepaper_october2004.pdf (accessed May 7th, 2009).

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. American Statistician 49, 327–335.

Davis, G.A., Yang, S., 2001. Bayesian identification of high-risk intersections for older drivers via Gibbs sampling. Transportation Research Record 1746, 84–89.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41 (4), 820–828.

Fitzpatrick, K., Park, E.S., 2009. Safety effectiveness of HAWK pedestrian treatment. Transportation Research Record, No. 2140.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall.

Harwood, D.W., Bauer, K.M., Potts, I.B., Torbic, D.J., Richard, K.R., Kohlman Rabbani, E.R., Hauer, E., Elefteriadou, L., 2002. Safety effectiveness of intersection left- and right-turn lanes, Report No. FHWA-RD-02-089. Federal Highway Administration (FHWA), Washington, D.C.

Hauer, E., 1986. On the estimation of the expected number of accidents. Accident Analysis and Prevention 18, 1–12.

Hauer, E., 1997. Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Pergamon Press, Elsevier Science, Ltd., Oxford, United Kingdom.

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S., 2002. Estimating safety by the empirical bays method: a tutorial. Transportation Research Record 1784, 126–131.

Lan, B., Persaud, B., Lyon, C., Bhim, R., 2009. Validation of a Full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions. Accident Analysis & Prevention, doi:10.1016/j.aap.2009.02.010.

Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008. The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. Accident Analysis & Prevention 40, 1531–1542.

Liu, J.S., 2001. Monte Carlo Strategies in Scientific Computing. Springer, New York.

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.

Park, E.S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019, 1–6.

Patel, R.B., Council, F.M., Griffith, M.S., 2007. Estimating safety benefits of shoulder rumble strips on two-lane rural highways in Minnesota: empirical Bayes observational before-and-after study. Transportation Research Record 2019, 205–211.

Pawlovich, M.D., Wen, L., Carriquiry, A., Welch, T., 2006. Iowa's experience with road diet measures: use of Bayesian approach to assess impacts on crash frequencies and crash rates. Transportation Research Record 1953, 163–171.

Persaud, B.N., Hauer, E., 1984. Comparison of two methods for debiasing before-and-after accident studies. Transportation Research Record 975, 43–49.
Persaud, B.N., 1988. Do traffic signals affect safety? Some methodological issues. Transportation Research Record 1185, 37–47.
Persaud, B.N., Hauer, E., Retting, R., Vallurupalli, R., Mucsi, K., 1997. Crash reductions related to traffic signal removal in Philadelphia. Accident Analysis & Prevention 29, 803–810.
Persaud, B.N., Retting, R.A., Lyon, C.A., 2004. Crash reduction following installation of centerline rumble strips on rural two-lane roads. Accident Analysis & Prevention 36, 1073–1079.
Persaud, B., Lan, B., Lyon, C., Ghim, R., 2009. Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. Accident Analysis & Prevention, doi:10.1016/j.aap.2009.06.028.
Persaud, B., Lyon, C., 2007. Empirical Bayes before–after safety studies: lessons learned from two decades of experience and future directions. Accident Analysis & Prevention 39, 546–555.
Schluter, 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.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B 64, 583–639.
Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.