Robust Bayesian Methods for Monitoring Clinical Trials

Joel B. Greenhouse
and
Larry Wasserman

Department of Statistics
Carnegie Mellon University
Pittsburgh, PA. 15213-3890
(412) 268-8872

May, 1994
Revised: July 1994

Robust Bayesian Methods for Monitoring Clinical Trials

Summary

Bayesian methods for the analysis of clinical trials data have received increasing attention recently as they offer an approach for dealing with difficult problems that arise in practice. A major criticism of the Bayesian approach, however, has focused on the need to specify a single, often subjective, prior distribution for the parameters of interest. In an attempt to address this criticism, we describe methods for assessing the robustness of the posterior distribution to the specification of the prior. The robust Bayesian approach to data analysis replaces the prior distribution with a *class* of prior distributions and investigates how the inferences might change as the prior varies over this class. The purpose of this paper is to illustrate the application of robust Bayesian methods to the analysis of clinical trials data. Using two examples of clinical trials taken from the literature, we illustrate how to use these methods to help a data monitoring committee decide whether or not to stop a trial early.

Key words: Clinical trials; Data monitoring committee; ϵ-contaminated class; Prior distribution; Robust Bayesian inference; Sensitivity analysis.

1 Introduction

The implementation and analysis of randomized controlled clinical trials is rarely easy. Although in theory the conduct of a trial should be relatively straightforward, the complexity of doing a trial in practice arises in part from the need to balance the ethics of human experimentation with the principles of scientific design. Procedures for monitoring clinical trials, for example, developed out of a desire to minimize the number of patients exposed to a potentially inferior or harmful treatment. Yet, these procedures often require relatively sophisticated mathematical and statistical methods for their implementation. Recently, interest in the application of Bayesian methods in clinical trials has grown because these methods offer an approach for dealing with many of the difficult practical problems that arise in the conduct and analysis of such studies. Some of the features of the Bayesian approach include the ability to utilize relevant historical information regarding the magnitude of the treatment effect of interest; the ability to monitor a trial without the need for adjustments according to the number of looks at the data; and the ability to make inferences from a trial based on posterior probabilities of clinically meaningful outcomes[1–4].

Even as an appreciation for the usefulness of the Bayesian approach has grown, a persistent criticism of Bayesian methods has focused on the need to specify a prior distribution for the parameters of interest. In particular, investigators are concerned that inferences based on a posterior distribution will be highly sensitive to the specification (or misspecification) of the prior distribution. The development of robust Bayesian methods is an attempt to address this criticism. Instead of worrying about the choice of a single "correct" prior distribution, the robust Bayesian approach replaces the prior distribution with a *class* of prior distributions. The goal of this approach is to see how sensitive inferences are to the choice of prior by studying how the inferences might change as the prior varies over this class[5–7]. The purpose of this paper is to illustrate the application of robust Bayesian methods to the analysis of clinical trials data. Specifically, we hope to demonstrate the usefulness of these methods as inputs to a data monitoring committee to help them decide whether or not to terminate a trial early.

In the next section, we review the basic approach and results for the application of robust Bayesian methods in practice based on the *ϵ-contaminated class* of prior distributions. Although a complete robust Bayesian analysis would consider the sensitivity of posterior inferences not only to the specification of the prior distribution but also to the specification of the likelihood function, in this paper our concern is only with sensitivity to the prior distribution. In sections 3, we illustrate how to use this approach to assist in the decision as to whether to stop a phase II trial early due to an unexpected toxic side effect. In section 4, we consider a robust sequential Bayesian analysis of an efficacy trial (the ECMO trial) to evaluate whether the accumulating evidence from that trial was sufficiently persuasive to stop early. Section 5 contains concluding remarks.

2 Robust Bayesian Data Analysis

2.1 Overview

In the usual Bayesian framework, we have data $x = (x_1, \ldots, x_n)$ from a statistical model $f(\cdot|\theta)$, $\theta \in \Theta$, and a single specified prior density π_0 for θ. We then compute the posterior density for θ, $\pi_0(\theta|x) = L(\theta)\pi_0(\theta)/m_{\pi_0}(x)$, where $L(\theta) = \prod_i^n f(x_i|\theta)$ is the likelihood function (assuming independence) and $m_{\pi_0}(x) = \int L(\theta)\pi_0(\theta)d\theta$. To be more precise, we should work with distributions not densities since the classes of priors we deal with do not necessarily have densities. But, for simplicity, we continue to write expressions in terms of density functions. In robust Bayesian inference, we replace the prior π_0 with a class of priors Γ and we compute the corresponding class of posteriors $\Gamma^x = \{\pi(\cdot|x); \pi \in \Gamma\}$. We then determine the range of the posterior quantity of interest as the prior varies over Γ^x. Specifically, if $g(\theta)$ is some quantity of interest we compute upper and lower bounds on the posterior expectation of $g(\theta)$:

$$\overline{E}(g(\theta)|x) = \sup_{\pi\in\Gamma} E_\pi(g(\theta)|x) = \sup_{\pi\in\Gamma} \int g(\theta)\frac{L(\theta)\pi(\theta)}{m_\pi(x)}d\theta,$$
$$\underline{E}(g(\theta)|x) = \inf_{\pi\in\Gamma} E_\pi(g(\theta)|x) = \inf_{\pi\in\Gamma} \int g(\theta)\frac{L(\theta)\pi(\theta)}{m_\pi(x)}d\theta. \tag{1}$$

Sometimes the posterior quantity of interest is the indicator for a set A in which case we write these bounds as $\overline{P}(A|x)$ and $\underline{P}(A|x)$. If $\overline{E}(g(\theta)|x) - \underline{E}(g(\theta)|x)$ or $\overline{P}(A|x) - \underline{P}(A|x)$ is small, we conclude that the inferences for $g(\theta)$ are not sensitive to the prior.

We assume that we can specify a prior distribution, π_0, that quantifies relevant historical or subjective information for the quantity of interest. We do not discuss how to specify or elicit such a prior distribution since others have addressed methods for elicitation of subjective prior information for the analysis of clinical trials data[3,8–10]. Since we may not have confidence in the choice of π_0, our approach, as noted earlier, is to consider a class of priors similar to π_0. A tractable class used by many authors, including Huber[11], Berger[6], Berger and Berliner[12] and Lavine, Wasserman and Wolpert[13], is the *ϵ-contaminated class* defined by

$$\Gamma_\epsilon = \{\pi = (1-\epsilon)\pi_0 + \epsilon q; q \in \mathcal{Q}\} \tag{2}$$

where $\epsilon \in [0,1]$ reflects the amount of uncertainty concerning the accuracy of π_0 and $\mathcal{Q}$ is a class of reasonable alternative priors.

The class $\mathcal{Q}$ is a set of probability distributions (or contaminations) that reflects the types of deviations from π_0 deemed plausible. An easy class with which to work is $\mathcal{Q}$={all prior distributions}. Although this class is large and contains many more distributions than we would consider reasonable in practice, if we find that the difference between the upper and lower bounds of the posterior quantity of interest is small relative to this class, then we should have considerable confidence in the posterior robustness to π_0. We focus here exclusively on this class that consists of all prior distributions because of this feature and because, as we shall see, it is computationally easy to work with. Other choices of $\mathcal{Q}$ considered in the literature are less conservative since they restrict the set of possible contaminations. For example, assuming that π_0 is symmetric and unimodal, a reasonable choice of $\mathcal{Q}$ is the class of all symmetric, unimodal distributions with the same mode as π_0[14].

2.2 Methods

In this section we summarize some relevant results for performing robust Bayesian analysis with the ϵ-contaminated class of priors Γ_ϵ in (2). The posterior distribution for θ with respect to the prior $(1-\epsilon)\pi_0 + \epsilon q$ can be shown to be

$$\pi(\theta|x) = \lambda(x)\pi_0(\theta|x) + (1-\lambda(x))q(\theta|x)$$

where $\pi_0(\theta|x)$ is the posterior distribution for θ with respect to the base prior π_0, $q(\theta|x)$ is the posterior distribution for θ with respect to q, and

$$\lambda(x) = \frac{(1-\epsilon)m_{\pi_0}(x)}{m_\pi(x)}$$

where $m_\pi(x) = (1-\epsilon)m_{\pi_0}(x) + \epsilon m_q(x)$. We note for future reference that both $\pi_0(\theta|x)$ and $m_{\pi_0}(x)$ are fixed. It follows then for $\pi \in \Gamma_\epsilon$ that

$$\begin{aligned} E_{\pi(\theta|x)}(g(\theta)|x) &= \int g(\theta)\pi(\theta|x)d\theta \\ &= \lambda(x)E_{\pi_0}(g(\theta)|x) + (1-\lambda(x))E_q(g(\theta)|x) \end{aligned}$$

where, for example, $E_{\pi_0}(g(\theta)|x) = \int g(\theta)L(\theta)\pi_0(\theta)d\theta/m_{\pi_0}(x)$.

Finally when $\mathcal{Q} = \mathcal{P}$, the set of all priors, it follows from Berger and Berliner[12] that

$$\overline{E}_{\pi(\theta|x)}(g(\theta)|x) = \sup_{P\in\Gamma_\epsilon} E_{\mathcal{P}}(g(\theta)|x) = \sup_\theta \frac{(1-\epsilon)m_{\pi_0}(x)E_{\pi_0}(g(\theta)|x) + \epsilon L(\theta)g(\theta)}{(1-\epsilon)m_{\pi_0}(x) + \epsilon L(\theta)}. \qquad (3)$$

We find the lower bound by minimizing the expression in (3). The minimization and maximization of $E_{\mathcal{P}}(g(\theta)|x)$ over $\mathcal{P} \in \Gamma_\epsilon$ depends only on variation in $m_q(x)$ and $E_q(g(\theta)|x)$ as q varies over $\mathcal{P}$ to yield the bounds given in (3). For the special case where g is the indicator function for a set A we can show that the bounds are

$$\overline{P}(A|x) = \frac{(1-\epsilon)m_{\pi_0}(x)P(A|x) + \epsilon L_A}{(1-\epsilon)m_{\pi_0}(x) + \epsilon L_A} \quad \text{and} \quad \underline{P}(A|x) = \frac{(1-\epsilon)m_{\pi_0}(x)P(A|x)}{(1-\epsilon)m_{\pi_0}(x) + \epsilon L_{A^c}} \qquad (4)$$

where $P(A|x)$ is the posterior probability of A under the base prior and $L_A = \sup_{\theta\in A} L(\theta)$. In practice, we compute these bounds for values of ϵ that vary from 0 to 1, and then we plot the bounds versus ϵ. The interpretation becomes clear when we consider the applications of these methods to the data from the clinical trials presented in sections 3 and 4.

To carry out these computations, we must find $E_{\pi_0}(g(\theta)|x)$, $m_{\pi_0}(x)$ and then either maximize or minimize the expression in (3). Finding $E_{\pi_0}(g(\theta)|x)$ is routine and can be done through numerical integration or, as has become more popular recently, by way of Markov Chain Monte Carlo. For details on the latter, see Tanner and Wong[15], Gelfand and Smith[16] and Tierney[17]. For examples of the sort considered here, simple numerical integration suffices. Computing $m_{\pi_0}(x)$ is more challenging. Usually this involves numerical integration, by importance sampling or by other related techniques. For the examples considered in this paper, simple numerical integration suffices. To compute the suprema and infima, it suffices in simple problems to maximize (or minimize) over a grid. For more complicated problems, we would need a numerical optimization routine. All these calculations are simple for low dimensional problems like those considered here.

3 Example I: Stopping a Trial Early Based on Unexpected Toxicity

3.1 Background

Korn, Yu and Miller[18] describe their experience with monitoring a phase II trial for safety. Briefly, the trial was designed to demonstrate whether the concurrent administration of a biologic agent, granulocyte colony-stimulating factor (G-CSF), given with a standard chemotherapy, 5-fluorouracil and low-dose leucovorin (5-FU/LV), would lessen the incidence of mucositis. It was thought that giving the agents concurrently would increase the effectiveness of the biologic agent in reducing the high incidence of mucositis caused by the chemotherapy. Thirty-five cancer patients with various tumor types, some of whom had received prior chemotherapy, were to be enrolled. Three of the first four patients in the trial developed grade 4 (life-threatening) leukopenia. Although the investigators were aware that leukopenia was a possible side-effect of the therapeutic regimen, there was no formal stopping rule specified in the design of the study. In a previously conducted randomized trial, 12 of 176 (6.8%) of the patients who received the same dose of the chemotherapy but without the biologic agent developed grade 4 leukopenia. After the third patient developed this toxic side effect, the statisticians for the trial were asked to advise on whether or not to stop the trial[18].

3.2 A Robust Bayesian Analysis

Let p denote the incidence rate of grade 4 leukopenia in the trial, which we refer to as the toxicity rate. As inputs into the recommendation on whether to stop the trial or not, we concentrate on two posterior quantities of interest: the posterior mean toxicity rate, $E(p \mid x)$, and the posterior probability that the toxicity rate is greater that 20%, $P(p > 0.20 \mid x)$. We have chosen these measures for illustrative purposes. In practice it is not difficult, and in fact, it is desirable to supply a data monitoring committee with a range of values for, say $P(p > \delta \mid x)$, for different reasonable values of δ (see, also, Spiegelhalter, Freedman, Parmar[4]).

We wish to use information about the toxicity rate from the previous trial to help specify a prior distribution for p. Since, however, the earlier study was historical and included previously untreated patients with different cancers than the current study, we do not consider the historical study and the current study as exchangeable. In other words, we wish to downweight the information from the previous trial. To this end, we take $\pi_0(p)$ as a Beta distribution with parameters $\alpha = 1.56$ and $\beta = 8.44$. This specification has mode equal to 0.07, mean equal to 0.16, and variance equal to 0.012. Therefore we locate the mode of this prior distribution at the observed toxicity rate found in the previous study, but we chose this specification so that the prior is relatively diffuse such that the effective prior sample size is $\alpha + \beta = 10$. Figure 1(a) presents a plot of this prior distribution, $\pi_0(p)$, the binomial likelihood function for the observed data based on three patients with grade 4 leukopenia among the first four patients, and the posterior distribution. The posterior mean with respect to $\pi_0(p)$ is 0.32 and the posterior probability that the toxicity rate is greater than 20% with respect to $\pi_0(p)$ is 0.84. Based on these posterior quantities, we might recommend stopping the trial since the evidence is pointing towards an elevated toxicity rate. This is, in fact, the recommendation that Korn, Yu, and Miller[18] and Emerson[19] make based on frequentist analyses. Since, however, there are only 4 patients in the trial our concern is that the prior distribution might overly influence our conclusions.

Figure 1 about here

To assess the sensitivity of these posterior quantities of interest to the specification of $\pi_0(p)$ we consider an ϵ-contaminated class of prior distributions as in equation (2) with

$\mathcal{Q} = \mathcal{P}$ taken as the set of all priors. Figure 2 is a plot of the upper and lower bounds for $P(p > 0.20 \mid x)$ as a function of ϵ. We find it useful to interpret ϵ as a weight denoting our uncertainty in the specified prior. Therefore $(1-\epsilon)$ reflects our degree-of-belief or confidence in the accuracy of π_0. For example, for $\epsilon = 0.50$, we are equally split between the belief in the specified prior and the class of all other priors. At this value of ϵ we see from the figure that the posterior probability that the toxicity rate is greater than 20% is between 0.45 and 0.95. Similarly, figure 3 plots the bounds for the posterior mean. For $\epsilon = 0.50$, the bounds for the expected toxicity rate are 0.30 and 0.80, respectively. Based on these analyses we have considerable confidence in supporting a recommendation to stop the trial, since even for modestly large amounts of uncertainty in the prior specification, the lower bounds for the expected toxicity rate and for the posterior probability that the toxicity rate is greater than 20% are fairly large.

Figures 2 and 3 about here

Both figures 2 and 3 illustrate how, as the uncertainty in $\pi_0(p)$ increases, the upper (lower) bound for the posterior quantities of interest increases (decreases) monotonically. Thus, as our uncertainty in the specified prior increases, the interval between the upper and lower bounds for the posterior quantities of interest increases, indicating that the inferences are sensitive to the prior. We would expect this for large ϵ and for trials with relatively small n. If for small ϵ, however, the interval between the upper and lower bounds for the posterior quantity of interest is large, we consider this to be a diagnostic that should raise concerns about the sensitivity of the inferences to the choice of prior. In this case, we should either collect more data, refine the specification of the prior distribution, or consider a more restricted class of priors $\mathcal{Q}$. For example, suppose the posterior quantity of interest is the probability that the toxicity rate is greater than 50%. The interval for the posterior bounds for this quantity based on the π_0 given above is quite large even for small ϵ (figure not shown). This high degree of sensitivity to the specified prior follows since 0.50 is in the tail of the likelihood and thereby sensitive to small changes in the prior distribution. Inferences based on *this* posterior quantity of interest, therefore, are sensitive to the specified $\pi_0(p)$ even though inferences based on $P(p > 0.20|x)$ are not.

3.3 Precision Versus Robustness

One might suspect that there is a relationship between sensitivity and precision. That is, we might expect that a posterior distribution with a small variance would indicate robustness. This is false. Consider the extreme case where the prior is a point mass at a point θ_0. Then the posterior is also a point mass at θ_0 and the posterior variance is 0. But clearly the posterior depends heavily on the prior.

The present example provides a more realistic illustration of this phenomenon. Suppose we use the information concerning the grade 4 leukopenia toxicity rate from the previous study without downweighting it, that is, treating that study as exchangeable with the current study. Recall that the previous study found that 12 of 176 patients developed grade 4 leukopenia. We use this information to specify a prior distribution for p which is Beta$(\alpha = 12, \beta = 164)$. This is the posterior obtained using the results from the previous study together with a flat prior. This prior has mean equal to 0.07 and variance equal to 0.0004 (compared to the prior variance in section 3.2 equal to 0.012). This is a very sharp prior distribution, located in the tail of the likelihood function. For this prior, $E(p|x)$ is 0.08 and the $P(p > 0.20|x) < 0.001$. It is clear from Figure 1(b), which is a plot of this prior, likelihood function and posterior, how tightly concentrated the posterior distribution is around its mean.

To assess the sensitivity of the posterior quantities of interest to this prior specification, figures 4 and 5 give the posterior bounds for $P(p > 0.20|x)$ and $E(p|x)$, respectively, with respect to the ϵ-contaminated class with $\mathcal{Q} = \mathcal{P}$. In figure 4, we see that even for ϵ very small the upper bound for $P(p > 0.20|x)$ is nearly 1 and the lower bound is 0. Similarly, in figure 5 the interval between the upper and lower bound for $E(p|x)$ is very large for all ϵ's. Clearly, this analysis indicates that inferences based on these posterior quantities are highly sensitive to the choice of prior despite the precision of the posterior. In this case the robust Bayesian analysis highlights the effect of the large disparity between the specified prior distribution and the data.

Figure 4 and 5 about here

4 Example II: A Sequential Robust Bayesian Analysis of an Efficacy Trial

4.1 Background

The Harvard ECMO clinical trial has received a great deal of attention in the statistics' literature (see Ware[20]). ECMO (extracorporeal membrane oxygenation) was a new therapy for the treatment of persistent pulmonary hypertension of the newborn (PPHN), a highly fatal illness. Based on anecdotal information, infants with PPHN treated with ECMO were reported at several centers to have a survival rate of 80% or better. A chart review of historical experience at the two hospitals participating in the trial showed that of 13 infants who had developed severe PPHN and received conventional medical therapy (CMT), 11 (85%) died. This rate was consistent with the mortality rate among infants with severe PPHN treated with CMT reported at other institutions. To assess the relative efficacy of ECMO to CMT, the Harvard trial used an adaptive two phase design to minimize the number of infants exposed to an inferior treatment. In the first phase of the trial, patients were randomly assigned to either ECMO or CMT using blocks of size four. As soon as four deaths occurred in either of the two treatment groups, randomization ceased, and all subsequent patients received the superior phase I therapy. Nineteen patients were enrolled during the first phase. In the CMT group, 6 of 10 survived and in the ECMO group, 9 of 9 survived (p-value=0.054, Fisher's Exact Test). See Ware[20] for more details.

Some have criticized this study believing that, given the historical experience with CMT, the study should either never have been done or at the very least the investigators should have stopped the trial early before the end of the randomization phase. Taking this point of view, Berry[21], using a single prior distribution, illustrated a sequential Bayesian analysis of the phase I data after having observed the treatment response of each patient. (He also incorporated covariate information into his analysis.) Here, we consider a sequential robust Bayesian analysis of these same data. Our aim is to illustrate how to use the sequential application of these methods to assist a data monitoring committee in a decision about early stopping based on the accumulated evidence. Table 1 gives the sequence of enrollment of patients in the ECMO trial, their treatment group assignments, and the outcome for each patient. We note that others have considered robust Bayesian analyses of these data at the end of the randomization phase but not sequentially (e.g., Kass and Greenhouse[3]; Lavine, Wasserman, and Wolpert[13]).

Table 1 about here

4.2 A Robust Bayesian Analysis

Let p_E and p_C denote the probability of success (i.e., survival) on ECMO and CMT, respectively. For the base prior π_0 on p_E and p_C, respectively, we take independent uniform distributions on [0,1]. As in Kass and Greenhouse[3], we choose a prior specification that reflects the view of a "cautious reasonable skeptic" who is unwilling to state a preference in favor of either treatment. In this case, we chose the specification of uniform prior distributions for p_E and p_C so that we assign a nontrivial probability to a very large treatment effect and at the same time we assign most of the probability to more moderate treatment effects. We consider an ϵ-contaminated class of prior distributions for p_E and p_C mixing over the class of all priors. The posterior quantities of interest after observing the ith patient are the posterior probability that ECMO is better than CMT,
$P(p_E > p_C \mid x_1 \ldots x_i)$, and the posterior mean treatment effect, $E(p_E - p_C \mid x_1 \ldots x_i)$. In the last two columns of Table 1, we present the value of these posterior quantities after the response of each patient with respect to the base uniform prior distribution. Figure 6 (a-d) presents for four specified values of ϵ ($\epsilon = 0.10, 0.20, 0.30, 0.40$) what we call a "picket-fence" plot for the bounds on $P(p_E > p_C \mid x_1 \ldots x_i)$. The horizontal axis for each plot gives the patient number in sequential order from patient 1 to patient 19. The upper (lower) end point of the vertical line at patient i is the upper (lower) bound for $P(p_E > p_C \mid x_1 \ldots x_i)$ for that specified value of ϵ.

Figure 6 and 7 about here

With the addition of each new patient, we see that the trend is for the value of the upper and lower bounds on $P(p_E > p_C \mid x_1 \ldots x_i)$ to increase for each specified ϵ and for the length of the interval to decrease. As we express more and more uncertainty in the specified uniform prior distributions, however, the lower bound for $P(p_E > p_C \mid x_1 \ldots x_i)$ for $\epsilon = 0.40$ straddles 0.50. It is only for relatively small ϵ (e.g. $\epsilon = 0.10$) that the lower bound for the posterior probability that ECMO is better than CMT consistently exceeds 50% early in the trial. Figure 8 (a-d) presents a picket-fence plot for $E(p_E - p_C \mid x_1 \ldots x_i)$ for $\epsilon = 0.10, 0.20, 0.30, 0.40$. We again see that as the number of patients in the trial increases there is a trend for the values of the bounds on the expected posterior treatment effect to increase for each specified value of ϵ. As ϵ increases, however, the length of the interval $\overline{E} - \underline{E}$ also increases. For $\epsilon = 0.40$, we see that starting with the tenth or eleventh patient, the lower bound for $E(p_E - p_C \mid x_1 \ldots x_i)$ is greater than 0 and remains positive.

In summary, for a small degree of uncertainty in the prior specification ($\epsilon = 0.10$), ECMO appears consistently superior to CMT starting with the tenth patient. The analyses based on slightly larger ϵ's, however, suggest that the evidence does not favor ECMO with sufficient strength to stop the trial early. As noted earlier, by taking $\mathcal{Q}$={all prior distributions}, we have chosen to be the most conservative about the class of contaminations over which we are mixing. Some might argue that from a data monitoring committee's perspective conservatism is desirable in such an analysis. Nevertheless, one can and should consider other choices of $\mathcal{Q}$.

5 Discussion

The decision to stop a trial is difficult and depends on many factors both objective and subjective. In particular, different individuals hold different opinions about the effectiveness of a treatment and thereby attach different degrees of importance to various pieces of information concerning the merits of a treatment. For example, to justify the ethics of randomization, Kass and Greenhouse[3] discuss priors based on a "cautious reasonable skeptic" who is unwilling to state a preference in favor of either of the treatments. Spiegelhalter, Freedman, and Parmar[4] suggest use of a "community of priors" in monitoring a trial that range from reference priors that represent minimal prior information to "enthusiastic" priors that reflect the belief of individuals "who are reluctant to stop a trial when results supporting the null hypothesis are observed." Thus, it is understandable that in applying Bayesian methods to the analysis of clinical trials, a major concern is the dependence of the inferences on the prior distribution. Even in the Bayesian literature a number of authors have commented on the importance of checking the sensitivity of the posterior to the choice of prior[4,19,21,22] As we have shown in this paper, the use of robust Bayesian methods to carry out such a sensitivity analysis is relatively straightforward.

The easiest method for assessing sensitivity is to use an ϵ-contamination class where the set of contaminations consists of all priors and then to plot the bounds of the posterior quantity of interest as a function of ϵ. We have introduced several graphical displays for this purpose. Although this class is too large in that it contains many unreasonable priors, if robustness obtains over this large class then we can be certain that robustness will obtain over a smaller set of reasonable priors. Even though we have emphasized the use of ϵ-contaminated classes of priors, there are other classes that are nearly as easy with which to work. Berger[7] and Wasserman[23] review many of these classes.

Edwards, Lindman and Savage[24] described the situation where the likelihood function is relatively concentrated in a region of the parameter space over which the prior is essentially constant. In this situation, which they called *stable estimation*, inferences are insensitive to the specification of the prior distribution. Thus, one might expect that sensitivity to the prior decreases as the sample size increases and the likelihood function becomes more concentrated. Indeed, this is the case for the ϵ-contaminated family of prior distributions in that, for fixed $\epsilon > 0$, $\overline{E}(g(\theta)|x) - \underline{E}(g(\theta)|x)$ tends to 0 as $n \to \infty$. However, the convergence is slower than the usual rate at which a standard error decreases. Specifically, the above quantity tends to 0 at rate $O(\sqrt{\log n/n})$ (Sivaganesan[25]) . In fact, it is interesting to note that this rate obtains for many other classes in addition to the ϵ-contaminated class[26].

We have not discussed sensitivity to the model. In principle, one can assess this with methods similar to those used here. But usually, assessment of sensitivity to the model is best handled by means of other methods. For example, one can compare competing parametric models by way of Bayes factors (see, Kass and Raftery[27]).

ACKNOWLEGEMENTS

This work was supported in part by grants from the National Institutes of Health, CA54852, MH15758, and MHCRC30915, and by the National Science Foundation, grant DMS-9005858.

References

1. Cornfield, J. The Bayesian outlook and its application. *Biometrics*, 24:617-657, (1969).

2. Freedman, L. and Spiegelhalter, D. Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Controlled Clinical Trials: Design, Methods, and Analysis*, 10:357-367, (1989).

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

4. Spiegelhalter, D.J., Freedman, L.S., and Parmar, M.K.B. Bayesian approaches to randomized trials. To appear in *Journal of the Royal Statistical Society, A*, (1994).

5. Good, I.J. Subjective probability as the measure of a non-measurable set. In *Logic, Methodology and the Philosophy of Science.* Stanford: University Press, 319-329, (1962).

6. Berger, J. The robust Bayesian viewpoint (with discussion). In *Robustness in Bayesian Statistics.* (J. Kadane ed.): North-Holland, Amsterdam, 1984.

7. Berger, J. Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inference* **25** 303-328, (1990).

8. Freedman, L. and Spiegelhalter, D. The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *The Statistician*, 32:153-160, (1983).

9. Kadane, J. Progress toward a more ethical method for clinical trials. *Journal of Medicine and Philosophy*, 11:385-404, (1986).

10. Chaloner, K., Church, T., Louis, T., and Matts J. Graphical elicitation of a prior distribution for a clinical trial. *Statistician*, 42:341-353, (1994).

11. Huber, P.J. The use of Choquet capacities in statistics. *Bulletin of the International Statistics Institute*, 45:181-191, (1973).

12. Berger, J. and Berliner, L. M. Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *Annals of Statistics*, 14:461-486, (1986).

13. Lavine, M., Wasserman, L. and Wolpert, R. Bayesian inference with specified prior marginals. *J. Amer. Statist. Assoc.* **86**, 964-971, (1991).

14. Sivaganesan, S. and Berger, J. Range of posterior measures for priors with unimodal contaminations. *Annalys of Statistics*, 17:868-889, (1989).

15. Tanner, M. and Wong, W. The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82** 528-550, (1987).

16. Gelfand, A. and Smith, A.F.M. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association,* **85**, 398-409, (1990).

17. Tierney L. Exploring posterior distributions using Markov chains. In *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface.* (ed. E.M. Keramidas), 563-570. Fairfax Station: Interface Foundation, 1991.

18. Korn, E. L., Yu, K.F., and Miller L. L. Stopping a clinical trial very early because of toxicity: Summarizing the evidence. *Controlled Clinical Trials*, 14:286-295, (1993).

19. Emerson, S. S. Stopping a clinical trial very early based on unplanned interim analyses: A group sequential approach. Submitted to *Biometrics*, (1994).

20. Ware, J. Investigating therapies of potentially great benefit: ECMO, *Statistical Science*, 4:31-317, (1989).

21. Berry, D. A. Monitoring accumulating data in a clinical trial. *Biometrics*, 45:1197-1211, (1989).

22. Hughes, M.D. Reporting Bayesian analyses of clinical trials. *Statistics in Medicine*, 12:1651-1663, (1993).

23. Wasserman, L. Recent methodological advances in robust Bayesian inference (with discussion). In *Bayesian Statistics IV: Fourth Valencia International Meeting on Bayesian Statistics.* p. 483-502, 1992.

24. Edwards, W., Lindman, H. and Savage, L.J. Bayesian statistical inference for psychological research. *Psychological Review*, 70:193-242, (1963).

25. Sivaganesan, S. Range of posterior measures for priors with arbitrary contaminations. *Comm. Stat. Theory and Methods.* **17**, 1591-1612, (1988).

26. Gustafson, P. *Local Sensitivity in Bayesian Statistics.* Ph.D. Dissertation, Department of Statistsics, Carnegie Mellon University, 1994.

27. Kass, R.E. and Raftery, A. Bayes Factors and Model Uncertainty. Technical report 571, Department of Statistics, Carnegie Mellon University, (1993).

Table 1. Sequence of enrollment of patients in the ECMO trial, treatment group assignments, and posterior quantities of interest computed after the outcome is observed for each patient.

Patient #	Treatment	Survive (Y/N)	$\mathrm{P}(p_E > p_C \mid x_1 \ldots x_i)$	$E(p_E - p_C \mid x_1 \ldots x_i)$
1	ECMO	Yes	0.66	0.16
2	CMT	Yes	0.49	0.00
3	ECMO	Yes	0.59	0.08
4	CMT	Yes	0.49	0.00
5	**CMT**	**No**	0.70	0.15
6	ECMO	Yes	0.78	0.20
7	ECMO	Yes	0.82	0.23
8	CMT	Yes	0.77	0.16
9	CMT	Yes	0.71	0.11
10	**CMT**	**No**	0.83	0.20
11	ECMO	Yes	0.87	0.23
12	ECMO	Yes	0.90	0.25
13	**CMT**	**No**	0.94	0.31
14	ECMO	Yes	0.96	0.33
15	ECMO	Yes	0.97	0.34
16	CMT	Yes	0.95	0.30
17	CMT	Yes	0.94	0.26
18	ECMO	Yes	0.95	0.27
19	**CMT**	**No**	0.97	0.32

Figure Captions

Figure 1. Beta prior distribution, binomial likelihood function and posterior distribution for p for Example I. (a.) Prior is Beta(1.56, 8.44); (b.) Prior is Beta(12,164).

Figure 2. The upper and lower bounds for $P(p > 0.20|x)$ as a function of ϵ with the base prior, $\pi_0(p)$, as Beta(1.56, 8.44).

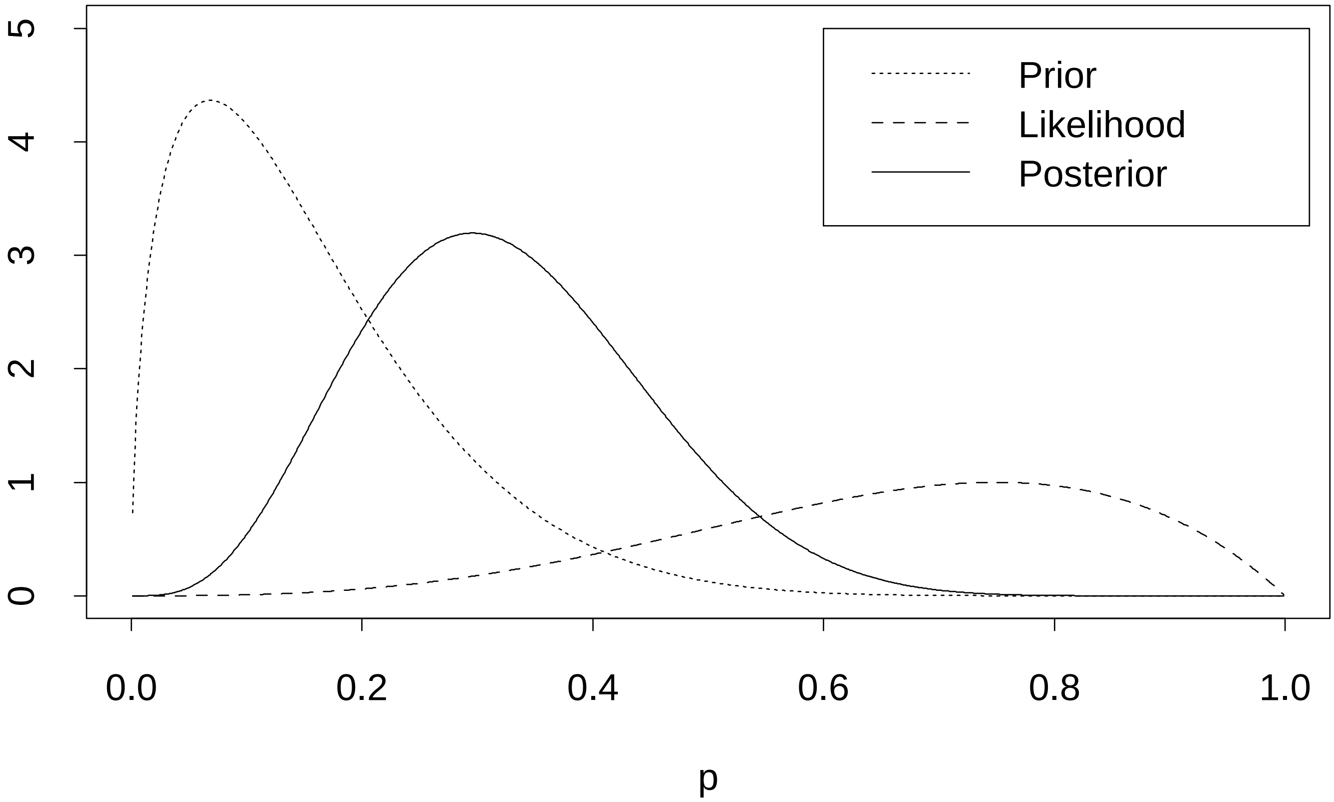
Figure 3. The upper and lower bounds for $E(p|x)$ as a function of ϵ with the base prior, $\pi_0(p)$, as Beta(1.56, 8.44).

Figure 4. The upper and lower bounds for $P(p > 0.20|x)$ as a function of ϵ with the base prior, $\pi_0(p)$, as Beta(12, 164).

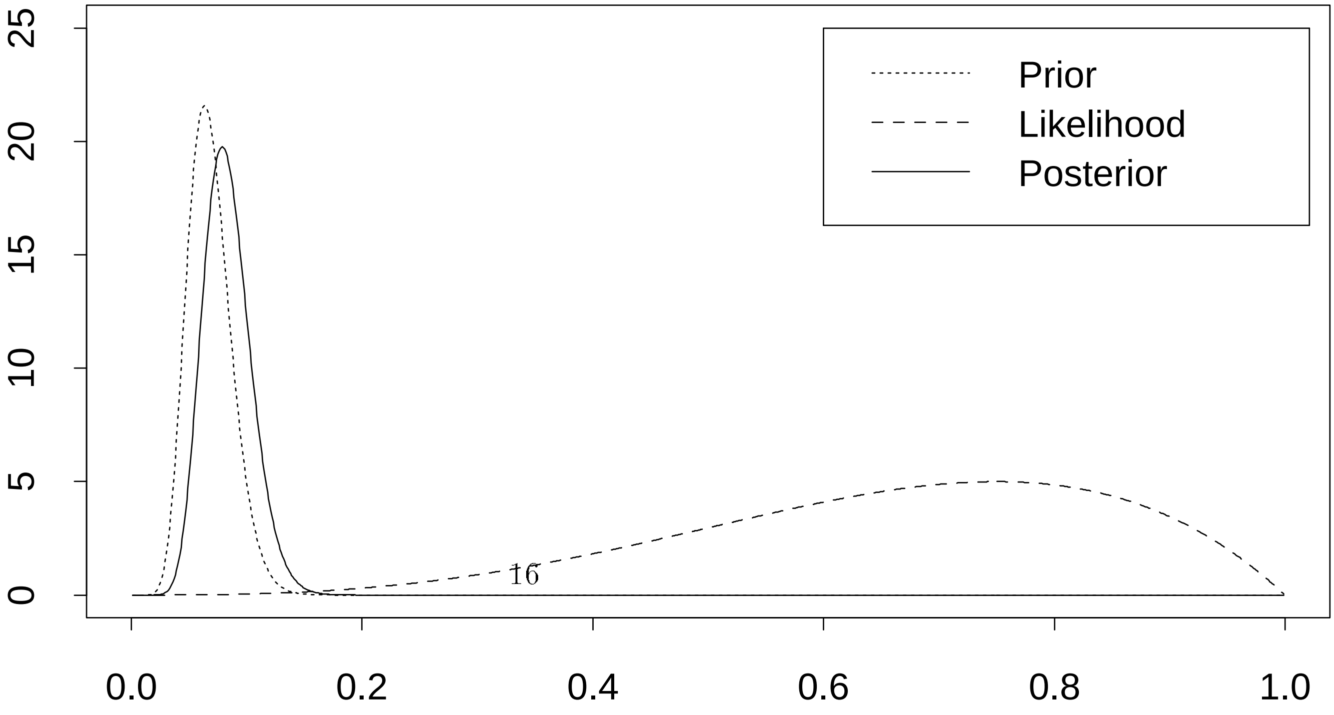
Figure 5. The upper and lower bounds for $E(p|x)$ as a function of ϵ with the base prior, $\pi_0(p)$, as Beta(12, 164).

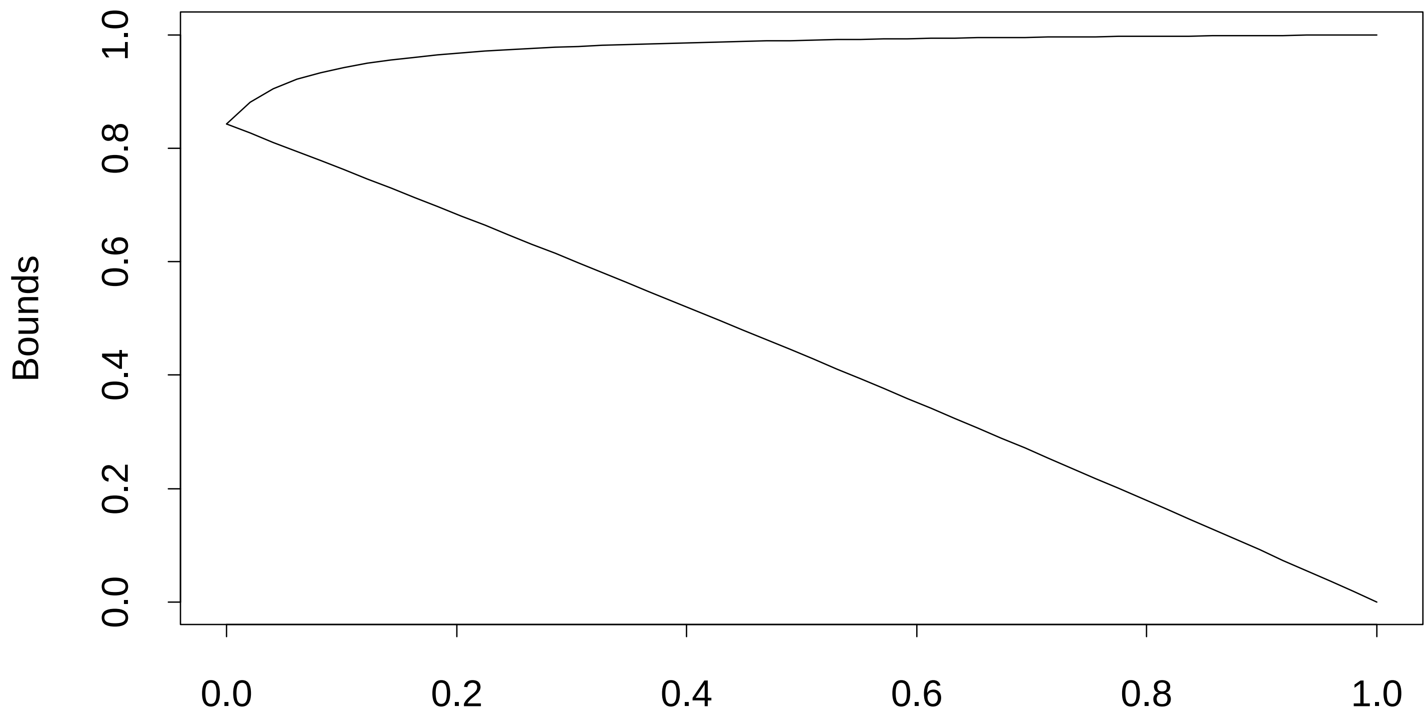
Figure 6. Picket fence plot of the upper and lower bounds for $P(p_E > p_C \mid x_1 \ldots x_i)$ for each patient computed sequentially for (a) $\epsilon = 0.10$, (b) $\epsilon = 0.20$, (c) $\epsilon = 0.30$, and (d) $\epsilon = 0.40$.

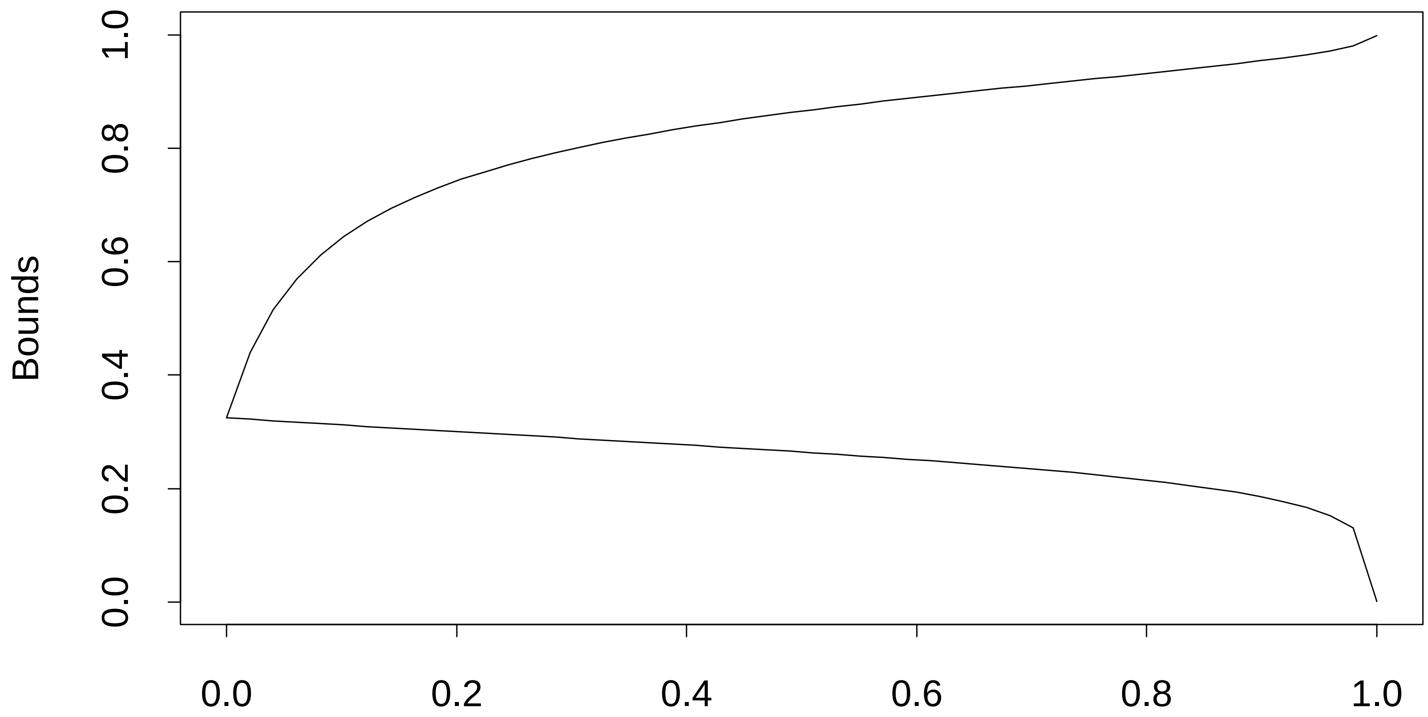
Figure 7. Picket fence plot of the upper and lower bounds for $E(p_E - p_C \mid x_1 \ldots x_i)$ for each patient computed sequentially for (a) $\epsilon = 0.10$, (b) $\epsilon = 0.20$, (c) $\epsilon = 0.30$, and (d) $\epsilon = 0.40$.

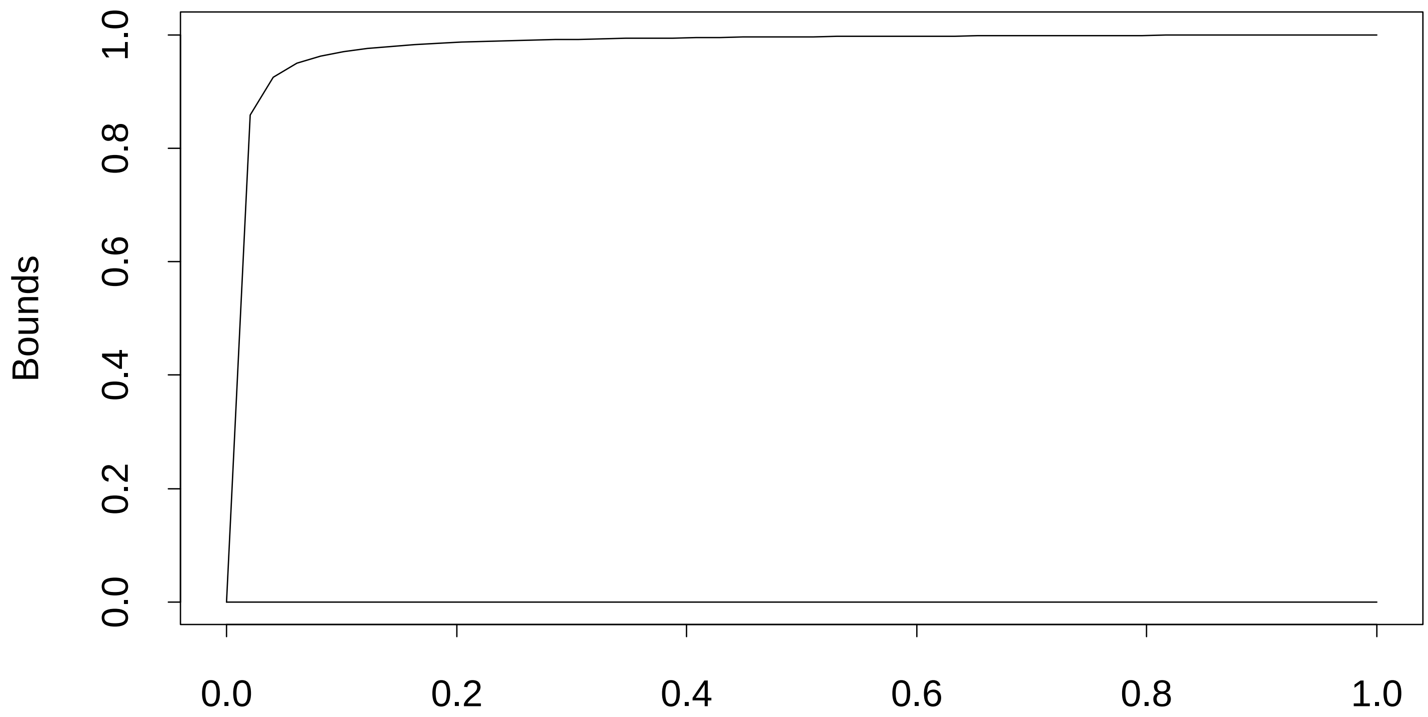


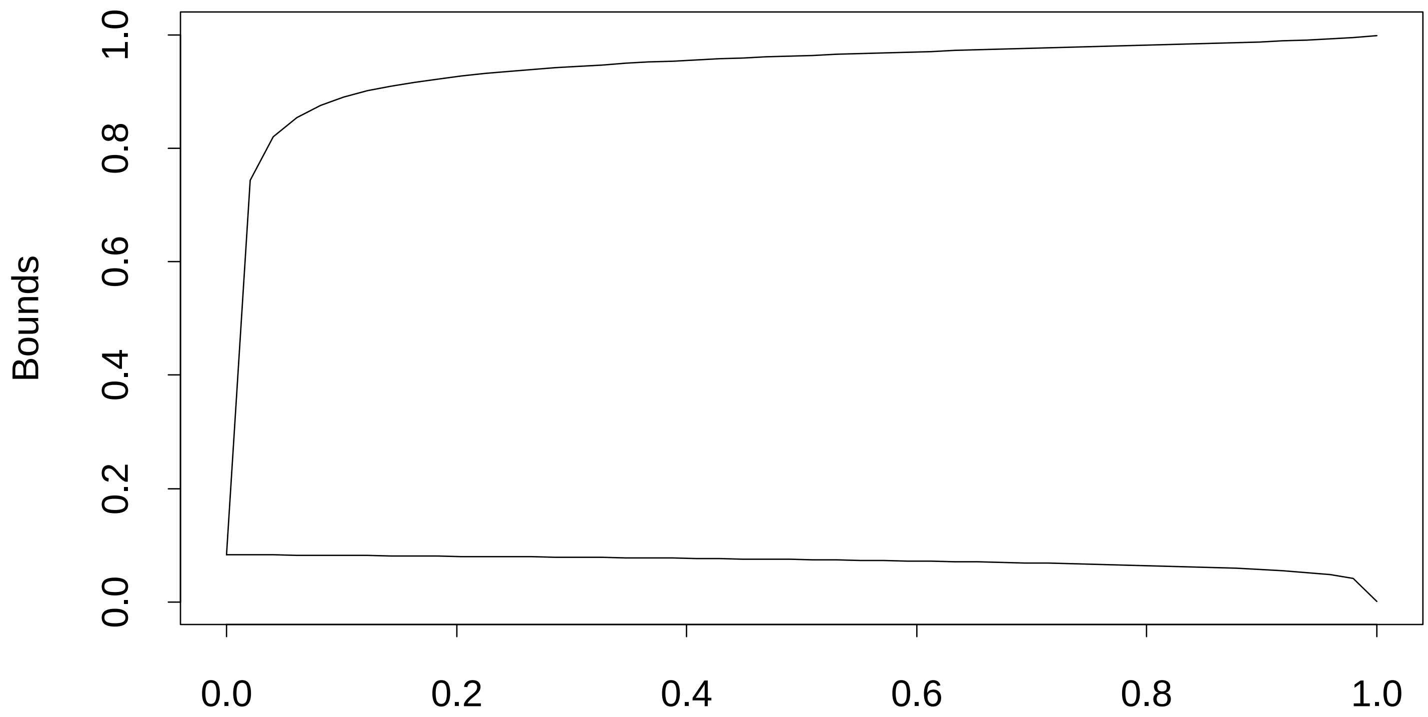
Prior is Beta (1.56,8.44)

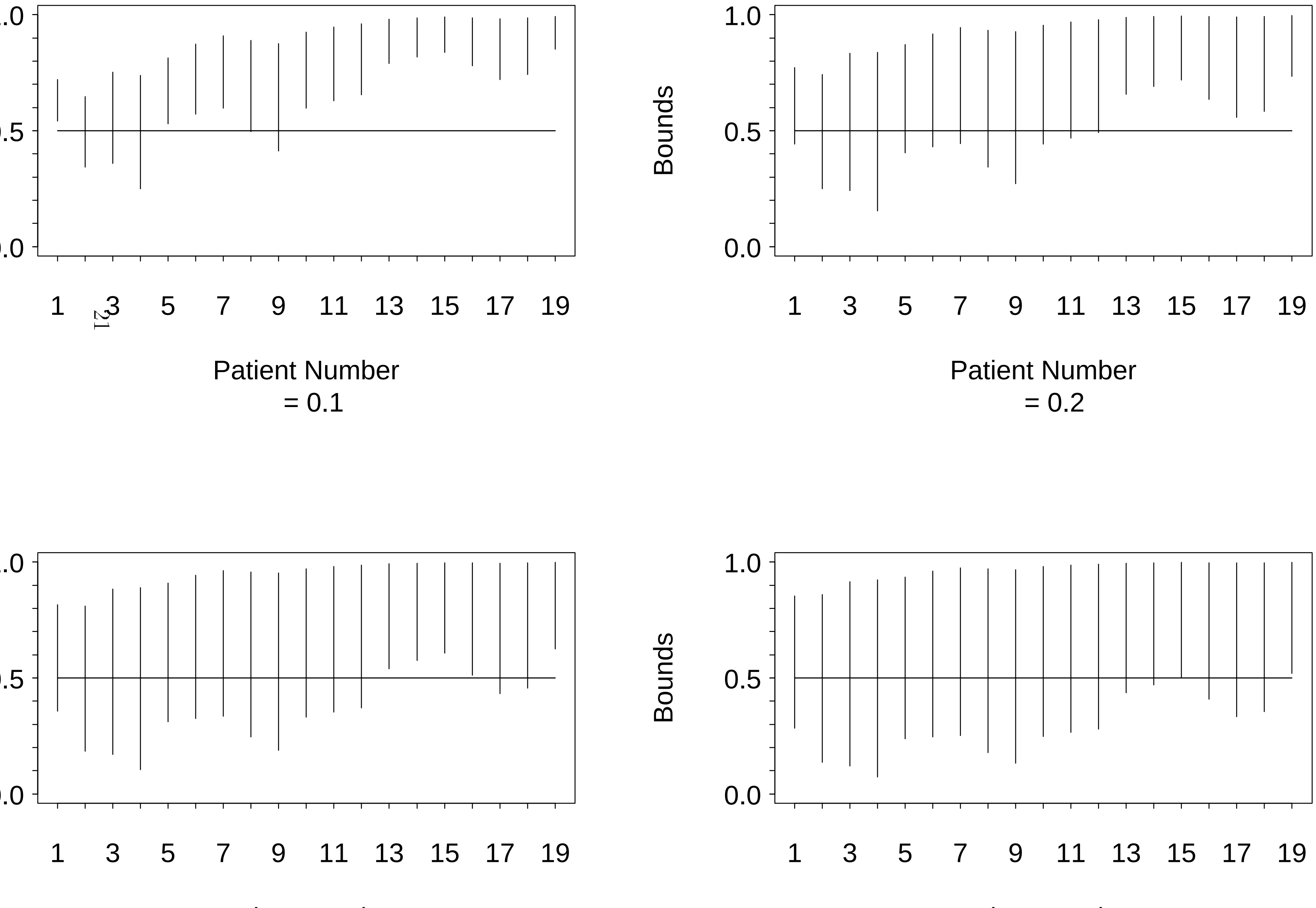


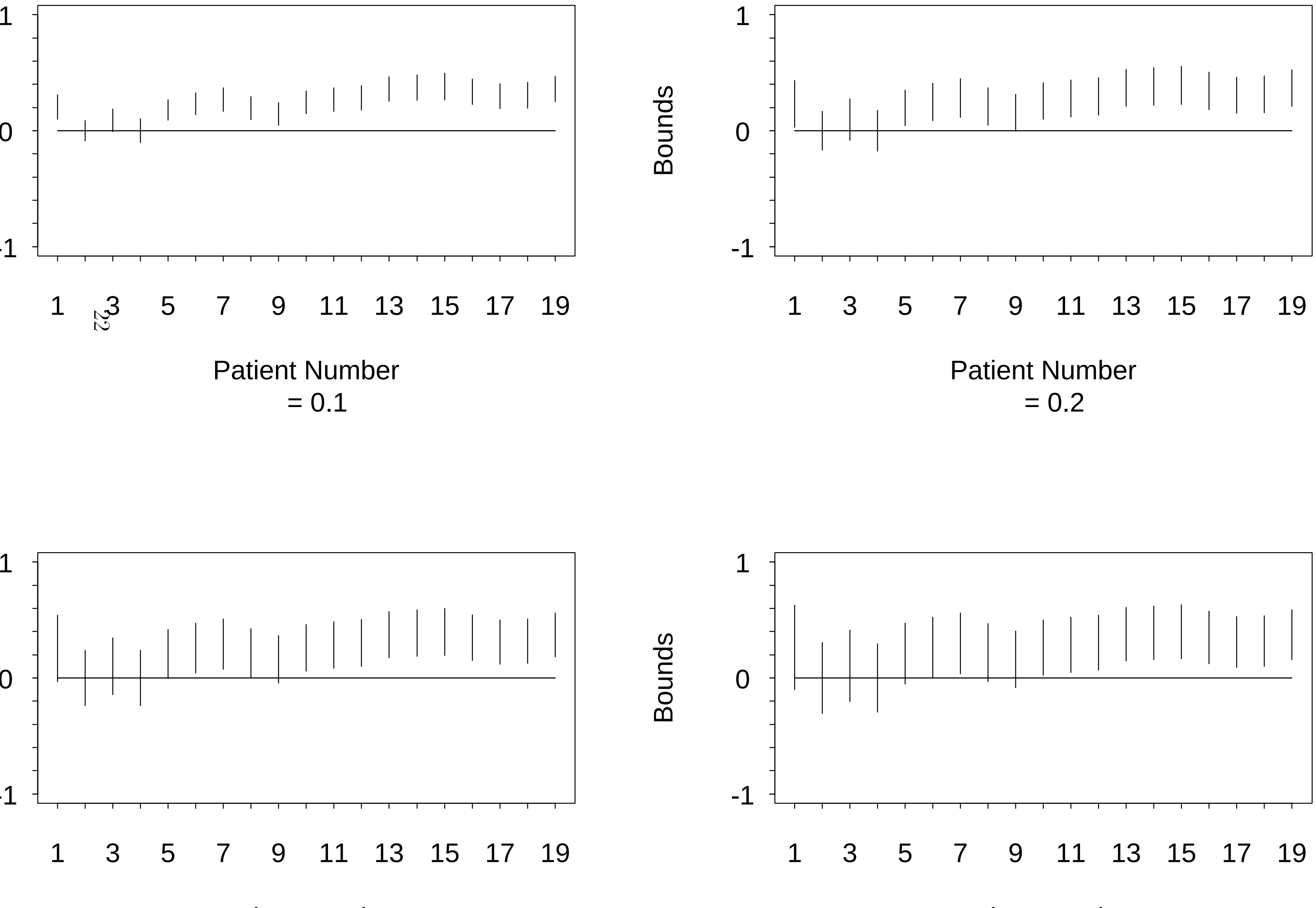

Bounds
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


Bounds
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


Bounds
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


Bounds
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


Bounds
1.0
0.5
0.0
1 3 5 7 9 11 13 15 17 19
Patient Number
= 0.1
= 0.2



Bounds
Bounds
1
0
-1
1 3 5 7 9 11 13 15 17 19
Patient Number
= 0.1
Patient Number
= 0.2