Robust Bayesian Approaches for Clinical Trial Monitoring

By Bradley P. Carlin and Daniel J. Sargent

Division of Biostatistics, School of Public Health, University of Minnesota,
Box 303 Mayo Memorial Building, Minneapolis, Minnesota 55455-0392, U.S.A.

Correspondence author: Bradley P. Carlin
telephone: (612) 624-6646
fax: (612) 626-0660
email: brad@muskie.biostat.umn.edu

August 3, 1995
Robust Bayesian Approaches for Clinical Trial Monitoring

Abstract

The interim monitoring and final analysis of data arising from a clinical trial require an inferential method capable of convincing a broad group of potential consumers: doctors, patients, politicians, members of the media, and so on. While Bayesian methods offer a powerful and flexible analytic framework in this setting, this need to convince a diverse community necessitates a practical approach for studying and communicating the robustness of conclusions to the prior specification. In this paper we attempt to characterize the class of priors leading to a given decision (such as stopping the trial and rejecting the null hypothesis) conditional on the observed data. We evaluate the practicality and effectiveness of this procedure over a range of smoothness conditions on the prior class. First, we consider a nonparametric class of priors restricted only in that its elements must have certain prespecified quantiles. We then obtain more precise results by further restricting the prior class, first to a nonparametric class whose members are quasiuniformal, then to a semiparametric normal mixture class, and finally to the fully parametric normal family. We illustrate all of our comparisons with a dataset from an AIDS clinical trial that compared the effectiveness of the drug pyrimethamine and a placebo in preventing toxoplastic encephalitis.
1 Introduction

As its recent appearance [1] in a major medical journal suggests, the Bayesian approach to the interim monitoring and final analysis of clinical trials has gained appreciation not only by statisticians, but by clinicians as well. The strengths of this methodology include the independence of the inference from the stopping rule, easy interpretability of the results, realistic sample size determination based on the full range of the experimenter's prior beliefs, and freedom from the need to prespecify the number of looks at the data or the form of an "\( \alpha \)-spending function" (see e.g. Lan and DeMets [2]). Bayesian methodology also blends easily with formal decision-theoretic tools in settings that require a specific conclusion, such as a therapy recommendation for a particular patient; see Berry [3] for an excellent discussion of this aspect. Thorough reviews of the use of Bayesian methodology in clinical trials have been provided recently by Spiegelhalter, Freedman and Parmar [4, 5].

Nevertheless, practitioners have been slow to embrace Bayesian methods, due in large measure to the occasionally pronounced dependence of the resulting inference on the particular choice of the prior distribution on the model parameters. This is of course a potential problem in any Bayesian setting, for which a standard remedy is to repeat the analysis using one or more different (but still plausible) priors, checking to see if one obtains a noticeable change in conclusions. Agreement among all the inferences drawn implies that the data are strongly informative, and hence the precise form of prior distribution is irrelevant. Disagreement precludes a single "correct" summary of the trial, but still serves to quantify the range of plausible treatment effects and the sensitivity of the conclusions to the prior. In the specific context of clinical trials, Spiegelhalter et al. [4, 5] suggest implementing such an approach using a "clinical" prior, representing the (typically optimistic) prior feelings of the trial’s investigators, a "skeptical" prior, reflecting the opinion of a person or regulatory
agency that doubts the treatment’s effectiveness, and a “noninformative” prior, a neutral position that leads to posterior summaries formally equivalent to those produced by standard maximum likelihood techniques.

An alternative to this “forward” approach to prior robustness (where one respecifies the prior and recomputes the result) is the “backward,” or prior partitioning approach of Carlin and Louis [6]. In the case of a point null hypothesis and a two-sided alternative for a treatment effect $\theta$, these authors attempt to characterize the class of priors that lead to a particular conclusion (e.g., stopping the trial and deciding in favor of the treatment) given the observed data. Since this class can be quite large, they suggest restricting attention to “plausible” priors, such as those that have a certain mean, a certain mean and variance, or certain quantiles (e.g. median, or $5^{th}$ and $95^{th}$ percentiles). Sargent and Carlin [7] extend this work to the case of an interval null hypothesis. This formulation is actually more natural for clinical trials work since such settings often involve an indifference zone, $[\theta_L, \theta_U]$, within which one is indifferent as to the use of treatment or placebo. For example, we might take $\theta_L = 0$ but $\theta_U > 0$ if there were increased costs or toxicities associated with the treatment.

Prior partitioning calculations are of a type often done in studies of Bayesian robustness, an area pioneered by Edwards, Lindman, and Savage [8]. There is a vast literature in this area, of which Berger [9] provides a comprehensive review; we mention only a few particularly important papers. In the point null setting, Berger and Sellke [10] and Berger and Delampady [11] show that we attain the minimum of $P(\theta = \theta_0|x)$ over all conditional priors $G$ for $\theta \neq \theta_0$ when $G$ places all of its mass at $\hat{\theta}$, the maximum likelihood estimate of $\theta$. Even in this case, where $G$ is working with the data against $H_0$, these authors showed that the resulting $P(\theta = \theta_0|x)$ values are typically still larger than the corresponding two-sided p-value, suggesting that the standard frequentist approach is biased against $H_0$ in this case. In the interval null hypothesis setting, prior partitioning is reminiscent
of the work of O’Hagan and Berger [12], who obtain bounds on the posterior probability content of each of a collection of intervals that form the support of a univariate parameter, under the restriction that the prior probability assignment to these intervals is in a certain sense unimodal. In the specific context of clinical trial monitoring, Greenhouse and Wasserman [13] compute bounds on posterior expectations and tail areas (stopping probabilities) over an $\epsilon$-contaminated class of prior distributions [14].

In Section 2 we review the basics of nonparametric prior partitioning, following the papers by Carlin and Louis [6] and Sargent and Carlin [7]. In Section 3, we then extend the approach by restricting to a set of priors that meet certain fundamental smoothness conditions. These conditions can increase the precision of our results while they still encompass an acceptably wide range of prior beliefs. In particular, we investigate the class of quasiunimodal priors with certain prespecified quantiles, a semiparametric class that treats the indifference zone separately from the remainder of the parameter space, and a fully parametric family chosen as conjugate with the likelihood. We apply our findings to monitoring an AIDS clinical trial in Section 4, and discuss the practicality of our results and their ability to convince a broad audience. Finally, in Section 5 we summarize our findings and suggest areas for future research.

2 Prior partitioning

Consider first the point null testing scenario investigated by Carlin and Louis [6], namely $H_0 \colon \theta = \theta_0$ versus $H_1 \colon \theta \neq \theta_0$. Without loss of generality, set $\theta_0 = 0$. Suppose we have an observation $x$ that has density $f(x|\theta)$, where $\theta$ is an unknown scalar treatment effect parameter. Let $\pi$ represent the prior probability of $H_0$, and $G(\theta)$ the prior cumulative distribution function (cdf) of $\theta$ conditional on $\{\theta \neq 0\}$. Then the complete prior cdf for $\theta$ is $F(\theta) = \pi I_{[0,\infty]}(\theta) + (1 - \pi)G(\theta)$,
where \( I_S \) is the indicator function of the set \( S \). The posterior probability of the null hypothesis is therefore
\[
P_G(\theta = 0|x) = \frac{\pi f(x|0)}{\pi f(x|0) + (1 - \pi) \int f(x|\theta) dG(\theta)}.
\]
For a given prior distribution \( G \) and some \( p \in (0,1) \), we stop the experiment and reject the null hypothesis if \( P_G(\theta = 0|x) \leq p \). Elementary calculations show that characterizing this class of priors \( \{G\} \) is equivalent to characterizing the set \( \mathcal{H}_c \), defined as
\[
\mathcal{H}_c = \left\{ G : \int f(x|\theta)dG(\theta) \geq c = \frac{1 - p}{p} \frac{\pi}{1 - \pi} f(x|0) \right\}.
\]
Carlin and Louis [6] 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.

Turning to the interval null hypotheses \( H_0 : \theta \in [\theta_L, \theta_U] \) and \( H_1 : \theta \notin [\theta_L, \theta_U] \), let \( \pi \) again denote the prior probability of \( H_0 \), and let \( G(\theta) \) now correspond to the prior cdf of \( \theta \) given \( \theta \notin [\theta_L, \theta_U] \). Making the simplifying assumption of a uniform prior over the indifference zone, the posterior probability of \( H_0 \) is computable by Bayes Rule as
\[
P_G(\theta \in [\theta_L, \theta_U]|x) = \frac{\int_{\theta_L}^{\theta_U} f(x|\theta) \left[ \frac{\pi}{\theta_U - \theta_L} I_{[\theta_L, \theta_U]}(\theta) + (1 - \pi) g(\theta) \right] d\theta}{\int f(x|a) \left[ \frac{\pi}{\theta_U - \theta_L} I_{[\theta_L, \theta_U]}(a) + (1 - \pi) g(a) \right] da}.
\]
In our decision paradigm, the priors \( G \) that lead to rejection of \( H_0 \) are those for which (1) is less than or equal to some prespecified probability \( p \). Since \( g(\theta) \) has no support on the interval \([\theta_L, \theta_U]\), this is equivalent to describing the set
\[
\mathcal{H}_c = \left\{ G : \int f(x|\theta)dG(\theta) \geq c = \frac{1 - p}{p} \frac{\pi}{1 - \pi} \left( \frac{1}{\theta_U - \theta_L} \int_{\theta_L}^{\theta_U} f(x|\theta)d\theta \right) \right\}.
\]
Sargent and Carlin [7] restrict the class of candidate $G$'s somewhat by considering only those for which $P_G(\theta \leq \xi_L) = a_L$ and $P_G(\theta > \xi_U) = a_U$ for some fixed $\xi_L$ and $\xi_U$, where $a_L$ and $a_U$ lie in the unit simplex. That is, they require that the prior cdf $G$ passes through the points $(\xi_L, a_L)$ and $(\xi_U, 1 - a_U)$. They further assume that $\max(\xi_L, \theta_L) \leq \min(\xi_U, \theta_U)$, and that $f(x|\theta)$ is a unimodal function of $\theta$ for fixed $x$ that vanishes in both tails. Due to the asymptotic normality of the observed likelihood function, this final assumption is at least approximately true for large datasets. These assumptions lead to expressions for $\sup_G \int f(x|\theta) dG(\theta)$ and $\inf_G \int f(x|\theta) dG(\theta)$, where the sup and inf are over the restricted class of $G$'s described above. Since $\mathcal{H}_c$ is empty if the sup does not exceed $c$, we can use the supremum expression to determine whether there are any $G$ that satisfy equation (2), i.e., whether any priors $G$ exist that enable stopping to reject the null hypothesis. Similarly, the infimum expression may be useful in determining whether any $G$ enable stopping to reject the alternative hypothesis, $H_1$.

Sargent and Carlin [7] consider searching over $(a_L, a_U)$ pairs for a fixed region $(\xi_L, \xi_U)$, as well as the reverse procedure of finding quantile pairs $(\xi_L, \xi_U)$ that permit stopping for fixed prior tail areas $(a_L, a_U)$. In all cases, however, the weak restrictions on the form of the prior lead to prior classes that, while plausible, are typically too broad for practical use. In the next section, we consider a sequence of increasingly tight restrictions on the shape and smoothness of permissible priors, which in turn enable increasingly informative results.

3 Further restricting the prior class

3.1 Quasiunimodal prior with prespecified quantiles

A first reasonable restriction to the class of allowable prior distributions is to consider only continuous and unimodal priors. Almost certainly, any prior a clinician would offer is of this form, and such a class still covers a very broad range of possible prior distributions. Berger and
O'Hagan [15] investigated this particular restriction, and found the restriction to the general class of unimodal priors computationally challenging (though see Lavine, Wasserman and Wolpert [16] for a linearization method that greatly eases the computational burden). An easier class to consider is the class of all quasiunimodal priors, where quasiunimodality is a slight relaxation of the unimodality restriction at the boundaries of the intervals over which one defines the prior. O'Hagan and Berger [12] hypothesize that the difference between optimizing over these two classes is minimal, and so we consider only the quasiunimodal class. Below, we apply their method to the clinical trial setting with an interval null hypothesis.

Our clinical trial setting requires specification of the prior probability content of only three intervals: two tail areas and an indifference zone. This precludes an analysis wherein one fixes the tail probabilities \( a_L \) and \( a_U \) at small values and then elicits \( \xi_L \) and \( \xi_U \) from the investigator. Such an approach fixes the prior probability for three segments of the real line, but leaves the distribution of prior mass on two remaining intervals unspecified. Thus to apply prior partitioning in this setting we must instead fix the indifference zone and search over values for the prior tail probabilities. For convenience, we set \( \xi_L = \theta_L \) and \( \xi_U = \theta_U \).

The restriction to quasiunimodal priors in our clinical trial setting proceeds in the following manner. Define intervals \( I_1, I_2, I_3 \) with endpoints \((a_0, a_1), (a_1, a_2), (a_2, a_3)\) as the regions \((\infty, \theta_L), (\theta_L, \theta_U), (\theta_U, \infty)\) respectively. Define the constants \( q_i \) by

\[
q_i = \frac{p_i}{a_i - a_{i-1}},
\]

where \( p_i \) is the prior probability on interval \( I_i \). Note that this definition assigns \( q_1 = q_3 = 0 \), thus these \( q \)'s satisfy the conditions for unimodality [12]. A key limiting assumption to the application of the quasiunimodality constraints is that we assume that the mode of the prior is in the interval with
the largest $q_i$, thus in the current application we allow only priors with modes in the indifference zone.

Given this framework, and assuming a unimodal likelihood $f(x|\theta)$, maximizing or minimizing the posterior probability of an interval reduces to optimizing the posterior probability of each interval independently. O’Hagan and Berger [12] give an algorithm for determining the prior that leads to the maximum or minimum posterior mass for each interval. This algorithm depends on the shape of the likelihood in the interval, the location of the prior mode (necessarily $I_2$ in our application), and the interval over which we are optimizing. Finally, we can easily calculate the quantities of interest, the maximum and minimum posterior probabilities of intervals $I_1, I_2$, and $I_3$ subject to quasimodality constraints, from the maximizing and minimizing priors as follows. Define $\omega_i = \int_{\omega_{i-1}}^{\omega_i} g_i(\theta) f(x|\theta) d\theta$, where $g_i$ isolates that part of the prior $g$ that is in $I_i$. Notice that we can calculate the posterior probability of the indifference zone $I_2$ as

$$P(I_2|x) = \left(1 + \frac{\omega_1 + \omega_3}{\omega_2}\right)^{-1}.$$ 

Thus

$$\sup P(I_2|x) = \overline{P}(I_2|x) = \left(1 + \frac{\omega_1 + \omega_3}{\omega_2}\right)^{-1} \quad (3)$$

and

$$\inf P(I_2|x) = \underline{P}(I_2|x) = \left(1 + \frac{\omega_1 + \omega_3}{\omega_2}\right)^{-1} \quad (4)$$

where $\overline{\omega}_i = \sup \omega_i$ and $\underline{\omega}_i = \inf \omega_i$, where the suprema and infima are over the class of quasimodal priors that assign mass $p_i$ to interval $I_i$, and have modes in $I_2$. O’Hagan and Berger [12] give expressions for $\overline{\omega}_i$ and $\underline{\omega}_i$, $i = 1, 2, 3$, based on the optimizing prior for each interval. We may then use equations (3) and (4) to determine which combinations of $p_1, p_2,$ and $p_3$ lead to $P(I_2|x) < p$, and
and which lead to $P(I_2|x) < p$. For the former class, rejection of $H_0$ is possible for at least one quasiunimodal prior with mode in $I_2$; for the latter, rejection of $H_0$ occurs for all quasiunimodal priors with modes in $I_2$.

### 3.2 Semiparametric prior

Yet another approach to obtain more specific results via prior partitioning is to retain the mixture form used in equation (1),

$$h(\theta) = \frac{\pi}{\theta_{U} - \theta_{L}}I_{\theta_{L}, \theta_{U}}(\theta) + (1 - \pi)g(\theta),$$

but now to restrict $g(\theta)$ to some particular parametric family. We refer to this 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), we could argue that “biparametric” or simply “mixture” are technically more accurate names.

Following equation (2), note that requiring $G \in \mathcal{H}_c$ is equivalent to requiring $B \leq \left(\frac{1}{\pi} \right) \left(\frac{1 - \pi}{\pi} \right)$, where

$$B = \frac{\int_{\theta_{L}}^{\theta_{U}} f(x|\theta) \, d\theta}{\int f(x|\theta) g(\theta) \, d\theta},$$

the Bayes factor in favor of the null hypothesis. Equation (6) expresses the Bayes factor as the ratio of the marginal densities under the competing hypotheses; it is also expressible as the ratio of posterior to prior odds in favor of the null. As such, $B$ gives the extent to which the data have revised our prior beliefs concerning the two hypotheses. Note that if we take $\pi = 1/2$ (equal prior weighting of null and alternative), then a Bayes factor of 1 suggests equal posterior support for the two hypotheses. In this case, we require a Bayes factor of 1/19 or smaller to insure that $P(H_0|x)$ does not exceed 0.05. Kass and Raftery [17] provide a comprehensive review of the usage and
computation of Bayes factors.

In practice, familiar models from the exponential family are often appropriate (either exactly or asymptotically) for the likelihood \( f(x|\theta) \). This naturally leads to consideration of the restricted class of conjugate priors \( g(\theta) \), to obtain a closed form for the integral in the denominator of (6). Since a normal approximation to the likelihood for \( \theta \) is often suitable for even moderate sample sizes, we illustrate in the case of a conjugate normal prior. The fact that \( g \) is defined only on the complement of the indifference zone presents a slight complication, but, fortunately, the calculations remain tractable under a renormalized prior with the proper support. That is, we take

\[
g(\theta) = \frac{N(\theta|\mu, \tau^2)}{1 - \Phi \left( \frac{\nu_\mu - \theta}{\tau} \right) - \Phi \left( \frac{\nu_\ell - \theta}{\tau} \right)}, \quad \theta \notin [\theta_L, \theta_U],
\]

where the numerator denotes the density of a normal distribution with mean \( \mu \) and variance \( \tau^2 \), and \( \Phi \) denotes the cdf of a standard normal distribution.

To obtain a computational form for equation (6), suppose we can approximate the likelihood satisfactorily with a \( N(\theta|\hat{\theta}, \hat{\sigma}^2) \) density, where \( \hat{\theta} \) is the maximum likelihood estimate (MLE) of \( \theta \) and \( \hat{\sigma}^2 \) is a corresponding standard error estimate. Probability calculus then shows

\[
\int f(x|\theta)g(\theta)d\theta = \frac{1}{\sqrt{2\pi(\hat{\sigma}^2 + \tau^2)}} \exp \left[ -\frac{(\mu - \hat{\theta})^2}{2(\hat{\sigma}^2 + \tau^2)} \right] \left\{ 1 - \left[ \Phi \left( \frac{\nu_\mu - \eta}{\nu} \right) - \Phi \left( \frac{\nu_\ell - \eta}{\nu} \right) \right] \right\}, \quad (7)
\]

where \( \eta = (\hat{\sigma}^2\mu + \tau^2\hat{\theta})/(\hat{\sigma}^2 + \tau^2) \) and \( \nu^2 = \hat{\sigma}^2\tau^2/(\hat{\sigma}^2 + \tau^2) \). (Note that \( \eta \) and \( \nu^2 \) are respectively the posterior mean and variance under the fully parametric normal/normal model, described in the next subsection.) Since \( \int_{\theta_L}^{\theta_U} f(x|\theta)d\theta = \Phi \left( \frac{\nu_\mu - \hat{\theta}}{\nu} \right) - \Phi \left( \frac{\nu_\ell - \hat{\theta}}{\nu} \right) \), we can now obtain the Bayes factor (6) without numerical integration, provided that there are subroutines available to evaluate the normal density and cdf.
3.3 Fully parametric prior

As a final approach, we might abandon the mixture prior form (5) in favor of a single parametric family $h(\theta)$, preferably chosen as conjugate with the likelihood $f(x|\theta)$. If such a choice is possible, we obtain simple closed form expressions for Bayes factors and tail probabilities whose sensitivity to changes in the prior parameters we can easily examine. For example, for our $N(\theta|\hat{\theta}, \sigma^2)$ likelihood under a $N(\theta|\mu, \tau^2)$ prior, the posterior probability of $H_0$ is nothing but

$$P(\theta \in [\theta_L, \theta_U]|x) = \Phi\left(\frac{\theta_U - \eta}{\nu}\right) - \Phi\left(\frac{\theta_L - \eta}{\nu}\right),$$

where $\eta$ and $\nu^2$ are again as defined beneath equation (7). Posterior probabilities that correspond to stopping to reject the hypotheses $H_L : \theta < \theta_L$ and $H_U : \theta > \theta_U$ arise similarly.

4 Application to a Toxoplastic Encephalitis Prophylaxis Trial

We now illustrate the methods of the previous two sections with data from an AIDS clinical trial originally reported by Jacobson et al. [18]. The data are from a double-blind randomized trial that compared the drug pyrimethamine with placebo for preventing toxoplastic encephalitis (TE), a major cause of morbidity and mortality among those with AIDS. In a Bayesian reanalysis of these data, Carlin et al. [19] employed a proportional hazards likelihood [20] 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 a treatment effect indicator (1 for active drug, 0 for placebo). Denoting the parameters that correspond to these two covariates as $\beta$ and $\theta$, respectively, we obtain a marginal partial likelihood for $\theta$ by numerically integrating $\beta$ out of the Cox partial likelihood.

Following the start of the trial in September of 1990, the trial’s Data Safety and Monitoring
Board met on three occasions to assess its progress and determine whether or not the trial should continue. At these three meetings there were 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 because the pyrimethamine group had not shown sufficiently fewer TE events up to that time, and a statistically significant difference seemed unlikely to emerge in the future due to the low rate of TE overall (essentially an informal stochastic curtailment stopping rule). The trial did not actually stop until 3/30/92, when patients were instructed to discontinue their study medication following unblinding and review of the data by the protocol chairperson. Figure 1 displays the posterior for \( \theta \) under a flat prior (i.e., the standardized marginal likelihood) at each of the four dates mentioned above. The parametrization of our Cox model implies that negative values of \( \theta \) correspond to an efficacious treatment; the shift towards positive \( \theta \) values evident in the third and fourth monitoring points reflects an excess of deaths in the treatment group. This emerging superiority of the placebo is quite surprising, suggesting that prior robustness is an especially important issue in this problem.

(Figure 1 about here)

Following Carlin et al. [19] we take the endpoints of the indifference zone as \( \theta_U = 0 \) and \( \theta_L = \log(.75) = -0.288 \). Hence, due to its increased cost and toxicity, the treatment is preferred only if it reduces the placebo hazard rate by at least 25\% (i.e., if \( \theta < \theta_L \)). On the other hand, any positive value for \( \theta \) favors the placebo. Using this indifference zone, Sargent and Carlin [7] set \( \xi_L = \theta_L \) and \( \xi_U = \theta_U \), which in turn implies that the lower and upper prior tail probabilities \( a_L \) and \( a_U \) sum to 1, since the conditional prior \( g(\theta) \) has no support over the indifference zone. They then used the methods of Section 2 to find pairs \((a_L, a_U)\) for which \( \mathcal{H}_\xi \) is non-empty given these fixed quantiles for \( g(\theta) \). For example, at the trial’s fourth monitoring point, they showed that these
\((a_L, a_U)\) pairs form the set

\[
\left\{ (a_L, a_U) : a_L f(x|\xi_L) + a_U f(x|\hat{\theta}) \geq \frac{1 - p}{p} \frac{1}{1 - \pi \xi_U - \xi_L} \int_{\xi_L}^{\xi_U} f(x|\theta) \, d\theta \right\}.
\]  

(8)

Values for \(f(x|\xi_L), f(x|\hat{\theta})\), and the integral in this expression are available from the data. Selecting \(p = .1\), replacing \(a_L\) by \(1 - a_U\), and solving for \(a_U\) as a function of \(\pi\) produce the inequality

\[
a_U \geq .536 \frac{\pi}{1 - \pi} - .016.
\]

(9)

Combining (9) with the constraints \(0 \leq a_U \leq 1\) and \(0 \leq \pi \leq 1\) produces the regions determined by the solid curve in Figure 2(a). For \((\pi, a_U)\) combinations above and to the left of this curve, there exist priors consistent with that combination that permit stopping to reject \(H_0\), while for combinations below and to the right of the curve, no such priors exist. That is, no prior with a \((\pi, a_U)\) combination lying below the curve would lead to rejection of the null hypothesis. We note that rejection of \(H_0\) is always possible (regardless of \(a_U\)) for \(\pi < .029\), while it is never possible for \(\pi > .655\). These are sensible results, since low prior weight on the null should encourage rejection, while sufficiently high prior weight on the null should prevent it.

(Figure 2 about here)

Figure 2(a) also shows the boundaries obtained in the same manner as formula (9) for monitoring points two and three. In the first case, all but the most extreme priors (those having \(\pi < .104\)) preclude stopping to reject \(H_0\). As the number of observed events \(n\) and \(\hat{\theta}\) increase over time, the potential stopping regions lying to the left of the curves also increase in size.

To ease the interpretability of our results, we might replace the conditional upper tail probability \(a_U\) with the corresponding unconditional probability, \(p_U \equiv a_U(1 - \pi)\). This converts (9) into a linear
inequality, but with the added constraint that \( p_U + \pi \leq 1 \). Figure 2(b) plots the \((\pi, p_U)\) pairs and their status relative to stopping and rejecting \( H_0 \) for the same three monitoring points shown in Figure 2(a). A clinician may find this plot easier to interpret, since it avoids the notion of \( a_U \), a probability that is conditional on the null hypothesis being false. Again, at this monitoring point, no prior corresponding to a region to the right of a boundary enables stopping to reject the null hypothesis.

The fact that Figure 2(b) shows a large region of \((\pi, p_U)\) pairs for which rejection is not possible for any prior suggests that the assumption of a uniform prior on the indifference zone leads to some degree of conservatism. A different restriction to the class of candidate priors may result in stronger recommendations, as well as a better understanding of how the restriction to a uniform prior on the indifference zone affects the results of the analysis. We thus now consider the restriction to quasiunimodal priors that have modes lying in the indifference zone, described in Subsection 3.1.

We use a \( N(\theta|\hat{\theta}_k, \hat{\sigma}_k^2) \) approximation to the likelihood at monitoring point \( k, k = 2, 3, 4 \), as suggested by Figure 1. Using equation (4) and the expressions for \( \overline{\omega}_i \) and \( \underline{\omega}_i, i = 1, 2, 3 \) from O'Hagan and Berger [12], we can determine whether there exists a prior in this class that leads to rejection of \( H_0 \). Figure 3 defines the \((\pi, p_U)\) pairs for which such rejection is possible, which again are those that lie to the left of the boundary for a given monitoring point.

(Figure 3 about here)

We notice that for the third and fourth monitoring points, the regions where rejection of \( H_0 \) is possible are somewhat larger in Figure 3 than in Figure 2(b). This is perhaps counterintuitive, since one might have expected quasiunimodality to be a more restrictive prior condition, thus eliminating some of the priors that had previously led to rejection of \( H_0 \). It is important, however, to remember that quasiunimodality is not a special case of the method of Sargent and Carlin [7], since the latter
imposes a uniform prior on the indi/erence zone. The Figure 3 boundaries thus indicate that for this dataset, the quasiunimodal prior is actually less restrictive.

Another counterintuitive feature of Figure 3 is the boundary between the regions of priors at the second monitoring point. For example, if we hold $\pi$ constant at .06, rejection of $H_0$ becomes impossible as $p_U$ increases from .4 to .6, whereas one might think that increasing the prior mass on the upper interval (the one supported by the data) should make rejection more likely. The restriction that the prior have its mode in the indi/erence region, however, means that the addition of prior mass to the upper region can only lengthen the tail of the distribution. At the second monitoring point the likelihood still allocates substantial support to the lower region, so that moving prior mass from the lower to the upper region decreases the supremum of the posterior mass outside of the indi/erence zone, thereby increasing the infimum of the posterior probability of $H_0$. This in turn results in rejection becoming impossible as one moves prior mass from the lower to the upper region.

The bounds obtained in Figures 2 and 3 are not fully satisfying, since the very broad classes of candidate priors considered lead to correspondingly broad prior partitions. The mere fact that a prior exists that permits rejection of the null hypothesis may have limited use in practice. As such, we now turn to a semiparametric approach as in Subsection 3.2, again using a normal approximation to the likelihood at the second, third, and fourth monitoring points. Since our analysis is targeted toward the members of the Data Safety and Monitoring Board, we restrict attention to “skeptical” conditional priors $g$ by setting $\mu = 0$. We may then compute the Bayes factor $B$ in favor of $H_0$ that arises from equations (6) and (7) as a function of the conditional prior standard deviation $\tau$ for $0 < \tau < 4$. To facilitate comparison with previous figures, Figure 4(a) plots $B$ not versus $\tau$, but, instead, versus the conditional prior upper tail probability $a_U$, a one-to-one function of $\tau$. Assigning a mass $\pi = .25$ to the null region (the indi/erence zone) and retaining our rejection threshold of
\( p = .1 \), we would reject \( H_0 \) for \( B < \frac{1}{3}; \) this value is marked on the figure with a dashed horizontal reference line. The message from this figure is clear: no uniform-normal mixture priors of the form (5) allow rejection of \( H_0 \) at monitoring points 2 or 3, but by monitoring point 4, rejection of \( H_0 \) is favored by all but the most extreme priors (i.e., those that are either extremely vague or essentially point masses at 0).

(Figure 4 about here)

Next, suppose we wish to investigate robustness as \( \pi \) varies. We reject \( H_0 \) if and only if

\[
B(\tau) \leq \left( \frac{\bar{P}}{1-\bar{P}} \right) \left( \frac{\bar{P}+\pi}{\pi} \right),
\]

or equivalently

\[
\pi \leq \left[ \left( \frac{1-\bar{P}}{\bar{P}} B(\tau) \right) + 1 \right]^{-1}.
\]

Again using the one-to-one relationship between \( a_U \) and \( \tau \), Figure 4(b) plots this boundary in \((\pi, a_U)\)-space for the final three monitoring points. Conditional on the data, all priors that correspond to points lying to the left of a given curve lead to rejection of \( H_0 \), while all those to the right do not. (Note the difference in interpretation between this graph and Figure 2, where each plotted point corresponded to infinitely many priors.) As in Figure 2(a), we see the mounting evidence against \( H_0 \) in the boundaries’ gradual shift to the right, increasing the number of priors that result in rejection. While the rejection regions are smaller than before, the implication of Figure 4(b) is actually more precise than that of Figure 2(a), since, before, “rejection” was guaranteed only for at least one prior with the given \((\pi, a_U)\) combination.

Finally, we turn to the fully parametric approach outlined in Subsection 3.3. Fixing the prior mean \( \mu \) at 0 again, the prior support for the indifference zone, \( \pi \), is now determined automatically once we specify \( \tau \). This loss of one degree of freedom in the prior means that bivariate plots like Figures 2, 3 and 4(b) are no longer possible, but univariate plots as in Figure 4(a) are still
sensible. To obtain the strongest possible results for our dataset, in Figure 5(a) we plot the posterior probability of $H_L : \theta < \theta_L$, rather than $H_0$, versus $\tau$. We see that rejection of this hypothesis is possible for all skeptical normal priors by the third monitoring point; by the fourth monitoring point, the evidence against this hypothesis is overwhelming. While this analysis considers a far smaller class of priors than considered in previous figures, it still provides compelling evidence by the third monitoring point that the treatment is not superior in terms of preventing TE or death.

(Figure 5 about here)

It is of interest to compare our results with those obtained using the $\epsilon$-contamination method of Greenhouse and Wasserman [13]. As mentioned in Section 1, this method enables computation of upper and lower bounds on posterior expectations for priors of the form $(1 - \epsilon)g_0 + \epsilon g$, where $g_0$ is some baseline distribution and $g$ can range across the set of all possible priors. Reasonably tight bounds for moderately large $\epsilon$ suggest robustness of the original result to changes in the prior. Figure 5(b) plots the Greenhouse and Wasserman bounds for the posterior probability of $H_L$ for the final three monitoring points, where we have taken $g_0$ as our $N(\theta|\mu, \tau^2)$ prior with $\mu = 0$ and $\tau = 1$. At the third monitoring point, the posterior probability is bounded below the rejection threshold of .1 only for $\epsilon$ roughly less than .17, suggesting that we cannot stray too far from the baseline normal prior and maintain total confidence in our conclusion. By the fourth monitoring point, however, the upper bound remains below the threshold for $\epsilon$ values as large as .65, indicating a substantial degree of robustness to prior specification in our stopping decision.

5 Discussion

As seen in the progression of displays in Section 4, the practical usefulness of a prior partitioning analysis often depends on the breadth of the class of priors considered. One must select
the proper class carefully: large classes may lead to overly broad posterior bounds, while narrow classes may eliminate many plausible prior candidates. The quasiunimodal, semiparametric, and fully parametric classes considered here constitute only a small fraction of those discussed in the Bayesian robustness literature. Other possibilities include density bounded classes, density ratio classes, and total variation classes; see Wasserman and Kadane [21] for a discussion and associated computational strategies.

Finally, while we have frequently employed a normal distribution for the likelihood or some portion of the prior, we made such assumptions purely for computational convenience. Thanks to advanced Monte Carlo integration algorithms such as the Gibbs sampler [22], one could perform prior partitioning in principle with any combination of likelihood and prior. Similarly, one could apply the method in settings where $\theta$ is a multivariate parameter, such as a multi-arm trial or a simultaneous study of effectiveness and toxicity. As such, prior partitioning offers an attractive complement to traditional prior elicitation and robustness methods in a wide array of clinical trial settings.

**Acknowledgements**

This work was supported in part by National Institute of Allergy and Infectious Diseases (NIAID) Contract NO1-AI05073. In addition, the work of the first author was supported in part by NIAID FIRST Award 1-R29-AI33466. The authors thank Jim Hodges, Michael Lavine, the editor, the deputy editor, and two conscientious anonymous referees for several corrections and editorial suggestions that greatly improved the manuscript.
References


Figure 1. Posterior for treatment effect under a flat prior, TE trial data

Endpoint = TE or death; Covariate = Baseline CD4 count
In both plots, for combinations to the left of each curve, priors exist that permit stopping to reject Ho.
Figure 3. Quasiunimodal prior partitions, TE trial data

For combinations to the left of each curve, quasiunimodal priors exist that permit stopping to reject $H_0$. 
Figure 4. Semiparametric prior partitions, TE trial data

a) Bayes factor versus conditional upper tail area

b) Conditional upper tail area versus prior mass on indifference zone; for combinations to the left of each curve, ALL priors result in stopping to reject Ho
Figure 5. Fully parametric prior partitions, TE trial data

a) Posterior probability of $H_L$ versus prior standard deviation

b) Bounds on posterior probability of $H_L$ versus prior contamination factor