Robust Bayesian Analysis in Medical and Epidemiological Settings

Bradley P. Carlin, University of Minnesota
María-Eglée Pérez, Universidad Simón Bolívar

ABSTRACT Many medical and epidemiological professionals cite their distaste for informative priors as a prime reason for their ongoing aversion to Bayesian methods. In this paper we attempt to ease these concerns by investigating Bayesian robustness in such settings. Past attempts in this regard have either demonstrated the range of results possible using a certain class of priors ("forward robustness"), or the range of priors leading to a particular result ("backward robustness"). Application areas of particular interest include longitudinal data studies, clinical trial monitoring, survival analysis, and spatial epidemiology. After a brief review of the relevant methodology we consider two specific application areas. First, in the context of AIDS clinical trials we analyze a dataset that compared the effectiveness of the drug pyrimethamine versus a placebo in preventing toxoplasmic encephalitis. Our method uses nonparametric classes of prior distributions which attempt to model prior neutrality regarding the effect of the treatment. The resulting classes of prior distributions are reasonably wide, so that a clear conclusion emerging therefrom should be regarded as convincing by a broad group of potential consumers. Turning to spatial disease mapping, we investigate the impact of changes in the "heterogeneity plus clustering" priors commonly used to model excess spatial variation. In particular, we use the notion of Bayesian learning about the proportion of excess variability due to clustering to see whether a prior can be determined that offers a "fair" prior balance between these two components while exerting little influence on the posterior.

1 Introduction

Over the last decade or so, the expansion of Bayesian methods into biostatistical practice generally (and medicine and epidemiology in particular) has been substantial. This expansion has been due in large part to the advent of Markov chain Monte Carlo (MCMC) methods and associated generalist software (e.g., the `BUGS` language; Spiegelhalter et al., 1995a), but also to increasing recognition by practitioners that traditional analytic techniques were inadequate. As such, in this chapter we make no attempt at an exhaustive review, but rather merely seek to elucidate some of the

ways in which robust Bayesian methods and practices have found application in the fields of medicine and epidemiology. Similarly, since many previous chapters of this volume have carefully elucidated the necessary key methodological tools, we do not review them here, but instead focus only on those tools most relevant in our context.

Naturally the most common approach to Bayesian robustness in biomedical settings is through *sensitivity analysis*, wherein we simply make various modifications to a particular modeling assumption (say, some aspect of the prior) and recompute the posterior quantities of interest. If our resulting interpretations or decisions are essentially unaffected by this change, we say the Bayesian model is robust with respect to the assumption in question. This "forwards" approach to robustness is conceptually and implementationally simple, but it does not free us from careful development of the original prior, which must still be regarded as a reasonable baseline. Unlike many applied settings (e.g., business decisionmaking), in biomedical work this approach can be impractical, since the prior beliefs and vested interests of the potential consumers of our analysis may be very broad. For example, the results of a Bayesian clinical trial analysis might ultimately be read by doctors in clinical practice, epidemiologists, government regulatory workers, legislators, members of the media, and of course, patients and patient advocate groups. These groups are likely to have widely divergent opinions as to what constitutes "reasonable" prior opinion, ranging from quite optimistic (e.g., a clinician who has seen a few patients respond well to the drug being tested) to rather pessimistic (e.g., a regulatory worker who has seen many similar drugs emerge as ineffective). What is needed is a method for communicating the robustness of our conclusions to *any* prior input a consumer deems appropriate.

Carlin and Louis (1996a) suggest an alternate, "backwards" approach to this problem. Suppose that, rather than fix the prior and compute the posterior distribution, we fix the posterior (or set of posterior distributions) that produce a given conclusion, and determine which prior inputs are consistent with this desired result, given the observed data. The reader would then be free to determine whether the outcome was reasonable according to whether the prior class that produced it was consistent with his or her own prior beliefs. Carlin and Louis (1996a) refer to this approach simply as *prior partitioning* since it subdivides the prior class based on possible outcomes, though it is important to remember that such partitions also depend on the data and the decision to be reached.

To illustrate the basic idea, consider the point null testing scenario $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Without loss of generality, set $\theta_0 = 0$. Suppose our data x has density $f(x|\theta)$, where θ is an unknown scalar parameter. Let π represent the prior probability of H_0, and $G(\theta)$ the prior cumulative distribution function (cdf) of θ conditional on $\{\theta \neq 0\}$. The complete prior cdf for θ is then $F(\theta) = \pi I_{[0,\infty)}(\theta) + (1-\pi)G(\theta)$, where I_S is the indicator function of the set S. Hence the posterior probability of the null hypothesis

is

$$P_G(\theta = 0|x) = \frac{\pi f(x|0)}{\pi f(x|0) + (1-\pi)\int f(x|\theta)dG(\theta)} \ . \tag{1.1}$$

Prior partitioning seeks to characterize the G for which this probability is less than or equal to some small probability $p \in (0,1)$, in which case we reject the null hypothesis. (Similarly, we could also seek the G leading to $P_G(\theta \neq 0|x) \leq p$, in which we would reject H_1.) Elementary calculations show that characterizing this class of priors $\{G\}$ is equivalent to characterizing the set $\mathcal{H}_c$, defined as

$$\mathcal{H}_c = \left\{ G : \int f(x|\theta)dG(\theta) \geq c = \frac{1-p}{p}\frac{\pi}{1-\pi}f(x|0) \right\} \ . \tag{1.2}$$

Carlin and Louis (1996a) establish results regarding the features of $\mathcal{H}_c$, and then use these results to obtain sufficient conditions for $\mathcal{H}_c$ to be nonempty for classes of priors that satisfy various moment and percentile restrictions. The latter are somewhat more useful, since percentiles and tail areas of the conditional prior G are transform-equivariant, and Chaloner et al. (1993) have found that elicitees are most comfortable describing their opinions through a "best guess" (mean, median or mode) and a few relatively extreme percentiles (say, the 5^{th} and the 95^{th}).

Sargent and Carlin (1996) extend this general approach to the case of an interval null hypothesis, i.e., $H_0 : \theta \in [\theta_L, \theta_U]$ versus $H_1 : \theta \notin [\theta_L, \theta_U]$. This formulation is useful in the context of clinical trial monitoring, where $[\theta_L, \theta_U]$ is thought of as an *indifference zone*, within which we are indifferent as to the use of treatment or placebo. For example, we might take $\theta_U > 0$ if there were increased costs or toxicities associated with the treatment. Let π again denote the prior probability of H_0, and let $G(\theta)$ now correspond to the prior cdf of θ given $\theta \notin [\theta_L, \theta_U]$. Making the simplifying assumption of a uniform prior over the indifference zone, the complete prior density for θ may be written as

$$p(\theta) = \frac{\pi}{\theta_U - \theta_L} I_{[\theta_L, \theta_U]}(\theta) + (1-\pi)g(\theta) \ . \tag{1.3}$$

Sargent and Carlin (1996) derive expressions similar to (1.1) and (1.2) under the percentile restrictions of the previous subsection. However, these rather weak restrictions lead to prior classes that, while plausible, are often too broad for practical use. As such, we might consider a sequence of increasingly tight restrictions on the shape and smoothness of permissible priors, which in turn enable increasingly informative results. For example, we might retain the mixture form (1.3), but now restrict $g(\theta)$ to some particular parametric family. Carlin and Sargent (1996) refer to such a prior as "semiparametric" since the parametric form for g does not cover the indifference zone $[\theta_L, \theta_U]$, although since we have adopted another parametric form over this range (the uniform) one might argue that "biparametric" or simply "mixture" would be better names.

In the remainder of our paper we consider two biomedical settings where it is important that any Bayesian method be not only robust, but also "prior neutral," in the sense that the informative content of any prior used should be symmetric with respect to the null and the alternative. First, Section 2 considers a new approach to the Bayesian analysis of clinical trials data, illustrating with an example from a recent AIDS clinical trial. Section 3 then turns to robust Bayesian methods for spatial disease mapping, illustrating with the well-known Scottish lip cancer data of Clayton and Kaldor (1987). Finally, Section 4 discusses our findings, and presents some possible avenues for future research.

2 Clinical trial monitoring and analysis

Phase III clinical trials are large scale trials designed to identify patients who are best treated with a drug whose safety and effectiveness has already been reasonably well-established. In order to ensure scientific integrity and ethical conduct, decisions concerning whether or not to continue a clinical trial based on the accumulated data are made by an independent group of statisticians, clinicians and ethicists who form the trial's *data and safety monitoring board*, or DSMB. These boards have been a standard part of clinical trials practice (and NIH policy) since the early 1970s. As a result, the trial's statisticians require efficient algorithms for computing posterior summaries of quantities of interest to the DSMB.

From a Bayesian point of view, this also implies the need to select a prior distribution (or a set of prior distributions) on which all further analyses will be based. Kass and Greenhouse (1989) point out that, "randomization is ethically justifiable when a *cautious reasonable skeptic* would be unwilling to state a preference in favor of either the treatment or the control." If we assume that it is ethical to conduct a clinical trial, then an appropriate baseline prior for its monitoring and analysis might be one that reflects the opinions of a cautious reasonable skeptic.

Suppose that the effect of the drug is summarized by means of a parameter θ, such that $\theta < 0$ if patients treated with the drug perform better than those receiving placebo, and $\theta > 0$ otherwise. $\theta = 0$ means that both treatments (drug and placebo) have the same behavior with respect to the outcome of interest (take, for example, the log-odds of the probabilities of survival for both groups). A posterior quantity of interest for the DSMB might be the posterior probability that $\theta < 0$ given the accumulated data, $P(\theta < 0 \mid \mathbf{x})$. Following the approach of Section 1, if this probability is either too small or too big, the trial should be stopped. Our objective in this section is proposing classes of priors for θ which we believe reflect "prior neutrality" with respect to the treatment and the control and which are wide enough so that a clear conclusion emerging therefrom should be

regarded as convincing by most observers.

In a similar situation arising in an age-discrimination trial setting, Kadane (1990) addresses the problem of representing prior neutrality, stating symmetry around zero and unimodality as reasonable features for a neutral prior for θ to have, and choosing the normal family with mean zero and standard deviations $1, 2, 4$ and ∞ for calculating posterior quantities. Pérez and Pericchi (1994) suggested other neutral and almost neutral priors for the same problem, but none of these reflect a class of priors, and so they do not address the issue of robustness in the inference. Though Pérez and Pericchi explored classes of prior distributions, these classes don't articulate well with the idea of prior neutrality.

A very natural class for representing the opinion of a cautious reasonable skeptic is the class of all unimodal priors symmetric around 0. However, Kadane et al. (1999) show that this class turns out to be too broad, in the sense that it leads to trivial bounds (one being the prior value $1/2$) for $P(\theta < 0 \mid \mathbf{x})$, provided the likelihood $f(x|\theta)$ goes to 0 exponentially fast as $|\theta| \to \infty$. After finding similar behavior in several subsets of the class of unimodal priors symmetric around zero (the ones obtained by fixing its height at 0, fixing one of its quantiles, or bounding its variance from below), Kadane et al. conclude that a subclass of all unimodal priors symmetric around zero can lead to a non-trivial bound only if it avoids putting too much probability close to zero, and avoids allowing too much probability to be put on extremely high and low values of θ.

For achieving such a class, Kadane et al. focus on predictive distributions. Berger (1994) points out that the predictive distribution is in fact the likelihood of the prior (for a fixed likelihood), and a limitation of some robust Bayesian analyses is that robustness might be missing due to the inclusion of priors which have a very low (predictive) likelihood. In other words, lack of robustness might be caused by priors which are ruled out by the data. As the set of possible data outcomes involved in the age-discrimination trial is discrete and finite, Kadane et al. state that neutrality might be considered in terms of not being too surprised at any way the data might come out. More formally, suppose that the prior is $\pi(\theta)$, and $f_i(\theta)$ is the likelihood corresponding to a hypothetical observed value i, where i runs over the set of all possible data outcomes I. Let

$$g_\pi(i) = \int_{-\infty}^{\infty} f_i(\theta)\pi(\theta)d\theta \ . \tag{1.4}$$

Then the restricted *neutral prior class* A can be defined as

$$\begin{aligned} A \quad = \{\pi(\theta) : \quad & \pi(\theta) \text{ is unimodal and symmetric around } 0, \\ & \text{and } g_\pi(i) \geq \varepsilon \text{ for all } i \in I\} \end{aligned} \tag{1.5}$$

The parameter ε of this class is then the minimum prior predictive probability of the possible data. The idea of this class is that it constrains the

neutral arbitrator to have probability at least $\varepsilon > 0$ on each possible data point. In other words, only priors which have a non-negligible likelihood, for all possible data, are allowed in this neutrality class. Kadane et al. determine restrictions on ε for getting a non empty class which leads to non-trivial posterior bounds for $P(\theta < 0|x)$, and show how to find those bounds in the case where I is a discrete finite set.

We now extend these ideas to the case where the set of possible data outcomes I is uncountably infinite. To handle this problem, we replace the true sample space I by a representative grid of plausible hypothetical values. To illustrate this, suppose $f_i(\theta)$ can be reasonably well-approximated by a normal density $\phi(\theta|m,s)$ having mean $m \in M$ and standard deviation $s \in S$, where M and S are discrete finite sets of grid points corresponding to plausible likelihoods. Define

$$g_\pi(m,s) = \int_{-\infty}^{\infty} \phi(\theta|m,s)\pi(\theta)d\theta \ . \tag{1.6}$$

A neutral prior class similar to the class A in (1.5) can be then defined as

$$\begin{aligned}\Gamma_\varepsilon \quad = \{\pi(\theta): \quad & \pi(\theta) \text{ is unimodal and symmetric around 0,} \\ & \text{and } g_\pi(m,s) \geq \varepsilon \text{ for all } m \in M, s \in S\}\end{aligned}$$

This is a class of neutral priors such that none of them is going to be ruled out by data leading to parameters (m,s), $m \in M$, $s \in S$ for the normal approximation to the likelihood of θ.

Bounds for $P(\theta < 0|\mathbf{x})$ over the class Γ_ε can be obtained using the following procedure, similar to the one indicated in Kadane et al. (1999). Using Khinchine's representation, the set of priors in Γ_ε can be expressed as the set of distribution functions F satisfying $\pi(\theta) = \int_\theta^\infty \frac{1}{a} dF(a)$, where F is a distribution function in the set

$$\mathcal{F} = \left\{ F(\cdot): \int_0^\infty dF(a) = \frac{1}{2}, \ \int_0^\infty f^*_{m,s}(a)dF(a) \geq \varepsilon > 0, \ \forall \ m \in M, \ s \in S \right\}$$

and

$$\begin{aligned} f^*_{m,s}(a) &= \frac{1}{a}\int_0^a (\phi(-\theta|m,s) + \phi(\theta|m,s))d\theta \\ &= \frac{1}{a}\left[\Phi\left(\frac{a-m}{s}\right) - \Phi\left(\frac{-a-m}{s}\right)\right] \end{aligned}$$

for $a > 0$. Here, $\Phi(.)$ is the distribution function of a standard normal variable, and $f^*_{m,s}(0)$ is defined by continuity. The quantity of interest is the probability that $\theta < 0$ when the observed data is $\mathbf{x}$ and the prior is $\pi(\theta) \in \Gamma_\varepsilon$. This can be rewritten as

$$P^\pi(\theta < 0 \mid \mathbf{x}) = \left[1 + \frac{\int_{\theta \geq 0} \phi(\theta|\mu_\mathbf{x}, \sigma_\mathbf{x})\pi(\theta)d\theta}{\int_{\theta \geq 0} \phi(-\theta|\mu_\mathbf{x}, \sigma_\mathbf{x})\pi(\theta)d\theta}\right]^{-1} \tag{1.7}$$

where $\phi(\theta|\mu_{\mathbf{X}}, \sigma_{\mathbf{X}})$ is the normal approximation to the likelihood with parameters $\mu_{\mathbf{X}}$ and $\sigma_{\mathbf{X}}$ corresponding to the observed data $\mathbf{x}$. Equation (1.7) can equivalently be written as

$$P^F(\theta < 0 \mid \mathbf{x}) = \left[1 + \frac{\int_0^\infty f^1_{\mathbf{X}}(a)dF(a)}{\int_0^\infty f^2_{\mathbf{X}}(a)dF(a)}\right]^{-1} \tag{1.8}$$

where

$$\begin{aligned} f^1_{\mathbf{X}}(a) &= \tfrac{1}{a}\int_0^a \phi(\theta|\mu_{\mathbf{X}}, \sigma_{\mathbf{X}})d\theta = \tfrac{1}{a}\left[\Phi\left(\tfrac{a-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right) - \Phi\left(-\tfrac{\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\right], \\ f^2_{\mathbf{X}}(a) &= \tfrac{1}{a}\int_0^a \phi(-\theta|\mu_{\mathbf{X}}, \sigma_{\mathbf{X}})d\theta = \tfrac{1}{a}\left[\Phi\left(-\tfrac{\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right) - \Phi\left(\tfrac{-a-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\right], \end{aligned}$$

and $F(\cdot) \in \mathcal{F}$. Then the supremum of the posterior probability that $\theta < 0$ can be written as

$$\sup_{\pi \epsilon A} P^\pi(\theta < 0 \mid \mathbf{x}) = \sup_{F \epsilon \mathcal{F}} P^F(\theta < 0 \mid \mathbf{x}) = \left[1 + \inf_{F \in \mathcal{F}} \frac{\int_0^\infty f^1_{\mathbf{X}}(a)dF(a)}{\int_0^\infty f^2_{\mathbf{X}}(a)dF(a)}\right]^{-1}. \tag{1.9}$$

By the linearization algorithm (Lavine, Wasserman and Wolpert, 1993), the infimum in (1.9) is the unique solution λ_0 of the equation in λ

$$\inf_{F \in \mathcal{F}} \int_0^\infty \left[f^1_{\mathbf{X}}(a) - \lambda f^2_{\mathbf{X}}(a)\right] dF(a) = 0. \tag{1.10}$$

Once λ_0 has been found, $\sup_{\pi \in \Gamma_\varepsilon} P^\pi(\theta < 0 \mid \mathbf{x}) = (1+\lambda_0)^{-1}$. Using Kemperman (1987) (see also Salinetti (1994) and Liseo, Moreno and Salinetti (1994)), equation (1.10) can be rewritten

$$\begin{aligned} 0 = \sup_{\substack{d_{m,s} \geq 0 \\ m \in M \\ s \in S}} \Bigg\{ & \varepsilon \sum_{\substack{m \in M \\ s \in S}} d_{m,s} \\ & + \inf_{F \in \mathcal{F}_0} \int_0^\infty \left[f^1_{\mathbf{X}}(a) - \lambda f^2_{\mathbf{X}}(a) - \sum_{\substack{m \in M \\ s \in S}} d_{m,s} f^*_{m,s}(a) \right] dF(a) \Bigg\} \end{aligned} \tag{1.11}$$

where $\mathcal{F}_0$ is the class $\{F(\cdot) : \int_0^\infty dF(a) = 1/2\}$.

This last equation has important consequences. First, the internal infimum occurs at an F that puts all its probability at a single point a. This means that the extremum will occur at a single uniform distribution for θ. Thus

$$\inf_{F \epsilon \mathcal{F}_0} \int_0^\infty \left[f^1_{\mathbf{X}}(a) - \lambda f^2_{\mathbf{X}}(a) - \sum_{\substack{m \in M \\ s \in S}} d_{m,s} f^*_{m,s}(a) \right] dF(a)$$

$$= \inf_{a \geq 0} \frac{1}{2} \left[f^1_{\mathbf{X}}(a) - \lambda f^2_{\mathbf{X}}(a) - \sum_{\substack{m \in M \\ s \in S}} d_{m,s} f^*_{m,s}(a) \right],$$

which permits reduction of (1.11) to

$$\sup_{\substack{d_{m,s} \geq 0 \\ m \in M \\ s \in S}} \inf_{a \geq 0} \left\{ f^1_{\mathbf{X}}(a) - \lambda f^2_{\mathbf{X}}(a) - \sum_{\substack{m \in M \\ s \in S}} (f^*_{m,s}(a) - 2\varepsilon) d_{m,s} \right\} = 0 \,. \quad (1.12)$$

From (1.12), if for any pair (m, s), $f^*_{m,s}(a) < 2\varepsilon$, allowing the corresponding $d_{m,s}$ to go to infinity results in a sup of infinity, so (1.12) cannot be satisfied. Hence the supremum is attained when $d_{m,s} = 0$ for all (m, s). Thus (1.12) further simplifies to finding a value λ_0 of λ such that

$$\inf_{a \geq 0} \left\{ (f^1_{\mathbf{X}}(a) - \lambda f^2_{\mathbf{X}}(a)) I_{\{a \geq 0 : f^*_{m,s}(a) \geq 2\varepsilon, \forall m \in M,\ s \in S\}}(a) \right\} = 0. \quad (1.13)$$

Now there are two cases to be considered separately. Since the supremum of $P^\pi(\theta < 0 \mid \mathbf{x})$ corresponds to a small value of λ, find the value of a for which $f^1_{\mathbf{X}}(a)/f^2_{\mathbf{X}}(a)$ is a minimum. If that value of a satisfies the constraint $\min_{\substack{m \in M \\ s \in S}} f^*_{m,s}(a) > 2\varepsilon$, then the supremum has been found. If not, then the constraint is binding. In this case, because $f^1_{\mathbf{X}}(a) - \lambda f^2_{\mathbf{X}}(a)$ is continuous in a, the infimum in (1.13) occurs when $f^*_{m,s}(a) = 2\varepsilon$ for some $m \in M$, $s \in S$.

Thus the search for a solution of (1.13) in the second case can be made at the points a at which

$$\min_{\substack{m \in M \\ s \in S}} f^*_{m,s}(a) = 2\varepsilon, \quad (1.14)$$

and then $\lambda(a) = f^1_{\mathbf{X}}(a)/f^2_{\mathbf{X}}(a)$. If there are several points a satisfying (1.14), the smallest $\lambda(a)$ in the set corresponds to the infimum in (1.9). This can be accomplished by a one-dimensional search over possible values a. Finally, to find $\inf_{\pi \in \Gamma_\varepsilon} P^\pi(\theta < 0 \mid \mathbf{x})$, we simply reverse the roles of inf and sup in (1.9). This can be done by reversing the roles of $f^1_{\mathbf{X}}(a)$ and $f^2_{\mathbf{X}}(a)$ in each of the subsequent formulas.

We now illustrate in the specific context of the Community Programs for Clinical Research on AIDS toxoplasmic encephalitis (TE) prophylaxis trial. When the degree of immune damage becomes sufficiently severe, an HIV-infected person may develop a specific subset of more than 20 infections, several cancers, a variety of neurological abnormalities including severe declines in mental function, and wasting. Among the most ominous infections is encephalitis due to *Toxoplasma gondii*. This infection is the cause of death in approximately 50% of persons who develop it and the median survival is approximately six months. Additional clinical and immunological background concerning TE is provided in the review paper by Carlin, Chaloner, Louis and Rhame (1995).

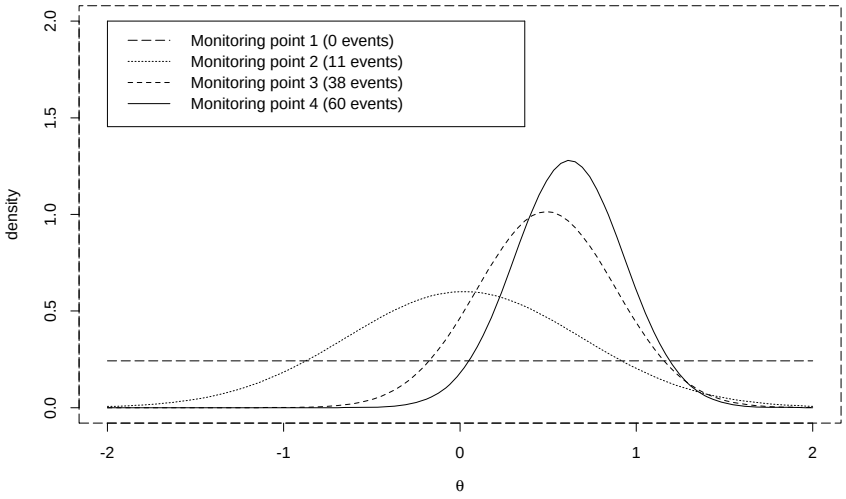

FIGURE 1. Normal approximations to the likelihood for the treatment effect, TE trial data. Endpoint is TE or death; covariate is baseline CD4 count.

Our study is a double-blind randomized TE prophylaxis trial comparing the drug pyrimethamine to placebo; a previous robust Bayesian analysis of this trial can be found in Carlin and Sargent (1996). All patients entered into the study had either an AIDS defining illness or a CD4 count of less than 200. In addition, all had a positive titre for *Toxoplasma gondii* and were therefore at risk for TE. As described in the report by Jacobson et al. (1994), the trial's data and safety monitoring board met on three occasions after the start of the trial in September of 1990 to assess its progress and determine whether it should continue or not. These three meetings analyzed the data available as of the file closing dates 1/15/91, 7/31/91, and 12/31/91, respectively. At its final meeting, the board recommended stopping the trial based on an informal stochastic curtailment rule: the pyrimethamine group had not shown significantly fewer TE events up to that time, and due to the low TE rate a significant difference was judged unlikely to emerge in the future. An increase in the number of deaths in the pyrimethamine group was also noted, but this was not a stated reason for the discontinuation of the trial (although subsequent follow-up confirmed this mortality increase). The recommendation to terminate the study was conditional on the agreement of the protocol chairperson after unblinding and review of the data. As a result, the trial did not actually stop until 3/30/92, when patients were instructed to discontinue their study medication.

In a Bayesian reanalysis of this data, Carlin et al. (1993) employed a proportional hazards likelihood using the time from randomization until development of TE or death as the response variable. Specifically, their model used two covariates for each patient: baseline CD4 cell count, and

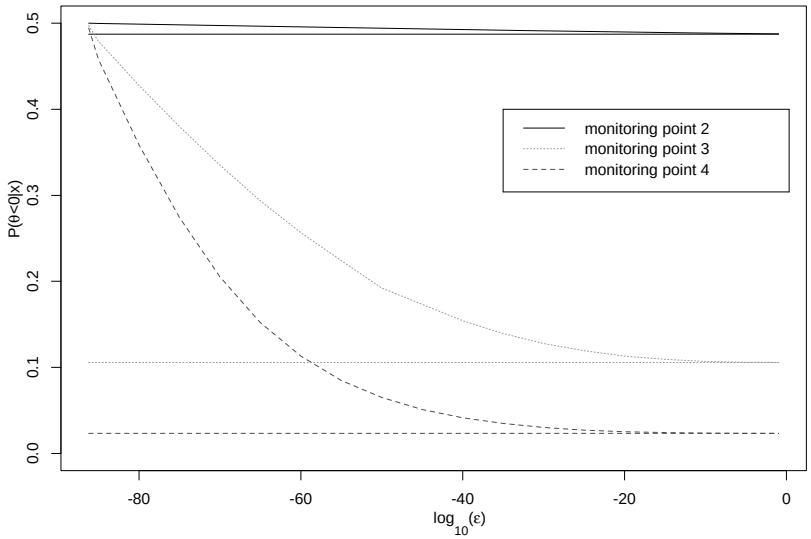

FIGURE 2. Bounds for the posterior probability of $\theta < 0$ over the neutral class Γ_ε for different values of ε, TE trial data.

a treatment effect indicator (1 for active drug, 0 for placebo). Denoting the parameters which correspond to these two covariates as β and θ, respectively, we obtain a marginal partial likelihood for θ by numerically integrating β out of the Cox partial likelihood. Negative values of θ correspond to an efficacious treatment; the relatively low support for these values at the final two monitoring points suggests that perhaps the trial should be stopped and the treatment rejected. Normal approximations to the standardized Cox partial likelihood at each of the four data monitoring points are shown in Figure 1. We will base our inference for θ on these normal approximations.

Each outcome of the trial produces different values for the parameters of the normal approximation. Stating that a cautious reasonable skeptic shouldn't be too surprised at any data outcome is equivalent to saying that he or she won't be too surprised at observing a wide set of parameters for this approximation. For this analysis we defined our likelihood grids as follows: for means, $M = \{-2.0, -1.9, -1.8, \ldots, 1.8, 1.9, 2.0\}$; for standard deviations, $S = \{0.1, 0.2, \ldots, 2.0\}$. Note that $\theta = -2.0$ corresponds to a reduction of 86% in hazard for the treatment relative to control, significantly greater than the target reduction of 50% specified in the trial protocol. In a similar way, $\theta = 2$ indicates that the hazard rate in the control group is 86% lower than the hazard for treatment group. So, this class states that a cautious reasonable skeptic shouldn't be surprised by results that support either the treatment or the placebo. Different grids can be defined on the basis of further design considerations.

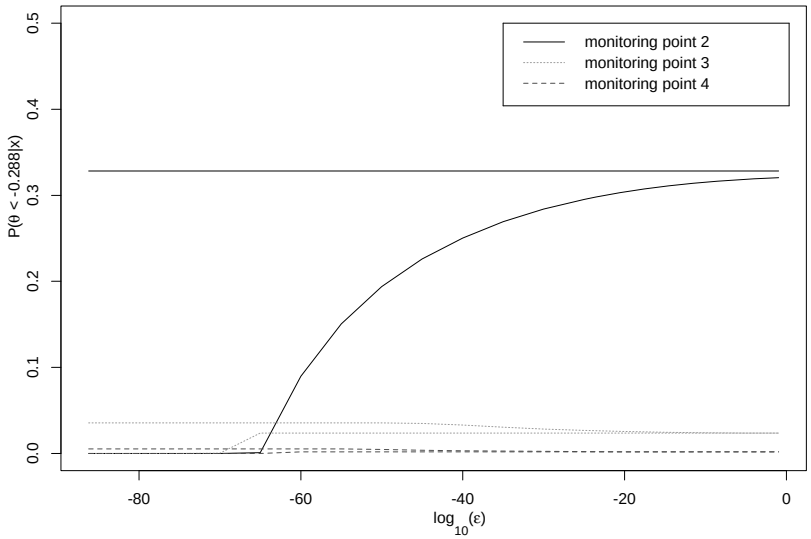

FIGURE 3. Bounds for the posterior probability of $\theta < -0.2880$ over the neutral class Γ_ε for different values of ε, TE trial data.

Figure 2 shows the bounds for $P^\pi(\theta < 0|\mathbf{x}_i)$, where $\mathbf{x}_i$ is the data collected at monitoring point i, $i = 2, 3, 4$ and $\pi \in \Gamma_\varepsilon$, for different values of $\log_{10}(\varepsilon)$. At monitoring points 3 and 4, these bounds are very near 0 even for very small values of ε, suggesting the need to stop the trial.

Other posterior probabilities are also of interest, especially if the treatment is preferred only if it reduces the hazard rate by some meaningful amount. Following Carlin et al. (1993), we can concentrate on $P(\theta < \theta_L|\mathbf{x})$, where $\theta_L = \log(0.75) = -0.288$ (corresponding to a reduction of at least 25% in relative hazard). Bounds for this posterior probability over the class Γ_ε can be obtained in a similar way, and are shown in Figure 3. The message of this graph is even clearer: regardless of the value of ε, the trial should be stopped at the penultimate monitoring point.

3 Spatial Epidemiology

Bayes and empirical Bayes models for spatial data aggregated by geographic region have begun to see widespread use in spatial epidemiology, especially in the creation of disease maps. Developed by Clayton and Kaldor (1987) and refined by Besag, York and Mollié (1991), these models typically assume the observed disease count in region i, Y_i, has a Poisson distribution with mean $E_i e^{\mu_i}$, where E_i is an expected disease count (perhaps obtained via reference to an external standard table) and μ_i is a log-relative risk of

disease, modeled linearly as

$$\mu_i = \mathbf{x}_i'\boldsymbol{\beta} + \theta_i + \phi_i,\ i = 1, \ldots, I\ . \tag{1.15}$$

Here the $\mathbf{x}_i$ are explanatory spatial covariates, while $\boldsymbol{\beta}$ is a vector of fixed effects. The θ_i capture *heterogeneity* among the regions via the mixture specification $\theta_i \stackrel{iid}{\sim} N(0\ ,\ 1/\tau_h)$, while the ϕ_i capture regional *clustering* by assuming that

$$\phi_i \mid \phi_{j \neq i} \sim N(\bar{\phi}_i\ ,\ 1/(n_i\tau_c))\ , \tag{1.16}$$

where n_i is the number of "neighbors" of region i, and $\bar{\phi}_i = n_i^{-1}\Sigma_{j \in \partial_i}\ \phi_j$ with ∂_i denoting the neighbor set of region i. The usual assumption is that regions are neighbors if and only if they are adjacent on the map, though other (e.g., distance-based) modifications are often considered. This distribution for $\boldsymbol{\phi} \equiv \{\phi_i\}$ is called an *intrinsically* or *conditionally autoregressive* specification, which for brevity we typically write in compact notation as $\boldsymbol{\phi} \sim CAR(\tau_c)$.

Model (1.15) with the CAR prior formulation has several quirks that cloud its complete specification. First, the CAR prior is translation invariant, since an arbitrary constant could be added to all of the ϕ_i without changing the joint probability specification. This necessitates the addition of a identifiability-preserving constraint (say, $\sum_{i=1}^{I} \phi_i = 0$), which is awkward theoretically but easy to implement "on the fly" during an MCMC algorithm by recentering the $\phi_i^{(g)}$ samples around their own mean at the end of each iteration g. Even with this correction to the prior, only the sum of the two random effects, $\eta_i \equiv \theta_i + \phi_i$, is identified by the datapoint Y_i, so the effective dimension of the full model is often much smaller than the actual parameter count.

This identifiability problem is in some sense a non-issue for Bayesians, since as observed by Besag et al. (1995) and others, even under improper priors for the θ_i and ϕ_i, MCMC algorithms may still operate on the resulting overparametrized space, with convergence still obtaining for the *proper embedded posterior* (i.e., the lower-dimensional parameter vector having a unique integrable posterior distribution). Indeed, Gelfand and Sahu (1999) show that noninformative priors are often optimal in such settings, producing immediate convergence for the well-identified embedded subset. Unfortunately, such an approach is less attractive here due to our genuine interest in the random effects $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, and in particular in the proportion of excess variability due to clustering, since this may help us identify missing covariates which vary spatially. As a specific measure of this proportion, Best et al. (1999) define the quantity

$$\psi = \frac{sd(\boldsymbol{\phi})}{sd(\boldsymbol{\theta}) + sd(\boldsymbol{\phi})}\ , \tag{1.17}$$

where $sd(\cdot)$ is the empirical marginal standard deviation of the random effect vector in question. A posterior for ψ concentrated near 1 suggests

most of the excess variation (i.e., that not explained by the covariates $\mathbf{x}_i$) is due to spatial clustering, while a posterior concentrated near 0 suggests most of this variation is mere unstructured heterogeneity. This genuine interest in the tradeoff between $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ forces these authors into a search for proper yet vague priors for these two components – a task complicated by the fact that the prior for the former is specified marginally, while that for the latter is specified conditionally. In the remainder of this section, we investigate whether a prior can be determined that offers a "fair" prior balance between heterogeneity and clustering, while remaining minimally informative. We can then check whether this in fact enables robust analyses within this "fair" class.

Eberly and Carlin (2000) show that Bayesian learning about ψ is indeed possible (i.e., that its prior and posterior can be determined and shown to be distinct). As such, the class of priors for which $\psi \approx 1/2$ seems an appropriate "fair" class to which we may restrict our attention. To proceed with this line of inquiry, we reconsider the Scottish lip cancer data originally presented by Clayton and Kaldor (1987) and reanalyzed by many others since. This dataset provides observed and expected cases of lip cancer in the 56 districts of Scotland for 1975-1980; the expected cases are based on MLEs of the age effects in a simple multiplicative risk model, and are thought of as fixed and known. For each district i we also have one covariate x_i (the percentage of the population engaged in agriculture, fishing or forestry, or AFF) and a list of which other districts j are adjacent to i. The raw data and the AFF covariate are mapped in Figure 4.

Since Gibbs sampler code for analyzing these data and model is readily available as an example in the `BUGS` software package (Spiegelhalter et al., 1995b), we use this language to carry out our investigation. The newest version of `BUGS` for Windows, `WinBUGS 1.2`, automatically imposes the sum-to-zero constraint $\sum_{i=1}^{I} \phi_i = 0$ numerically by recentering the ϕ_i samples around their own mean at the end of each iteration (Best et al., 1999). All older versions of the program do not, which in turn prohibits the inclusion of an intercept term in the log-relative risk model (1.15). Note that neither approach solves the Bayesian identifiability problem with the ϕ_i due to the continued presence of the covariate coefficient β and the θ_i.

In order to specify a "fair" prior balance between heterogeneity and clustering, we must first make a connection between the prior precision parameters τ_c and τ_h, and then see what prior for ψ they induce. Regarding the first problem, the primary difficulty is obviously that τ_c is a *conditional* prior precision, while τ_h is a precision in a standard, marginal prior specification. An investigation by Bernardinelli et al. (1995) suggests that the marginal standard deviation of ϕ_i is roughly proportional to the corresponding conditional expression; i.e., from (1.16) we have $sd(\phi_i) \approx 1/(K\sqrt{n_i \tau_c})$, where the authors propose $K = .7$ as a plausible variance inflation factor. A sensible "fair" prior might then be one which equates this expression to that for the marginal standard deviation of the θ_i, $1/\sqrt{\tau_h}$. Replacing n_i by

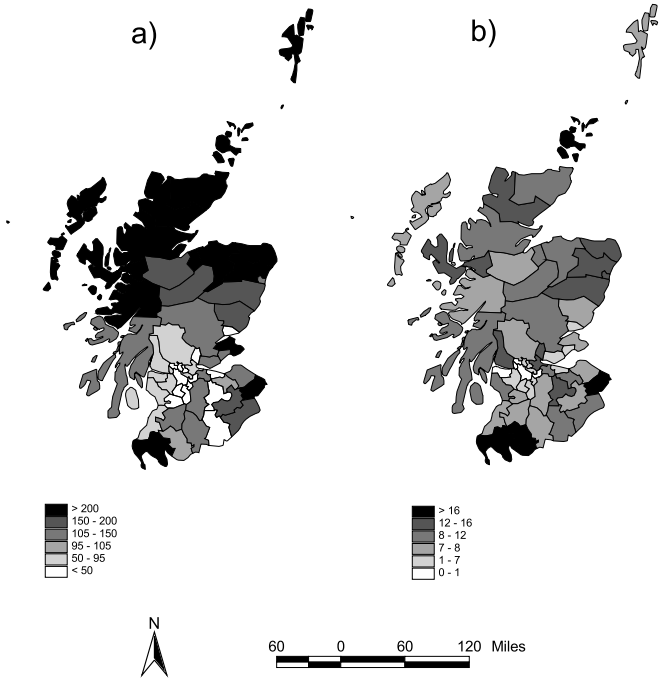

FIGURE 4. Scotland lip cancer data: a) crude standardized mortality ratios (observed/expected × 100); b) AFF covariate values.

$\bar{n} = 264/56 = 4.71$, the average number of neighbors across the Scotland map, we obtain the "rule of thumb"

$$\tau_h = K^2 \bar{n} \tau_c \ . \tag{1.18}$$

To check the accuracy of this formula, we employ a more direct approach made possible by `WinBUGS 1.2`. Freely available over the web at `http://www.mrc-bsu.cam.ac.uk/bugs/`, this program allows direct sampling from the centered version of the CAR prior (i.e., the version incorporating the sum-to-zero constraint) via its `car.normal` function. Running the `WinBUGS` code for a simplified version of our model that does not include the data produces draws $\psi^{(g)}$ via equation (1.17), hence an estimate of the induced prior for ψ. (All of our simulations use a single sampling chain, run for a 1000-iteration burn-in period followed by a 10,000-iteration "production" period.) After a bit of experimentation, we discovered that setting $K = .62$ in equation (1.18) produced ψ samples having empirical mean (and median) .50 and standard deviation .059, a suitably fair specification.

Note that the induced prior for ψ does not depend on the τ_c value selected, since ψ measures only the amount of excess variability due to clustering *relative to* the total amount present; changing the scale of the θ_i and ϕ_i does not affect ψ provided (1.18) still holds. Table 1.1 investigates the

	posterior for ψ			posterior for β		
τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
1, 1.81	.54	.041	.64	.42	.21	.97
0.1, 0.181	.50	.032	.64	−.022	.59	.99
0.001, 0.00181	.49	.017	.22	.043	3.00	.99
	posterior for η_1			posterior for η_{56}		
τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
1, 1.81	1.05	.46	.46	−1.31	.72	.06
0.1, 0.181	1.82	.99	.86	−2.85	1.85	.02
0.001, 0.00181	1.73	4.81	.95	−2.20	1.62	−.04

TABLE 1.1. Posterior summaries for spatial model with fixed values for τ_c and τ_h, Scotland lip cancer data; "sd" denotes standard deviation while "l1acf" denotes lag 1 sample autocorrelation. In each case, the prior for ψ has mean .50 and standard deviation .059.

robustness of our conclusions under three such priors, namely, those obtained by setting $\tau_c = 1$, 0.1, and 0.001 in (1.18), respectively. We see that the posterior for ψ is rather robust to these changes, remaining close to 0.50 albeit with a slightly decreasing posterior standard deviation ("sd" in the table). However, the picture is less reassuring for the other parameters summarized (the AFF covariate, β, and the sums of the random effects for the counties with the highest and lowest observed rates, η_1 and η_{56}). Vaguer priors appear to lead to larger random effects, and a corresponding collapse in the significance of the AFF covariate. Worse, the very high lag 1 sample autocorrelations ("l1acf") for β suggest that the posterior summaries for this parameter are likely to be quite unreliable; far more Monte Carlo samples would be required to obtain effective sample sizes large enough for reliable inference.

The lack of robustness and convergence problems evident in Table 1.1 are not terribly surprising, given that fixing τ_c and τ_h at small values essentially precludes borrowing of strength across counties in an overparametrized model setting where such borrowing is badly needed. Indeed, hyperpriors for these two parameters are commonly used in practice, in order to reduce the prior specification burden and allow the data to play a bigger role in determining the posterior. However, adding these hyperpriors also complicates specification of a fair prior balance between heterogeneity and clustering. To see this, consider the usual conjugate hyperprior specification

$$\tau_c \sim Gamma(a_c, b_c) \quad \text{and} \quad \tau_h \sim Gamma(a_h, b_h)\ .$$

If we require $E(\tau_c) = 1.0$ and $Var(\tau_c) = \sigma_c^2$, then we must take $a_c = b_c = 1/\sigma_c^2$. If we then follow (1.18) and similarly insist $E(\tau_h) = 1.81$ and $Var(\tau_h) = \sigma_h^2$, it follows that $a_h = (1.81)^2/\sigma_h^2$ and $b_h = 1.81/\sigma_h^2$. Clearly taking σ_c^2 and σ_h^2 very small would essentially reproduce the first line in

	posterior for ψ			posterior for β		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.57	.058	.80	.43	.17	.94
G(.1, .1), G(.32761, .181)	.65	.073	.89	.41	.14	.92
G(.1, .1), G(.001, .001)	.82	.10	.98	.38	.13	.91
	posterior for η_1			posterior for η_{56}		
priors for τ_c, τ_h	mean	sd	l1acf	mean	sd	l1acf
G(1.0, 1.0), G(3.2761, 1.81)	.92	.40	.33	−.96	.52	.12
G(.1, .1), G(.32761, .181)	.89	.36	.28	−.79	.41	.17
G(.1, .1), G(.001, .001)	.90	.34	.31	−.70	.35	.21

TABLE 1.2. Posterior summaries for spatial model with Gamma hyperpriors for τ_c and τ_h, Scotland lip cancer data; "sd" denotes standard deviation while "l1acf" denotes lag 1 sample autocorrelation.

Table 1.1. The first two lines of Table 1.2 summarize results obtained from two nontrivial hyperpriors of this form, the first setting $\sigma_c^2 = \sigma_h^2 = 1$, and the second setting $\sigma_c^2 = \sigma_h^2 = 10$. These two hyperpriors actually do *not* produce priors for ψ having means of .50, but rather .62 and .99, respectively, suggesting that our moment-matching approach is not sufficient to ensure our previous definition of "fairness". The third line of Table 1.2 reports results obtained under the specification recommended by Best et al. (1999), namely $a_c = b_c = 0.1$ and $a_h = b_h = 0.001$ (i.e., hyperpriors having mean 1 and variance 10 and 1000, respectively). This specification actually leads to a ψ prior mean of .09, so it is not particularly "fair" either. Obviously we could continue searching for induced ψ priors centered near .50 via ad-hoc experimentation, but we instead move on to the summarization of results under these three since they are reasonably encouraging even in the absence of a rigorously imposed prior-fairness constraint.

The results in Table 1.2 are to some extent the "mirror image" of those in Table 1.1, since the results for ψ are disappointing but those for the remaining parameters are reassuring. Under all three priors, it appears that the excess variability in the data is mostly due to clustering ($E(\psi|\mathbf{y}) > .50$), but the posterior distribution for ψ does not seem robust to changes in its prior. In fact, $E(\psi|\mathbf{y})$ is actually largest under the prior for which $E(\psi)$ was the smallest! Apparently the extreme one-tailed shapes of our hyperpriors for τ_c and τ_h can make both the prior and posterior distributions for ψ difficult to anticipate and interpret.

Despite these difficulties, we do see a reasonable degree of robustness for the remaining parameters. While autocorrelation in the β chain is still fairly high, it is reduced to a level enabling fairly precise estimation (and in particular, a 95% credible interval excluding 0 under all three priors). The random effect sums are much more modest than those in Table 1.1, and are well-estimated from essentially uncorrelated MCMC chains.

4 Discussion and future directions

In this paper we have investigated the use of robust Bayesian methods in medical research, focusing on clinical trials and spatial disease mapping, two application areas where the methods are increasingly popular. Our methods emphasized use of priors which are not only minimally informative, but also neutral or "fair" in the sense that they do not favor any particular hypothesis (drug versus placebo, unstructured heterogeneity versus spatial clustering).

Our findings in both cases suggest a bright future for practitioners seeking to apply Bayesian methods in the absence of overly informative or "unfair" priors. In the clinical trial setting of Section 2, the bounds we obtain seem helpful, especially in Figure 3 where the range produced is quite narrow and nearly constant across ε for monitoring points 3 and 4. Our results in the spatial epidemiology setting of Section 3 are similarly encouraging, and consistent with those of Gelfand and Sahu (1999) and Eberly and Carlin (2000). That is, they indicate that well-identified subsets of the parameter space tend to converge quickly and can be robustly estimated under vague prior specifications, but less well-identified parameters (such as the θ_i, ϕ_i, and ψ) may converge poorly and in any case produce posterior estimates that are difficult to interpret. To the extent that spatial epidemiologists and statisticians wish to focus on such quantities, there would appear to be much work remaining to do in the specification of fair, informative priors for ψ and the determination of more efficient associated computational algorithms.

In many model settings, Bayesian inferences will depend on the precise form of prior selected for variance components. This is somewhat surprising, since such components are typically thought of as nuisance parameters, and all priors contemplated for them are typically "minimally informative" in some sense. Still, the results for ψ in Table 1.2 reemphasize this point. While the $Gamma(\epsilon, \epsilon)$ prior (i.e., having mean 1 but variance $1/\epsilon$) currently seems to enjoy widespread use, recent work by Hodges and Sargent (1998) and Natarajan and Kass (2000) shows that such priors can actually have significant impact on the resulting posterior distributions. And while this prior is proper, it is "nearly improper" for suitably small ϵ, potentially leading to MCMC convergence failure – or worse, the appearance of MCMC convergence when in fact the joint posterior is also improper. More work is needed to determine priors for variance components that have minimal impact on the resulting posterior while still allowing MCMC algorithms with acceptable convergence properties. Alternatively, Carlin and Louis (2000) suggest that reverting to an empirical Bayes approach (i.e., replacing an unknown variance component by a point estimate, rather than attempting to pick a hyperprior for it) may well produce a estimated posterior that produces improved estimates while at the same time is safer to use and easier to obtain.

Acknowledgements

The research of the first author was supported in part by National Institute of Allergy and Infectious Diseases (NIAID) Grant R01-AI41966 and by National Institute of Environmental Health Sciences (NIEHS) Grant 1-R01-ES07750, while that of the second author by the National Council for Science and Technology of Venezuela (CONICIT) Grant G97-000592. This research was completed during the second author's sabbatic visit to the Division of Biostatistics at the University of Minnesota. The authors are grateful to Drs. Lynn Eberly and Erin Conlon for their assistance in the development of Section 3.

References

Berger, J.O. (1994). An overview of robust Bayesian analysis (with discussion). *Test*, **3**, 5–124.

Bernardinelli, L., Clayton, D.G., and Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, **14**, 2411–2431.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3–66.

Besag, J., York, J.C. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.

Best, N.G., Waller, L.A., Thomas, A., Conlon, E.M., and Arnold, R.A. (1999). Bayesian models for spatially correlated disease and exposure data (with discussion). In *Bayesian Statistics 6*, J.M. Bernardo et al., eds., Oxford: Oxford University Press, pp. 131–156.

Carlin, B.P., Chaloner, K., Church, T., Louis, T.A., Matts, J.P. (1993). Bayesian approaches for monitoring clinical trials with an application to toxoplasmic encephalitis prophylaxis. *The Statistician*, **42**, 355–367.

Carlin, B.P., Chaloner, K., Louis, T.A. and Rhame, F.S. (1995). Elicitation, monitoring, and analysis for an AIDS clinical trial (with discussion). In *Case Studies in Bayesian Statistics, Volume II*, eds. C. Gatsonis, J.S. Hodges, R.E. Kass and N.D. Singpurwalla, New York: Springer-Verlag, pp. 48–89.

Carlin, B.P. and Louis, T.A. (1996a). Identifying prior distributions that produce specific decisions, with application to monitoring clinical trials. In *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, eds. D. Berry, K. Chaloner, and J. Geweke, New York: Wiley, pp. 493–503.

Carlin, B.P. and Louis, T.A. (1996b). *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press.

Carlin, B.P. and Louis, T.A. (2000). Empirical Bayes: past, present, and future. To appear *J. Amer. Statist. Assoc.*

Carlin, B.P. and Sargent, D. (1996). Robust Bayesian approaches for clinical trial monitoring. *Statistics in Medicine*, **15**, 1093–1106.

Chaloner, K., Church, T., Louis, T.A. and Matts, J.P. (1993). Graphical elicitation of a prior distribution for a clinical trial. *The Statistician*, **42**, 341–353.

Clayton, D.G. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671–681.

Eberly, L.E. and Carlin, B.P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. To appear *Statistics in Medicine*.

Gelfand, A.E. and Sahu, S.K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *J. Amer. Statist. Assoc.*, **94**, 247–253.

Hodges, J.S. and Sargent, D.J. (1998). Counting degrees of freedom in hierarchical and other richly-parameterised models. Technical report, Division of Biostatistics, University of Minnesota.

Jacobson, M.A., Besch, C.L., Child, C., Hafner, R., Matts, J.P., Muth, K., Wentworth, D.N., Neaton, J.D., Abrams, D., Rimland, D., Perez, G., Grant, I.H., Saravolatz, L.D., Brown, L.S., Deyton, L., and the Terry Beirn Community Programs for Clinical Research on AIDS (1994). Primary prophylaxis with pyrimethamine for toxoplasmic encephalitis in patients with advanced human immunodeficiency virus disease: Results of a randomized trial. *J. Infectious Diseases*, **169**, 384–394.

Kadane, J.B. (1990). A statistical analysis of adverse impact of employer decisions. *Journal of the American Statistical Association*, **85**, 925–933.

Kadane, J.B., Moreno, E., Pérez, M.-E., and Pericchi, L.R. (1999). Applying nonparametric robust Bayesian analysis to non-opinionated judicial neutrality. Technical Report No. 695, Department of Statistics, Carnegie Mellon University.

Kass, R. and Greenhouse, J. (1989). Comment on "Investigating therapies of potentially great benefit: ECMO" by J. Ware. *Statistical Science*, **4**, 31–317.

Kemperman, J. (1987). Geometry of the moment problem. *Proceedings of Symposia in Applied Mathematics*, **37**, 16–53.

Lavine, M., Wasserman, L., and Wolpert, R. (1993). Linearization of Bayesian robustness problems. *Journal of Statistical Planning and Inference*, **37**, 307–316.

Liseo, B., Moreno, E., and Salinetti, G. (1996). Bayesian robustness of the class with given marginals: an approach based on moment theory (with discussion). In *Bayesian Robustness*, IMS Lecture Notes – Monograph Series, **29**, eds. J.O. Berger et al., Hayward, CA: Institute of Mathematical Statistics, pp. 101–118.

Natarajan, R. and Kass, R.E. (2000). Reference Bayesian methods for generalized linear mixed models. To appear *J. Amer. Statist. Assoc.*

Pérez, M.-E. and Pericchi, R. (1994). A case study on the Bayesian Analysis of 2×2 tables with all margins fixed. *Revista Brasileira de Probabilidade e Estatistica*, **1**, 27–37.

Salinetti, G. (1994). Discussion of "An overview of robust Bayesian analysis," by J.O. Berger, *Test*, **3**, 1–125.

Sargent, D.J. and Carlin, B.P. (1996). Robust Bayesian design and analysis of clinical trials via prior partitioning (with discussion). In *Bayesian Robustness*, IMS Lecture Notes – Monograph Series, **29**, eds. J.O. Berger et al., Hayward, CA: Institute of Mathematical Statistics, pp. 175–193.

Spiegelhalter, D.J., Thomas, A., Best, N. and Gilks, W.R. (1995a). BUGS: Bayesian inference using Gibbs sampling, Version 0.50. Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

Spiegelhalter, D.J., Thomas, A., Best, N. and Gilks, W.R. (1995b). BUGS examples, Version 0.50. Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.