# A comparison of statistical methods in interrupted time series analysis to estimate an intervention effect

Wang[a,b], J.J.J., Walter[c], S., Grzebieta[a], R. & Olivier[b], J.

[a] Transport and Road Safety, University of New South Wales, Australia , [b] School of Mathematics and Statistics, University of New South Wales, Australia, [c] Centre for Health Systems and Safety Research, University of New South Wales, Australia

## Abstract

Since the introduction of mandatory helmet legislation (MHL) in Australia, debate on the effect of MHL on cyclist head injuries has been ongoing. The debate sometimes revolves around the statistical methodology used to assess intervention effectiveness. Supporters of rescinding the MHL thereby encouraging cyclists to ride without helmets, regularly dismiss statistical evaluations as being flawed for various reasons. In a more general context, researchers want to estimate whether and how a policy intervention changed an outcome of interest. Quasi-experimental interrupted time series (ITS) is the most appropriate design to evaluate the longitudinal effects of policy interventions and segmented regression analysis is often used as a powerful statistical method for ITS. Recent research has employed a log-linear regression model for the hospital admission counts of head and limb injuries from New South Wales, Australia, from a 36 month period centred at the time of legislation. Estimation of the model was done using a frequentist approach. In this paper, we re-analyse this data using empirical Bayes and full Bayesian methods, since the use of these methods has become popular in road safety studies. In particular, we show how a full Bayesian method can be readily implemented in WinBUGS software. We discuss the advantages and disadvantages of each method and describe and compare the different estimation methods in terms of parameter estimates. The results show that all three estimation methods give consistent conclusions regarding the positive effect of compulsory helmet wearing on cyclist head injuries in New South Wales.

## Introduction

For the before-after evaluation of the effects of implemented policy interventions, interrupted time series (ITS) is the strongest quasi-experimental approach (Wagner *et al.*, 2002). An appropriate statistical method for analysing effects of interventions in ITS data is segmented regression analysis. In a segmented regression analysis or a change-point model, each segment of the time series is allowed to exhibit different levels and trends. In other words, the segmented regression model allows the outcome of interest to evolve differently over the time periods before and after the intervention. A change in level of the outcome after the intervention may constitute an abrupt intervention effect due to policy implementation. Segmented regression analysis also allows analysts to control for variables other than the intervention that can potentially cause the change in level and/or trend of the outcome of interest. A range of statistical approaches can be applied to estimate pre-intervention level and trend and estimate post-intervention changes in level and trend. Despite the flexibility of a change-point model for analysing before-after studies, it has not yet received much attention in road safety literature.

The use of general Bayesian techniques in road safety studies has become very popular over the past two decades, which began with the introduction of empirical Bayes (EB) by Persaud and Hauer (1984). The EB approach has been used widely and successfully for before-after evaluation of the effects of implemented treatments. Typically, a Poisson or negative binomial crash model is used for the untreated reference site and it is combined with pre-intervention data at the treated site to estimate the expected number of crashes that would have occurred without intervention. This estimate is then compared with the observed crash count after intervention to assess the effect of policy implementation. This approach effectively copes with several threats to validity such as the regression-to-the-mean bias. Extensive literature on EB methods are available (Hauer, 1986; Persaud, 1988; Pendelton *et al.*, 1991; Hauer, 1997) and these authors argue that EB methods have advantages over the classical approach. Beginning with computer and methodological advances in the early 1990s, the feasibility of implementing full Bayesian (FB) approaches has been explored by road safety researchers. The FB approach has been suggested as a useful alternative to the EB approach due to several advantages such as the ability to fully account for uncertainty in data. However, there have not been many applications of the FB methods in before-after studies mainly due to complexity in modelling and estimation. Out of the few authors who have used FB methods, Carriquiry and Pawlovich (2004) demonstrated the feasibility of implementing a FB framework and compared it with the EB approach using a hypothetical example. Pawlovich *et al.* (2006) introduced a Bayesian hierarchical Poisson regression model with a change point to assess whether road diets appear to result in crash reduction. Persaud *et al.* (2010) compared the EB and FB using two empirical applications and concluded that the two approaches provide similar results. However, the FB approach in Persaud *et al.* (2010) (and Lan *et al*., 2009) was close to the EB framework. Park *et al*. (2010) developed a FB multivariate approach to model crash counts for a before-after evaluation.

In this paper, we perform segmented regression analysis for a particular ITS data set and estimate the model using FB and EB approaches. We compare the estimation procedures as well as empirical results with a previous study that used only classical frequentist method of estimation. The particular ITS data we consider is the hospital admission counts of head and arm injuries of cyclists from New South Wales (NSW), Australia. Australia was the first country to introduce mandatory cycle helmet legislation and debate on the efficacy of the legislation on reducing cyclist head injuries has been ongoing. In NSW, the study by Walter *et al.* (2011) was the first to assess the effect of mandatory helmet legislation on cyclist head injuries employing an ITS approach. More specifically, a negative binomial regression model was estimated to identify differential changes in head and limb injury rates at the time of legislation and model inference was performed using classical maximum likelihood method. The authors identified evidence of a positive effect of mandatory helmet legislation on cyclist head injuries at a population level and concluded that repealing the law cannot be justified. Rissel (2012) however, made some criticisms about the data and methodology used in the study, thus questioning the validity of the study and suggested re-analysis of the data. Walter *et al.* (2013) responded to those criticisms and to which demonstrated that the original analysis is robust. In this paper, we aim to compare the FB and EB methods for assessment of road safety interventions using the NSW cyclist hospitalisation data. In addition to practical benefits and disadvantages of each approach, we will also consider the implications for the conclusions drawn from the original frequentist analysis.

## Methodology

### *Data*

The data we used in this study is taken from Walter *et al.* (2013). Hospital admissions were recorded by the NSW Admitted Patients Data Collection where external causes and diagnoses were classified exclusively using the International Classification of Diseases, 9$^{th}$ Revision, Clinical Modification (ICD 9-CM). Head injury admissions were defined as all injuries to the skull, face and scalp while arm injuries were defined as all injuries to the shoulder girdle, arm, wrist and hand. In our study, eighteen months of pre- and post-law head and arm hospital admissions were included. Since the mandatory helmet law in NSW was introduced at separate times for adults and children, the 36 month analysis period was from July 1989 to June 1992 for adults and January 1990 to December 1992 for children. To adjust for cyclical effects, monthly counts of hospital admissions were adjusted using the X11 method separately for adults and children and then combined into a single time series. The X11 method (Shiskin *et al.,* 1967) uses a series of moving averages and smoothing calculations to decompose the original series into trend, seasonal and irregular components.

***Model specification***

Injury counts are usually modelled as Poisson or negative binomial random variables. We adopt the same model as the one described in Walter *et al.* (2011) where the log-mean is expressed as a function of time period, injury type and law indicator. However to facilitate Bayesian estimation, we formulate a hierarchical Poisson-Gamma mixture model where the Poisson mean is modelled as a Gamma random variable. This additional parameter accommodates extra variability in the data and effectively overcomes the problem with overdispersion. The negative binomial log-linear regression of admission counts is as follows:

$$\begin{aligned} Y_i \mid \eta_i &\sim Poisson\,(\eta_i \mu_i) \\ \eta_i &\sim Gamma(\alpha, \alpha) \\ \log(\mu_i) &= \beta_0 + \beta_1 TIME + \beta_2 INJURY + \beta_3 LAW + \beta_4 TIME \times INJURY + \beta_5 TIME \times LAW \\ &\quad + \beta_6 INJURY \times LAW + \beta_7 TIME \times INJURY \times LAW + \log(\text{exposure}), \qquad (1) \end{aligned}$$

where $Y_i$ are the admission counts of head and arm injuries. In this approach, the mean of the Poisson distribution is written as a product of two parameters in which the first parameter $\eta_i$ is a random variable drawn from a Gamma distribution with mean 1. This formulation is convenient since we can explicitly model the mean of the negative binomial distribution $\mu_i$. Arm injury counts are used as a comparative time series which acts as a 'proxy' for the unmeasured time-dependent confounding since the cycling arm injuries are subjected to the same threats to internal validity but are unaffected by the intervention. Thus any changes in level and/or slope after the intervention is an adjusted effect of the intervention after accounting for these threats. More information on choosing appropriate comparative time series can be found in Olivier *et al.* (2013). Three covariates, TIME, INJURY and LAW are included in the above model. TIME is treated as a continuous variable which represents the 18 monthly intervals pre-law and 18 monthly intervals post-law. The indicator variable INJURY takes a value of one for a head injury and zero for an arm injury. The variable LAW has a value of zero prior to legislation and one post-legislation. This model is able to capture temporal trend in injury counts and the interaction term between TIME and INJURY allows for different time trends for arm and head injuries. Any difference in the post-law trend and the pre-law trend is captured in the interaction between TIME and LAW. The interaction between INJURY and LAW indicates any differential changes in head injuries as compared to arm injuries. Finally, the three-way interaction accounts for the possible difference in the rate of change of head and arm injuries between pre- and post-law time periods.

***Full Bayesian Estimation***

The basic concept of a Bayesian estimation procedure involves combining both the observed data information and some prior belief to generate the posterior distribution of the unknown parameters. Prior information is incorporated into the model through the use of a prior distribution, which is a distribution of likely values for a parameter. In the absence of specific prior information on a parameter, a non-informative prior distribution with large variance is usually adopted. As compared to the classical maximum likelihood approach, the FB approach offers a number of advantages. Firstly, while the maximum likelihood methods depend on their asymptotic properties, the small sample properties of Bayesian estimation may allow estimation of models with smaller sample sizes. For a large sample size, Bayes estimators are asymptotically equivalent to maximum likelihood estimators under appropriate regularity conditions. Secondly, the ability to include prior knowledge on the values of parameters in the model before observing any data appears attractive when the investigator wants to utilise findings from previous or similar studies. Thirdly, a Bayesian approach allows one to specify very complex model forms where the likelihood function is intractable, particularly in the context of hierarchical models. Lastly, in the classical maximum likelihood approach, model parameters are considered as fixed quantities and the emphasis is on obtaining point estimators for the parameters and their standard errors. The Bayesian approach on the other hand, is able to provide the entire posterior distributions of outcomes based on which inference can be made. For example, a common Bayes estimator for a parameter is the mean of its posterior distribution. Hence in this respect, a Bayesian approach enables representation and takes fuller account of the uncertainties related to models and parameter values.

A major limitation for implementing of Bayesian approaches in the past was that the posterior distributions were analytically tractable for only a small number of simple models. In most cases, the posterior distribution does not have a closed form expression and requires the evaluation of high-dimensional integrals. Simulation-based MCMC methods became widely known among statisticians since the early 1990's and have now become the routine tool for Bayesian computation for a wide range of complicated statistical models. Due to the advancement of computational power, the emergence of the statistical software WinBUGS (Lunn *et al*. 2000) enables a more straightforward and feasible implementation of the Bayesian methodology. WinBUGS adopts the Gibbs sampler to simulate realisations of a Markov chain whose limiting distribution is the posterior distribution of the parameters. This consists of sampling each parameter in turn from their full conditional posterior distributions. WinBUGS has a system for choosing the best sampling method to draw from these univariate distributions. The system first checks for conjugacy, in which case the full conditional distribution reduces analytically to a well-known distribution and direct sampling using standard algorithms can be applied. If conjugacy is not detected, but instead the density is log-concave, then the adaptive rejection sampling algorithm is employed. If the target density is not log-concave, then a Metropolis-Hastings algorithm will be used.

We implement the FB approach in WinBUGS. To complete the Bayesian estimation procedure, we need to specify a prior distribution for each of the model parameters. For the regression coefficient parameters, we assign a normal prior with mean 0 and a large variance, that is

$$\beta_i \sim Normal\,(0, 1000), \qquad\qquad i = 0, 1, \ldots, 7,$$

to reflect *a priori* none of variables in the model may be associated with the injury count. We assign a uniform prior to the dispersion parameter α of the Poisson distribution, that is

$$\alpha \sim Uniform\,(0.5, 200)$$

The hyper-parameters are chosen to reflect ignorance about possible values of parameters. We ran three parallel Markov chains initiated from different starting values for 75,000 iterations. The initial 20,000 iterations are discarded as the burn-in period. We then take simulated values every $20^{th}$ iteration to reduce autocorrelation. A final sample of size 9,000 (pooled over three chains) almost independent draws were obtained for each parameter in the model and are used for posterior inference. Trace plots, autocorrelation plots and posterior density plots are examined carefully to ensure convergence is achieved for each parameter. Convergence of the parameters is also extensively examined using the Brooks, Gelman and Rubin convergence plots (Gelman and Rubin, 1992; Brooks and Gelman, 1998).

***Empirical Bayesian Estimation***

EB is a related approach that has been used extensively in the analysis of traffic safety data, particularly for the before-after evaluation of road safety treatments. An EB approach is related to the full Bayesian approach in the sense that prior information is combined with current data to obtain an estimate. However, different to the FB approach, the parameters of the prior distribution are estimated from existing data and are used in subsequent steps assuming there is no uncertainty related to the parameter. For example, in evaluating the effect of some treatments in the number of road crashes, the prior information is obtained from comparison sites similar to those treatment sites. An estimate of the sample mean or a function of the mean (commonly known as the Safety Performance Functions (SPF)) can be calculated and is then combined with the site-specific crash count to acquire an improved estimate of a site's expected crash count. However, this implies that in the EB approach, population-level estimates do not contribute to the uncertainty in the site specific estimate and as a result, intersection-level estimates may have lower standard errors. The FB approach on the other hand, better accounts for this uncertainty by recognising that the population-level estimates follow certain distributions which in turn depend on higher-level parameters. Also, implementing a FB approach does not require the estimation of SPF, which is one of the largest limitations of an EB approach. French and Heagerty (2008) applied a particular EB procedure which involves fitting a regression model prior to policy intervention and use the resulting model to form a trajectory of outcomes in time periods subsequent to the policy intervention. This is followed by contrasting post-intervention observations with their expected outcomes (predicted by the regression model) under the absence of a policy intervention. Regression-based models have the flexibility to account for temporal trends as well as adjust for important covariates and their interactions. The differences between these expected outcomes and the data observed after the policy change are then averaged to estimate the policy effects. Formal statistical tests such as the two-sample *t*-test, can be used to examine whether there exists a significant difference in policy effect between the primary target group of interest and its comparison group.

For EB methods, we follow the approach described in French and Heagerty (2008) since it is an EB procedure specifically designed for drawing inference from longitudinal partial crossover data. The idea is to first fit a regression model to capture any temporal trend prior to policy intervention and then use this model to forecast the expected outcome subsequent to the policy implementation. More specifically we fit the following model to the data prior to the intervention

$$\begin{aligned} Y_i &\sim Poisson\ (\eta_i \mu_i) \\ \eta_i &\sim Gamma\ (\alpha, \alpha,) \\ \log(\mu_i) &= \delta_0 + \delta_1 TIME_{PRE} + \delta_2 INJURY + \delta_3 TIME_{PRE} \times INJURY + \log(\text{exposure}) \end{aligned} \quad (2)$$

where $Y_i$ are the admission counts of head and arm injuries prior to the helmet legislation, $\text{TIME}_{\text{PRE}}$ is 18 monthly intervals pre-law and INJURY is again an indicator variable taking a value of one for a head injury and zero for an arm injury. These expected outcomes predicted from the model are then subtracted from the observed data after the policy change to create contrasts, $\Delta_i$. That is,

$$\Delta_i = \log(Y_i) - \log(\hat{\mu}_i)$$

Simple statistical analysis such as a *t*-test can be used to assess the significance of a policy effect. Moreover, we fit a simple linear regression to these contrasts to obtain estimates of baseline level and slope change after the intervention as well as differential changes in head injuries as compared to arm injuries after the law. That is,

$$\Delta = \delta_4 + \delta_5 TIME_{POST} + \delta_6 INJURY + \delta_7 TIME_{POST} \times INJURY \quad (3)$$

where $\text{TIME}_{\text{POST}}$ is the 18 monthly intervals post-law. $\delta_4$ and $\delta_5$ represent the baseline level for arm injuries and slope change after the policy intervention respectively, while $\delta_6$ and $\delta_7$ effectively capture differential changes in head and arm injuries post-law. A negative value for $\delta_6$ suggests that the head injuries decreased more than arm injuries after the law.

## Results

Monte Carlo estimates of the posterior mean, standard deviation, $2.5^{th}$ and $97.5^{th}$ percentiles of the marginal posterior distribution of each parameter were obtained from the draws and they are reported in Table 1. Maximum likelihood estimates obtained from Walter *et al.* (2011) are included in Table 1 as a comparison. There is little difference between the estimates obtained under the two methods. The full Bayesian 95% credible intervals are reasonably close to the classical 95% confidence intervals, although they are generally wider than the corresponding confidence interval for all parameters. This is consistent with the Bayesian concept as the prior distributions take into account the variability of the unknown parameters and the Bayesian credible interval reflects both parametric and sampling uncertainty, whereas the classical confidence interval reflects only sampling uncertainty. Similar to the maximum likelihood estimate, we obtained a negative estimate for the coefficient $\beta_3$, which suggested that overall injury count decreased following the helmet law.

***Table 1. Negative binomial model estimates using frequentist MLE, FB and EB methods***

| | **Frequentist MLE** | **Full Bayesian** | **Empirical Bayes** |
|---|---|---|---|
| **Variable** | Estimate (95% CI) | Estimate (95% CI) | Estimate (95% CI) |
| **Intercept ($\beta_0$ or $\delta_0$)** | -11.470 (-11.613,-11.326) | -11.470 (-11.630,-11.320) | -11.470 (-11.601, -11.337) |
| **TIME ($\beta_1$ or $\delta_1$)** | -0.005 (-0.019,0.009) | -0.005 (-0.020,0.009) | -0.005 (-0.018,0.007) |
| **INJURY ($\beta_2$ or $\delta_2$)** | 0.072 (-0.128,0.272) | 0.071 (-0.136,0.287) | 0.077 (-0.111,0.255) |
| **LAW ($\beta_3$ or $\delta_4$)** | -0.112 (-0.318,0.093) | -0.111 (-0.330,0.106) | -0.145 (-0.314,0.024) |
| **TIME×INJURY ($\beta_4$ or $\delta_3$)** | -0.003 (-0.022,0.016) | -0.003 (-0.023,0.017) | -0.003 (-0.021,0.014) |
| **TIME×LAW ($\beta_5$ or $\delta_5$)** | 0.015 (-0.005,0.034) | 0.015 (-0.006,0.036) | 0.017 (0.000,0.033) |
| **INJURY×LAW ($\beta_6$ or $\delta_6$)** | -0.322 (-0.618,-0.027) | -0.323 (-0.635,-0.014) | -0.302 (-0.514,-0.064) |
| **TIME×INJURY×LAW ($\beta_7$ or $\delta_7$)** | 0.010 (-0.018,0.038) | 0.010 (-0.021,0.040) | 0.007 (-0.016,0.030) |

We also obtained a significant negative estimate for $\beta_6$, which indicates differential changes in head injuries as compared to arm injuries after the helmet legislation. In this case, a significantly negative estimate suggested that the head injuries dropped by more than arm injuries at the time of legislation, providing evidence for a legislation attributable benefit. Head and arm injury counts and predicted counts estimated using the FB method are shown in Figure 1. The graph looks almost identical to the one presented in Walter *et al.* (2011), since very similar parameter estimates were obtained. The most noticeable characteristic is the downward 'step' in head injury counts having a larger magnitude than the corresponding arm injury counts.

Results from the EB approach are quite similar to those obtained using classical MLE and FB methods, especially in terms of parameter estimates. Apart from the intercept, only $\delta_5$ and $\delta_6$ are significant and $\delta_5$ is only marginally significant. The significant negative estimate of $\delta_6$ has similar implication as $\beta_6$, that is, the decrease in head injury counts is more than that for arm injury with the change in helmet wearing legislation. The drop in arm injury after the law is non-significant since the 95% confidence interval or credible interval cover the value 0 for the estimate of $\beta_3$ or $\delta_4$, respectively. Since the mandatory helmet law is a safety intervention aimed specifically at head injuries but not limb injuries and we assume that other changes to the cycling environment would affect both head and arm injuries, the significant differential decrease in injury counts is most likely due to the compulsory helmet legislation. Using FB and EB methods, the estimated legislation attributable decrease in head injury are 27.6% (1-exp(-0.323)) and 26.1% respectively, which are comparable to the estimated 27.5% found in the original study (Walter *et al.*, 2011). Therefore, using two different estimation methods, namely FB and EB, we confirmed the original finding, that there is evidence for a positive effect of the compulsory cycle helmet legislation on head injuries. Note that the confidence intervals of all parameters are the tightest in the EB method among three estimation methods. However, the parameter standard errors estimated in the EB method are not directly comparable to those in classical MLE and FB methods since simpler regression models (2) and (3) (with less parameters) are fitted separately to pre- and post- intervention data as opposed to the full model (1). Figure 2 shows the plot of the fitted log-linear regression model using pre-law data (solid line) and predicted outcomes (dashed lines) after the policy intervention separately for arm and head injuries. At first sight, the post-law arm injuries seem to fluctuate around the trend formed by the expected outcomes. Result from the *t*-test suggests that the mean contrast for arm injuries is not significantly different from zero ($\hat{\delta}_4$=-0.145, s.e.=0.086, $p$=0.102). On the other hand, the observed post-law head injury counts are, in general, less than the predicted outcomes and the one sample *t*-test for the head contrast rejects the null hypothesis that the mean is zero ($\hat{\delta}_4 + \hat{\delta}_6 = -0.447$, s.e.=0.086, $p=1.14\times10^{-5}$). The difference between head and arm contrasts, $\hat{\delta}_6$, is also statistically significant ($\hat{\delta}_6$=-0.302, s.e.=0.122, $p$=0.019). Hence, relative to arm injuries, there is an estimated 26.1% drop in head injuries.

***Figure 1. Cyclist head vs. arm injury counts and fitted model for 18 months prior and post helmet legislation***

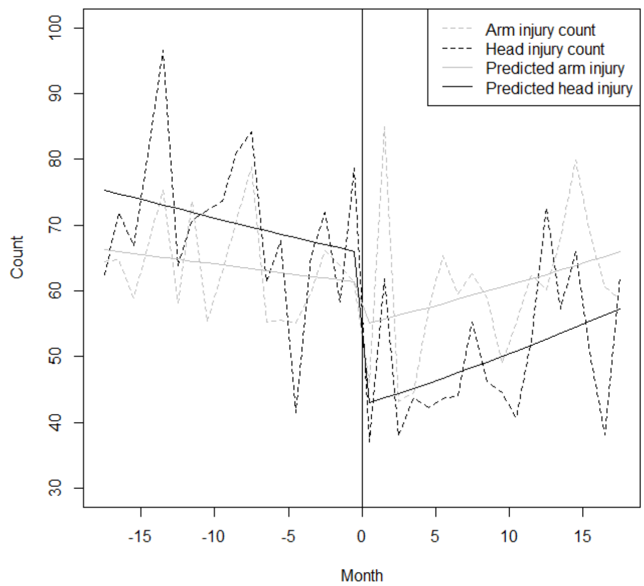
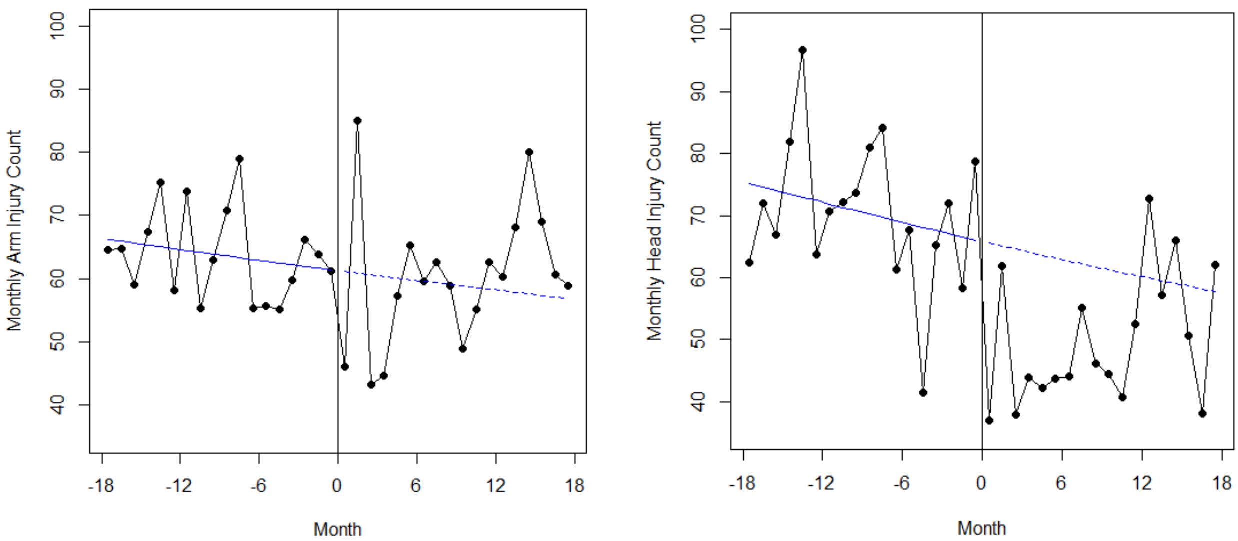



***Figure 2. Cyclist head and arm injury counts with pre-policy estimation (solid line) and post-policy prediction(dashed line)***




## Discussion

In this study, we fit a hierarchical negative binomial regression model to the hospital admission of head and arm injury counts, from a 36 month period centred at the mandatory helmet legislation in NSW. The explanatory variables in the segmented regression included injury type, time, indicator for the policy intervention and all two-way and three-way interactions between the variables. The data and the model was previously analysed in Walter *et al.* (2011) in which they estimated significant decreased head injury rates relative to arm injury rates at the time of legislation among cyclists and concluded this additional benefit was attributed to compulsory helmet legislation. The model estimation was done using a classical maximum likelihood method. In this study, we re-analysed the data using two different estimation methods, namely full Bayesian and empirical Bayes approaches. For the FB methods, noninformative priors were used for all parameters in the model, and sampling from the posterior distribution was accomplished through MCMC methods implemented in WinBUGS. Parameters estimated using a FB method are very similar to those obtained under MLE, although the 95% credible intervals are (as expected) wider than the corresponding 95% confidence intervals as to account for the uncertainty in the parameters. For the EB method, we adopted the procedure described in French and Heagerty (2008) specifically designed for evaluating longitudinal effects of interventions. A negative binomial regression model was fit to head and arm injury counts prior to the intervention and the resulting model was used to predict expected outcomes under the absence of the policy change. These expected outcomes w then compared with the injury counts observed after the policy change to compute contrasts. These contrasts were then analysed by a linear regression model to estimate the (differential) changes in the level and slope of head and arm injuries after the law. Parameter estimates using the EB method are comparable to classical MLE and FB estimates, Persaud *et al.* (2010) also found that EB and FB for treatment evaluation studies give comparable results. As pointed out in French and Heagerty (2008), the standard error calculation for the contrasts $\Delta$ is not straightforward. Our *t*-tests are performed based on the assumption that variation due to estimation of pre-policy model parameters do not contribute to post-policy parameter estimation. Hence, standard error computation is simplified and an estimate of the variance of contrast $\Delta$ is simply var($\Delta_i$). A more complicated estimator for the variance of contrast can be found in the Supplementary Material of French and Heagerty (2008). Regardless of the estimation limitation in EB, we estimated a significant decrease in head injuries more than arm injuries at the time of legislation and the estimated effect had a similar magnitude to that estimated using MLE and FB approaches.

Generally speaking, when we have a large amount of information available, the choice of whether model parameters are estimated using a classical frequentist, FB or EB framework makes little difference. However, when the number of observations is small, both classical and Bayesian paradigms can result in invalid conclusions, and it is exacerbated in EB since only the pre-intervention data are used to construct the model which is then used to predict expected outcomes. Furthermore, the FB method is more flexible than the classical frequentist and EB methods as prior distributions are used to better account for uncertainty in the sample. In our study, the classical MLE and EB approaches rely on the assumption of a negative binomial distribution of injury counts in the estimation process. The FB approach on the other hand, is able to model this as a hierarchical Poisson-Gamma mixture distribution. This hierarchical structure allows other distributions, such as the Poisson-Lognormal distribution, to be implemented and investigated in a similar fashion. Major criticisms for the FB approach include the choice of prior distributions and the computational cost involved in sampling from the posterior distribution. Typically, parallel Markov chains are run for a large number of iterations to ensure all parameters have converged and successive iterates are not correlated.

It is worth mentioning that the EB method employed in our study is not the same approach commonly used for analysing traffic safety data such as in Carriquiry and Pawlovich (2004). Generally, in a parametric EB procedure, the model is formulated in the same way as a Bayesian hierarchical model and prior distributions are assigned to model parameters. The prior distributions are assumed to belong to a family of distributions indexed by a low-dimensional parameter (hyperparameter). An estimate of this hyperparameter can be computed using the marginal distribution of all data, usually as a MLE. For our full model specified in (1), the parameter space is $\underline{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)$ and the prior distributions are

$$\beta_i \sim Normal\ (\mu_{\beta_i}, {\sigma_{\beta_i}}^2), \qquad i = 0, 1, \ldots, 7,$$
$$\alpha \sim Uniform\ (a_\alpha, b_\alpha)$$

Hence to obtain MLE estimates for the hyperparameters $\mu_{\beta i}$, $\sigma_{\beta i}^2$, $a_\alpha$, $b_\alpha$ for each model parameter,

we need to maximise the marginal distribution of the data which is 9-dimensional integral since we need to integrate out model parameters $\underline{\theta}$. Since the FB method is readily implementable in this case, we feel that the use of a parametric EB method is not justified for our segmented regression model.

Since the data and model were the same as in Walter *et al*. (2013), some limitations associated with the data and the analysis method also apply in this case. For instance, due to lack of population level exposure and helmet wearing data, we used arm injuries as a comparison group to cope with this problem. However, there may be more than one such comparison group (for example, cyclist leg injuries) and it remains unclear how to choose the most ‘appropriate’ comparison group. Furthermore, our study is limited by the amount of data available prior to helmet legislation and thus conducting a similar study using data from a jurisdiction with more pre-law data is highly recommended. Lastly, it is worthwhile to empirically assess the relative performance of these methods when they are applied to more complicated time series models and state space models.

This comparison of three estimation methods for road safety intervention assessment showed that MLE, FB and EB give similar results for large samples. The agreement between the methods for the example data also confirmed the original findings of Walter *et al.* (2011, 2013). Generally speaking, the FB method incorporates parameter uncertainty via prior distributions and does not depend on asymptotic properties, which are the advantages over the classical MLE. However, the FB method may be computationally costly to implement, especially when WinBUGS is not used. The EB method is a relatively simpler method but issues may arise for small samples. Hence in performing detailed policy change evaluations, analysts must be mindful of the important differences between these estimation approaches both in terms of the implementation procedure and inference.

**Acknowledgements**

Data for this study were provided under agreement by the NSW Ministry of Health, Centre for Epidemiology and Evidence.

**References**

Brooks, S., Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434-456.

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

French, B., Heagerty, P.J. (2008) Analysis of longitudinal data to evaluate a policy change. *Statistics in Medicine* 27, 5005-5025.

Gelman, A., Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457-511.

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.

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 and Prevention* 41, 574-580.

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

Olivier, J., Wang, J.J.J., Walter, S., Grzebieta, R. (2013). On the use of empirical Bayes for comparative interrupted time series with an application to mandatory helmet legislation. *Submitted to Road Safety, Research, Policing and Education Conference, 2013.*

Park, E.S., Park, J., Lomax, T.J. (2010). A fully Bayesian multivariate approach to before-after safety evaluation. *Accident Analysis and Prevention* 42, 1118-1127.

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.

Pendelton, O.J., Ofelia, G., Horacio, D. (1991). Application of new accident analysis methodologies Volume II A Users Manual for BEATS. FHWA-RD-91-014. Federal Highway Administration (FHWA), McLean, VA.

Persaud, B.N. (1988). Do traffic signals affect safety? Some methodological issues. *Transportation Research Record* 1185, 37-47.

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., Lan, B., Lyon, C., Bhim, R. (2010). Comparison of empirical Bayes and full Bayes approaches for before-after road safety evaluations. *Accident Analysis and Prevention* 42, 38-43.

Persaud, B., Lan, B., Lyon, C., Bhim, R. (2010). Comparison of empirical Bayes and full Bayes approaches for before-after road safety evaluations. *Accident Analysis and Prevention* 42, 38-43.

Rissel, C. (2012). The impact of compulsory cycle helmet legislation on cyclist head injuries in New South Wales, Australia: A rejoinder. *Accident Analysis and Prevention* 45, 107-109.

Shiskin, J., Young, A.H., Musgrave, J.C. (1967). The X-11 Variant of the Census Method II Seasonal Adjustment Program. In: Technical Report 15. US Department of Commerce, Bureau of the Census.

Walter, S., Olivier, J., Churches, T., Grzebieta, R. (2011). The impact of compulsory cycle helmet legislation on cyclist head injuries in New South Wales, Australia. *Accident Analysis and Prevention* 43, 2064-2071.

Walter, S., Olivier, J., Churches, T., Grzebieta, R. (2013). The impact of compulsory cycle helmet legislation on cyclist head injuries in New South Wales, Australia: A response. *Accident Analysis and Prevention* 52, 204-209.

Wagner, A.K., Soumerai, S.B., Zhang, F., Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics* 27, 299-309.