Towards 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 3, 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 enthuastic 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 clincial trials is well devloped, 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 Monitoring versus Estimation Priors

2.1.1 Bayesian hypothesis testing based on posterior probabilities

The Bayesian paradigm provides direct inference on a parameter of interest through specification of a model for the data and prior distributions for unknown quantities. 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 $\theta \in \Theta$.

Suppose the hypothesis for the trial is $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$. These hypotheses are judged based on posterior probabilities of θ by evaluating the marginal likelihood

$$P(\theta \in \Theta_i|\mathbf{D}) = \int_{\Theta_i} p(\theta|\mathbf{D})d\theta \text{ for } i \in \{0, 1\}, \tag{1}$$

where the posterior distribution of θ depends on the choice of prior distribution $\pi(\theta)$ since $p(\theta|\mathbf{D}) = p(\mathbf{D}|\theta)\pi(\theta)/p(\mathbf{D})$ by Bayes rule.

2.1.2 Prior elicitation

It has been said that "the purpose of a trial is to collect data that bring to conclusive consensus at termination opinions that had been diverse and indecisive at the *outset*" (Kass and Greenhouse (1989), emphasis added). These opinions manifest as priors $\pi(\theta)$ based on their relation to $P(\theta \in \Theta_i|\pi(\theta))$ $i \in \{0, 1\}$. Note this quantity does not depend on the data $\mathbf{D}$ and therefore reflects a-priori opinion.

The specification of the prior distribution depends on the research objective. An *inference prior* is a prior that is used when the research objective is to make final analysis after data collection is complete. A *monitoring prior* is a prior that is used when the research objective is to see if there is a persuasive result based in the interim data. Stopping for efficacy is ceasing enrollment due to a promising interim result (one that is consistent with H_1), and stopping for futility is ceasing enrollment due to a discouraging interim result (one that is consistent with H_0).

Define $1 - \epsilon \in (0, 1)$ as a threshold for *a compelling level of evidence* as it relates to θ. We say that an individual is "all but convinced" that H_i is true given the observed data if

$$P(\theta \in \Theta_i|\mathbf{D}) > 1 - \epsilon \text{ for } i \in \{0, 1\}. \tag{2}$$

The quantity ϵ reflects *residual uncertainty* of H_i being true relative to the competing hypothesis.

A enthuastic prior is an informative prior that gives preference to H_1 such that it is "all but convinced" that H_1 is true a-priori. This prior $\pi_E(\theta) \equiv \pi_E$ has the property that

$$P(\theta \in \Theta_1|\pi_E) > 1 - \epsilon \tag{3}$$

(equivalently $P(\theta \in \Theta_0|\pi_E) \leq \epsilon$). The choice of $1 - \epsilon \in (0, 1)$ is motivated by a *compelling level of evidence* as it relates to θ, although in this setting the "evidence" reflects a theoretical opinion rather than empirical judgement. For example, if $1 - \epsilon = 0.95$, then this choice of enthaustic prior places 95% prior probability that $\theta \in \Theta_1$.

A skeptical prior is an informative prior that does not give strong preference to H_1. This prior $\pi_S(\theta) \equiv \pi_S$ could have the property that $P(\theta \in \Theta_0|\pi_S) > 1 - \epsilon$, in which case it is "all but convinced" that H_0 is true a-priori, however, this demonstrates such an extreme disbelief in the possibility of a positive effect that conducting the trial at all would be viewed as dubious. Consider a region $\Theta_A \subset \Theta_1$ that demonstrates a substantial positive effect. The skeptical prior is then constructed such that a substantial positive effect is unlikely, that is,

$$P(\theta \in \Theta_A|\pi_S) \leq \epsilon. \tag{4}$$

2.1.3 Sequential monitoring

The use of monitoring based on changing the opinion of skeptical and enthuastic 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 appriopriate if diverse prior opinions about θ would be in agreement given the interim data (e.g. the skeptical and enthuastic person reach the same conclusion).

Promising interim result

In order for interim evidence showing H_1 is true to be persuasive, it has to cause the skeptic, who initially held that $P(\theta \in \Theta_A \subset \Theta_1|\pi_S) \leq \epsilon$ to conclude

$$P(\theta \in \Theta_1|\mathbf{D}, \pi_S) > 1 - \epsilon. \tag{5}$$

Disillusioning interim result

Recall that $\theta \in \Theta_A \subset \Theta_1$ represents a substantial positive effect. A disillusioning interim result not only demonstrates that a substantial positive effect is unlikely, but furthermore demonstrates that a moderate or intermediate positive effect is also unlikely. For this reason, consider $\theta \in \Theta_I \subset \Theta_A$ to demonstrate a moderate positive effect. In order for interim evidence showing H_1 is false to be persuasive, it has to cause the enthusiast, who initially held that $P(\theta \in \Theta_1|\pi_E) > 1 - \epsilon$ to conclude that

$$P(\theta \in \Theta_I|\mathbf{D}, \pi_E) \leq \epsilon. \tag{6}$$

2.1.4 Final inference

An inference prior $\pi_I(\theta) \equiv \pi_I$ is often less divisive than the skeptical and enthaustic priors, and can be viewed as a balance of the more divisive opinions. Consider a mixture prior constructed from the monitoring process as the inference prior:

$$\pi_I = \omega \cdot \pi_S + (1 - \omega) \cdot \pi_E \tag{7}$$

for $\omega \in [0, 1]$.

The choice of ω will be based on posterior model probabilities. In particular,

$$\omega = p(\pi_S|\mathbf{D}) = \frac{p(\mathbf{D}|\pi_S)p(\pi_S)}{p(\mathbf{D}|\pi_S)p(\pi_S) + p(\mathbf{D}|\pi_E)p(\pi_E)} \tag{8}$$

where $p(\pi_S) + p(\pi_E) = 1$. The quantities $p(\pi_E)$ and $p(\pi_S)$ reflect prior belief in the distribution of θ. A default option is $p(\pi_S) = p(\pi_E) = \frac{1}{2}$. The inference prior is used to evaluate the hypotheses in (1), 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.

The inference prior will be used at the point of enrollment stoppage due to a persuasive monitoring result, and again at the end of data collection (once those in active follow-up have completed outcomes).

2.1.5 Default selection of priors for response proportions

The conjugate prior for binomially distributed data is the beta prior, however, here we consider using generalized normal priors truncated to the unit interval for its flexibility and adaptability to higher dimensions. Consider the univariate generalized normal kernel $\exp\{-(\frac{|\theta-\mu|}{\alpha})^\beta\}$ 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. When truncated to the unit interval, this density becomes

$$\pi(\theta) \propto \exp\left\{-\frac{|\theta - \mu|^\beta}{\alpha}\right\} I(\theta \in [0, 1]). \tag{9}$$

The parameters μ, α, and β create enthuastic and skeptical priors that satisfy (3) and (4).

This prior naturally extends to higher dimensions. For example, 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.

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

$$\pi(\theta_0, \theta_1) = \pi(\theta_0) \times \pi(\theta_1|\theta_0) \tag{10}$$

$$\pi(\theta_0) \propto \exp\left\{ \left(\frac{|\theta_0 - \mu_0|}{\alpha_0} \right)^{\beta_0} \right\} I(\theta_0 \in [0,1]) \tag{11}$$

$$\pi(\theta_1|\theta_0) \propto \exp\left\{ \left(\frac{|(\theta_1 - \theta_0) - \delta|}{\alpha_1} \right)^{\beta_1} \right\} I(\theta_1 \in [0,1]) \tag{12}$$

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 $\theta_1 - \theta_0$.

Notes on computation and simulations

3 Examples

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

3.1.1 Motivating example

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 enthuastic priors will be of the form (9) 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 (3) and (4), 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 (3) and (4). 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 enthuastic prior can be flattened around the modal value of θ_1 by decreasing the mass located in the interval $(\frac{\theta_0+\theta_1}{2}, \theta_1]$.

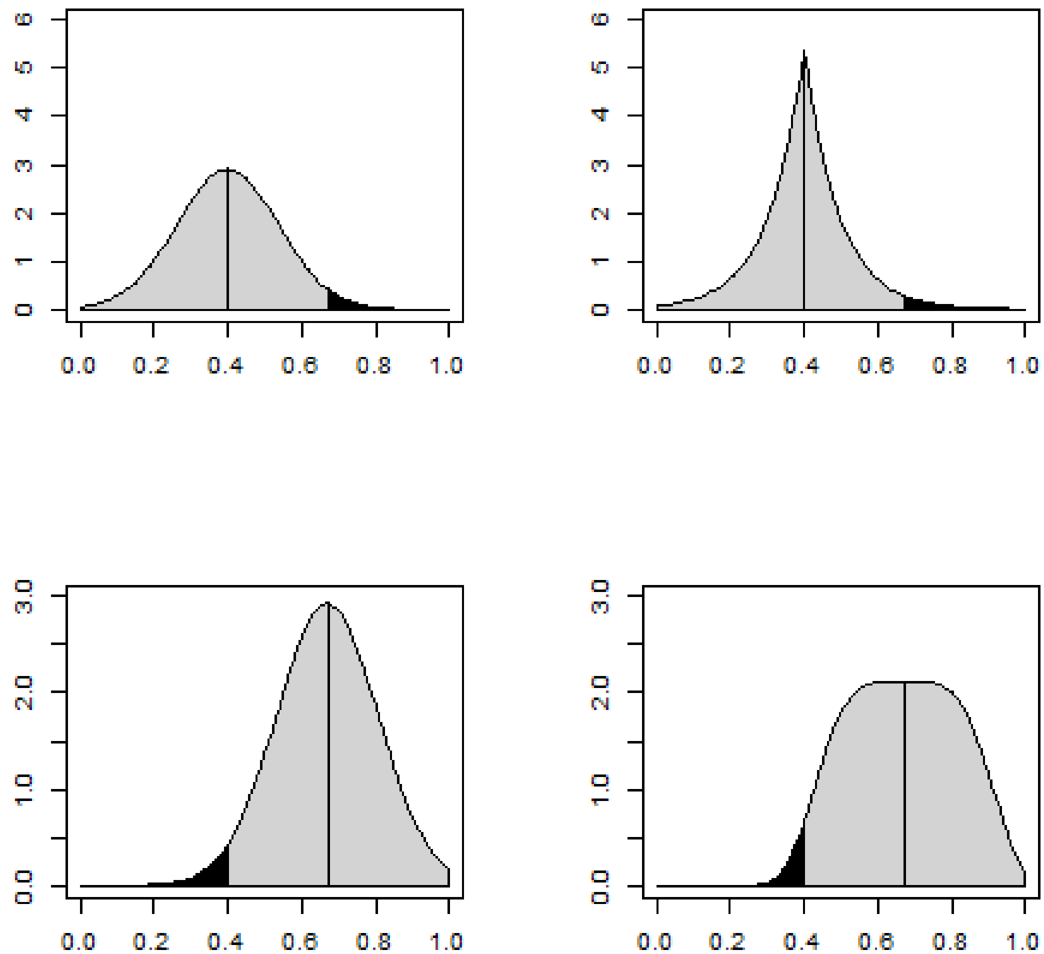


Figure 1: (a) Default skeptical prior ($\beta = 2$) (b) Concentrated skeptical prior
(c) Default enthuastic prior ($\beta = 2$) (d) Flattened enthaustic prior

The following analyses are done with the concentrated skeptical prior and the default enthuastic prior.

3.1.3 Sequential monitoring

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

$$\text{Efficacy criteria (EFF): } P(\theta > \theta_0 | \mathbf{D}, \pi_S) \geq 0.975$$

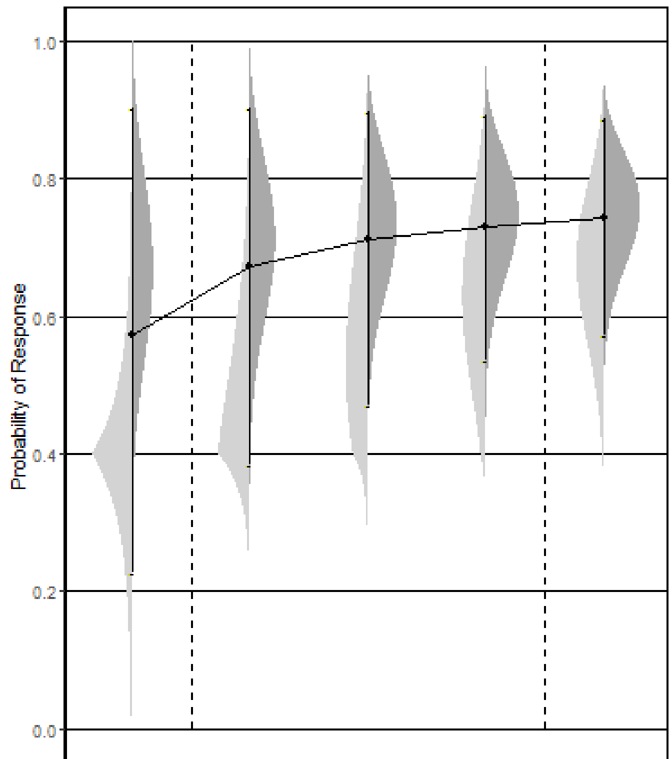
$$\text{Futility criteria (FUT): } P\left(\theta \leq \frac{\theta_0+\theta_1}{2} \middle| \mathbf{D}, \pi_E\right) \geq 0.975$$

$$\text{Maximum sample size: } N = 112 \text{ patient outcomes obtained}$$

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

To demonstrate the monitoring procedure, two example trials are considered.



Sample Size	0	5	10	15	20
# Successes	0	4	8	12	16
# Missing	0	0	0	0	0
$P(\theta > (\theta_1 + \theta_0)/2 \mid D, \pi_S)$	0.503	0.791	0.931	0.982	0.996
$P(\theta \leq \theta_0 \mid D, \pi_E)$	0.165	0.072	0.031	0.014	0.006

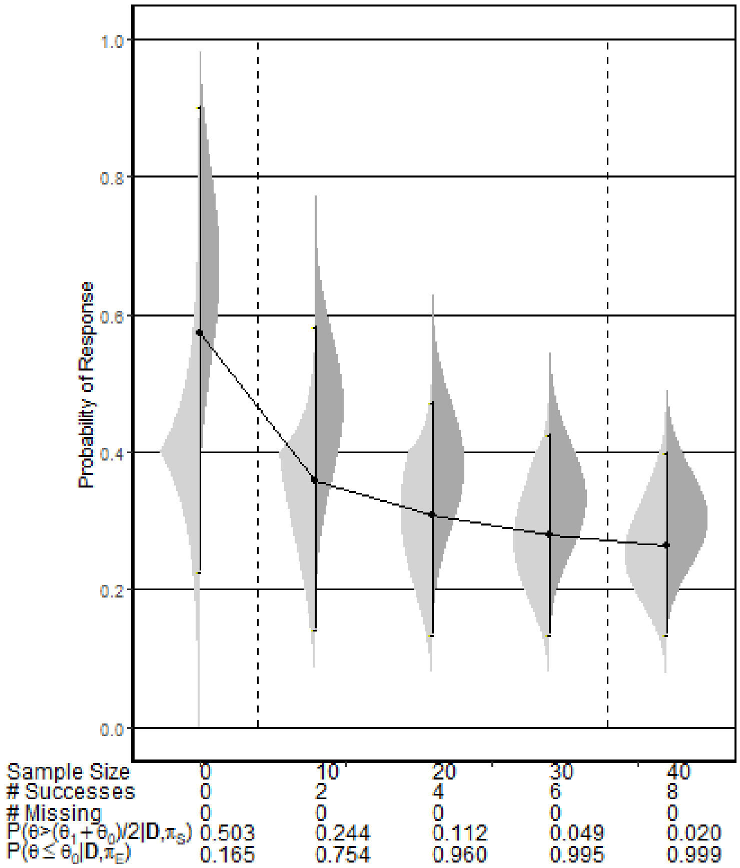
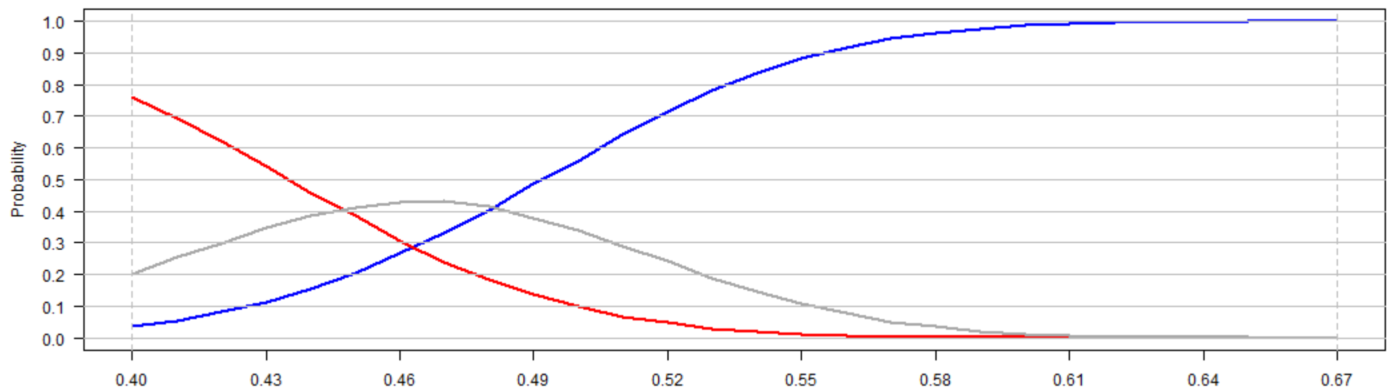


Figure 3: (a) Early stopping for efficacy (b) Early stopping for futility

3.1.5 Preposterior Analysis of Operating Characteristics

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

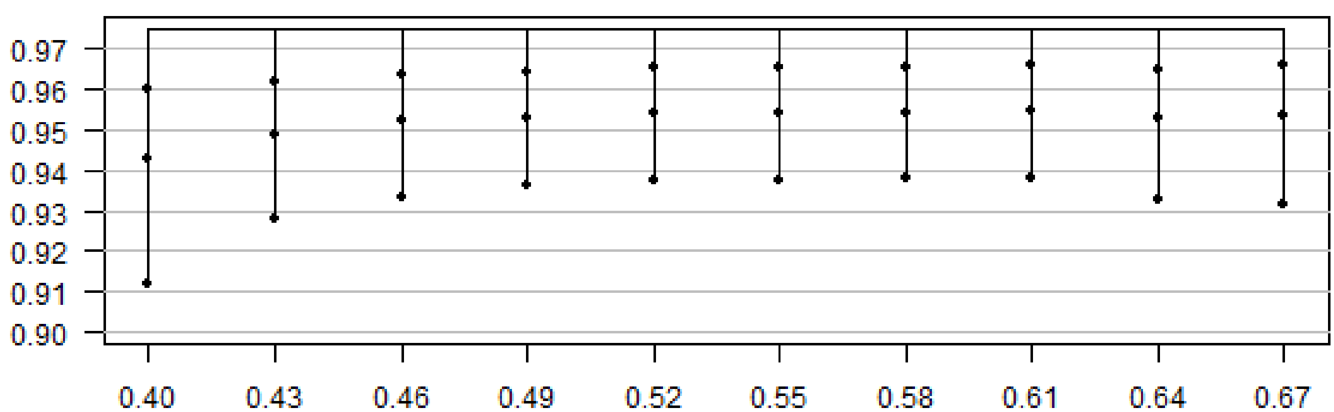


EFF	0.037	0.112	0.266	0.486	0.712	0.882	0.963	0.993	0.999	1.000
FUT	0.762	0.541	0.306	0.136	0.047	0.012	0.003	0.000	0.000	0.000
INC	0.201	0.347	0.428	0.378	0.241	0.106	0.034	0.007	0.001	0.000
SS	73.1 + 6.1 = 79.2	83.3 + 4.9 = 88.2	87.2 + 4.2 = 91.4	83.5 + 4.6 = 88.1	73.5 + 5.8 = 79.3	60.4 + 6.9 = 67.4	49.1 + 7.6 = 56.7	39.7 + 7.9 = 47.6	32.9 + 8.0 = 40.9	27.8 + 8.0 = 35.8
PM	(I) 0.400 (F) 0.402	(I) 0.431 (F) 0.432	(I) 0.469 (F) 0.468	(I) 0.510 (F) 0.505	(I) 0.549 (F) 0.541	(I) 0.582 (F) 0.574	(I) 0.610 (F) 0.601	(I) 0.635 (F) 0.627	(I) 0.658 (F) 0.650	(I) 0.678 (F) 0.672
CP	(I) 0.958 (F) 0.958	(I) 0.937 (F) 0.940	(I) 0.929 (F) 0.930	(I) 0.893 (F) 0.911	(I) 0.903 (F) 0.916	(I) 0.925 (F) 0.936	(I) 0.967 (F) 0.956	(I) 0.990 (F) 0.972	(I) 0.988 (F) 0.976	(I) 0.988 (F) 0.980

Figure 4a: Sequential design properties of proof-of-activity trial

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 enthuastic priors.

3.1.6 Agreement between interim and final result



AGR	0.484	0.580	0.644	0.699	0.750	0.791	0.830	0.870	0.898	0.920

Figure 4b: Percent agreement between interim and final result is listed in AGR. The lower dot is the 10th percentile, middle dot is 25th percentile, upper dot is 50th percentile/median

3.1.7 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.

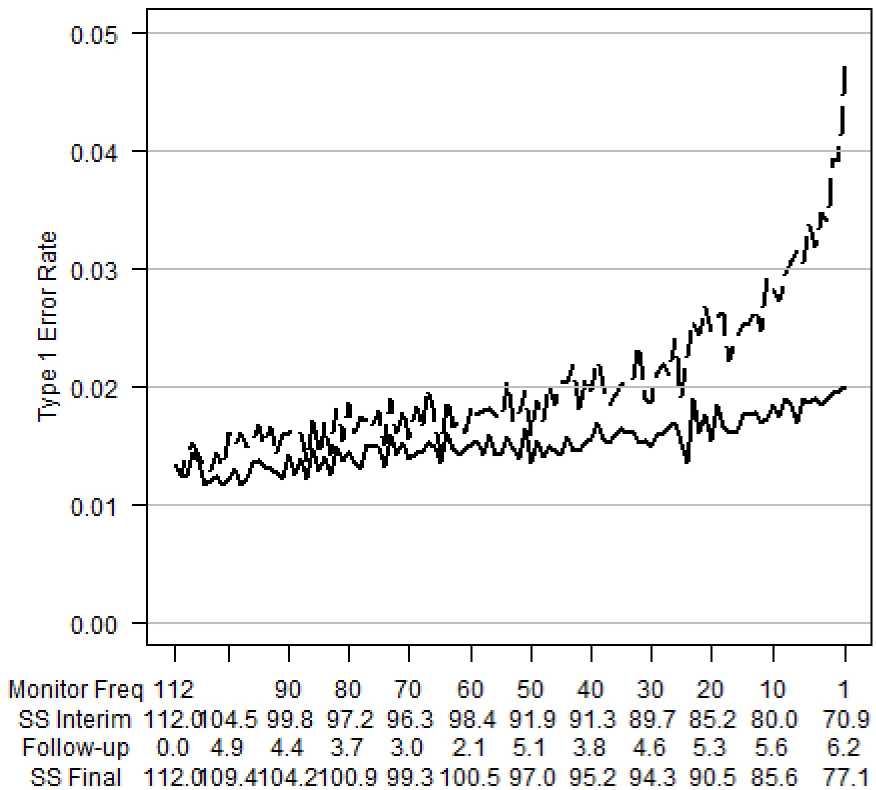


Figure 4: Type 1 error rate depending on frequency of sequential monitoring

Monitoring Freq 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

3.2.1 Motivating example

The Pediatric Lupus Trial of Belimumab Plus Background Standard Therapy (PLUTO) trial, was a multi-center 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 evaulated 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 completetition date yielding an average enrollment rate of one enrollment per 17.2 days.

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 (10). First, a prior on the response probability for the placebo group is given in the form of (11). This prior is chosen to be flat in the region 0.39 ± 0.10.

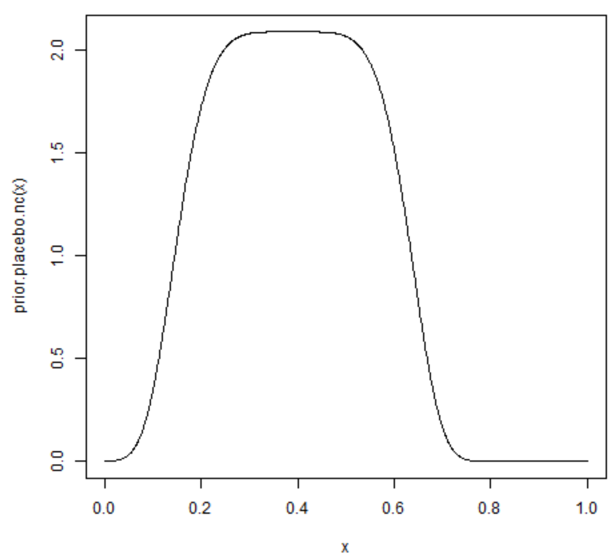


Figure 5a: Prior on placebo response probability in the form of (11) ($\mu_0 = 0.39, \alpha_0 = 0.26, \beta_0 = 5.3$)

To complete the joint specification of (10), skeptical and enthuastic priors of the form (12) will be parameterized as to satisfy (3) and (4). 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.

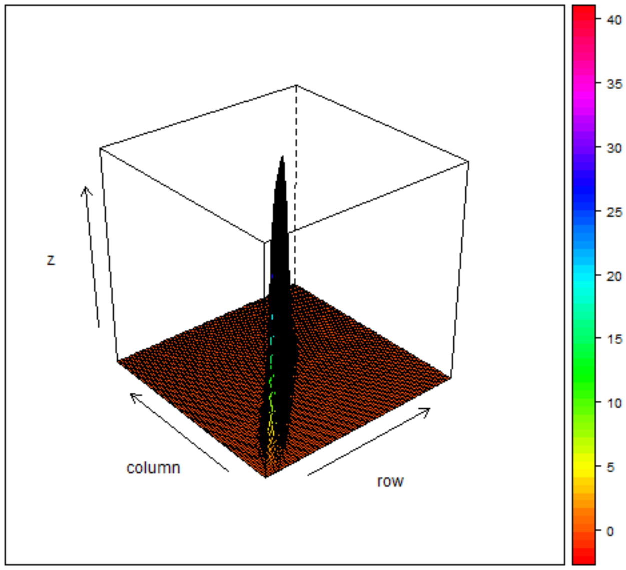


Figure 5b: Skeptical prior on response difference in form of (12) ($\delta = 0, \alpha_1 = 0.03, \beta_1 = 1.6$)

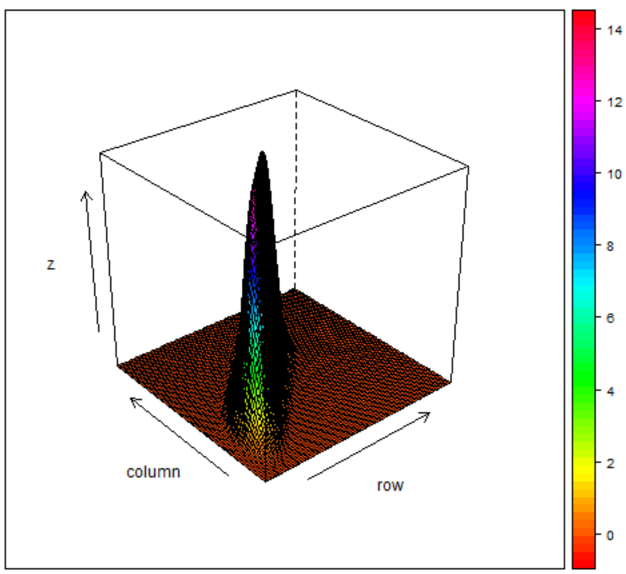


Figure 5c: Enthuastic prior on response difference in form of (12) ($\delta = 0.12, \alpha_1 = 0.087, \beta_1 = 2$)

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

Efficacy criteria (EFF): $P(\theta_1 - \theta_0 > 0|\mathbf{D}, \pi_S) \geq 0.975$

Futility criteria (FUT): $P(\theta_1 - \theta_0 \leq 0.06|\mathbf{D}, \pi_E) \geq 0.975$

Maximum sample size: $N = 100$ patient outcomes

An interim analysis is compelted 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 critera, an inference prior π_I of the form (7) is used where the choice of ω in (8) is determined based on varying $p(\pi_S)$ and $p(\pi_E)$ at the outset.

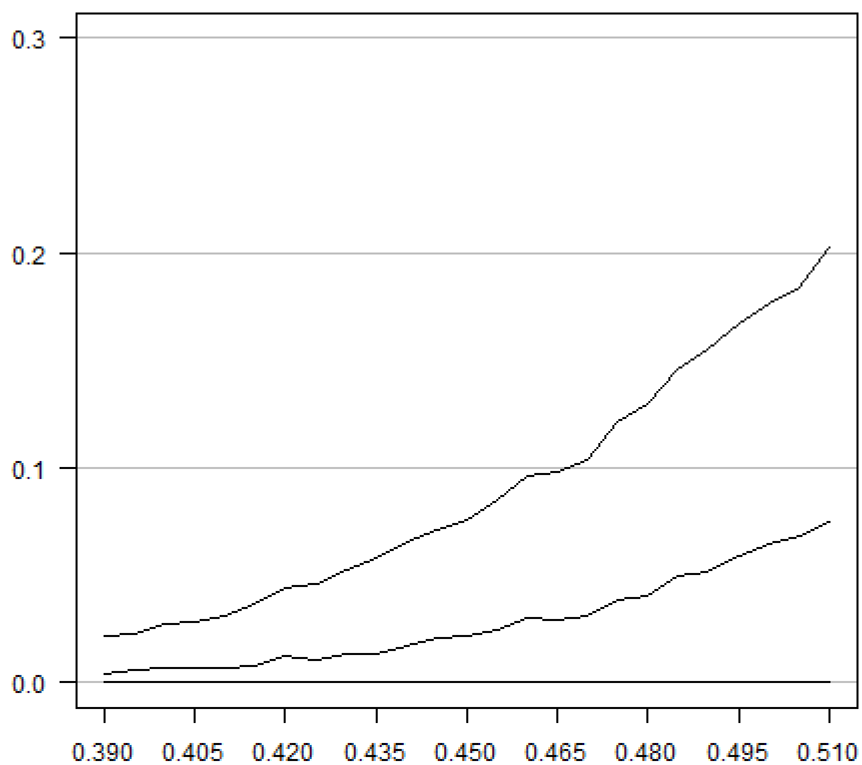


Figure 6: Probability of efficacy with three specifications of inference prior:

(1) $p(\pi_S) = 1$, $p(\pi_E) = 0$ (usual skeptical prior, lower line)

(2) $p(\pi_S) = 0.75$, $p(\pi_E) = 0.25$ (middle line)

(3) $p(\pi_S) = p(\pi_E) = 0.5$ (upper line)

4 Discussion

5 Supplementary material

5.0.1 Type 1 error rate depending on enrollment schemes

Consider the same trial but with a longer follow-up length of 8 months rather than 4 months.

Table 2: Comparison of efficacy stopping, Type 1 Error Rate
(efficacy criteria with full data), and sample size by follow-up length and frequency of sequential monitoring.

	4 month FU				8 month FU			
# Monitoring Freq	EFF	T1E	SS Final	Ongoing	EFF	T1E	SS Final	Ongoing
1	0.046	0.018	77.1	8.1%	0.045	0.014	94.0	24.6%
2	0.039	0.018	79.2	7.7%	0.039	0.014	95.4	23.4%
3	0.037	0.019	79.9	7.5%	0.038	0.014	96.0	22.8%
5	0.034	0.017	82.2	7.2%	0.035	0.014	97.3	22.1%
10	0.030	0.019	85.3	6.7%	0.030	0.013	99.5	19.8%
20	0.024	0.017	90.6	5.9%	0.021	0.012	102.7	17.2%
61	0.019	0.015	98.3	2.6%	0.019	0.014	106.3	9.8%
112	0.012	0.012	112.0	0.0%	0.012	0.012	112.0	0.0%

Monitoring Freq is 1 for fully sequential design and 112 when the only analysis is with all completed outcomes.

Note that the probability of efficacy stopping and Type 1 error rate increase monotonically for both specifications of follow-up length. The Type 1 error rate is lower for the 8-month follow-up design since there are more subjects in the final sample size.

5.0.2 Robustness of parameterizations of monitoring priors

Model "1" corresponds to default skeptical and enthuasic priors, "2" is concentrated skeptical and default enthuastic (the case considered in the main text), "3" is default skeptical and flattened enthaustic, and "4" is concentrated skeptical and flattened enthaustic.

Table 3: Sequential design properties using alternative monitoring priors

		0.4	0.4675	0.535	0.6025	0.67
1	EFF	0.073	0.418	0.869	0.993	1.000
1	FUT	0.755	0.255	0.026	0.001	0.000
1	INC	0.172	0.326	0.105	0.005	0.000
1	SS	71.0 + 6.3 = 77.3	79.3 + 5.1 = 84.4	56.9 + 7.0 = 63.9	34.4 + 8.0 = 42.4	23.0 + 8.0 = 31.0
2	EFF	0.039	0.313	0.805	0.988	1.000
2	FUT	0.765	0.252	0.026	0.001	0.000
2	INC	0.196	0.435	0.169	0.011	0.000
2	SS	72.9 + 6.2 = 79.1	87.5 + 4.1 = 91.7	67.1 + 6.3 = 73.5	41.4 + 7.8 = 49.3	27.7 + 8.0 = 35.7
3	EFF	0.071	0.415	0.871	0.994	1.000
3	FUT	0.811	0.314	0.036	0.001	0.000
3	INC	0.118	0.271	0.093	0.005	0.000
3	SS	66.5 + 6.8 = 73.4	76.6 + 5.5 = 82.1	56.3 + 7.1 = 63.3	34.5 + 8.0 = 42.5	23.0 + 8.0 = 31.0
4	EFF	0.040	0.313	0.803	0.987	1.000
4	FUT	0.815	0.321	0.038	0.002	0.000
4	INC	0.144	0.366	0.159	0.011	0.000
4	SS	68.8 + 6.7 = 75.5	84.1 + 4.7 = 88.9	66.5 + 6.5 = 72.9	41.9 + 7.9 = 49.8	27.7 + 8.0 = 35.7

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*