Mobile safety cameras: Estimating casualty reductions and the demand for secondary healthcare

Lee Fawcett and Neil Thorpe

Newcastle University, Newcastle Upon Tyne, UK.

Summary. We consider a fully Bayesian analysis of road casualty data at 56 designated mobile safety camera sites in the Northumbria Police Force area in the UK. It is well–documented that regression to the mean (RTM) can exaggerate the effectiveness of road safety measures, and, since the 1980s, an empirical Bayes estimation framework has become the standard tool for separating real treatment effects from those of RTM. In this paper we show that, relative to a fully Bayesian treatment, the empirical Bayes method is over–optimistic when quantifying the variability of estimates of casualty frequency. More realistically, a fully Bayesian analysis allows population–level estimates to contribute to the uncertainty in such estimates of casualty frequency. Implementing a fully Bayesian analysis via Markov chain Monte Carlo also provides a more flexible and complete inferential procedure. We assess the sensitivity of estimates of treatment effectiveness, as well as the expected monetary value of prevention owing to the implementation of the safety cameras, to different model specifications, which include the construction of informative priors for some parameters.

Keywords: Markov chain Monte Carlo, mobile safety cameras, negative binomial distribution, Northumbria Safety Camera Partnership, regression to the mean.

1. Background

In 2007, official statistics revealed that 222,146 people were reported as injured as a result of road traffic accidents in Great Britain (Department for Transport (DfT), 2010). Of these, 2,222 people were killed and 24,690 were seriously injured placing a huge economic and human cost on society. Road casualty reduction is therefore a key aim of government transport policy with new road safety measures continually being tried and tested in an attempt to reduce the number and severity of casualties. Implementing a road safety measure, whether this is a new junction layout, education programmes for young children or new technology to assist in the enforcement of traffic laws, clearly comes at a financial cost. The ability to isolate out only the effects of the measure on changes in the pattern of casualties is therefore vital for assessing the potential impacts of alternative accident remedial measures, selecting locations that might benefit from treatment and for evaluating the actual performance of measures once implemented.

1.1. Sources of error in road safety scheme evaluation

To assess the effects of accident remedial measures, road safety practitioners often use the observed number of casualties at a site before and after the implementation of a scheme. However, casualty counts from such before–after studies are prone to numerous sources of error, most of which are well–known and well–documented (e.g. Hauer (1980); Maher (1991); Jarrett (1994)). Above all, these include exaggerated estimates of the effects of a scheme due to *regression to the mean* (RTM; see, for example, Hauer *et al.*, 2001). Sites considered as candidates for road safety schemes are often those that have observed, over some pre–determined baseline period, an unusually high number of casualties. In any subsequent time interval, these sites will probably have fewer recorded casualties anyway, even if left untreated, simply because the casualty frequency recorded in the baseline period was abnormally high. The main consequence of ignoring this phenomenon and simply examining before and after figures is usually an exaggerated treatment effect and possibly unjustified financial investment. Most published work in the road safety literature now acknowledges the contribution of RTM to any observed before–after casualty reduction and, since the early 1980s, a standard procedure (the empirical Bayes approach; see Section 2.1) has been adopted to quantify this. Although the effect is variable, studies typically show a reduction in casualty frequency owing to RTM of between 20 and 30% (Hirst *et al.*, 2004). In fact, so widespread is the acceptance of this 'bias–by–selection' phenomenon that most studies do very little – if anything at all – to actually check whether or not it is present in the data; similarly, little effort is usually expended to check the appropriateness of the standard approach to quantify the RTM effect.

Scenarios 1, 3 and 4 in Figure 1 serve to illustrate the possible contributions of RTM to any reduction in casualty frequency. Scenario 1 represents a 'blip' in the number of casualties at a site, with an intervention at the peak of this blip – the suggestion is that the number of casualties returns to the 'norm' anyway, creating a 'benefit illusion'; here, the reduction owing to RTM is 100%. Advocates of road safety schemes, who acknowledge the role of RTM, might be more likely to believe scenario 3 in which there is a regression to the norm, but a further decrease in casualty frequency indicating that the intervention has given a 'true benefit'. Scenario 4 indicates RTM with 'disbenefit'; here, an intervention has been applied at the peak of the blip, but then the casualty frequency has failed to return to the norm – those against the intervention might argue that the intervention itself has actually resulted in an overall *increase* in casualties. Scenario 2 represents the situation if we have complete faith in a simple before–after comparison; any reduction is completely attributed to the intervention itself, meaning the high number of casualties is not a result of an abnormal blip. Of course, all four scenarios are possible across any road network, and it would be wise to investigate this (perhaps using historical records for the treated site) before assuming RTM is/is not present and applying statistical procedures based on this assumption. Underneath these hypothetical scenarios we see some real casualty data from a section of the Tullamarine Freeway in Victoria, Australia. In a report (Australian Government DIT, 2001), the investigators claim a 76% reduction in the number of casualties (21$\rightarrow$5) as a result of the safety scheme implemented at the start of 1997 (in this case, a fixed speed camera). However, it is clear that we might actually have something like scenarios 1 or 3, with the quoted treatment effect being an exaggeration of the true effect.

Of course, RTM is not the only non–treatment effect that can distort the apparent effectiveness of road safety schemes. For example, trends in risk might decline over time as a response to a road safety programme; changes in *exposure* to risk (e.g. as a side effect of the treatment, such as drivers bypassing sections of the network where safety camera schemes are known to operate) might also play a part in any change in observed casualty frequency between a before and after period. The aim of the evaluation is therefore to isolate the true effect of the safety scheme by properly accounting for RTM and other such 'nuisance' factors.

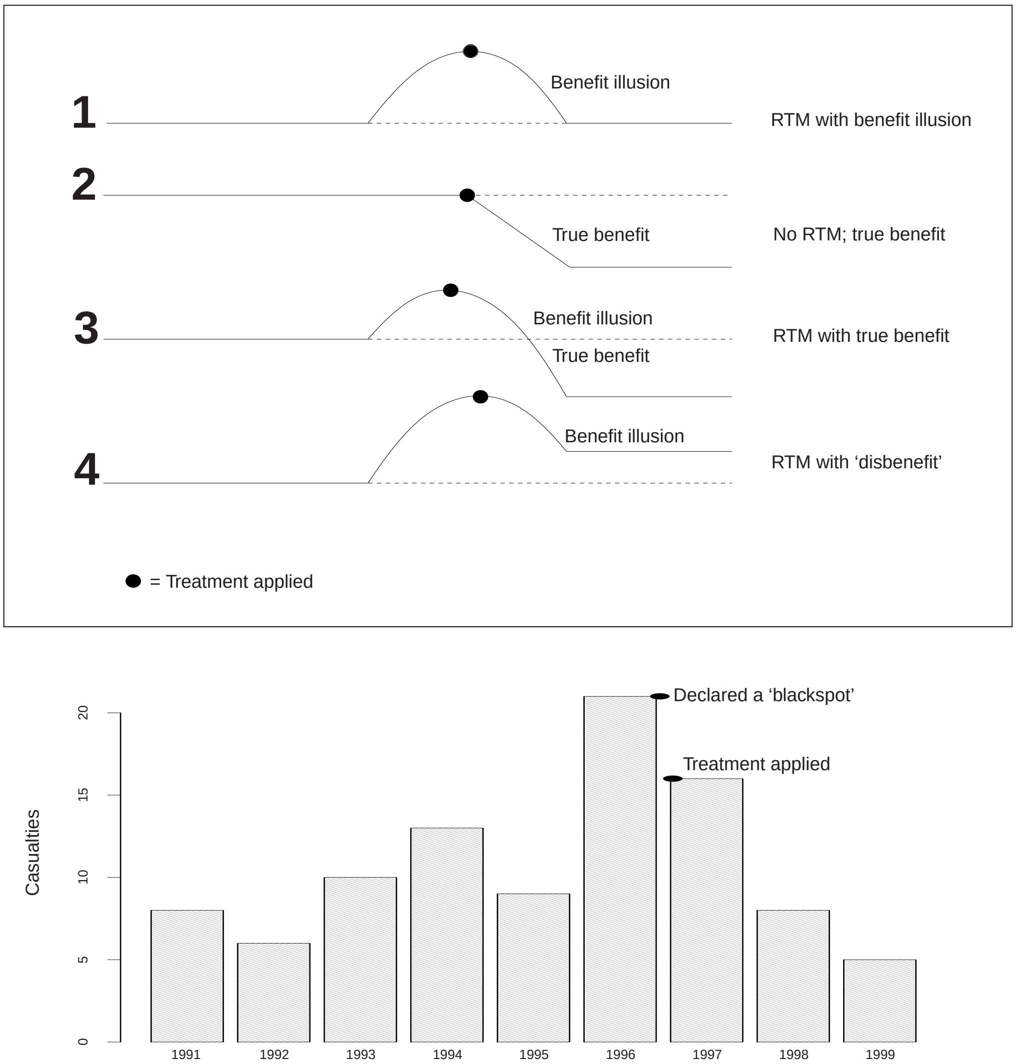

Fig. 1. Top: Hypothetical scenarios showing the effects of RTM, and a road safety scheme intervention, on the number of casualties. Bottom: Number of casualties, per anunum, on a section of the Tullamarine Freeway in Australia (1991–1999).

1.2. Road safety camera policy in the U.K.

In 1996, a government report concluded that road safety cameras (notably speed cameras) could be an effective weapon in reducing casualty frequencies (Hooke *et al.*, 1996). However, the relatively high implementation and running costs were felt to prohibit their widespread deployment under prevailing funding mechanisms. In 1998, the government took the decision to allow traffic authorities to recover the cost of installing and operating speed cameras from the revenues generated from speeding offences detected by the cameras. As a result of the 1996 report and the introduction of the cost recovery approach, the government viewed speed cameras as an important part of its strategy to achieve its casualty reduction targets for 2010 (DfT, 2000). In April 2000, a two year pilot programme commenced involving eight road safety camera partnerships. Results at the end of the first year prompted the government to take an earlier–than–expected decision to introduce legislation in 2001 to enable national roll–out of safety camera partnerships across Great Britain. By 2004, almost the entire area of England, Scotland and Wales was covered by 42 safety camera partnerships operating under the rules of the cost recovery programme introduced in 2001.

The rapid growth in speed camera activity, and subsequent increase in members of the public being punished for speeding offences, prompted a vigorous and detailed debate over the value of speed cameras in the national media. Opponents trying to discredit the operation of speed cameras focussed on a range of issues in an attempt to have the scheme abandoned, in particular disputing the claimed effectiveness of speed cameras as a casualty reduction measure. Mention *SafeSpeed* here? This tactic brought RTM effects to the forefront of the public debate to the extent that the impact of RTM was discussed in at least one daily national newspaper and eventually was incorporated (in 2005) in the official calculations of casualty reductions at speed camera sites. The recent cuts announced in local authority spending have refocussed attention on the claimed effectiveness – and value for money – of speed cameras, as road safety initiatives generally come under close scrutiny *vis–a–vis* other public spending priorities.

1.3. The Northumbria Safety Camera Partnership (NSCP)

The Northumbria Safety Camera Partnership (NSCP) joined the national programme in April 2003. In February 2004, the Partnership commissioned a team of researchers to investigate specifically the impact of operating mobile road safety cameras on the demand for secondary health care at the region's hospitals. The study group collected data from 67 mobile camera sites in the region from a 'before' period (April 2001–March 2003) and an 'after' period (April 2004–March 2006) to investigate changes in the number and severity of casualties and 'before' and 'after' cost–of–treatment estimates.

Following other published studies whose aims were to evaluate the effectiveness of road safety schemes, the NSCP adopted a standard empirical Bayes procedure (as outlined in Section 2.1 and demonstrated in Section 2.3.1) to separate real treatment effects at each of the 67 sites from the effects of RTM. The final report (Colligan *et al.*, 2008), in line with many other studies in the literature, shows that simple comparisons between 'before' and 'after' casualty frequencies do not account for bias in site selection, and as such tend to overstate the effectiveness of the safety cameras. After identifying the contribution of genuine treatment effects and RTM to the change in casualty frequencies at each site, Colligan *et al.* (2008) then attempt to link police accident records to local hospital databases to quantify the cost savings to local National Health Service (NHS) secondary healthcare providers as a result of the implementation of the safety camera scheme. In tariff terms, Colligan *et al.* (2008) estimate the cost of not having to treat the casualties 'saved' by the introduction of the safety cameras at about £30,000; of course, grossing up to the national level would significantly increase this figure.

1.4. Aims of this paper

The primary objectives of the current paper are: (1) to consider the appropriateness of current methodology for estimating the contribution of RTM to casualty reduction, and (2) to find improvements over the standard empirical Bayes procedure for modelling casualty frequencies. In Section 2 we describe, generally, the standard empirical Bayes approach for quantifying RTM, and outline the limitations of this approach as it is usually applied. The remainder of this paper then focuses on the safety camera data used in the NSCP–commissioned study.

In Section 3, we consider some exploratory investigations to check for the presence of RTM in the casualty records at sites treated with safety cameras; we also consider some pre–analysis checks of the assumptions implicit in the standard approach for quantifying RTM. We then apply the standard empirical Bayes procedure to the safety camera data and compare this to a fully Bayesian treatment, paying particular attention to the estimates of variability of expected casualty frequencies at each site. We also consider the resulting estimates of cost savings to the NHS, and the wider society, as a result of implementing the safety camera scheme, again comparing both empirical and fully Bayesian approaches.

In Section 4, we consider some alternative model strctures. In particular, we: assess the sensitivity of estimates of RTM, treatment effects and cost to the NHS/the wider society to different prior distributions for the mean casualty rate at each safety camera site; attempt to construct more realistic prior distributions for the regression coefficients used to predict the mean casualty rate, and identify the contribution of trend to any change in casualty frequencies at safety camera sites.

2. Statistical modelling: accounting for RTM

2.1. Empirical Bayes approach

To date, most road safety scheme evaluation studies that have attempted to quantify the effects of RTM have made use of an empirical Bayes (EB) procedure; see, for example, Hauer *et al.* (1988), Pendleton (1991), Mountain *et al.* (1992), Persaud and Dzbik (1993) and Kulmala (1994) Give more examples here?. The model formulation is rather simple and, given a specific choice of prior distribution for the mean casualty rate at each treated site j, the resulting posterior mean for this rate is – conveniently – a weighted sum of the abnormally high *observed* casualty frequency at site j and what we might usually *expect* to see at this site. Assuming a Poisson distribution with mean m_j for the casualty frequency y_j at site j in any period *before* the implementation of a road safety scheme, and a gamma distribution for m_j itself, with mean μ_j and variance μ_j^2/γ, i.e.

$$\begin{aligned} y_j|m_j &\sim \text{Poisson}(m_j) \qquad \text{and} \\ m_j &\sim \text{Gamma}(\gamma, \gamma/\mu_j), \end{aligned}$$

the posterior distribution of $m_j|y_j$ is also of gamma form. Specifically,

$$m_j|y_j \sim \text{Gamma}\left(\gamma + y_j, \gamma/\mu_j + 1\right). \tag{1}$$

The mean of this posterior is then used as the EB estimate of casualty frequency, i.e.

$$\mathbb{E}\left(m_j|y_j\right) = \alpha_j\mu_j + (1-\alpha_j)y_j, \qquad \text{where} \tag{2}$$
$$\alpha_j = \gamma/(\gamma+\mu_j) \tag{3}$$

and $0 \leq \alpha_j \leq 1$. Thus, the EB estimate of casualty frequency at site j is a weighted sum of the prior mean number of casualties at that site (μ_j) and the actual number of casualties observed in the before period (y_j). In the studies we refer to above, evaluation of the effectiveness of a particular road safety scheme at site j is based on a comparison of the observed number of casualties at that site *after* scheme implementation ($y_{j,\text{after}}$), not with the number of casualties in any before period (y_j) but with the EB estimate of casualty frequency for that site given by Equation (2). The percentage change from y_j to $\mathbb{E}(m_j|y_j)$ is taken to be the percentage change that would have happened anyway, even without the implementation of any road safety measure, i.e. the RTM effect.

Generalised linear modelling is usually adopted to obtain an estimate of μ_j, i.e.

$$\hat{\mu}_j = \exp\left\{\hat{\beta}_0 + \sum_{p=1}^{n_p}\hat{\beta}_p x_{p_j}\right\}, \tag{4}$$

where x_{p_j} are variables associated with attributes at site j that could have an effect on the mean number of casualties at that site (e.g. traffic flow, road type (single/dual carriageway), averaged observed speed of vehicles etc.) and n_p is the number of such variables used. The estimated regression coefficients $\hat{\beta}_i$, $i = 0, \ldots, n_p$, are obtained from a set of reference sites that are representative of the sites at which the road safety scheme has been implemented – in terms of the explanatory variables x_p, but *not* in terms of their casualty frequency. Indeed, sites that have been chosen for a road safety scheme have usually been done so on the basis of their unusually high casualty frequency during some pre–determined observation period; for μ_j we require a model that will give us a more representative prediction of mean casualty frequency at each treated site j. Thus, the inclusion of y_j in Equation (2) takes into account specific site characteristics not included in the predictive model estimate of μ_j, while μ_j itself smooths out any random variation. This provides the basis of any EB analysis – a frequentist approach is adopted within the Bayesian framework (here, maximum likelihood can be used to estimate μ_j from the data) to obtain a posterior summary of interest (here, the posterior mean).

Estimation of the weight α_j also requires an estimate of the prior shape parameter γ. The unconditional distribution of y_j is negative binomial with mean μ_j and variance $\mu_j + \kappa\mu_j^2$, where $\kappa = 1/\gamma > 0$ is the negative binomial 'over–dispersion' parameter. Thus, if we assume a negative binomial error structure in model (4), maximum likelihood estimates of β_i, $i = 0, \ldots, n_p$ and γ can be obtained, leading to estimates of μ_j and α_j via Equations (4) and (3) (respectively) and hence EB estimates of casualty frequency at each site j via Equation (2).

2.2. Limitations of the EB approach

The EB approach for estimating RTM, as outlined in Section 2.1, has become the standard tool for practitioners who wish to evaluate the true effectiveness of a road safety scheme. However, the approach is flawed if the set of reference sites (from which the estimated regression coefficients in Equation (4) are obtained) are not exchangeable with the sites at which the road safety scheme has been implemented (to which Equation (4) is applied). At best, some studies compare simple summaries of the covariates for the reference and treated sites (references), but more rigorous testing of the assumption of exchangeability is not commonplace. In this paper, we consider some pre–analysis exploratory checks of this assumption. Even if such checks can be shown to lend support to the assumption of exchangeability (as they do in our example), there are several limitiations to the EB approach as it stands; limitations whic can be overcome by working within a fully Bayesian framework.

2.2.1. Choice of prior for m_j

Although the use of the gamma distribution as a prior for the mean casualty rate m_j gives a convenient and appealing expression for $\mathbb{E}(m_j|y_j)$, this choice of prior distribution is borne out of mathematical convenience, the gamma distribution being the conjugate prior for the Poisson. Developments in computer–based simulation procedures since the first use of the EB method in the 1980s have revolutionised Bayesian modelling, with the result that there is no longer any need nor advantage of sticking with conjugate or other rather artificial forms of prior distribution for m_j. Using MCMC, we can now simulate directly from the posterior distribution of interest no matter what the choice of prior distribution for m_j, giving us a more complete inferential procedure from which we can summarise the posterior in whichever way we deem appropriate.

2.2.2. Over–optimistic quantification of variability

By substituting μ_j in Equation (2) with point estimates obtained from Equation (4) it is implied that population–level estimates do not contribute to the uncertainty in the estimate of casualty frequency for a specific site. Indeed, this is bound to lead to unrealistically low posterior standard deviations for EB estimates of casualty frequency. A fully Bayesian analysis would assign hyper–prior distributions to the regression coefficients β_i and the prior shape parameter γ; doing so explicitly recognises that population–level estimates of casualty frequencies are also uncertain and thus contribute to the variance of the site–level estimates of m_j. A fully Bayesian analysis would thus unify the entire estimation procedure, seamlessly integrating prior information and all available data into posterior distributions on which practitioners can base their inferences. In this paper, we compare the standard EB approach to a corresponding fully Bayesian analysis using rather vague prior distributions; we then check the sensitivity of our estimates of RTM to the choice of prior distribution for m_j, as well as investigate the use of more informative priors.

2.2.3. More sophisticated modelling

There are more sophisticated modelling procedures available to practitioners when analysing count data than the EB approach (or an analagous fully Bayesian approach) as outlined in Section 2.1. Generalised linear mixed models (GLMMs), for example, provide a natural extension to generalised linear models by adding random terms in the linear predictor to account for overdispersion, correlation and/or heterogeneity in the data. For example, writing the linear predictor in matrix form, we might replace the terms inside the braces in Equation (4) with

$$\boldsymbol{\eta} \quad = \quad \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U}; \tag{5}$$

here, the design matrix $\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_{n_p})$, where $\mathbf{1}$ is a column vector of 1's and $\mathbf{x}_p$, $p = 1, \ldots, n_p$ are column vectors of observations from covariates $x_1, \ldots, x_{n_p}$, and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{n_p})^T$. The matrix $\mathbf{Z}$ includes some – or all – of the covariates in $\mathbf{X}$, whilst $\mathbf{U}$ is a vector of unobservable random effects associated with the covariates in $\mathbf{Z}$. Commonly, $\mathbf{U}$ is taken to be multivariate Normal with mean $\mathbf{0}$ and some covariance $\boldsymbol{\Sigma}$, the elements of which are estimated. To account for over–dispersion, in which we might expect the variance to exceed the mean, we could use

$$\boldsymbol{\eta} + \mathbf{e}, \tag{6}$$

where $\mathbf{e}$ is a 'residual' for which a variance is estimated.

Count data, such as casualty frequencies, often display excess zeroes relative to what might be expected using a Poisson model, for example. Zero–inflated models (see, for example, Lord *et al.*, 2005) can provide a way of modelling the excess zeroes. In particular, for each observation, there are two possible data generation processes; the result of a Bernoulli trial determines which process used. For each site j, process 1 is chosen with probability φ and process 2 with probability $1 - \varphi$. Process 1 generates only zero counts, whereas process 2, say $g(\cdot)$, generates counts from the usual Poisson model. Thus, we might have

$$y_j \quad = \quad \begin{cases} 0 & \text{with probability } \varphi_j \\ g(y_j|m_j) & \text{with probabiity } 1 - \varphi_j, \end{cases} \tag{7}$$

where $g(y_j|m_j)$ might be defined as in Equation (1) and the prior mean μ_j might have a linear predictor with random effects (as in (5)) and/or an over–dispersion term (as in (6)).

Analytical results for such non–Gaussian GLMMs are generally not available; although restricted maximum likelihood (REML; see Patterson and Thompson, 1971) can be used, REML–based procedures use approximate likelihood methods which may not work well. MCMC within a Bayesian framework (once sensible priors have been chosen) is also an approximation but one whose accuracy increases the longer the analysis is run for (being exact in the limit). REML also uses large sample theory to derive approximate confidence intervals that may have poor coverage; MCMC measures of confidence are exact, up to Monte Carlo error, and provide an easy way of obtaining measures of confidence on derived statistics. Making inference on models with linear predictors given by (5) or (6) within the Bayesian framework can be readily done in R using the MCMCglmm package (Hadfield, 2010).

The purpose of our work in the current paper is to compare the standard EB approach for quantifying RTM to a fully Bayesian approach, assessing the effects of different priors for the mean casualty rate and associated regression coefficients on estimates of RTM and corresponding estimates of cost savings to the NHS owing to the implementation of the safety camera scheme. Although we do consider a zero–inflated model formulation in Section 4, more exotic model formulations includin the use of GLMMs, will be the focus of future work.

3. Application to the NSCP safety camera data

In this Section, we apply the EB method for estimating the RTM effect, as outlined in Section 2.1, to the safety camera data collected by the NSCP. We then compare these results to those from a fully Bayesian analysis wherein all sources of variability have been accounted for through the use of prior distributions for the regression coefficients in Equation (4) and the prior shape parameter for the gamma distribution.

At each treated site the number of casualties before and after the implementation of the safety cameras was observed, as well as various explanatory variables collected by the NSCP as part of the standard DfT reporting procedure for all safety camera partnerships: the average observed speed (x_1 miles per hour); the percentage of drivers exceeding the speed limit (x_2); daily traffic flow (x_3); speed limit (x_4 miles per hour); the eighty–fifth percentile speed (x_5 miles per hour); the percentage of drivers at least 15 miles per hour over the limit (x_6); and road classification and road type (x_7 = A/B/C/Unclassified roads, and x_8 = single/dual/mixed carriageway, respectively). The total number of casualties in the before period was 436; in the period after the implementation of safety cameras, this reduced to 298.

For the purpose of the present paper, **after removing sites with outliers and/or other data issues**, we work with a subset of 56 of the 67 mobile camera sites in the original study. These sites possess a wide variety of road characteristics, for example: some sites are urban, some rural, giving a mixture of road types and classifications; speed limits range from 30 mile–per–hour roads in towns to 70 mile–per–hour stretches of dual carriageways; some roads have heavy traffic flows whilst for others the flow is relatively light. What all 56 sites have in common is their unusually high casualty records from the observational period. To formulate a regression equation for μ_j to be used at each treated site j (Equation (4); often known as a "Predictive Accident Model", or PAM, in the road safety literature), we also have the same data available for a set of 67 reference sites in the Northumbria police force area.

3.1. Checking the RTM assumptions

As discussed in Section 1.1, most published studies make use of the EB procedure without even checking whether or not we can expect RTM to be present. **Figure 2 shows total casualty frequencies at the safety camera sites prior to the before period (April 2001–March 2003). More here once the data comes in... can we compare to the hypothetical scenarios in Figure 1 to support our claim that RTM is present? Tests for "blips"? e.g. peaks & troughs test?**

We also discussed in Section 2.2 that the standard procedure for estimating RTM is flawed if the sites treated with safety cameras cannot be considered exchangeable with sites in the reference set. Here, we consider some simple methods to check this assumption. Consider the matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_8)$, where $\mathbf{x}_p$, $p = 1, \ldots, 8$, are column vectors of observations from the explanatory variables $x_1, \ldots, x_8$ (described above) of length 123, with each of the 67 treated and 56 reference sites having a row entry. Combining treated and reference sites into a single data matrix, we can then perform a principal components analysis on $\mathbf{X}$; plotting the scores for the first two principal components against each other, using different

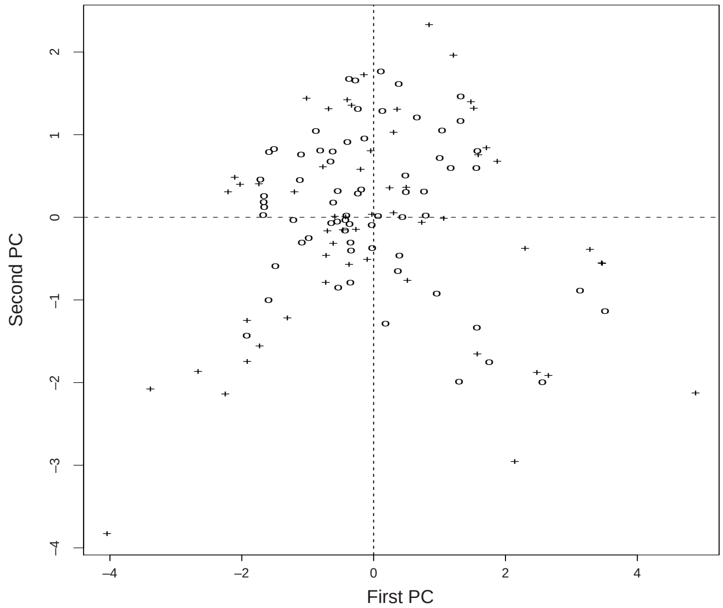

Fig. 2. Principal components analysis on $\mathbf{X}$: scores from first PC plotted againt those from second PC for reference sites (**o**) and treated sites (**+**).

plotting symbols for treated and reference sites, could help determine whether or not the assumption of exchangeability is plausible. Figure 2 shows this plot: as can be seen, there is no clear distinction between scores for the reference and treated sites.

To further investigate the plausibility of our assumption of exchangeability, we can perform permutation tests for various statistics that serve to compare the treated and reference sites. For example, suppose we wish to compare average values of our explanatory variables at the treated sites to the corresponding averages observed in the reference set. The absolute differences are given by

$$\delta_p \quad = \quad |\bar{x}_p^{\text{TRT}} - \bar{x}_p^{\text{REF}}|, \qquad p = 1, \ldots, 8, \tag{8}$$

where TRT and REF denote the treated and reference sets respectively. If the treated and reference sites *are* exchangeable with respect to the explanatory variables, then the values calculated from (5) would not be significantly different to those obtained after a random allocation of 67 sites to the reference set and 56 to the treatment set. More specifically, in a permutation test we find the statistic of interest for every possible allocation of sites to each of the reference and treatment sets; the exact p–value P for a test of the null hypothesis H_0 : sites are exchangeable, is then found as the proportion of allocations for whom the statistic of interest is at least as extreme as that found under the real allocation. In our example, there are $\binom{123}{56} \approx 4.68 \times 10^{35}$ permutations: thus, a Monte Carlo permutation test can be used to obtain a sample of size N from the permutation distribution of δ_p by randomly choosing N permutations from all of those available. Let Π_k be one such random permutation, and let $\delta_p^{(\Pi_k)}$ be the absolute mean difference for variable x_p, as defined in Equation (5), for the treatment and control sets allocated under permutation Π_k. Let I_k be an indicator variable such that

$$I_k = \begin{cases} 1 & \text{if } \delta_p^{(\Pi_k)} \geq \delta_p \\ 0 & \text{otherwise;} \end{cases}$$

then an estimate of P is given by

$$P_{\text{est}} = \sum_{k=1}^{N} I_k/N,$$

giving $\mathbb{E}[P_{\text{est}}] = P$ and $\mathsf{var}(P_{\text{est}}) = P(1-P)/N$. From this we can, for example, obtain a confidence interval for P. Performing such Monte Carlo permutation tests with $N = 10^6$ shows only a marginally significant difference between the reference and treated sets for x_5, the eighty–fifth percentile speed: here, the 95% confidence interval for P is (0.047, 0.050). For all other explanatory variables, $P \gg 0.05$.

Considering all explanatory variables together, we also perform a Monte Carlo permutation test on the mean Mahalanobis distance of each site in the treatment set to sites in the reference set with mean $\bar{\mathbf{M}}^{\text{REF}} = (\bar{x}_1^{\text{REF}}, \ldots, \bar{x}_8^{\text{REF}})$ and covariance matrix $\mathbf{\Sigma}$, whose (s,t)–th entry is given by $\mathsf{cov}(x_s^{\text{REF}}, x_t^{\text{REF}})$, $s, t = 1, \ldots, 8$; that is, we compare

$$\bar{D} = \frac{1}{56} \sum_{j=1}^{56} \sqrt{(\mathbf{X}_j^{\text{TRT}} - \bar{\mathbf{M}}^{\text{REF}})^T \mathbf{\Sigma}^{-1} (\mathbf{X}_j^{\text{TRT}} - \bar{\mathbf{M}}^{\text{REF}})},$$

where $\mathbf{X}_j^{\text{TRT}}$ is the jth row of $\mathbf{X}$ for those sites treated with a safety camera, to a sample of size N from the permutation distribution of $\bar{D}$. Doing so, we get (0.165, 0.173) as the 95% confidence interval for P, further supporting the assumption of site exchangeability.

3.2. Empirical Bayes analysis

Having checked the assumption of exchangeability between the reference and treated sites, we now apply the EB method, as outlined in Section 2.1, to the data collected by the NSCP. Using a backwards elimination procedure for the selection of suitable explanatory variables, we obtain the following model for data at the 67 reference sites, using a negative binomial error structure:

$$\hat{\mu} = \exp\{1.93 - 0.04x_1 - 0.01x_2 + 0.44x_3 + 0.67x_{4I} + 0.85x_{5I} + 1.06x_{6I}\}, \tag{9}$$

where x_1, x_2 and x_3 correspond to the average observed speed (miles per hour), the percentage of drivers exceeding the speed limit and traffic flow (respectively, as defined earlier), and x_{4I}, x_{5I} and x_{6I} are indicator variables associated with road classification (variable x_7), where: $x_{4I} = 1$ for road classification 'A', $x_{5I} = 1$ for road clssification 'B' and $x_{6I} = 1$ for road classification 'C', each taking the value 0 otherwise. The maximum likelihood estimate for the negative binomial over–dispersion parameter is $\hat{\kappa} = 0.401$, giving $\hat{\gamma} = 1/0.401 = 2.494$. The usual diagnostic tools for assessing the fit of such regression models can be used to confirm the adequacy of this model for data collected at the reference group of sites. Although we can also obtain standard errors for estimates of the

Table 1. Results of EB analysis to account for RTM for four sites treated with safety cameras, as well as totals for all 56 safety camera sites.

	Empirical Bayes method						Difference	
	y_j	μ_j	α_j	$\mathbb{E}(m_j\|y_j)$	SD$(m_j\|y_j)$	$y_{j,\text{after}}$	Observed	After RTM
Site 2	4	1.43	0.63	2.38	0.936	0	–4	–2.38
Site 13	12	1.71	0.59	5.95	1.564	2	–10	–3.95
Site 39	7	1.31	0.65	3.29	1.069	2	–5	–1.29
Site 47	16	7.84	0.24	14.06	3.273	5	–11	–9.06
Total	**436**			**321**		**298**	**–138**	**–23**

regression coefficients and the estimate of the shape parameter γ, the EB procedure uses Equation (6) on data x_{1j}, x_{2j}, x_{3j}, x_{4Ij}, x_{5Ij} and x_{6Ij} collected at each safety camera site j, $j = 1, \ldots, 56$, and treats each resulting $\hat{\mu}_j$ as the 'true value', substituting this into Equation (2) along with the observed casualty frequency in the before period (y_j) to obtain the EB estimate of casualty frequency at each site j.

Table 1 shows numerical results for some of the individual safety camera sites, as well as overall totals for all sites in the NSCP study. For example, at site 13 there was an observed reduction in casualties from 12 in the before period to 2 in the after period; however, the EB estimate suggests that this would have reduced to about 6 anyway, giving a more realistic reduction, after RTM, of just 4 casualties. Across all 56 safety camera sites, the total observed reduction of 138 casualties between the before and after periods is reduced to just 23 after taking RTM into account. This suggests an RTM effect of $100(\sum_{\forall j} \mathbb{E}(m_j|y_j) - y_j)/\sum_{\forall j} y_j = -26.4\%$. This is towards the middle of the range reported by Hirst *et al.* (2004), discussed in Section 1.1.

3.3. *Fully Bayesian analysis*

We now formulate a fully Bayesian (FB) modelling framework to assess the effectiveness of the safety cameras introduced at the 56 sites in the NSCP study. In this section, we work with exactly the same Poisson–Gamma hierarchy as outlined in the EB approach in Section 2.1. However, we now unify the entire modelling procedure by assigning prior distributions to the regression coefficients β_i, $i = 0, \ldots, 6$, and the negative binomial over–dispersion parameter κ. We adopt non–informative, independent priors; specifically, we use

$$\beta_i \sim N(0, 100), \qquad i = 0, \ldots, 6 \text{ and}$$

$$\rho = \log(\kappa) \sim N(0, 100),$$

working with the natural logarithm of the negative binomial over–dispersion parameter to retain the positivity of κ. Inference proceeds by initialising each of the regression coefficients β_i and $\rho = \log(\kappa)$ at their prior means, and then using a random walk Metropolis–Hastings scheme (with data from the reference set) to update the chains. At each iteration R in the MCMC, the current values of the regression coefficients $\beta_i^{(R)}$ are used to obtain the posterior draw $\mu_j^{(R)}$ for each safety camera site j via Equation (4); the current values $\mu_j^{(R)}$ and $\gamma^{(R)} = \exp\{-\rho^{(R)}\}$ are then used as the mean and shape (respectively) of the gamma prior distribution for m_j. Since the gamma distribution is the conjugate prior for the Poisson distribution, Gibbs sampling can then be used for straightforward sampling from the full conditional distribution for m_j. After initial pilot runs to tune the efficiency of the sampler (i.e. by adjusting the variances of the random walk innovations to achieve acceptance probabilities of around 23%), the MCMC was allowed to run for 500,000 iterations. The entire procedure was repeated using a range of starting values for β_i and ρ to check for convergence.

Table 2 shows some posterior summaries for the regression parameters and the gamma shape parameter $\gamma = \exp\{-\rho\}$ after the removal of the burn–in period (the first 5000 iterations). Also shown are the corresponding posterior summaries for μ_j for the four safety camera sites we reported in the EB analysis (Table 1), as well as posterior summaries for m_j for these sites. At each iteration R we have also computed the total expected casualties $T^{(R)}$ across all 56 sites, given by

$$T^{(R)} = \sum_{j=1}^{56} m_j^{(R)}|y_j.$$

The posterior means for the regression coefficients (β_i) match up quite closely to the MLEs for these parameters as given in Equation (6); none of the 95% credible intervals for these parameters include zero, which also agrees with the earlier frequentist analysis in which the regression coefficients were all significant at, or beyond, the 5% level. The posterior means for β_i are all quite close to the medians, indicating fairly symmetric marginal posteriors.

The posterior means for m_j for the four sites reported here also compare quite closely to the posterior means for m_j obtained in the EB analysis and reported in Table 1; however, as can be seen when comparing the standard deviations for m_j in Table 2 to those from the EB analysis in Table 1, posterior variability is substantially greater in the FB analysis. This is, of course, because we have now acknowledged the variability of the regression parameters β_i and hence the mean casualty rate μ_j, $j = 1, \ldots, 56$, through the prior distributions for β_i. Although our prior distributions for the regression coefficients are rather vague, and so probably act to *over*–state these sources of variability, this is still potentially more attractive than assuming that the MLEs for β_i (and hence μ_j) are the true values, as is the case in the EB approach.

The positive skew of the marginal posteriors for m_j is also noticeable when we inspect the posterior draws for m_j from the FB analysis; for four sites in particular, this gives a substantial difference between the posterior mean and median (site 47, as reported in Table 2, is one such site). Most studies using the standard EB framework for estimating RTM rely wholly on Equation (2) – essentially the analytical posterior mean. Clearly, for locations such as site 47, the mean might not be the most representative summary to use. *Not so strong now... obviously, you can get the median of the gamma in EB via simulation!*

Figure 4 (top) shows posterior summaries for m_j for *all* 56 safety camera sites; sites 5, 33, 36 and 47 can clearly be seen to have the largest mean–median difference. The bottom plot in Figure 1 shows the posterior means and medians (circles and crosses respectively, as in the top plot) plotted against the EB estimates of casualty frequency for each of the 56 safety camera sites. For most sites there is little difference between either FB posterior summary (mean/median) and the EB estimates of casualty frequency, though again we can clearly see discrepancies from the line of equality when using the posterior median at sites 5, 33, 36 and 47. The effect of using the mean at such sites (as is done in a standard EB

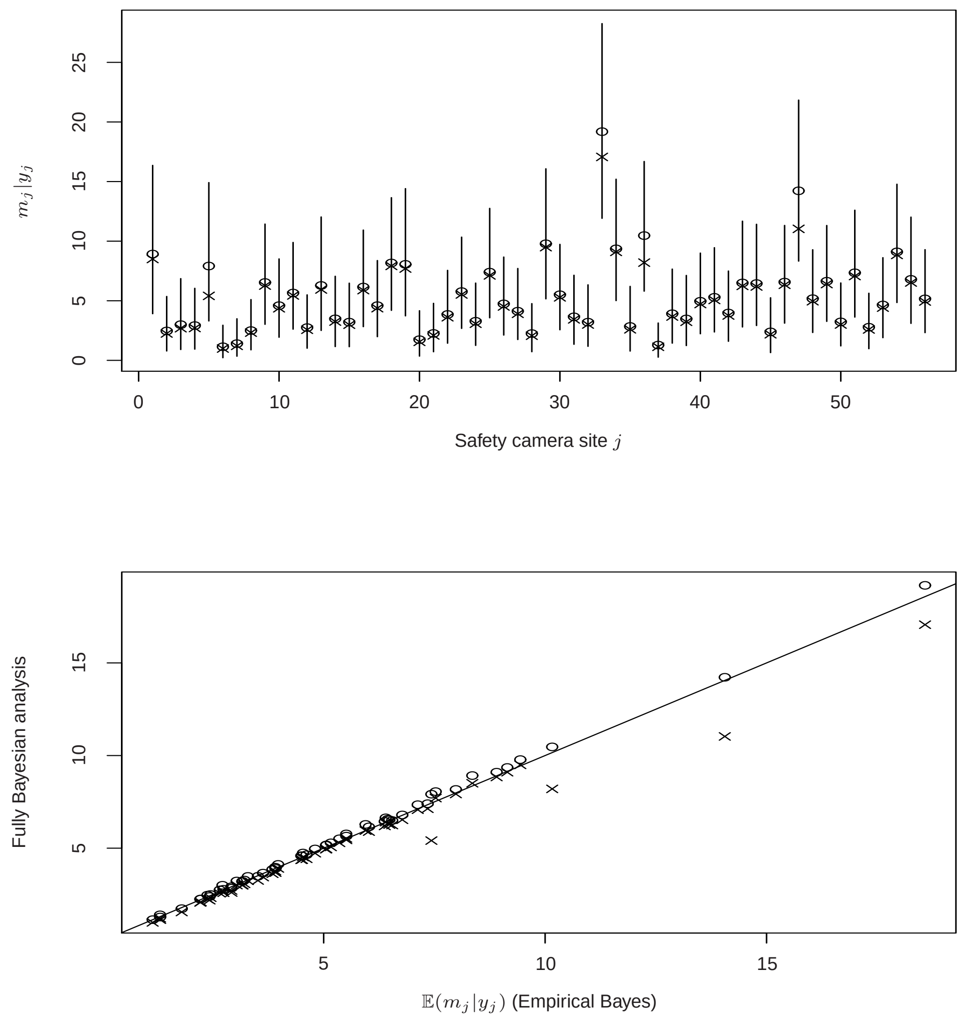

Fig. 3. Posterior means (circles) and medians (crosses) from the fully Bayesian analysis: along with 95% credible intervals, for each safety camera site (top); plotted against the corresponding empirical Bayes estimate of casualty frequency (bottom).

Table 2. Posterior summaries for a fully Bayesian analysis to account for RTM. The posterior means and medians of the regression parameters β_i can be compared to MLEs from the empirical Bayes analysis (Equation 6); posterior summaries for μ_j and m_j for the four sites reported here can be compared to those from the empirical Bayes analysis given in Table 1.

		Posterior			
		Mean	St. dev.	Median	95% credible interval
	β_0	1.981	0.544	1.981	(0.900, 3.044)
	β_1	–0.042	0.016	–0.042	(–0.074, –0.012)
	β_2	–0.013	0.004	–0.013	(–0.021, –0.005)
	β_3	0.476	0.218	0.474	(0.063, 0.912)
	β_4	0.648	0.437	0.651	(0.012, 1.506)
	β_5	0.840	0.451	0.839	(0.043, 1.733)
	β_6	1.059	0.400	1.058	(0.284, 1.845)
	$\gamma = \exp\{-\rho\}$	2.281	0.756	2.146	(1.201, 4.111)
μ_j	Site 2	1.534	0.639	1.416	(0.649, 3.105)
	Site 13	1.903	0.908	1.716	(0.725, 4.209)
	Site 39	1.411	0.601	1.301	(0.586, 2.907)
	Site 47	8.195	1.767	7.986	(5.355, 12.245)
m_j	Site 2	2.465	1.188	2.255	(0.780, 5.369)
	Site 13	6.283	2.447	5.946	(2.502, 12.039)
	Site 39	3.475	1.527	3.239	(1.232, 7.133)
	Site 47	14.225	3.466	11.032	(8.320, 21.842)
	Total T	322	23.833	308	(289.92, 369.97)

approach) instead of a more appropriate posterior summary will be to understate the effect of RTM: as the bottom row of Table 2 shows, when we sum casualty frequencies for all sites at each iteration in the MCMC, the posterior median total expected casualty frequency is just 308, compared to a mean of 322 (and an analytical mean, as found in the EB analysis, of 321, as given in Table 1). Thus, using the median gives a larger RTM effect than under the standard EB analysis (–29.4% *c.f.* –26.4%). *take out?*

3.4. Implications for the demand on secondary healthcare

One of the aims of the original NSCP study (Colligan *et al.*, 2008) was to estimate the financial consequences of the implementation of the safety cameras for local NHS secondary healthcare providers. If the safety cameras helped 'save' a total of y casualties, what *would* these y casualties have cost the NHS, in terms of treatment, if they *had* occurred? A large, multi–stage data linking exercise took place in the NSCP study to link police collision data from the before period to NHS casualty data. This was attempted using unique identifiers which are collected by both the police, at the scene of the accident, and the hospital involved. Rather disappointingly, it was only possible to match about 44% of the 436 casualties in the before period to corresponding hospital records. However, as Colligan *et al.* (2008) point out, this is in line with matching rates obtained in other studies that have attempted to link police records to hospital databases; in a similar exercise, Simpson (1996) achieved a 46% success rate and Tunbridge *et al.* (1988) matched 56% of cases. Cryer *et al.* (2001) suggest reasons for differences between hospital and police data which can lead to low matching rates.

NHS hospital Accident and Emergency (A&E) patients fall into one of eight Health Resource Group (HRG) categories depending on the severity of their injuries and the extent of treatment required. These A&E HRG categories range from 'high cost' category V13.1 which includes computerised tomography (CT) scans and magnetic resonance imaging (MRI) scans, to relatively 'low cost' categories such as V063.1, involving more routine urine/bacteriological/haematological investigations. Most 'high cost' allocations (and possibly some 'low cost' allocations) would then probably require admission to hospital for inpatient treatment. Inpatients are allocated to one of some 700 inpatient HRGs (see Colligan *et al.* (2008), Section 3.9.2, for full details of all A&E and inpatient HRGs). Each A&E HRG has an associated financial tariff, as does each inpatient HRG; however, overall inpatient tariffs are also computed as a function of time, whereby the longer a patient spends in hospital the greater the overall inpatient tariff for that patient.

Since overall individual inpatient tariffs are calculated as a function of time, instead of considering the 700 inpatient HRGs themselves we consider inpatient tariffs in groups of £500, and so have inpatient tariff categories of £0 (no inpatient treatment), £1–£500, £501–£1000, £1001–£1500, ... Each inpatient falls into one of our tariff categories, and would have first been an A&E casualty falling into one of the eight A&E HRG categories. Thus, we consider each A&E HRG/inpatient tariff category combination t, with associated combined financial tariff £C_t, as a multinomial outcome whose probabilities p_t are just the observed proportions (of the 44% matched cases) falling into each t in the 'before' period.

The estimated saving to the NHS by implementing the safety cameras in the Northumbria region can then be obtained, in the EB analysis, by multiplying the total change in casualty frequency after RTM by each p_t; these expected frequencies are then converted into expected financial savings by multiplying by the financial tariff associated with each corresponding category t, and so the total expected financial saving £S is found as

$$\mathbb{E}(S) = \left(\sum_{j=1}^{56} \mathbb{E}(m_j|y_j) - y_{j,\mathrm{after}} \right) \sum_{\forall t} p_t C_t. \tag{10}$$

The A&E contributions to each C_t take a fixed value; however, since we have partitioned the inpatient tariffs into groups of £500, we use the midpoint of each inpatient tariff to obtain the combined tariff £C_t for each category t.

In the FB analysis we can obtain posterior draws for the expected number of casualties falling into each category t, by multiplying each posterior draw for the total change in casualty frequency by the corresponding p_t. Then, posterior draws for the expected financial saving to the NHS can be obtained by multiplying each expected frequency by the associated financial tariff for category t, giving, at each iteration R,

$$S^{(R)} = \left(T^{(R)} - \sum_{j=1}^{56} y_{j,\mathrm{after}} \right) \sum_{\forall t} p_t C_t. \tag{11}$$

Table 3 compares results from the standard EB analysis to posterior summaries from the FB approach. Again, since we have a sample of posterior draws in the FB analysis, we can summarise our findings in whichever way we deem appropriate, which can include a

Table 3. Empirical Bayes estimates of total expected financial savings to the NHS (£S thousand), and total value of prevention including human costs and lost output (£S^* thousand) owing to the implementation of the safety camera scheme in Northumbria (left); posterior summaries from the fully Bayesian analysis (right).

Thousand £		Empirical Bayes	Posterior Mean	St. dev.	Median	95% credible interval
	Midpoint	25.6	24.9	13.2	24.4	(0.3, 57.5)
S	Minimum	23.5	22.8	12.1	22.3	(0.1, 52.5)
	Maximum	27.7	27.1	14.4	26.5	(0.6, 62.5)
S^*		1215.6	1529.8	786.3	1479.8	(45.6, 4122.3)

95% credible set for S. Also shown in Table 3 are the corresponding results when using the minimum and maximum bounds for each inpatient tariff class.

Since 1993, the valuation of casualties has been based on a consistent willingness to pay approach, which encompasses *all* aspects of the valuation of casualties, including human costs (representing pain, grief and suffering to the casualty, relatives and friends) and loss of output due to injury. When combined with the direct cost of medical treatment, the UK DfT (2009) puts the average total cost of a road casualty at £52,850. Thus, replacing $\sum_{\forall t} p_t C_t$ in Equations (7) and (8) with 52850 will give an EB estimate/posterior draws (respectively) of S^*, the total (average) value of prevention owing to the implementation of the safety cameras in the Northumbria region, now accounting for human costs and lost output, as well as the cost of hospital treatment. The value of £52,850 per prevented casualty is, of course, an average; this will vary depending on the severity of the casualty's injuries, and the type of road user (e.g. car occupant, goods vehicle occupant, motorcycle user etc.). Posterior summaries for S^* are given in the bottom row of Table 3.

4. Further modelling considerations

In the previous Section, the standard EB procedure was used to estimate the contribution of RTM to the reduction in casualty frequency at 56 sites treated with safety cameras. Relative to an FB analysis, the standard EB approach was shown to be over–optimistic in its estimation of the variability of casualty frequency at each site. Further, an FB analysis gives us a posterior sample which we can summarise in whichever way we deem appropriate; as demonstrated with the NSCP safety camera data, for some sites, the posterior mean – as provided by the EB approach – might not be the most suitable summary to use.

The previous section provided a like–for–like comparison between the typical EB approach for estimating the RTM effect and an FB analysis. We now investigate the sensitivity of our results from this FB treatment to the choice of prior distributions for m_j. We investigate the use of more informative priors for the regression coefficients in Equation (4), giving some thought to the issue of variable selection in the Bayesian framework. We also quantify the contribution of trend to the reduction in casualty figures after the implementation of the safety cameras.

4.1. Sensitivity to other forms of prior distribution

We now assess the sensitivity of the results in Section 3 to the choice of prior distribution for m_j. Recall that the gamma prior mean and variance for m_j were μ_j and μ_j^2/γ respectively. We now use the lognormal (mean $= \lambda_j$, variance $= \sigma^2$) and Weibull (shape $= \omega$, scale $= \nu_j$) distributions as priors for m_j, keeping the mean and variances the same as in the original gamma prior to allow relative comparisons of the effects of using these different priors. This gives

$$\begin{aligned} \lambda_j &= \log(\mu_j) - \frac{1}{2}\log(1+\gamma^{-1}) \qquad \text{and} \\ \sigma^2 &= \log(1+\gamma^{-1}) \end{aligned}$$

for the lognormal prior; for the Weibull prior, we solve

$$\omega\frac{\Gamma(2\omega^{-1})}{\Gamma^2(\omega^{-1})} = \frac{1}{2}(1+\gamma^{-1})$$

for ω and then use

$$\nu_j = \frac{\mu_j}{\Gamma(1+\omega^{-1})}.$$

A summary of results based on the lognormal and Weibull priors for m_j is shown in Table 4, along with the results for the gamma prior from Section 3. There is clearly some agreement between results obtained using the original gamma prior and the Weibull prior. However, using the lognormal prior gives a considerably higher total number of expected casualties. For example, the posterior mean total casualties is 355; comparing this with the number of observed casualties in the after period (298) would suggest the safety cameras had been more effective than if we had used the gamma or Weibull priors, with less contribution to any change in casualty frequency attributed to RTM. In fact, the 95% credible interval for the effect of RTM when using the lognormal prior does not include the value given in the EB analysis, or indeed the posterior summaries of average from the analyses using the other two forms of prior for m_j. This is reflected in the estimation of S and S^* with, for example, the median total value of prevention due to the safety cameras, when using the lognormal prior (about £2.8 million), being almost twice that when using the gamma prior (£1.5 million) and even greater still than when using the Weibull prior (less than £1 million).

The Deviance Information Criterion (DIC), as discussed in Spiegelhalter *et al.* (2002), can be used to compare the three model formulations used. The DIC is akin to the Akaike information criterion and is based on an estimate of the log–likelihood, but includes a penalty for the number of parameters; hence, it can be used to compare alternative models. For the Poisson–gamma, Poisson–lognormal and Poisson–Weibull, we have a DIC of 693.3, 787.2 and 645.6 (respectively), suggesting the Weibull prior for m_j might be the most appropriate distribution (of the three tried) to use here. **Need some justification for the Weibull here – see referee 2's comments**

4.2. Choice of prior distribution for the regression coefficients

The FB analysis in Section 3.3 used independent Normal priors with large variances for the regression coefficients β_i. We now consider more appropriate prior specifications for

Table 4. Posterior summaries for: the total expected number of casualties (T); the regression to the mean effect (RTM); the expected financial savings to the NHS secondary healthcare providers as a result of implementing the safety cameras (S); and the total average value of prevention including human costs and lost output (S^*).

	EB	Gamma Mean	Gamma Median	Lognormal Mean	Lognormal Median	Weibull Mean	Weibull Median
		(95% CI)		(95% CI)		(95% CI)	
T	321	322	308	355	338	317	303
		$(290, 370)$		$(309, 394)$		$(296, 371)$	
RTM (%)	–26.4	–26.5	–29.7	–18.9	–22.8	–27.6	–30.8
		$(-35.6, -14.2)$		$(-26.3, -9.0)$		$(-39.3, -15.3)$	
S (thousand £)	25.6	24.9	24.4	29.3	29.3	25.3	24.9
		$(0.3, 57.5)$		$(6.1, 73.5)$		$(0.7, 70.9)$	
S^* (thousand £)	1215.6	1529.8	1479.8	2803.0	2801.0	986.3	951.3
		$(45.6, 4122.3)$		$(581.4, 5126.5)$		$(69.1, 4910.9)$	

the regression coefficients that more suitably capture the variability of these parameters, as well as any dependencies between them. We consider two forms of prior for the regression coefficients: a 'genuinely' informative prior developed by eliciting priors for the mean number of casualties at observed levels of covariates (known as a *conditional mean prior*), and a *data augmentation prior* making use of the covariate values themselves.

4.2.1. *Using a data augmentation prior*

An attractive method for augmenting a Bayesian analysis in the absence of any external or expert prior information is to adopt a data augmentation prior (DAP). One such prior for the general linear model is the reference informative prior introduced by Zellner (1986). Bové and Held (2010) discuss extensions of the reference informative prior to generalised linear models, including the unit information prior as suggested by Ntzoufras (2009) for Poisson regression. Here, the following informative prior for $\boldsymbol{\beta}_{\backslash 0,n_p}$ is used, where $\boldsymbol{\beta}_{\backslash 0,n_p}$ is the n_p–dimensional parameter vector of regression coefficients β_i excluding β_0:

$$\boldsymbol{\beta}_{\backslash 0,n_p} \quad \sim \quad N_{n_p}\left(\mathbf{0}, n(\mathbf{X}_{n_p}^T\mathbf{X}_{n_p})^{-1}\right), \tag{12}$$

where $\mathbf{X}_{n_p}$ is the design matrix without the first column that corresponds to the constant β_0, $\mathbf{0}$ is the zero vector (of length n_p) and n is the sample size. Ntzoufras (2009) suggests using a vague prior for β_0 such as that previously used, e.g. a zero mean Normal with large variance. Using the prior for $\boldsymbol{\beta}_{\backslash 0,n_p}$ given in (9), and using gamma, lognormal and Weibull priors for m_j, $j = 1, \ldots, 56$, gives DIC values of 733.4, 813.7 and 685.4 (respectively), indicating, as before, that the Poisson–Weibull structure provides the best fit. The columns in Table 6 labelled "DAP" summarise the MCMC runs for the Poisson–Weibull model; as before, the chains were allowed to run for 500,000 iterations, and the first 5000 were discarded as burn–in. Comparing inferences for the regression coefficients β_i to those in Table 2, where independent zero–mean Normal priors with large variances were used, we see a reduction in posterior variability. This follows through to inferences for the total number of expected casualties (T), as well as the RTM effect, both of these having narrower 95% credible intervals than when the non–informative priors were used (see the right–hand–side of Table 4); similarly, the 95% credible intervals for S and S^* are narrower when using the DAP.

Table 5. Elicited prior parameters for $\tilde{M}_p$.

			Gamma		Lognormal		Weibull	
	a_p	b_p	Shape	Scale	Mean	Variance	Shape	Scale
$\tilde{M}_1$	9.93	11.63	8.48	0.85	2.24	0.11	3.20	11.09
$\tilde{M}_2$	2.77	1.64	4.68	1.69	0.92	0.19	2.29	3.13
$\tilde{M}_3$	3.59	3.10	4.16	1.16	1.17	0.22	2.15	4.05
$\tilde{M}_4$	3.12	1.87	5.21	1.67	1.05	0.18	2.43	3.52
$\tilde{M}_5$	8.19	6.25	10.73	1.31	2.06	0.09	3.64	9.08
$\tilde{M}_6$	5.42	3.79	7.75	1.43	1.63	0.12	3.04	6.07

Table 6. Posterior summaries for: the regression coefficients β_p; the total expected number of casualties across all safety camera sites T; the regression to the mean effect (RTM); the estimated financial savings to NHS secondary healthcare providers as a result of implementing the safety cameras (£S thousand); and the total average value of prevention including human costs and lost output (£S^* thousand). Results are shown for the data augmentation prior (DAP) and the conditional mean prior (CMP) for $\boldsymbol{\beta}$, using Weibull priors on M_j.

	Posterior							
	Mean		St. dev.		Median		95% credible interval	
	DAP	CMP	DAP	CMP	DAP	CMP	DAP	CMP
β_0	1.785	1.543	0.463	0.521	1.790	1.549	(0.86, 2.69)	(0.51, 2.58)
β_1	–0.038	–0.025	0.014	0.013	–0.038	–0.025	(–0.07, –0.01)	(–0.05, 0.00)
β_2	–0.012	–0.013	0.004	0.004	–0.012	–0.013	(–0.02, –0.01)	(–0.02, –0.01)
β_3	0.464	0.318	0.217	0.188	0.458	0.313	(0.06, 0.90)	(0.04, 0.70)
β_4	0.701	0.799	0.426	0.412	0.701	0.802	(0.01, 1.55)	(0.03, 1.61)
β_5	0.886	0.913	0.442	0.432	0.882	0.913	(0.03, 1.77)	(0.06, 1.76)
β_6	1.101	1.142	0.385	0.395	1.097	1.140	(0.35, 1.87)	(0.37, 1.93)
γ	2.194	2.083	0.725	0.671	2.072	1.969	(1.16, 3.95)	(1.12, 3.71)
T	327	327	23.316	25.528	314	313	(286, 352)	(285, 354)
RTM %	–25.3	–25.3	6.265	5.855	–28.3	–28.5	(–35.3, –26.0)	(–36.6, –23.6)
S	31.4	30.9	15.128	14.092	31.0	30.9	(0.9, 61.9)	(0.9, 58.5)
S^*	1584.9	1541.5	448.4	349.1	1547.7	1541.7	(73.4, 4507.3)	(72.6, 4183.2)

4.2.2. *A conditional mean prior*

Need to reconsider ntation here: use of lowercase m (already used), PLUS should mean vector notation not be in line with Section 3.1? We now attempt to construct a truly informative prior for the vector of regression coefficients $\boldsymbol{\beta}$. Following Bedrick *et al.* (1996), we elicit a prior on $\tilde{\mathbf{M}} = (\tilde{M}_1, \ldots, \tilde{M}_{n_p})$ where the $\tilde{M}_p$'s are mean responses at covariates $\mathbf{x}_p$, $p = 1, \ldots, n_p$. We denote by $\tilde{\mathbf{X}}$ the $n_p \times n_p$ non–singular matrix with $\mathbf{x}_p^T$ in the pth row. Following the notation of Bedrick *et al.* (1996), $\mathbf{G}$ and $\mathbf{G}^{-1}$ are vector transformations that apply g and g^{-1} to each element; for example, $g(\cdot) = \log(\cdot)$, $g(\cdot) = \mathrm{logit}(\cdot)$ and $g(\cdot) = \Phi^{-1}(\cdot)$, where Φ is the distribution function of the standard Normal, for Poisson, logistic and probit regression respectively. Assessing the $\tilde{M}_p$'s independently, the conditional mean prior is

$$\pi_0(\tilde{\mathbf{M}}) \quad = \quad \prod_{p=1}^{n_p} \pi_{0_p}(\tilde{M}_p). \tag{13}$$

Writing

$$\tilde{\mathbf{M}} = \mathbf{G}^{-1}(\tilde{\mathbf{X}}\boldsymbol{\beta}) \qquad \text{and} \qquad \boldsymbol{\beta} = \tilde{\mathbf{X}}^{-1}\mathbf{G}(\tilde{\mathbf{M}})$$

induces a prior on $\boldsymbol{\beta}$ of the form

$$\pi(\boldsymbol{\beta}) \quad = \quad \prod_{p=1}^{n_p} \pi_{0_p} g^{-1}(\tilde{\mathbf{x}}_p^T \boldsymbol{\beta}) \Bigg/ |\tilde{\mathbf{X}}^{-1}| \prod_{p=1}^{n_p} \dot{g}(\tilde{M}_p)$$

where, generically, $\dot{f}(x) = \partial f(x)/\partial x$.

Using the same four covariates as in the original analyses in Section 3 (average observed speed, percentage of drivers exceeding the speed limit, traffic flow and road classification) requires us to elicit priors on $\tilde{\mathbf{M}} = (\tilde{M}_1, \ldots, \tilde{M}_6)$ since road classification is a factorial variable requiring three indicators. A regression analysis from a previous study of casualty frequencies at another group of sites in the Northumbria region gives a regression equation of the form in (6); covariates $\mathbf{x}_p$, $p = 1, \ldots, 6$, at six of these sites can be used to suggest means a_p and variances b_p for each $\tilde{M}_p$. Suitable priors for $\tilde{M}_p$ can then be proposed – for example, gamma distributions with means and variances a_p and b_p (respectively), or indeed lognormal or Weibull distributions as used in Section 4.1. For six sites used in this previous study, we obtain

$$\begin{aligned} \mathbf{a} &= (9.93, 2.77, 3.59, 3.12, 8.19, 5.42)^T \qquad \text{and} \\ \mathbf{b} &= (11.63, 1.64, 3.10, 1.87, 6.25, 3.79)^T. \end{aligned}$$

Using these means and variances we can obtain hyper–parameters for the priors on $\tilde{M}_p$: Table 5 reports these hyper–parameters when using gamma, lognormal and Weibull priors.

Using the DIC to assess fit when using gamma, lognormal and Weibull distributions for π_{0_p} and m_j, $p = 1, \ldots, 6$ and $j = 1, \ldots, 56$, once again suggests the Poisson–Weibull structure is best. Thus, Table 6 also reports posterior summaries from a Poisson–Weibull analysis using the conditional mean prior for $\boldsymbol{\beta}$ (columns labelled "CMP"). As when using the data augmentation prior, we see that we have smaller posterior standard deviations for the regression coefficients than in the analysis using non–informative priors, and again this has followed through to give greater posterior precision to our estimates of the effects of RTM and T, S and S^*.

4.2.3. Issues of model selection

In all the analyses performed so far, only four of the original eight predictor variables have been used to estimate μ_j. These four were selected in the original EB analysis using frequentist regression techniques; a backwards elimination procedure indicated that only the average observed speed, the percentage of drivers exceeding the speed limit, traffic flow and road classification were significant in the model – regression coefficients associated with speed limit, 85th percentile speed, percentage of drivers at least fifteen miles per hour over the speed limit and road type were deemed not to be significantly different from zero. In a Bayesian setting, the data augmentation prior, given by (9), can be used to aid variable selection. If the full model, defined by design matrix $\mathbf{X}_{n_p}$, has the prior for $\boldsymbol{\beta}_{\backslash 0, n_p}$ given in (9), then for any submodel defined by a reduced design matrix $\mathbf{X}_{n_p'}$,

$$\boldsymbol{\beta}_{\backslash 0, n_p'} \quad \sim \quad N_{n_p'}\left(\mathbf{0}, n(\mathbf{X}_{n_p'}^T \mathbf{X}_{n_p'})^{-1}\right)$$

where $n'_p < n_p$. Then models including/excluding each covariate can be compared by the computation of associated predictive probabilities.

Our full model has eleven possible covariates (recall that two of the original eight covariates – road classification and road type – require three and two indicators respectively) giving $2^{11} = 2048$ possible models. The marginal log–likelihoods are calculated for each model; the model with the largest marginal log–likelihood is then identified as the 'best'. The R package LearnBayes (Albert, 2009) can be used to implement this method of model selection. Doing so reveals that the model with the largest associated marginal log–likelihood is that which includes exactly the same covariates as identified in the EB analysis – average observed speed, the percentage of drivers exceeding the speed limit, traffic flow and road classification, with a marginal log–likelihood of –199.01. Alternatively, Bayes Factors (see Lee (1989), for example) could be used to aid model selection.

Some might argue that in a truly Bayesian analysis, there is no fundamental reason why *any* of the predictors should be removed. In fact, it might be argued that it is not coherent for a Bayesian to believe that their predictions will be improved by ignoring some information. Thus, the default Bayesian position might be to use all covariates in the log–linear model for μ_j. In fact, repeating the analyses in Sections 4.2.1 and 4.2.2 (results for which are summarised in Table 6) but using information on all covariates barely changes inferences on the effects of RTM, S and S^*.

4.3. Accounting for trend

In the original NSCP report (Colligan *et al.*, 2008), the EB procedure, as outlined in Section 2.1, was employed to assess the contribution of RTM to the change in casualty frequency owing to the implementation of safety cameras in the Northumbria region. However, nothing was done to allow for other non–treatment effects such as general trends in casualty frequencies over time. To account for this, we now specify the following modified form for μ_j:

$$\mu_j = \xi \exp\left\{\beta_0 + \sum_{p=1}^{n_p} \beta_p x_{p_j}\right\}, \tag{14}$$

where ξ is a trend effect constant across all sites j. Reports detailing the number of road traffic accident casualties in the Northumbria region as a whole suggest that, since the mid–1970s, the occurrence of casualties has been in steady decline. In fact, overall road traffic accident casualties have fallen by around 2% per year since 2001 (the start of the before period in the NSCP study). Since the difference between the mid–point of the before and after periods is three years (2002–2005) we adopt the following prior for ξ:

$$\xi \sim U(0.94, 1.015),$$

the uniform upper bound being slightly greater than 1 to reflect the fact that, since 2005, the number of (reported) 'slight' casualties has increased, on average, by about 0.5% per year.

Table 7 reports posterior summaries for the total number of expected casualties across

Table 7. Comparison of posterior distributions for: the total number of expected casualties at sites treated with safety cameras (T); expected financial saving to the NHS as a result of implementing the safety cameras (£S thousand); and the total average value of prevention (£S^* thousand), when accounting for RTM both with and without trend. Results are shown for an analysis using the conditional mean prior for $\boldsymbol{\beta}$ and Weibull priors for m_j.

		Posterior			
		Mean	St. dev.	Median	95% credible interval
T	*without trend*	327	25.528	313	(285, 354)
	with trend	323	25.709	309	(273, 340)
S	*without trend*	30.9	14.092	30.9	(0.9, 58.5)
	with trend	28.7	14.192	28.8	(0.9, 56.6)
S^*	*without trend*	1541.5	349.136	1541.7	(72.6, 4183.2)
	with trend	1318.5	358.725	1324.5	(68.5, 3979.6)

the safety camera sites (T), the expected financial saving to the NHS as a result of implementing the safety camera scheme in the Northumbria region (S) and the total average value of prevention (S^*), now accounting for trend through the modified form for μ_j given in Equation (11). A Poisson–Weibull model structure was used, with the conditional mean prior for $\boldsymbol{\beta}$ as outlined in Section 4.2.2. Also shown in Table 7, for comparison, are the corresponding posterior summaries when trend has *not* been accounted for, as given in Table 6. The effect of including the trend parameter ξ is obvious: the median number of expected total casualties decreases from 313 to 309 (a decrease from 327 to 323 if the mean is used), with corresponding decreases in expected value of prevention (both S and S^*) owing to the implementation of the safety cameras.

5. Conclusions

In this paper, we have illustrated the shortcomings of the standard tool for assessing the effectiveness of road safety remedial schemes using a case–study of mobile safety cameras in the UK. We have shown the EB method to be over–optimistic in its assessment of variability of estimates of casualty frequency at individual sites treated with mobile safety cameras. A fully Bayesian treatment can be used to appropriately account for all sources of variability by specifying prior distributions for the regression coefficients β_i, whereas the standard EB procedure treats the estimates of these coefficients from a frequentist regression analysis as the true values. Thus, a fully Bayesian analysis also leads to more realistic estimates of variability of other quantities of interest – in particular, we investigate the value of prevention, in monetary terms, owing to the implementation of the safety cameras.

Using MCMC techniques within a fully Bayesian framework also gives the practitioner much more flexibility – both in terms of the prior distributions used and the way in which the resulting posteriors are summarised. For too long, assessment of the effectiveness of road safety schemes has relied on a Poisson–gamma model structure, using the resulting closed–form expression for the associated posterior mean as the estimate of casualty frequency at each site *had no safety camera been used*. In fact, as we show, estimates of casualty frequency and RTM can be sensitive to the choice of prior used for the Poisson mean; other prior distributions can provide a better fit to the data, and the posterior mean might not always be the most appropriate summary of casualty frequency to use. We have also shown that the extra variability induced by a fully Bayesian treatment can be reduced by implementing more informative priors in our analysis.

After accounting for trend, we have shown that, at our 56 sites treated with mobile safety cameras, we might expect the total number of casualties to have reduced from 436 to about 309 anyway (posterior median in Table 7), even if the safety cameras had not been used. Comparing the posterior distribution for the total number of expected casualties T with the number of casualties in the period *after* implementation gives a total median value of prevention of just over £1.3 million for the 56 sites in this study; of course, grossing up to the national level would greatly increase this figure.

Working within a fully Bayesian framework combines both steps of the original EB method (estimating μ_j and then estimating the posterior distributions for m_j) into one seamlessly integrated procedure. Applying fully Bayesian techniques opens up a field of opportunities for those interested in – or required to – assess the effectiveness of road safety schemes. More complicated, but realistic, model structures can be used that could otherwise prove very difficult to draw inferences from. Clearly, obtaining these more accurate estimates of the impacts of road safety measures is crucial to guiding increasingly limited investment opportunities.

Acknowledgements
We would like to thank Ian Abernethy (Gateshead Council) and the NSCP for providing the data used in this work. We would also like to thank Professor Mike Maher for providing feedback on an earlier version of this paper.

References

Albert, J. (2009). *Bayesian Computation with R.* Springer, London.

Bedrick, E.J., Christensen, R. and Johnson, W. (1996). A new perspective on priors for generalized linear models. *J. Am. Statist. Ass.*, **51**, pp. 211—218.

Bové, D.S. and Held, L. (2010). Hyper g–priors for generalised linear models. *Preprint.*

Colligan, J., Goulbourne, L.E., Thorpe, N., Fawcett, L. and McNay, A. (2008). Mobile Speed Cameras and the Demand for Secondary Health Care in Northumbria. *Northumbria Safety Camera Partnership.*

Cryer, P.C., Westrup, S., Cook, A.C., Ashwell, V., Bridger, P. and Clarke, C. (2001). Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports. *Injury Prevention*, **7**, pp. 234—241.

Department for Transport (2010). *Reported Road Casualties in Great Britain 2009.* The Stationery Office, London.

Department for Transport (2009). Benefits to Society Arising from Prevention of Road Accidents and Casualties. Department for Transport, London. (Available from http://www.dft.gov.uk/webtag/documents/expert/unit3.4.php)

Department for Transport (2000). *Tomorrow's roads – safer for everyone.* The Stationery Office, London.

Hauer, E., Harwood, D.W., and Griffith, M.S. (2001). Estimating Safety by the Empirical Bayes Method: A Tutorial. *Transportation Research Record*, 1784, pp. 126—131.

Hauer, E. (1980). Bias–by–selection: overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment. *Acci. Anal. & Prev.*, **12**, 2, pp. 113—117.

Hauer, E., Ng, J. and Lovell, J. (1988). Estimation of safety at signalised intersections. *Transport. Res. Rec.*, 1467, pp. 30—37.

Hirst, W.M., Mountain, L.J. and Maher, M.J. (2004). Sources of error in road safety scheme evaluation: a quantified comparison of current methods. *Acci. Anal. & Prev.*, **36**, pp. 705—-715.

Hooke, A., Knox, J. and Portas, D. (1996). Cost benefit analysis of traffic light and speed cameras. Police research series, paper 20. The Stationery Office, London.

Jarrett, D.F. (1994). Statistical analysis of road accident frequencies. In *Proc. Symposium on Safety in the Road Environment*, LNEC, Lisbon.

Kulmala, R. (1994). Measuring the safety effect of road accident measures at junctions. *Acci. Anal. & Prev.*, **6**, 781—794.

Lee. P.M. (1989). *Bayesian Statistics: An Introduction.* Wiley, Chichester.

Li, W., Carriquiry, A., Pawlovich, M. and Welch, T. (2008). The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. *Acci. Anal. & Prev.*, **40**, pp. 1531—1542.

Lord, D., and Miranda–Moreno, L.F. (2008). Effects of low sample mean values and small sample sizes on the parameter estimation of hierarchical Poisson models for motor vehicle crashes: a Bayesian perspective. *Journal of Safety Science*, 46, pp. 751—770. m Maher, M.J. (1991). The evaluation of road safety schemes: problems with statistical appraisal. In *Proc. County Surveyors' Society/University Transport Studies Group Joint Seminar on the Evaluation of Road Safety Schemes*, London.

Maher M.J. and Mountain L.J. (2009). The sensitivity of estimates of regression to the mean. *Acci. Anal. & Prev.*, **41**(4), pp. 861—868.

Miaou, S.P. and Lord, D. (2003). Modeling traffic crash–flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. *Transport. Res. Rec.*, 1840, pp. 1310—1340.

Mountain, L., Fawaz, B. and Sineng, L. (1992). The assessment of changes in accident frequencies on link segments: a comparison of four methods. *Traffic Engineering and Control*, **33**, 2, pp. 429—431.

Pendleton, O.J. (1991). Application of new accident analysis methodologies. Vol.1: New methodologies. *US Department of Transportation.* Report FHWA–RD–90–091.

Persaud, B. and Dzbik, L. (1993). Accident prediction models for freeways. *Transport. Res. Rec.*, 1401, pp. 55—60.

Simpson, H.F. (1996). Comparison of Hospital and Police Casualty Data: A National Study, TRL Report No. 173. Transport Research Laboratory, Crowthorne.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit (with Discussion), *J. R. Statist. Soc.*, B, **64**(4), pp. 583—616.

Tunbridge, R.J., Everest, J.T. and Wild, B.R. and Johnstone, R.A. (1988). An in–depth study of road accident casualties and their injury pattern, TRRL Report No. 136. Transport and Road Research Laboratory, Crowthorne.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with *g*–prior distributions, in: Goel, P.K. & Zellner, A. (Eds), Bayesian Inference and Decision Techniques: essays in honour of Bruno de Finetti, pp. 233—243. Amsterdam, North Holland.

Ztoufras, I. (2009). *Bayesian Modelling Using WinBUGS*. Wiley, New Jersey.