Accident Analysis and Prevention 41 (2009) 574–580

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Validation of a Full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions

Bo Lan[a,*], Bhagwant Persaud[a,1], Craig Lyon[a,2], Ravi Bhim[b,3]

[a] *Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada*
[b] *iTRANS Consulting Inc., 100 York Boulevard, Suite 300, Richmond Hill, Ontario L4B 1J8, Canada*

ARTICLE INFO

Article history:
Received 11 April 2008
Received in revised form 12 February 2009
Accepted 15 February 2009

Keywords:
Bayesian
DIC
Road safety
Treatment effect
Before–after evaluations
WinBUGS

ABSTRACT

The objective of the study on which the paper is based was to explore the application of fully Bayesian methods for before–after road safety studies. Several variations of the methodology were evaluated with a simulated dataset in which hypothetical treatments with no safety effect were randomly assigned to high accident locations to mimic the common site selection process in road jurisdictions. It was confirmed that the fully Bayesian method by estimating no safety effect can account for the regression-to-the-mean that results from this biased site selection process. The fully Bayesian method was then applied to California rural intersection data to evaluate the safety effect of conversion from stop to signalized control. The results were then compared with those from the empirical Bayesian method, currently the accepted approach for conducting unbiased before–after evaluations. This comparison was generally favorable in that FB can provide similar results as EB.

1. Introduction

One of the most important tasks for road safety analysts is the before–after evaluation of treatment effects. For the past two decades, the empirical Bayesian (EB) method has been used successfully to perform this evaluation. A recent paper by Persaud and Lyon (2007) has summarized experience to date with this methodology. With the availability of the software package WinBUGS (Spiegelhalter et al., 2003; Lunn et al., 2000), interest in the fully Bayesian (FB) approach for treatment effect analysis has increased significantly.

A Bayesian calculation combines prior information and current information to derive an estimate for the expected safety of a site that is being evaluated. In the context of accident analysis, the prior information is the expected accident frequency from a group of similar sites and the current information is the site-specific observed accident frequency. Empirical Bayes and Full Bayes are two related approaches to combine prior and current information.

In the empirical Bayes approach, the prior information comes from using a reference group of sites similar to those under evaluation to calculate a sample mean and variance or from a calibrated safety performance function that relates the crash frequency of the reference sites to their characteristics. The point estimates of the expected mean and the variance are then combined with the site-specific crash count to obtain an improved estimate of a site's long-term expected crash frequency.

In the Full Bayes approach, the procedure is integrated. It directly combines the information at reference sites and before period information at treated sites to develop the model for obtaining the estimate of the long-term expected crash frequency. In addition, instead of a point estimate of the expected mean and its variance, a distribution of expected crash frequency is obtained, enabling more precise statements of uncertainty in the results.

It is believed the FB approach has a number of advantages over the EB method. For example, the FB approach is believed to better account for uncertainty in data used, to require less data, and to provide more detailed causal inferences and more flexibility in selecting accident count distributions. In the latter regard, the FB method can accommodate distributions such as hierarchical Poisson-Gamma distribution, Poisson-LogNormal distribution, etc. (Carriquiry and Pawlovich, 2005; Pawlovich et al., 2006; Miaou and Lord, 2003; Wang et al., 2006; Congdon, 2003, 2001, 2005), while EB approach usually relies on the assumption of a negative binomial (NB) distribution of accident counts to facilitate the use of the NB dispersion parameter directly in weighting the prior and posterior information in the estimation process.

* Corresponding author. Tel.: +1 416 979 5191; fax: +1 416 979 5122.
E-mail addresses: blan@ryerson.ca, bolancan@hotmail.com (B. Lan), bpersaud@ryerson.ca (B. Persaud), craig.lyon@rogers.com (C. Lyon), rbhim@itransconsulting.com (R. Bhim).
[1] Tel.: +1 416 979 5000x6464; fax: +1 416 979 5122.
[2] Tel.: +1 416 979 5345; fax: +1 416 979 5122.
[3] Tel.: +1 905 882 4100; fax: +1 905 882 1557.

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.02.010

One recent study of treatment effects using the FB method was conducted by Pawlovich et al. (2006) to evaluate the installation of two-way left turn lanes. This study introduced treatment effect coefficients into model and employed matched pairs or treated and untreated sites. The authors developed a crash rate (crashes per unit of exposure) model to estimate expected crashes in the after period for both the treated sites and the matched comparison sites. A 25% reduction in crash frequency per mile averaged over several sites was found in their study, which is close to the 24% reduction obtained from the Naïve before–after method. The approach employed by Pawlovich et al. (2006) uses principles that are similar to those of the conventional comparison group (C-G) study (Hauer, 1997).

For this paper a before–after FB study was conducted, similar in principle to Aul and Davis (2006). The before–after FB approach is similar to the EB approach in that untreated reference group data are used to make inferences and to account for possible effects unrelated to the treatment. This FB method also includes data on the treated sites in the before period to develop inferential models. Then the developed model is used to predict the crashes for treatment sites in the after period, had the treatment not been implemented. On the other hand, the EB approach only uses data from reference sites for this purpose.

This paper discusses how to use the Bayesian framework, in conjunction with Markov Chain Monte Carlo (MCMC) methods, to derive the posterior distribution of treatment effects in models. The objective of the study on which the paper is based was to explore the application of fully Bayesian methods for before–after road safety studies. The results of a three-part study are presented. The first part of the study evaluated variations of the before–after FB method by analyzing a hypothetical treatment with no effect at stop-controlled intersections whose safety performance was simulated with the use of WinBUGS. The second part applied the FB method to conduct an analysis of the effects of converting rural intersections from stop to signal control, while the third part compares the results from FB method with those obtained from an EB evaluation.

2. Before–after FB method methodology

Crash counts are typically time series data across years and therefore can be represented by the following simple model structure (Congdon, 2001):

$$\textit{Observed series} = \textit{trend} + \textit{regression term} + \textit{random effects} \qquad (1)$$

where the "regression term" is of the same form as safety performance functions (SPFs) used in EB studies (Hauer et al., 2002; Vogt, 1999; Persaud et al., 2002), and "random effects" accounts for latent variables across the sites. There are two basic forms of regression models with random effects that are considered—the Poisson-Lognormal model and the Poisson-Gamma model. The two models can be described as follows:

$$Y_{i,t} \sim Poisson(\varepsilon_i \lambda_{i,t}),$$

where $Y_{i,t}$ = observed number of crashes at site i in year t, $\lambda_{i,t}$ = expected number of crashes at intersection i in year t, ε_i = multiplicative random effect at site i.

For the Poisson-Lognormal model, $\varepsilon_i \sim \mathrm{Log}\,N(0, \sigma^2)$.

For the *Poisson-Gamma model*, $\varepsilon_i \sim Gamma(\varphi, (1/\varphi))$ with the mean having a value of 1 and where φ is the dispersion parameter (for the gamma distribution, more generally, $\varepsilon \sim gamma(a, b)$, with mean and variance defined as $E(\varepsilon) = ab$ and $\mathrm{Var}(\varepsilon) = ab^2$, respectively. When $E(\varepsilon) = 1$ and $\mathrm{Var}(\varepsilon) = 1/\varphi$ (i.e., when $a = \varphi$ and $b = 1/\varphi$), the Poisson-Gamma function becomes NB distribution (Lord, 2006; Cameron and Trivedi, 1998).

Given the observed crash count $y_{i,t}$ in the "after" period at treated site i, the major task of treatment effect analysis is to compare this count with the level of safety, $\pi_{i,t}$, that would have been expected in the after period had the treatment not been implemented. The procedure for predicting the expected number of crashes $\pi_{i,t}$ in the after period without treatment includes two steps (Aul and Davis, 2006):

Step 1: Assuming $Y_{i,t} \sim Poisson(\varepsilon_i \lambda_{i,t})$, posterior distributions of the parameters (α, β_1, β_2, β_3) are calibrated by MCMC methods using the data from reference sites and the before period of treated sites.
Step 2: The corresponding expected total crashes $\varepsilon_i \lambda_{i,t}$ without treatment can then be obtained and used as an estimate of $\pi_{i,t}$, given the traffic volumes at each treated site in the after period. The change in safety is the difference between the predicted $\pi_{i,t}$ in the after period without treatment and the actual safety, $y_{i,t}$, usually the observed count of crashes in the same period with the treatment in place. The treatment effects can then be calculated, either in terms of the actual change in safety or in terms of a percentage change.

3. Evaluation of the FB method

Variations of the before–after FB method were validated using simulated data for stop-controlled intersections for a hypothetical treatment known to have no effect. In this investigation, if any variation estimates a safety effect of zero then the FB methodology is considered validated.

3.1. Simulated data

In deriving the simulated data, it was assumed, as is common, that the crash count over "similar" sites follows a negative binomial distribution (NBD). The NBD may be derived by "heterogenous Poisson sampling", which assumes that the crash count Y_i at a site over time is Poisson distributed with unknown mean λ_i per unit of time at site i and that these means λ_i follow a Gamma distribution over similar sites, such that

$$E(Y) = E(\lambda), \text{ and}$$

$$Var(Y) = E(\lambda) + \frac{E^2(\lambda)}{\varphi} \qquad (2)$$

where φ is the dispersion parameter of the NBD.

The data used to evaluate the FB methods were generated from a Poisson-Gamma distribution (Lord, 2006). The simulation process to derive the rural stop-controlled intersection dataset used for the evaluation is as follows:

Step 1: For each of 6 years for one hypothetical intersection, randomly generate average daily entering traffic volumes (AADT) on the major road in the range from 5000 to 40,000 with random yearly variation within 5%. This was such that most traffic volumes would be around a mean AADT value of 20,000, which is typical of traffic volumes entering a rural stop-controlled intersection from the major road. For convenience, it is assumed that there is no trend in traffic volumes.
Step 2: For the intersection, randomly generate traffic volumes on the minor road in the range of 500 to 4000 AADT with random yearly variation within 5% over the 6 years. For convenience, it is assumed that these volumes are not correlated with the major road volumes.
Step 3: Input safety performance function (SPF) parameters. These were developed for rural stop-controlled intersections from Cali-