Accident Analysis and Prevention 40 (2008) 1531–1542

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.

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 $y_{it} \sim \text{Poisson}(\theta_{it})$.