A Strategy for Structured Use of Bayesian Sequential Monitoring in Clinical Trials

Evan Kwiatkowski[†], Eugenio Andraca-Carrera[‡],
Mat Soukup[‡], Matthew A. Psioda[†*]

† Department of Biostatistics, University of North Carolina,
McGavran-Greenberg Hall, CB#7420,
Chapel Hill, North Carolina, U.S.A.

‡ Division of Biometrics VII, Office of Biostatistics
Center for Drug Evaluation and Research,
US Food and Drug Administration,
Silver Spring, Maryland, USA

January 30, 2020

Abstract

The text of your abstract. 200 or fewer words.

Keywords: 3 to 6 keywords, that do not appear in the title

*The authors gratefully acknowledge *please remember to list all relevant funding sources in the unblinded version*

1 Introduction

Things to discuss:

- 21st Century Cures Act (MATT)
- PDUFA VI reauthorization (MATT)
- Expansive work already done on sequential monitoring (EVAN – draft on 6/21)
- Our majors contribution (EVAN – as early as possible in introduction without having the flow appear weird – draft on 6/21)
- Outline for the remaining section of the paper (EVAN – draft on 6/21)

The theoretical foundations for the Bayesian clinical trials has been long established Cornfield (1966*a*) Cornfield (1966*b*) Neyman & Greenhouse (1967). These methods were not widely used in practice until a comprehensive framework for interpretation of results was developed through specifying prior distributions that were naturally and intuitively related to the research objectives (e.g. skeptical and enthusiastic priors) Freedman & Spiegelhalter (1989) Freedman & Spiegelhalter (1992) Spiegelhalter et al. (1993) Spiegelhalter et al. (1994) Fayers et al. (1997). (*Rewrite paragraph.*)

There is still potential for further utilization of Bayesian methods in the clinical trial setting. While the framework for interpretation of Bayesian clinical trials is well developed, the details of specifying prior distributions in a natural and intuitive way is lacking. This paper presents a structured or default way to determine prior distributions based on the trial design. Our major contribution is to present methods for the default or automatic selection of prior distributions in a way that is applicable to a wide array of clinical trial designs.

1. Bayesian methodology is widely developed.
2. It has been applied (cite).
3. The current perspective is that Bayesian methodology is only valid when Frequentist methods are insufficient, including where enrollment is challenging (rare diseases, pediatric studies)
4. Our contribution is to show that Bayesian methods are applicable to all clinical trials. This is shown by highlighting their improved interpretation and showing their use in varied and complicated situations.

2 Methods

2.1 Preliminaries

2.1.1 Bayesian hypothesis testing using posterior probabilities

Consider a one-arm trial with a single measured outcome per patient. The objective is to make inference on a single unknown quantity. Let $\mathbf{D}$ be a random variable representing the data collected in the trial with density $p(\mathbf{D}|\theta)$, where θ is the parameter of interest with sample space $\theta \in \Theta$. For example, in the context of a simple binary outcome (e.g., objective response to treatment) based on sample size n, $\mathbf{D} = (y_1, ..., y_n)$ may correspond to the n indicators for whether or not subjects responded to a treatment and $\theta \in [0, 1]$ may correspond to an assumed common probability of response.

Formal Bayesian hypothesis testing requires the specification of prior probabilities on the hypotheses (e.g., $p(H_i)$ for $i = 0, 1$) and prior distributions for θ for each hypothesis (e.g. $\pi\left(\theta\middle|H_i\right)$ for $i = 0, 1$). Consider the hypotheses $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ where $\Theta_0 \bigcup \Theta_1 = \Theta$ and $\Theta_0 \bigcap \Theta_1 = \emptyset$. The posterior probability of hypothesis H_i is

$$p(H_i|\mathbf{D}) = \frac{p(\mathbf{D}|H_i) \cdot p(H_i)}{p(\mathbf{D}|H_0) \cdot p(H_0) + p(\mathbf{D}|H_1) \cdot p(H_1)} \tag{1}$$

$$= \frac{\int_{\Theta_i} p(\mathbf{D}|\theta, H_i)\pi(\theta|H_i)d\theta \cdot p(H_i)}{p(\mathbf{D}|H_0) \cdot p(H_0) + p(\mathbf{D}|H_1) \cdot p(H_1)}. \tag{2}$$

The posterior probability of the *event defining* H_i is

$$P(\theta \in \Theta_i|\mathbf{D}) = \int_{\Theta_i} p(\theta|\mathbf{D})d\theta \tag{3}$$

$$= \frac{\int_{\Theta_i} p(\mathbf{D}|\theta)\pi(\theta)d\theta}{\int_{\Theta} p(\mathbf{D}|\theta)\pi(\theta)d\theta} \tag{4}$$

$$= \frac{\int_{\Theta_i} p(\mathbf{D}|\theta)\pi(\theta|\theta \in \Theta_i)d\theta \cdot \int_{\Theta_i} \pi(\theta)d\theta}{\int_{\Theta} p(\mathbf{D}|\theta)\pi(\theta)d\theta} \tag{5}$$

Denote $P(\theta \in \Theta_i) = \int_{\Theta_i} \pi(\theta)d\theta$. When $p(H_i) = P(\theta \in \Theta_i)$ and $\pi\left(\theta\middle|H_i\right) = \pi\left(\theta\middle|\theta \in \Theta_i\right)$ for $i = 0, 1$, it follows that the posterior probability of the event defining the alternative hypothesis and the posterior probability of the alternative hypothesis are equivalent. Consistent with the specifications of $p(H_i)$ and $\pi\left(\theta\middle|H_i\right)$ as described above, in what follows we will refer to the quantity $P(\theta \in \Theta_1|\mathbf{D})$ as the posterior probability of H_i for ease of exposition.

2.1.2 Compelling level of evidence

Define ϵ as the *residual uncertainty* of H_i being true relative to the competing hypothesis, and define $1 - \epsilon \in (0, 1)$ as the threshold for *a compelling level of evidence* that some claim about θ is true. We say

that an individual is *all but convinced* that H_i is true given the observed data if they believe that

$$P(\theta \in \Theta_i|\mathbf{D}) > 1 - \epsilon \text{ for } i = 0, 1. \tag{6}$$

Furthermore, $1 - \epsilon \in (0, 1)$ can be used as a threshold for *a compelling a priori belief* that some claim about θ is true before the beginning of the trial. We say that an individual is *all but convinced* that H_i is true a priori if they believe that

$$P(\theta \in \Theta_i) > 1 - \epsilon \text{ for } i = 0, 1. \tag{7}$$

2.1.3 Skeptical and Enthusiastic Monitoring Priors

Continuing the trial setup from Section 2.1, suppose every subject receives active treatment during the follow-up period spanning enrollment to study completion. The length of follow-up is the same for each subject. Subjects are enrolled progressively such that there is a time when there are both subjects with completed outcomes and subjects undergoing follow-up. In this paper we introduce two types of prior: monitoring and inference priors.

The purpose of monitoring priors to is help answer the question "Is the evidence compelling enough to end the trial early?" Monitoring priors are used for interim analyses on the data from subjects who completed follow-up while additional outcomes are pending. A promising interim result that shows clear efficacy of the treatment would justify stopping the enrollment of additional subjects, while enrolled subjects would continue receiving the presumed effective treatment during the follow-up period. A discouraging interim result that shows clear futility of the treatment would justify stopping the enrollment of additional subjects, and may call for enrolled patients undergoing follow-up to stop receiving the presumed ineffective treatment.

To have either efficacy or futility to be "clear" the evidence must be convincing to individuals with diverse opinions of θ. The judgment of efficacy must be convincing even to an individual who was initially skeptical about the benefit of the treatment, likewise the judgment of futility must be convincing to an invidual was was initially optimistic. For this reason monitoring priors generally represent relatively extreme (but still plausible) beliefs about θ. If a skeptic has become confident that the treatment is efficacious, that implies essentially anyone with some degree of equipoise would also be convinced, hence individuals with diverse opinions have a consensus. Alternatively, if an enthusiast has become convinced that the treatment is ineffective, then essentially everyone will share that opinion.

Monitoring priors will be defined based on the concepts of compelling a prior belief defined in Section 2.1. Continuing the trial setup from Section 2.1, suppose that θ corresponds to a probability of desired

response, and the null hypothesis is $H_0 : \theta \leq \theta_0$ where θ_0 is the expected null response among untreated subjects. Let $\theta_1 > \theta_0$ be a highly efficacious response probability.

We define an enthusiastic prior $\pi_E(\theta)$ as a prior that is centered around θ_1 and reflects the belief of an individual that is *all but convinced* that H_i is true a priori, that is,

$$P(\theta > \theta_0 | \pi_E) > 1 - \epsilon. \tag{8}$$

Consider a similarly defined skeptical prior $\pi_S(\theta)$ that is centered around θ_0 and reflects the belief of an individual that is *all but convinced* that H_0 is true a priori, that is, $P(\theta \leq \theta_0 | \pi_S) > 1 - \epsilon$. Having this belief demonstrates such an extreme disbelief in the possibility of a positive effect that conducting the trial at all would be viewed as dubious. This is analogous to specifying a prior model probability on H_0 that is $1 - \epsilon$ and that is not consistent with their being clinical equipoise about the hypotheses, that is to say, such extreme skepticism is not rational if a trial is to be conducted. Instead, we define the skeptical prior $\pi_S(\theta)$ to be centered around θ_0 and reflects the belief of an individual that is *all but convinced* that a highly efficacious response probability is unlikely, that is,

$$P(\theta < \theta_1 | \pi_S) > 1 - \epsilon. \tag{9}$$

2.1.4 Criteria for Early Stoppage

The use of monitoring based on changing the opinion of skeptical and enthusiastic priors has been described as overcoming a handicap (Freedman & Spiegelhalter (1989)) and providing a brake (Fayers et al. (1997)) on the premature termination of trials, or constructing "an adversary who will need to be disillusioned by the data to stop further experimentation" (Spiegelhalter et al. (1994)). Early termination of enrollment is appropriate if diverse prior opinions about θ would be in agreement given the interim data (e.g. the skeptical and enthusiastic person reach the same conclusion).

The skeptic, whose prior belief is reflected in (9), becomes convinced the treatment is effective if there is compelling evidence that $\theta > \theta_0$ is true (6), that is,

$$P(\theta < \theta_1 | \pi_S) > 1 - \epsilon \Rightarrow P(\theta > \theta_0 | \mathbf{D}, \pi_S) > 1 - \epsilon. \tag{10}$$

Consider $\theta < \frac{\theta_0 + \theta_1}{2}$ to be a response that is less than the highly efficacious response probability θ_1. The enthusiast, whose prior belief is reflect in (8), becomes convinced the treatment is ineffective, or not as effective as anticipated, if there is compelling evidence that $\theta < \frac{\theta_0 + \theta_1}{2}$, that is,

$$P(\theta > \theta_0 | \pi_E) > 1 - \epsilon \Rightarrow P\left(\theta < \frac{\theta_0 + \theta_1}{2} \middle| \mathbf{D}, \pi_E\right) > 1 - \epsilon. \tag{11}$$

2.2 Monitoring Prior Specification: The Generalized Normal Family of Distributions

The enthuastic and skeptical priors each have a required center (modal value) and tail area probability, however, there are still many ways to parameterize such distributions. Consider the univariate generalized normal kernel

$$\exp\left\{-\left(\frac{|\theta-\mu|}{\alpha}\right)^{\beta}\right\} \tag{12}$$

where $\mu \in \mathbb{R}$ is a location parameter, $\alpha > 0$ is a scale parameter, and $\beta > 0$ is a shape parameter. Note that $\beta = 2$ corresponds to the normal distribution. In the case of $\beta = 2$, this distribution is symmetric around μ and the combination of location and scale parameter can accommodate the required modal value and tail area probability requirement (i.e. (8), (9)). The modification of β can be used to concentrate or flatten the distribution around the modal value. Examples of skeptical and enthusiastic priors for a single parameter θ based on the trial from Section 2.1 are displayed in Figure 1. The methods for concentrating or flattening the priors are described in Section 5.1.

The generalized normal distribution can be used to parameterize skeptical and enthuastic priors for trials with multiple unknown quantities of interest. Suppose the trial from Section 2.1 has added a control arm, and let θ_0 and θ_1 be the response proportions for a control and treatment group respectively. Suppose that the risk difference $\theta_1 - \theta_0$ is of interest. Let δ_S denote a null risk difference and δ_E denote a highly efficacious risk difference.

Consider the following representation of the joint prior for θ_1 and θ_0:

$$\pi(\theta_0, \theta_1) = \pi(\theta_0) \times \pi(\theta_1|\theta_0) \tag{13}$$

where

$$\pi(\theta_0) \propto \exp\left\{-\left(\frac{|\theta_0-\mu_0|}{\alpha_0}\right)^{\beta_0}\right\} \tag{14}$$

$$\pi(\theta_1|\theta_0) \propto \exp\left\{-\left(\frac{|(\theta_1-\theta_0)-\delta|}{\alpha_1}\right)^{\beta_1}\right\} f(\theta_0) \tag{15}$$

The component $\pi(\theta_0)$ reflects prior opinion about the response rate in the placebo group, and the component $\pi(\theta_1|\theta_0)$ can be used to express pessimism or optimism in the difference in proportions $\theta_1 - \theta_0$.

The skeptical prior condition is

$$P(\theta_1 - \theta_0 < \delta_E) = 1 - \epsilon. \tag{16}$$

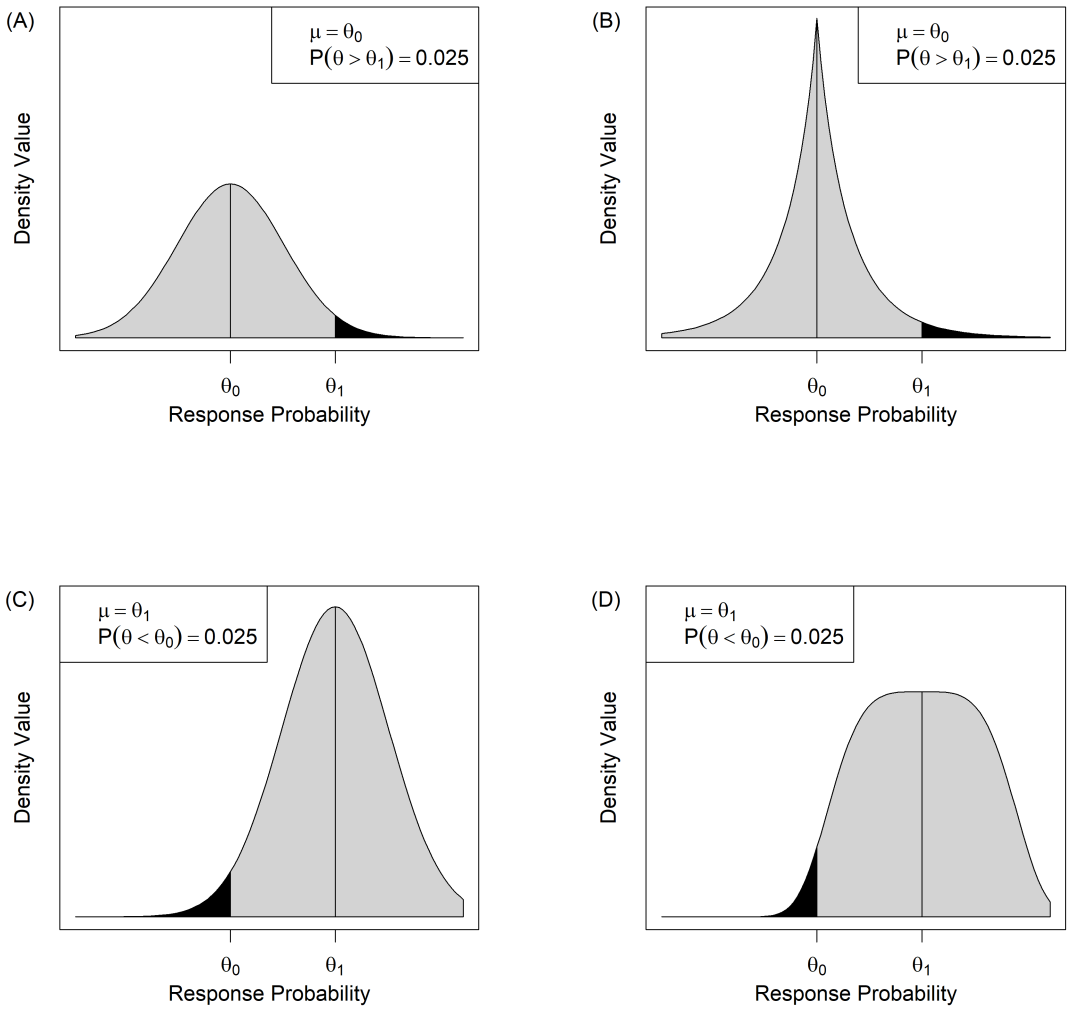


Figure 1: A, Skeptical prior, normal distribution. B, Skeptical prior, generalized normal distribution with added concentration around modal value. C, Enthusiastic prior, normal distribution. D, Enthusiastic prior, generalized normal distribution with added flattening around modal value.

The enthusiastic prior condition is

$$P(\theta_1 - \theta_0 > \delta_S) = 1 - \epsilon. \tag{17}$$

Examples of skeptical and enthusiastic priors are displayed in Figure 2. Additional details about the parameterization are described in Section 5.2.

2.3 Mixture Inference Priors

The purpose of the inference prior is to synthesize the posterior inferences from the a priori diverse perspectives to facilitate interpretation of the data once it has been obtained. In this paper we propose using an inference prior that is a combination of the skeptical and enthusiastic priors that were used for monitoring. The skeptical and enthusiastic monitoring priors defined in Section 2.1.3 represent extreme but plausible beliefs about θ. While analysis with these priors provides a rational perspective from which one can determine whether interim data are sufficient to cease enrolling patients, the a priori belief of most stakeholders will likely fall somewhere in between. Thus, when it interpreting the final data once in hand, intermediate perspectives should be considered. To that end, we define an inference prior as a mixture prior constructed by mixing the monitoring priors. Formally, the inference prior associated with mixing weight ω is given by

$$\pi_I(\theta) = \omega \cdot \pi_S(\theta) + (1 - \omega) \cdot \pi_E(\theta), \tag{18}$$

where $\omega \in [0, 1]$. The specification of ω is done a priori, and the value $\omega = 1/2$ will be referred to as an agnostic inference prior.

The posterior distribution for θ using (18)

$$p(\theta|\mathbf{D}, \pi_I) = \hat{\omega} \cdot p(\theta|\mathbf{D}, \pi_S) + (1 - \hat{\omega}) \cdot p(\theta|\mathbf{D}, \pi_E) \tag{19}$$

where the mixing weight is

$$\hat{\omega} = \frac{\omega \cdot p(\mathbf{D}|\pi_S)}{\omega \cdot p(\mathbf{D}|\pi_S) + (1 - \omega) \cdot p(\mathbf{D}|\pi_E)} \tag{20}$$

where $p(\mathbf{D}|\pi_S) = \int p(\mathbf{D}|\theta)\pi_S(\theta)d\theta$ and $p(\mathbf{D}|\pi_E) = \int p(\mathbf{D}|\theta)\pi_S(\theta)d\theta$.

The inference prior is used to evaluate the hypotheses in (3), and distribution of θ using the inference prior, $p(\theta|\mathbf{D}, \pi_I)$, will be used to compute summaries of θ such as the posterior mean and quantiles.

2.3.1 Incorporating Prior Information in the Monitoring Priors

We don't have anything in the paper about using a mixture of the skeptical and enthusiastic prior to define a "less" skeptical monitor prior that incorporates information from another source using the notion

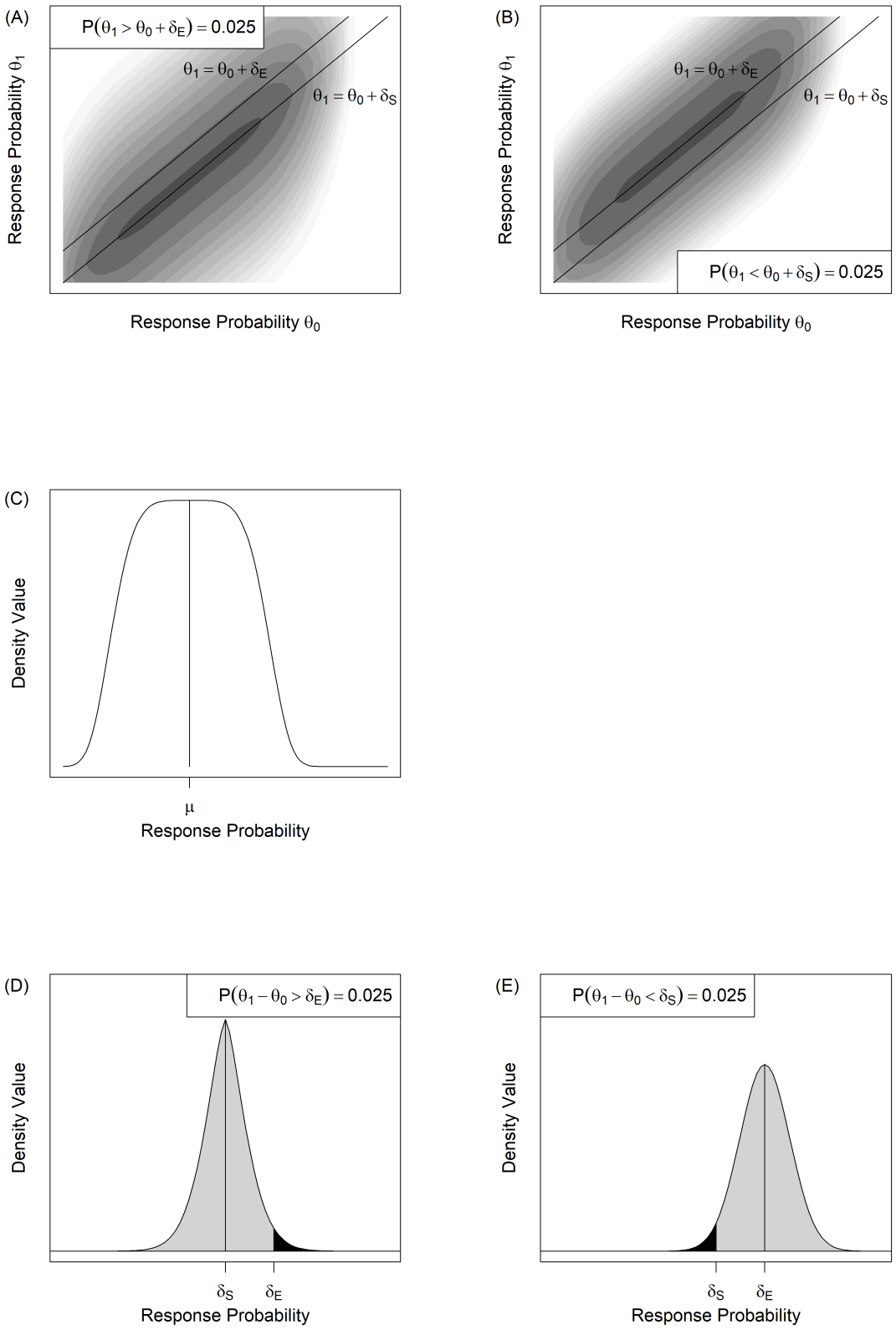


Figure 2: A, Skeptical joint prior. B, Enthusiastic joint prior.

of applicability (i.e., a fixed value of ω_0 that is not 1.). This is fundamental to the novelty of the paper. I think it would be good to show how the GN can be used to inform a prior for the control group with flexibility here (e.g., flattening the prior over an interval) AND talking about the mixture skeptical monitoring prior. In the case of using "no prior information" the enthusiastic prior is simply used for monitoring. But it can be incorporated directly into the skeptical prior when it is informed by data using the concept of applicability. Isn't that what was done for the second example? Lets talk if we need to. This is a big part of the novelty.

2.4 Notes on computation

3 Examples

3.1 Single-Arm Proof-of-Activity Trial with Binary Endpoint

3.1.1 Motivating example

You need more details on the motivating example. Describe the trial, cite papers about the trial. The current writeup leaves out way to much background. What was the outcome – currently the writeup does not even say? At what time point was the primary outcome assessed? Start by describing the trial(s) pediatric and adult that are used for this motivation. That should come first. If that is done adequately the first paragraph below will be unneeded as it will have been made clear by the motivating example description.

Consider a single-arm proof-of-activity trial with a binary endpoint. The data $\mathbf{D}$ are binomially distributed and the response rate θ is the parameter of interest, with higher values of θ being indicative of proof-of-activity.

An example application is based on the drug iniximab, which is FDA approved for the treatment of several diseases, including ulcerative colitis (UC). The goal of the trial is to test the hypothesis: $H_0 : \theta \leq \theta_0 = 0.40$ versus $H_1 : \theta > \theta_0$. From adult data $\theta_1 = 0.67$. Based on the 54-week follow-up period, we can infer enrollment took place over approximately 33 months (approximately 1 patient per 0.55 months).

3.1.2 Model formulation & prior elicitation

The default skeptical and enthusiastic priors will be of the form (**??**) with $\beta = 2$ corresponding to truncated normal distributions

$$\pi_S(\theta) \propto \exp\left\{-\frac{(\theta-\theta_0)^2}{\alpha_S}\right\} I(\theta \in [0,1])$$
$$\pi_E(\theta) \propto \exp\left\{-\frac{(\theta-\theta_1)^2}{\alpha_E}\right\} I(\theta \in [0,1])$$

with α_S and α_E chosen satisfy (8) and (9), where $\Theta_1 = (\theta_0, 1]$, $\Theta_A = [\theta_1, 1]$, and $\epsilon = 0.025$.

Alternative specifications of the priors will be used to concentrate or flatten the distribution around the modal value, which still satisfy conditions (8) and (9). In particular, the skeptical prior can be concentrated around the modal value θ_0 by increasing the mass located in the interval $[\theta_0, \frac{\theta_0+\theta_1}{2})$ and the enthusiastic prior can be flattened around the modal value of θ_1 by decreasing the mass located in the interval $(\frac{\theta_0+\theta_1}{2}, \theta_1]$.

All figures should be referenced in text and discussed.

The following illustrations and analyses were performed using the concentrated skeptical prior and the default enthusiastic prior. Question from reviewer: Then why are the others shown? Why are you recommending practitioners use these? Are results for other combinations somewhere? If so, you should have: A comparison of the various combinations of the monitoring priors in Figure **??** is provided in Supplemental Appendix X.

3.1.3 Sequential monitoring

Enrollment will proceed until one of the following three conditions are satisfied:

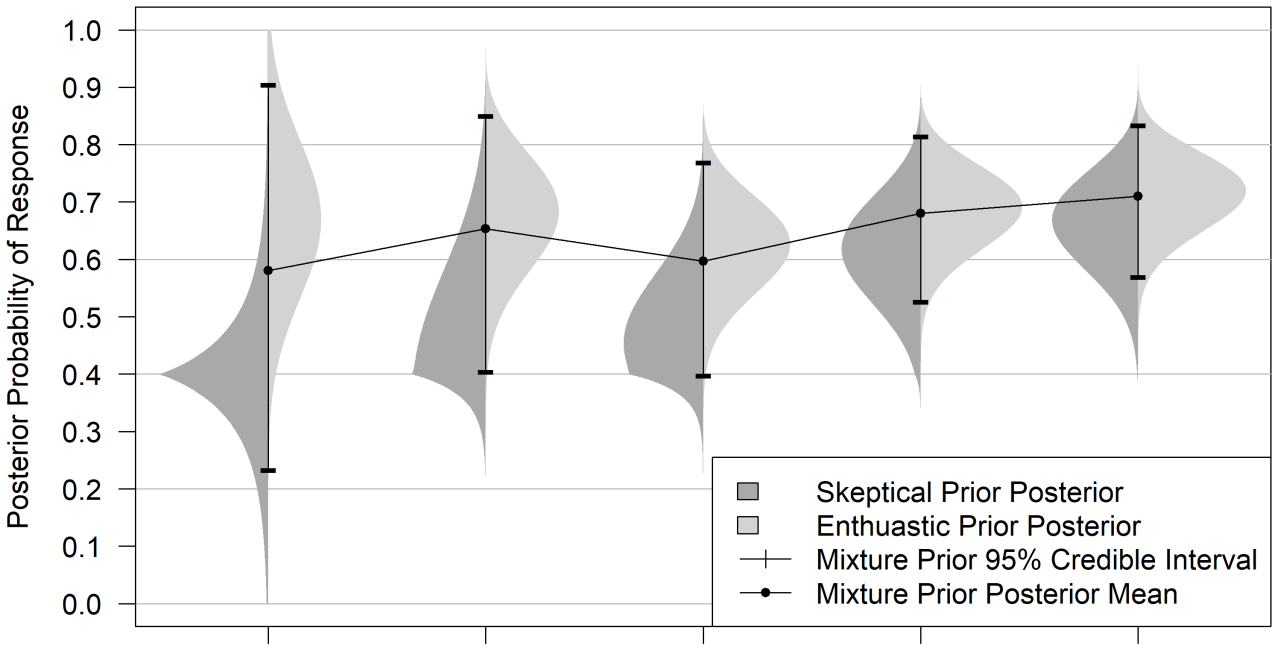
$$\text{Efficacy criteria (EFF): } P(\theta > \theta_0 | \mathbf{D}, \pi_S) \geq 0.975$$
$$\text{Futility criteria (FUT): } P\left(\theta \leq \frac{\theta_0+\theta_1}{2} \middle| \mathbf{D}, \pi_E\right) \geq 0.975$$
$$\text{Maximum sample size: } N = 112 \text{ patient outcomes obtained}$$

Assume that the outcomes are ascertained after approximately 4 months of follow-up and 2 patients per month on average are enrolled. If enrollment is terminated due to the efficacy or futility criteria being satisfied, those subjects who are still undergoing follow-up will still have their outcomes considered in the final analysis.

3.1.4 Example paths

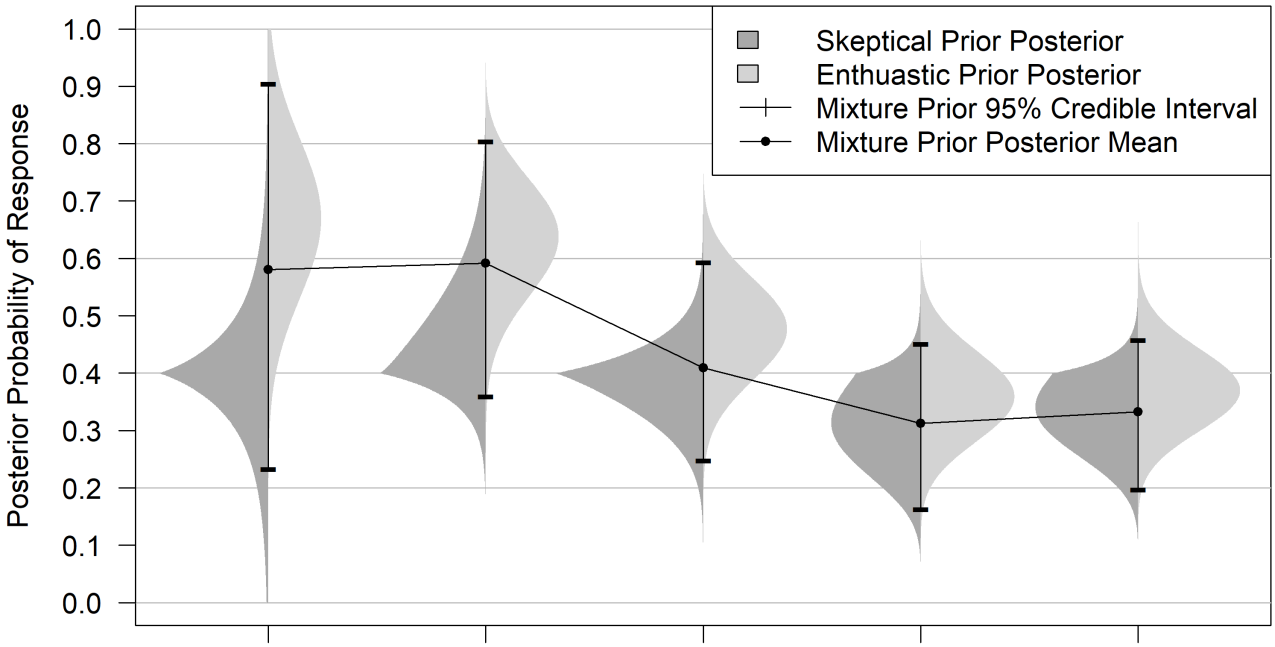
Figure 3a shows the results of a hypothetical trial with the initial prior specification (first panel), three interim analyses (middle panels), and a final analysis (last panel). This will never fly. There is no guidance

on how to interpret the results presented in the graphs. Also if the figure are going to be presented one on top of the other, then why make them so thin? Every abbreviation in a figure must be explained in the caption for the figure – see my paper recently published in biometrics (biom.13163). That level of detail is needed on all graphics.



Enrolled	0	12	23	34	34
Completed	0	10	20	30	34
Response Rate	n/a	70.0%	60.0%	70.0%	73.5%
Efficacy Criteria	0.496	0.148	0.124	0.006	0.001
Futility Criteria	0.165	0.079	0.148	0.020	0.006

(a) An example path - early stopage for efficacy.



Enrolled	0	12	24	37	37
Completed	0	10	20	30	37
Response Rate	n/a	60.0%	35.0%	23.3%	27.0%
Efficacy Criteria	0.496	0.250	0.603	0.904	0.873
Futility Criteria	0.165	0.163	0.754	0.988	0.989

(b) An example path - early stoppage for futility.

3.1.5 Preposterior Analysis of Operating Characteristics

An interim analysis will be completed after every 2 subjects complete follow-up.

Let INC be the probability of reaching the maximum sample size without a conclusive monitoring result, let SS be the average sample size at the definitive interim analysis (I) and at the end of follow-up (F), let CP be the coverage probability using the mixture prior, and let PM be the posterior mean an inference prior which is a 50/50 mixture of the skeptical and enthusiastic priors.

3.1.6 Type 1 error rate by the frequency of data monitoring

As expected, the probability of stopping enrollment due to a promising interim trial result and the Type 1 error rate at the final analysis increase with the frequency of interim monitoring, however, the increase is very slight at the final analysis. Regardless of frequency of monitoring there are good Type 1 error rates.

Monitoring Freq never abbreviate words is 1 for fully sequential design and 112 when the only analysis is with all completed outcomes.

3.2 Parallel Two-Group Design with Binary Endpoint

don't revise this much yet... I am going to look for a better example).

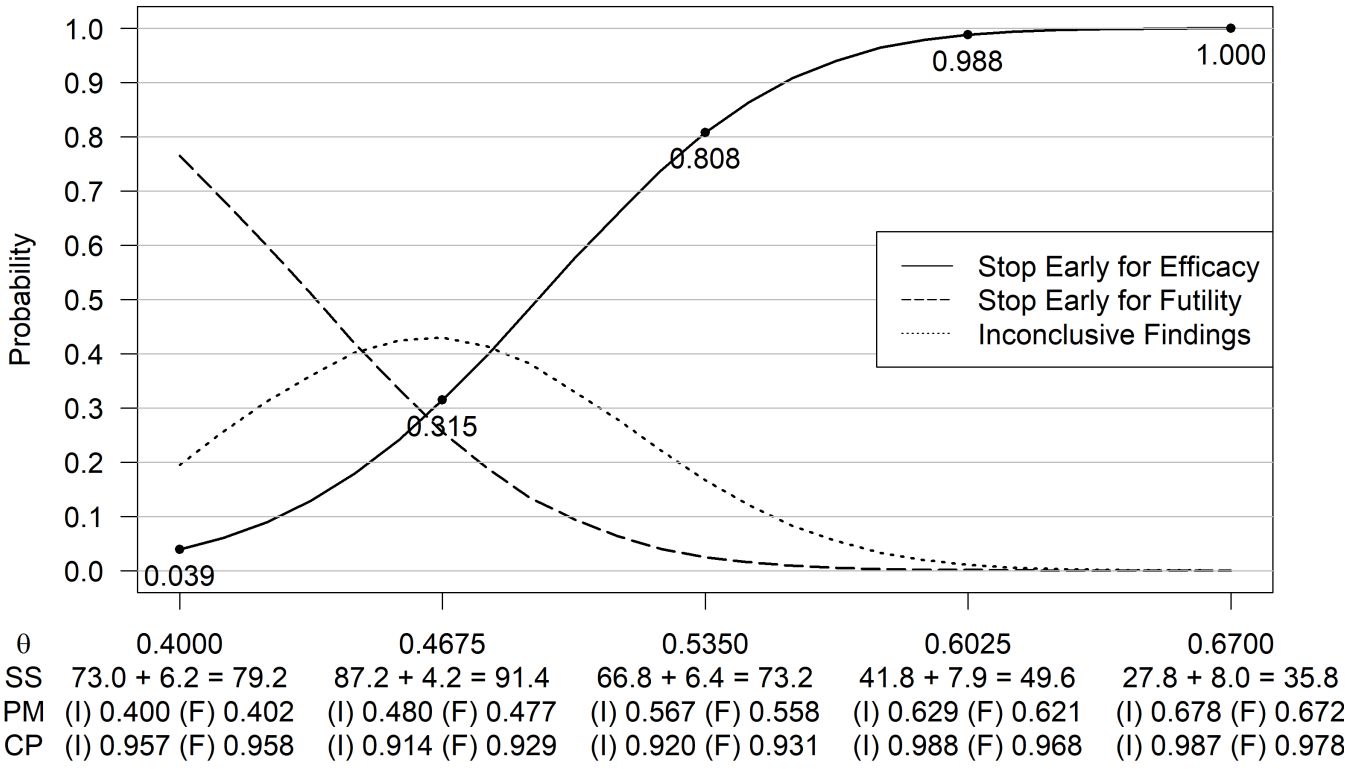
3.2.1 Motivating example

The Pediatric Lupus Trial of Belimumab Plus Background Standard Therapy (PLUTO) trial, was a multicenter study to evaluate the safety, pharmacokinetics, and efficacy of belimumab intravenous (IV) in pediatric patients 5 to 17 years of age with active systemic lupus erythematosus.

The goal was to test for superiority of belimumab to placebo. Based on adult studies, a response rate of 0.51 was expected for belimumab and based on previous research a response rate of 0.39 was expected for placebo.

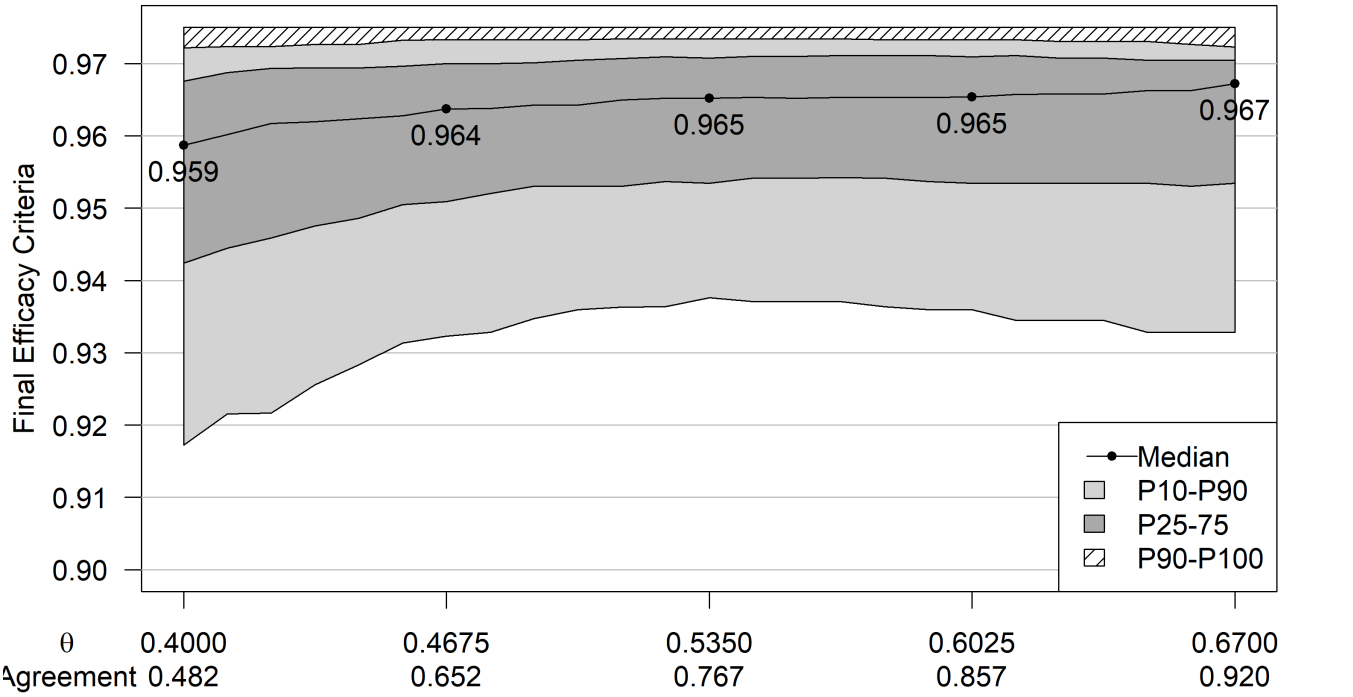
The study design included enrollment of 100 subjects, the first 24 subjects randomized in a 5:1 ratio (belimumab:placebo) and the remaining 76 subjects would be randomized in a 1:1 allocation ratio. Therefore, 58 subjects would be randomized to belimumab and 42 to placebo. The sample size was based on feasibility constraints rather than a power calculation.

The binary response endpoint was evaluated at 52 weeks post enrollment. The study start date was September 7, 2012, and the primary completion date was January 24, 2018. Since the follow-up period is 52 weeks the last enrollment is estimated to be a year prior to the primary competition date yielding an average enrollment rate of one enrollment per 17.2 days.



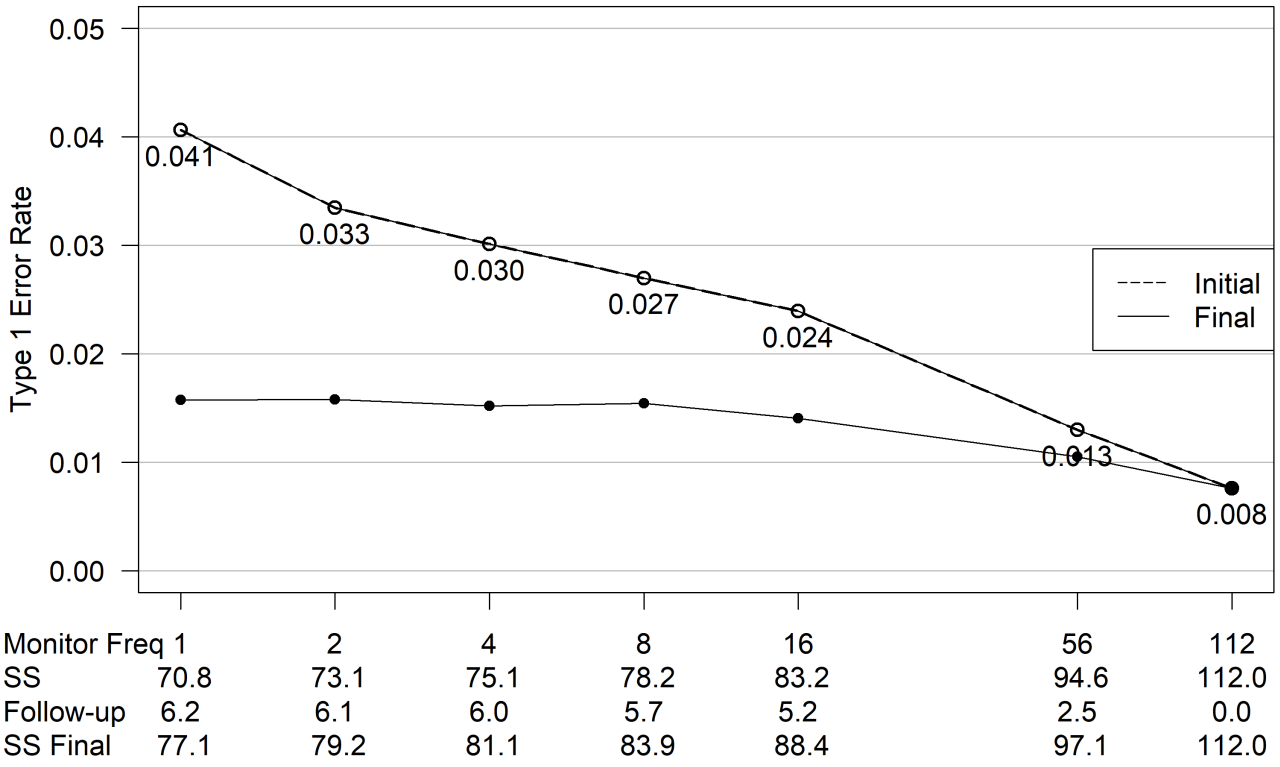
θ	0.4000	0.4675	0.5350	0.6025	0.6700
SS	73.0 + 6.2 = 79.2	87.2 + 4.2 = 91.4	66.8 + 6.4 = 73.2	41.8 + 7.9 = 49.6	27.8 + 8.0 = 35.8
PM	(I) 0.400 (F) 0.402	(I) 0.480 (F) 0.477	(I) 0.567 (F) 0.558	(I) 0.629 (F) 0.621	(I) 0.678 (F) 0.672
CP	(I) 0.957 (F) 0.958	(I) 0.914 (F) 0.929	(I) 0.920 (F) 0.931	(I) 0.988 (F) 0.968	(I) 0.987 (F) 0.978

(a) Caption text 1



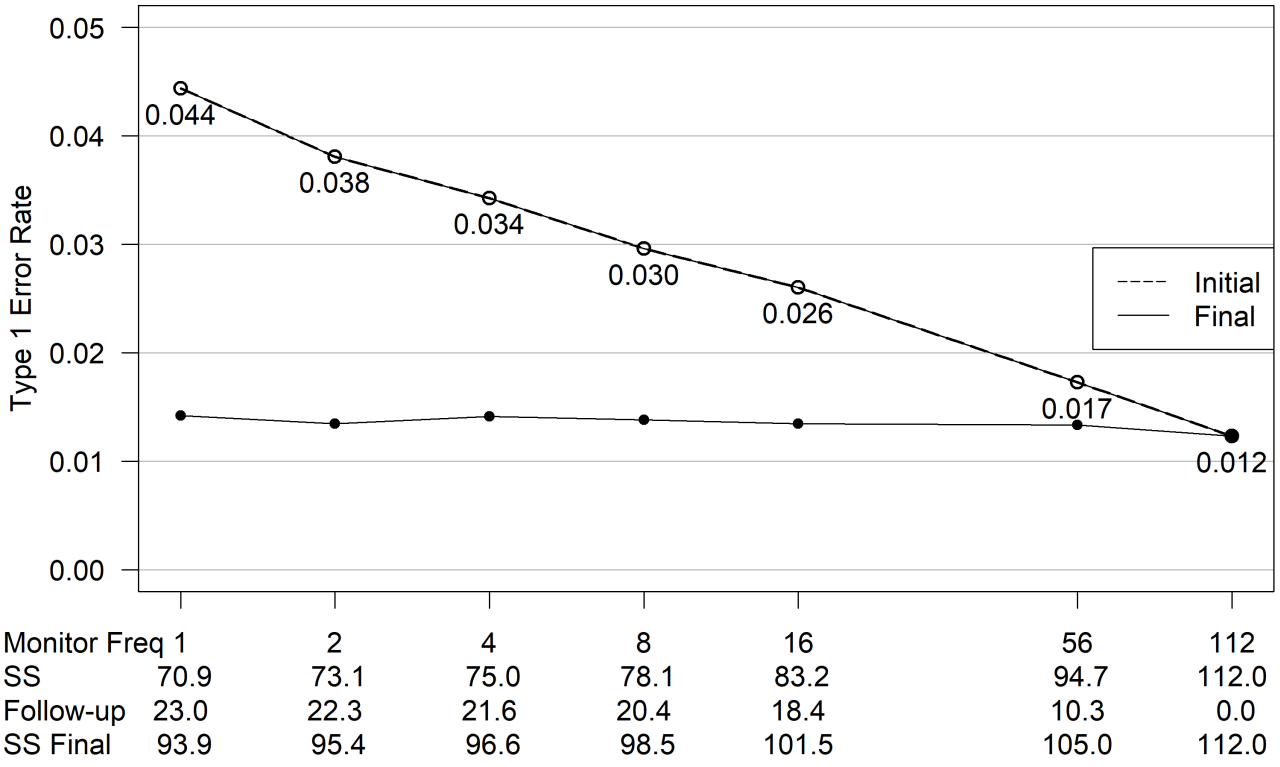
θ	0.4000	0.4675	0.5350	0.6025	0.6700
Agreement	0.482	0.652	0.767	0.857	0.920

(b) Caption text 2



Monitor Freq	1	2	4	8	16	56	112
SS	70.8	73.1	75.1	78.2	83.2	94.6	112.0
Follow-up	6.2	6.1	6.0	5.7	5.2	2.5	0.0
SS Final	77.1	79.2	81.1	83.9	88.4	97.1	112.0

(a) Caption text 1



Monitor Freq	1	2	4	8	16	56	112
SS	70.9	73.1	75.0	78.1	83.2	94.7	112.0
Follow-up	23.0	22.3	21.6	20.4	18.4	10.3	0.0
SS Final	93.9	95.4	96.6	98.5	101.5	105.0	112.0

(b) Caption text 2

3.2.2 Model formulation

Let θ_0 represent the response rate the control group and θ_1 represent the response probability for the investigational product (IP) group. Consider the hypothesis testing of IP superiority to control

$$H_0 : \theta_1 - \theta_0 \leq 0 \text{ vs. } H_1 : \theta_1 - \theta_0 > 0.$$

The priors will be chosen based on the joint specification in (13). First, a prior on the response probability for the placebo group is given in the form of (14). This prior is chosen to be flat in the region 0.39 ± 0.10.

To complete the joint specification of (13), skeptical and enthusiastic priors of the form (15) will be parameterized as to satisfy (8) and (9). The skeptic believes there is no difference in response rates by treatment group, and the enthusiastic person believes the IP group will have a response rate probability that is 0.12 higher than the placebo group.

Enrollment will proceed until one of the following three conditions are satisfied:

$$\text{Efficacy criteria (EFF): } P(\theta_1 - \theta_0 > 0|\mathbf{D}, \pi_S) \geq 0.975$$
$$\text{Futility criteria (FUT): } P(\theta_1 - \theta_0 \leq 0.06|\mathbf{D}, \pi_E) \geq 0.975$$
$$\text{Maximum sample size: } N = 100 \text{ patient outcomes}$$

An interim analysis is competed after every 10 subjects have completed outcomes.

3.2.3 Design properties: Results

Simulations were run fixing the placebo response rate at $\theta_0 = 0.39$ and varying the treatment response rate $\theta_1 \in [0.39, 0.51]$. Due to the low maximum sample size, no simulations resulted in the efficacy criteria ($P(\theta_1 - \theta_0 > 0|\mathbf{D}, \pi_S) \geq 0.975$) being satisfied. Instead of the skeptical prior π_S being used for the efficacy criteria, an inference prior π_I of the form (18) is used where the choice of ω in (**??**) is determined based on varying $p(\pi_S)$ and $p(\pi_E)$ at the outset.

4 Discussion

Text.

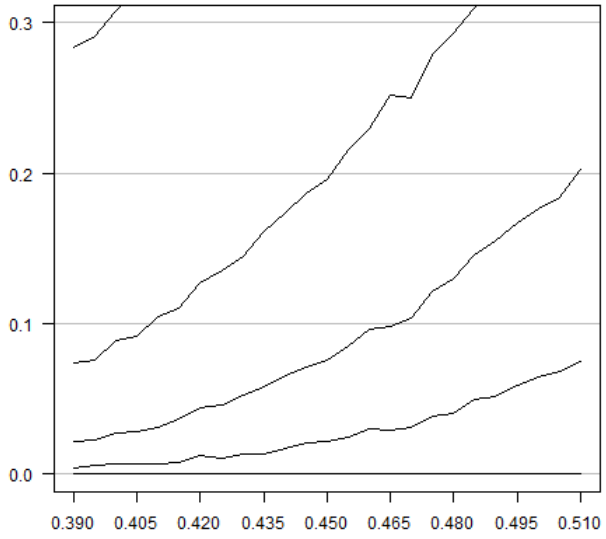


Figure 6: Caption text 2

5 Supplementary material

5.1 Shape parameter specification

5.2 Risk difference prior parameterization

5.2.1 Type 1 error rate depending on enrollment schemes

5.2.2 Robustness of parameterizations of monitoring priors

6 BibTeX

References

Cornfield, J. (1966*a*), 'A Bayesian Test of Some Classical Hypotheses, with Applications to Sequential Clinical Trials', *Journal of the American Statistical Association* **61**(315), 577.
URL: *https://www.jstor.org/stable/2282772?origin=crossref*

Cornfield, J. (1966*b*), 'Sequential Trials, Sequential Analysis and the Likelihood Principle', *The American Statistician* **20**(2), 18.
URL: *https://www.jstor.org/stable/2682711?origin=crossref*

Fayers, P. M., Ashby, D. & Parmar, M. K. B. (1997), 'Tutorial in Biostatistics: Bayesian Data Monitoring in Clinical Trials', *Statistics in Medicine* **16**(12), 1413–1430.
URL: *http://doi.wiley.com/10.1002/%28SICI%291097-0258%2819970630%2916%3A12%3C1413%3A%3AAID-SIM578%3E3.0.CO%3B2-U*

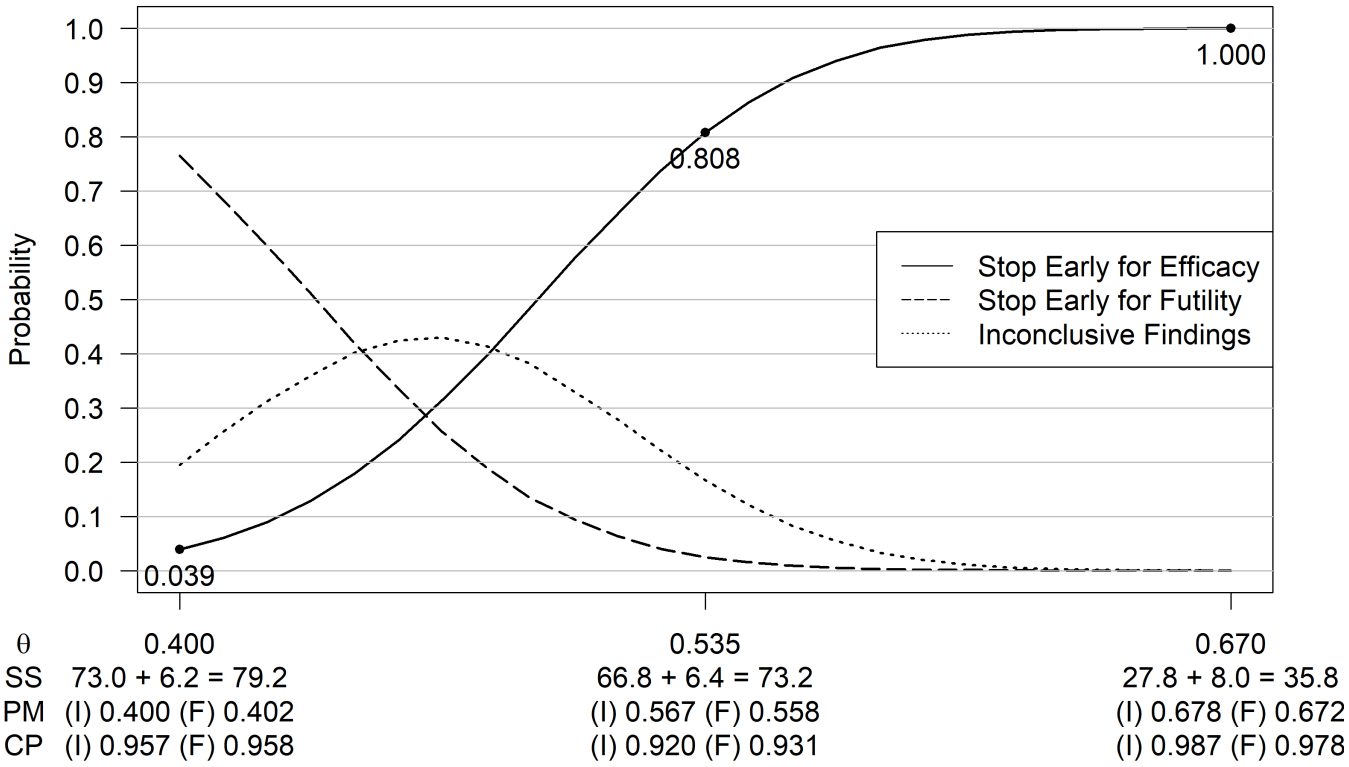
Freedman, L. S. & Spiegelhalter, D. J. (1989), 'Comparison of Bayesian with group sequential methods for monitoring clinical trials', *Controlled Clinical Trials* **10**(4), 357–367.
URL: *https://www.sciencedirect.com/science/article/pii/0197245689900019?via%3Dihub*

Freedman, L. S. & Spiegelhalter, D. J. (1992), 'Application of bayesian statistics to decision making during a clinical trial', *Statistics in Medicine* **11**(1), 23–35.
URL: *http://doi.wiley.com/10.1002/sim.4780110105*

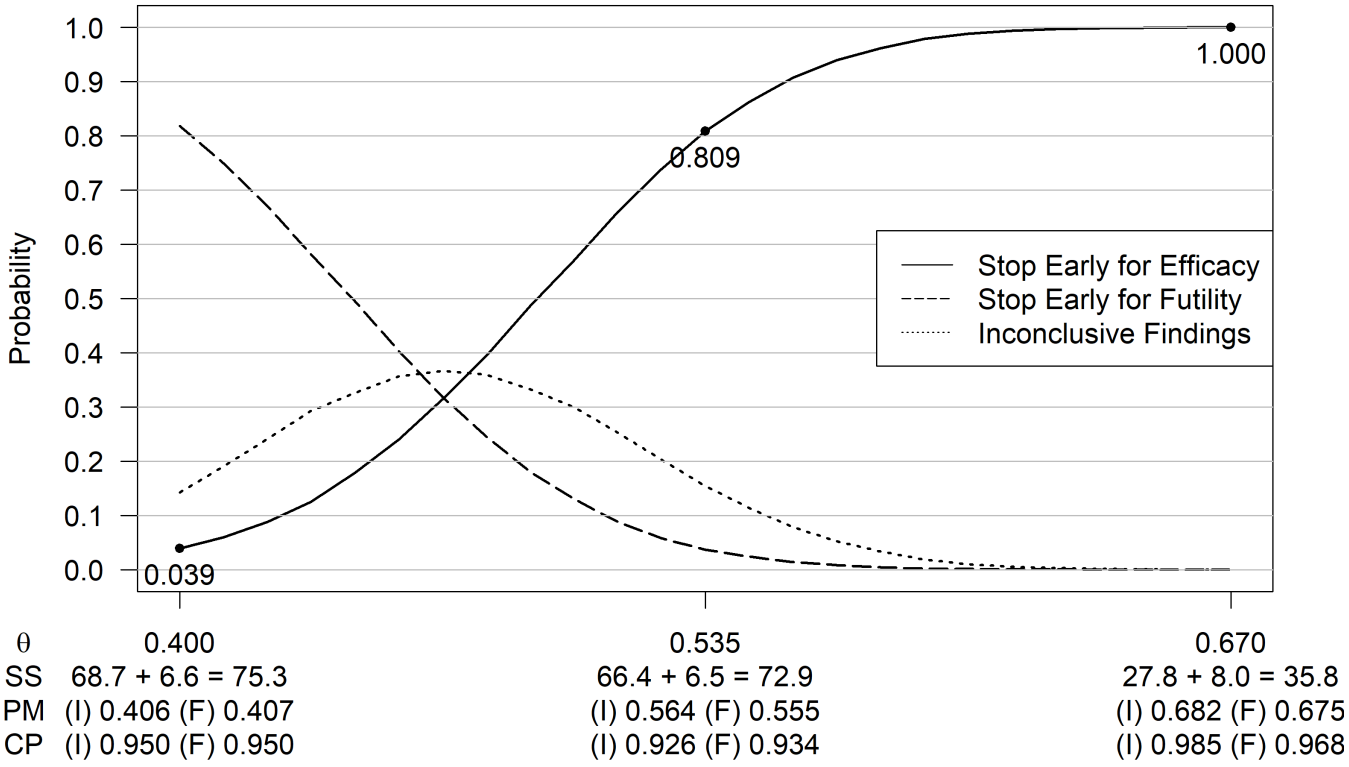
Neyman, J. & Greenhouse, S. W. (1967), *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability.*, University of California Press.
URL: *https://projecteuclid.org/euclid.bsmsp/1200513830*

Spiegelhalter, D. J., Freedman, L. S. & Parmar, M. K. B. (1993), 'Applying Bayesian ideas in drug development and clinical trials', *Statistics in Medicine* **12**(15-16), 1501–1511.
URL: *http://doi.wiley.com/10.1002/sim.4780121516*

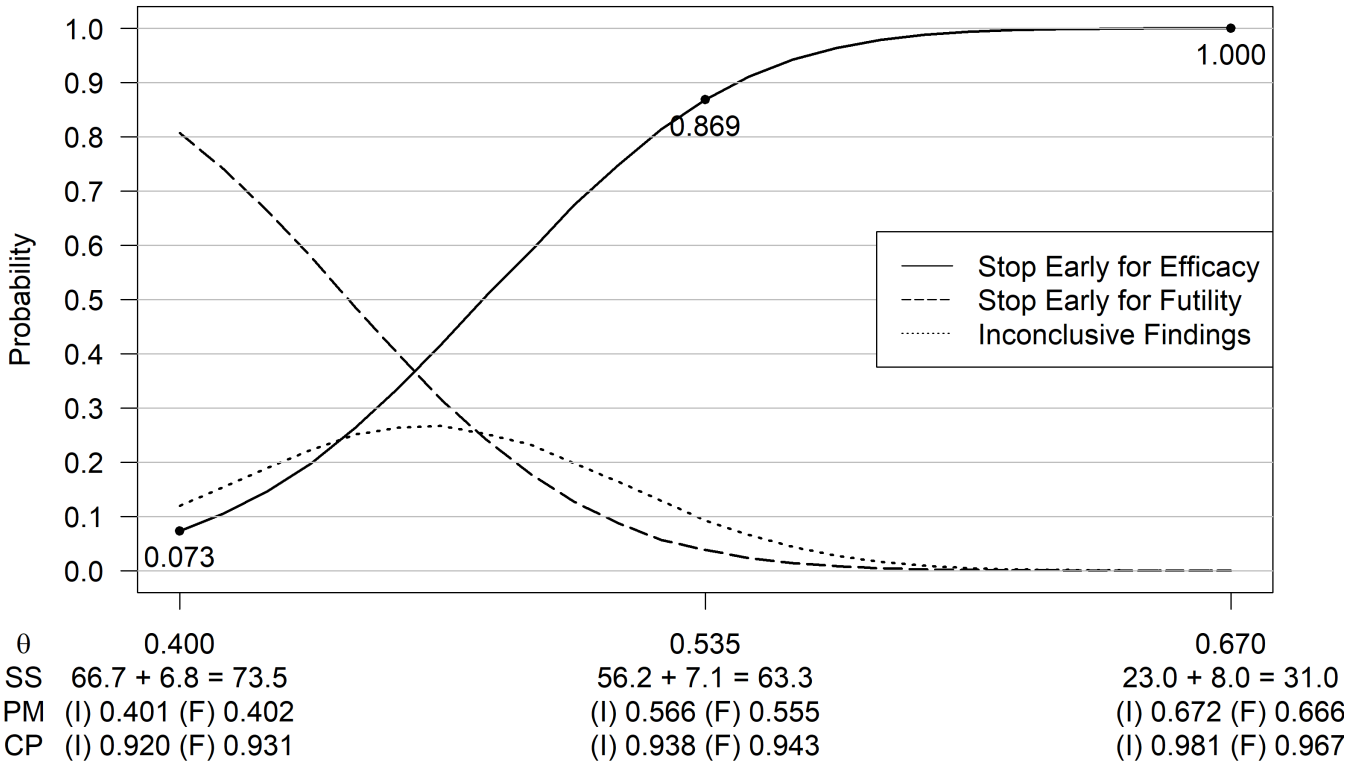
Spiegelhalter, D. J., Freedman, L. S. & Parmar, M. K. B. (1994), 'Bayesian Approaches to Randomized Trials', *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **157**(3), 357.
URL: *https://www.jstor.org/stable/10.2307/2983527?origin=crossref*



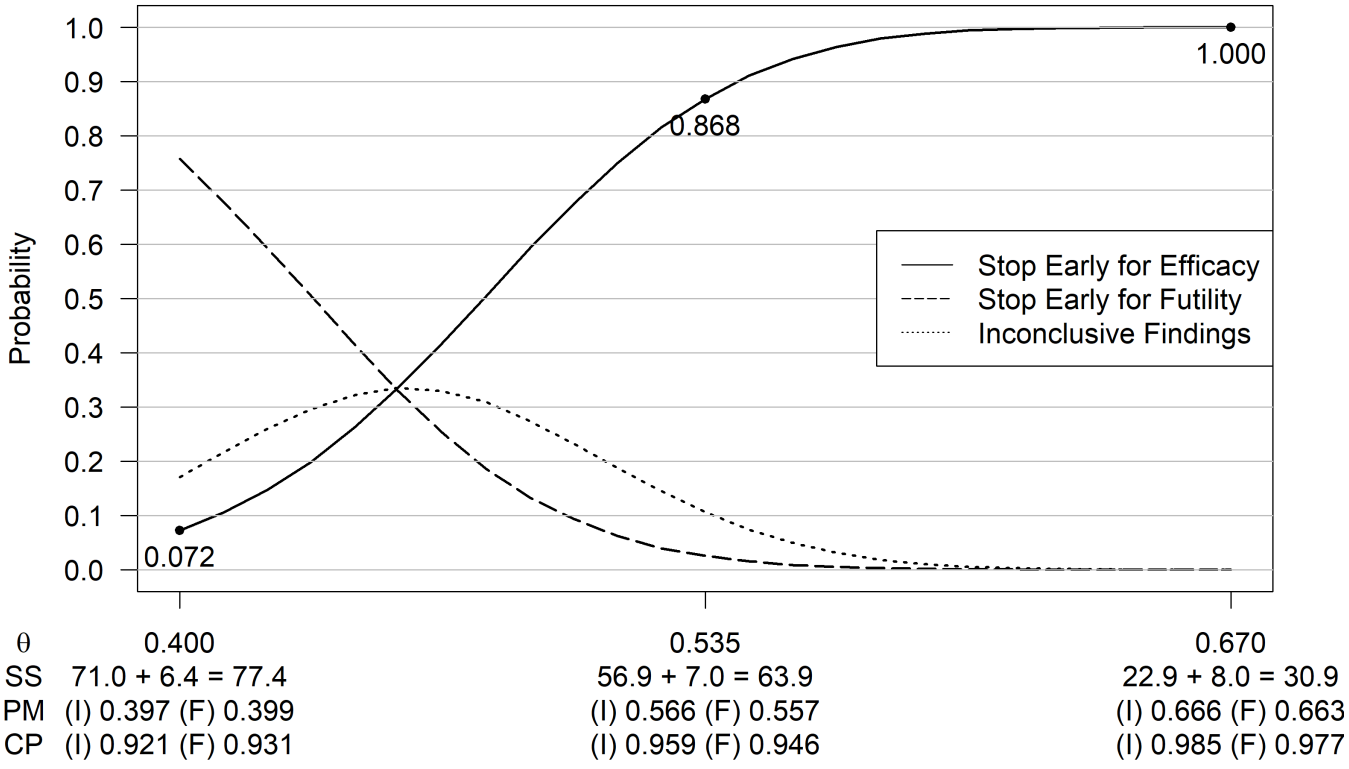
(a) Caption text 1



(b) Caption text 2



(a) Caption text 1



(b) Caption text 2