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

September 20, 2019

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 generating mechanism and prior distributions for unknown quantities. Let $\mathbf{D}$ be a random variable representing the data collected in the trial with density $p(\mathbf{D}|\theta, \psi)$ where θ and ψ are the unknown quantities. Let θ be the parameter of interest and ψ be the unknown quantities that are not of primary importance (nuisance parameters). Define the sample spaces for the unknown quantities as $\theta \in \Theta$ and $\psi \in \Psi$.

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 its marginal likelihood

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

where $p(\theta|\mathbf{D}) = \int_\Psi p(\theta, \psi|\mathbf{D})d\psi$ is marginalized over the nuisance parameters.

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 $\pi(\theta, \psi)$ for which their relation to $P(\theta \in \Theta_i|\pi(\theta, \psi))$ $i \in \{0, 1\}$ is examined. Note this quantity does not depend on the data $\mathbf{D}$ and therefore reflects a-priori opinion.

The posterior distribution of θ depends on the choice of prior distribution $\pi(\theta, \psi)$ since $p(\theta, \psi|\mathbf{D}) = p(\mathbf{D}|\theta, \psi)\pi(\theta, \psi)/p(\mathbf{D})$ by Bayes rule. 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 $\delta \in (0, 1)$ 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 $P(\theta \in \Theta_i|\mathbf{D}) \geq \delta$ for $i \in \{0, 1\}$. The quantity $1 - \delta$ 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 $\pi_E(\theta, \psi) \equiv \pi_E$ has the property that $P(\theta \in \Theta_1 | \pi_E) \geq \delta$ (equivalently $P(\theta \in \Theta_0 | \pi_E) < 1 - \delta$). The choice of $\delta \in (0, 1)$ 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 $\delta = 0.95$, 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 $\pi_S(\theta, \psi) \equiv \pi_S$ could have the property that $P(\theta \in \Theta_0 | \pi_S) \geq \delta$, in which case it is "all but convinced" that H_0 is true a-prior, 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 $P(\theta \in \Theta_A | \pi_S) < \delta$.

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 $P(\theta \in \Theta_A \subset \Theta_1 | \pi_S) < 1 - \delta$, to conclude that $P(\theta \in \Theta_1 | \pi_S) \geq \delta$.

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 $P(\theta \in \Theta_1 | \pi_E) \geq \delta$, to conclude that even a moderate positive effect is unlikely, that is, $P(\theta \in \Theta_I | \pi_E) < 1 - \delta$

Example

Consider the hypothesis $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$. The skeptic initially held that $P(\theta > \theta_A|\pi_S) < 1 - \delta$ and a promising interim result would be $P(\theta > \theta_0|\pi_S) \geq \delta$. The enthusiast initially held that $P(\theta > \theta_0|\pi_E) \geq \delta$ and a disillusioning interim result would be $P(\theta > \theta_I|\pi_E) < 1 - \delta$

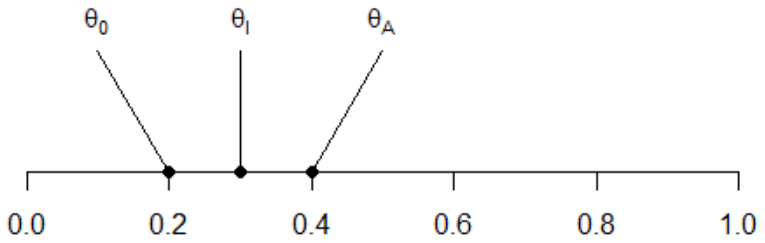

Figure 1: Graph of null response value (θ_0), intermediate response value (θ_I) and substantial response value (θ_A).

Probability of Success

As an alternative strategy to futility analysis, one can monitor the probability of success (POS) for the trial. The probability of getting a convincing result at the end of the trail can be computed using the interim data. Let $p(\theta|\mathbf{D}, \pi_I)$ denote the posterior distribution for θ based on the inference prior π_I and the current data $\mathbf{D}$. Let ξ denote the POS which is given as follows:

$$\begin{aligned}\xi &= P[\mathbf{D}_1 \in \mathbb{R}^{dim(\mathbf{D}_1)}|P(\theta \in \Theta_1|\mathbf{D}_1, \mathbf{D}, \pi_I) \geq \delta] \\ &= E[1\{P(\theta \in \Theta_1|\mathbf{D}_1, \mathbf{D}, \pi_I) \geq \delta\}]\end{aligned}$$

where the expectation is taken with respect to the posterior predictive distribution $p(\mathbf{D}_1)$ for future data $\mathbf{D}_1$ (which includes subjects yet to enroll):

$$p(\mathbf{D}_1) = \int p(\mathbf{D}_1|\theta) \cdot \pi(\theta|\mathbf{D}) d\theta.$$

One may stop the enrollment if ξ is sufficiently small (i.e. $\xi < 0.05$).

2.1.4 Final inference

Final inference on the parameter of interest is made once all data has been collected. Enrollment was either stopped based on a persuasive interim result or based on the maximum sample size. An inference prior $\pi_I(\theta, \psi) \equiv \pi_I$ is often less divisive than the skeptical and enthaustic priors, and can be viewed as a balance of the more divisive opinions. We propose use of a mixture prior constructed from the monitoring process as the inference prior:

$$\pi_I = \omega \cdot \pi_S + (1 - \omega) \cdot \pi_E$$

for $\omega \in [0,1]$. Choosing $\omega = 1/2$ for an equal mixture of π_S and π_E corresponds to an inference prior that equally weights the skeptical and enthuastic opinions. Define $p(\mathbf{D}|\pi(\theta,\psi)) = \int p(\mathbf{D}|\theta)\pi(\theta,\psi)d(\theta,\psi)$ to be the marginal likelihood for the data given the prior $\pi(\theta,\psi)$. Choosing ω based on posterior model probabilities of the null and alternative hypotheses yields $\omega = p(\mathbf{D}|\pi_S)/(p(\mathbf{D}|\pi_S) + p(\mathbf{D}|\pi_E))$.

All relevant information about θ can be derived from its marginal posterior distribution with an inference prior (e.g. posterior mean, credible intervals). For example, the posterior mean using the inference prior will be a two-part mixture of the posterior means using the skeptical and enthuastic priors:

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

3 Examples

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

3.1.1 Model formulation & prior elicitation

Consider a single-arm oncology 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. The formulation is discussed in Section 2.1.3 with $\theta_0 = 0.2$, $\theta_I = 0.3$, $\theta_A = 0.4$.

Beta priors for θ will be used to provide closed-form expressions of the posterior distributions via Beta-Binomial conjugacy. In particular, let y_1 be the number of successes and y_0 be the number of failures. If the skeptical prior is $\pi_S(\theta) \sim \mathcal{B}(\alpha_S, \beta_S)$ then the associated posterior is $p(\theta|\mathbf{D}, \pi_S) \sim \mathcal{B}(\alpha_S + y_1, \beta_S + y_0)$. Similarly, if the enthuastic prior is $\pi_E(\theta) \sim \mathcal{B}(\alpha_E, \beta_E)$ then the associated posterior is $p(\theta|\mathbf{D}, \pi_E) \sim \mathcal{B}(\alpha_E + y_1, \beta_E + y_0)$.

The skeptical prior is Beta distributed with expected value $\theta_0 = 0.2$ and has 0.045 prior probability that $\theta > \theta_A = 0.4$. The enthuastic prior has expected value $\theta_A = 0.4$ and has 0.05 prior probability that $\theta < \theta_0 = 0.2$. The quantities 0.045 and 0.05 in this setting represent the residual uncertainty that the skeptic and enthusiast have in their beliefs, and are shaded black in Figure 2. The inference prior will be at an equal mixture of the skeptical and enthuastic prior ($\omega = 0.5$).

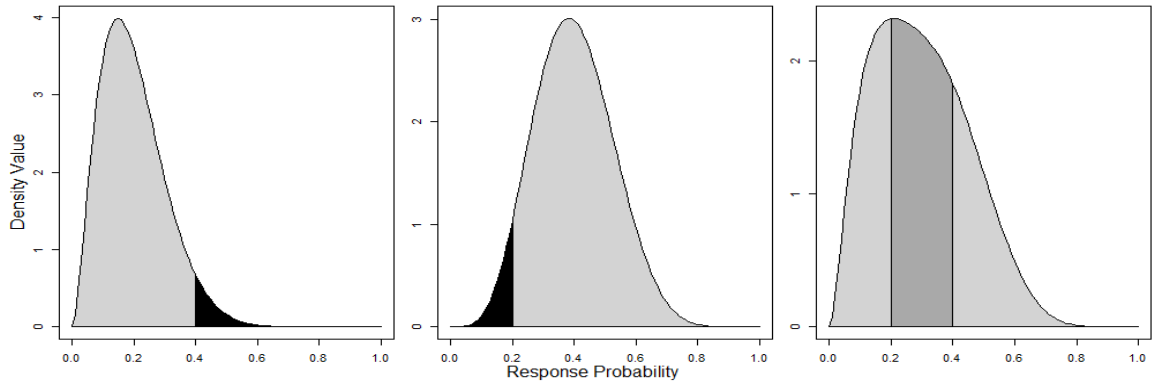

Figure 2: (a) Skeptical Beta prior (b) Enthuastic Beta prior (c) 50/50 mixture of skeptical and enthuastic Beta priors

3.1.2 Sequential monitoring

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

$$\text{Efficacy criteria (EFF): } P(\theta > 0.20|\mathbf{D}, \pi_S) \geq 0.95$$

$$\text{Futility criteria (FUT): } P(\theta \leq 0.30|\mathbf{D}, \pi_E) \geq 0.85$$

$$\text{Maximum sample size: } N = 76 \text{ patient outcomes obtained}$$

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.

3.1.3 Example paths

To demonstrate the monitoring procedure, two example trials are considered. As seen in Figure 3(a), at the second interim analysis the efficacy condition $P(\theta > 0.20|\mathbf{D}, \pi_S) \geq 0.95$ is satisfied and enrollment is terminated. As shown in Figure 3(b), at the fourth interim analysis the futility condition $P(\theta \leq 0.30|\mathbf{D}, \pi_E) \geq 0.85$ is satisfied and enrollment is terminated. In both examples there are subjects with missing outcomes, which are those who are still undergoing follow-up. The final analysis incorporates the final sample.

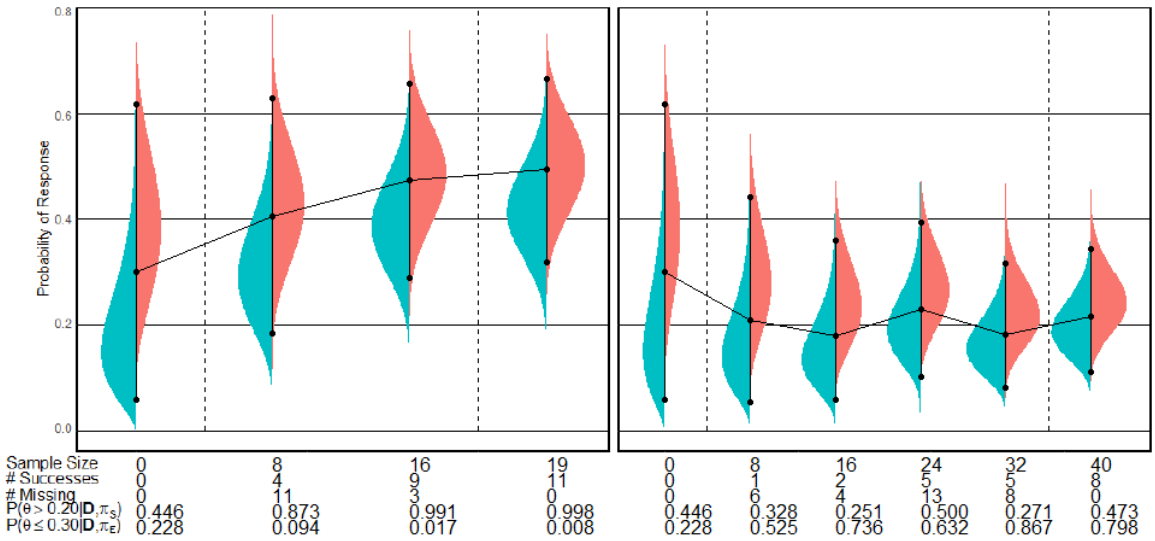

Figure 3: (a) Early stopping for efficacy (b) Early stopping for futility

3.1.4 Design properties: Results

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

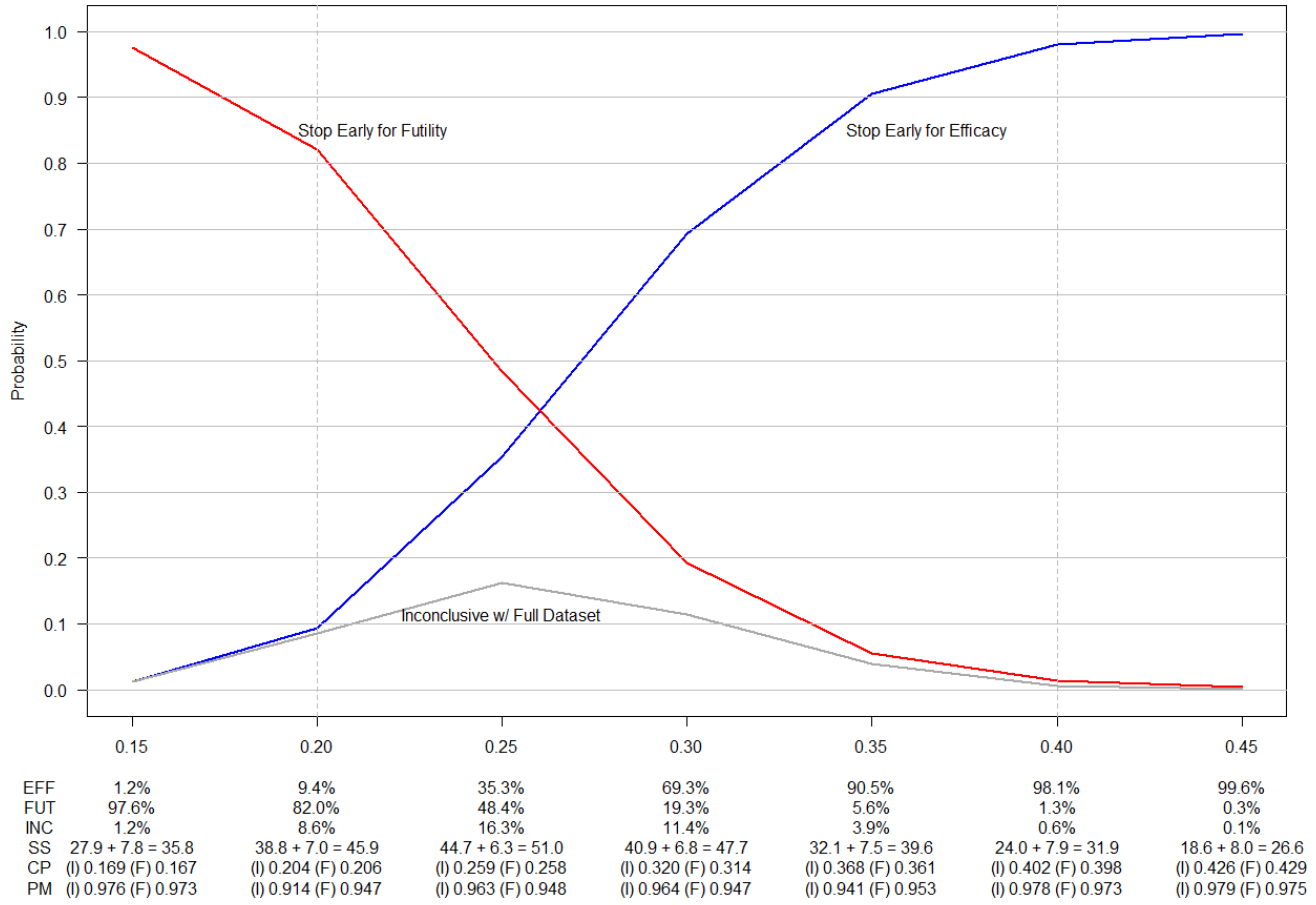

Figure 4: Sequential design properties of proof-of-activity trial

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.

3.1.5 Agreement between interim and final result

Table 1: Distribution of final posterior probability given interim stoppage (interim $P(\theta > 0.20|\mathbf{D}, \pi_S) \geq 0.95$) and evidence decrease.

	0.15	0.20	0.25	0.30	0.35	0.40	0.45
Final $P(\theta > 0.20|\mathbf{D}, \pi_S) \geq 0.95$	0.293	0.510	0.645	0.753	0.832	0.894	0.932
Final $P(\theta > 0.20|\mathbf{D}, \pi_S) < 0.95$	0.707	0.490	0.355	0.247	0.168	0.106	0.068
Conditional Median	0.91	0.92	0.92	0.93	0.93	0.93	0.93
Conditional 25th percentile	0.87	0.89	0.90	0.91	0.91	0.91	0.91
Conditional 10th percentile	0.83	0.86	0.87	0.88	0.88	0.88	0.88
Conditional 1st percentile	0.67	0.78	0.81	0.81	0.82	0.82	0.82

For example, at a true response rate of $\theta = 0.40$, there is an 89.4% that the threshold for a significant result is maintained after the additional subjects complete follow-up, and in the 10.6% of cases that the evidence decreases, the median posterior probability is 0.93 and only in 10% of cases is the posterior probability lower than 0.88. Thus there is a slight attenudation with respect to the dichotomous threshold, but little change in the posterior probability overall.

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. Even at the extreme case where an interim analysis is conducted after every outcome, the probability of stopping at the interim due to a promising result when the true response is at the null level is only 0.108 (about double the nominal rate), and even in this situation the Type 1 error rate once follow-up is complete does not exceed 0.05. Thus Bayesian sequential monitoring has good frequentist properties even with frequent interim analyses.

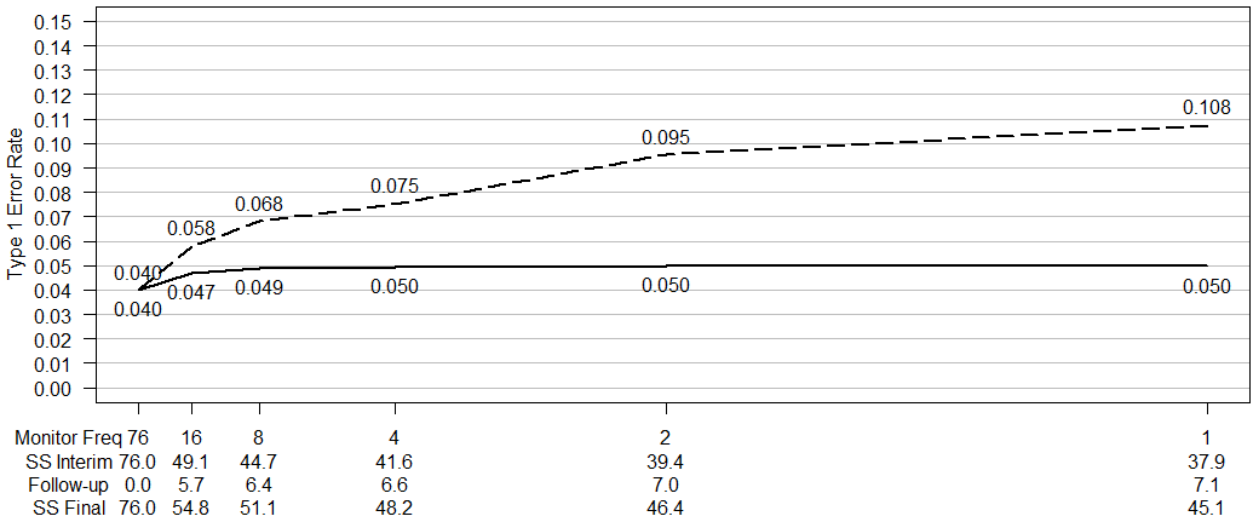

Figure 5: Type 1 error rate depending on frequency of sequential monitoring

3.1.7 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			
# Interim Analyses	EFF	T1E	SS Final	Ongoing	EFF	T1E	SS Final	Ongoing
1	0.108	0.05	45.1	15.7%	0.107	0.043	51.7	26.8%
2	0.095	0.05	46.4	15.1%	0.094	0.043	52.8	25.5%
4	0.075	0.05	48.2	13.7%	0.075	0.043	54.1	23.4%
8	0.068	0.049	51.1	12.5%	0.067	0.043	56.7	21.3%
16	0.058	0.047	54.8	10.4%	0.056	0.042	60.0	18.1%
76	0.04	0.04	76	0.0%	0.039	0.039	76	0.0%

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.

4 Robustness of parameterizations of monitoring priors

The generalized normal distribution is used to create priors with the same expected value and tail area as the default beta priors, but with densities that are either spike/slab around the expected value or flattened.

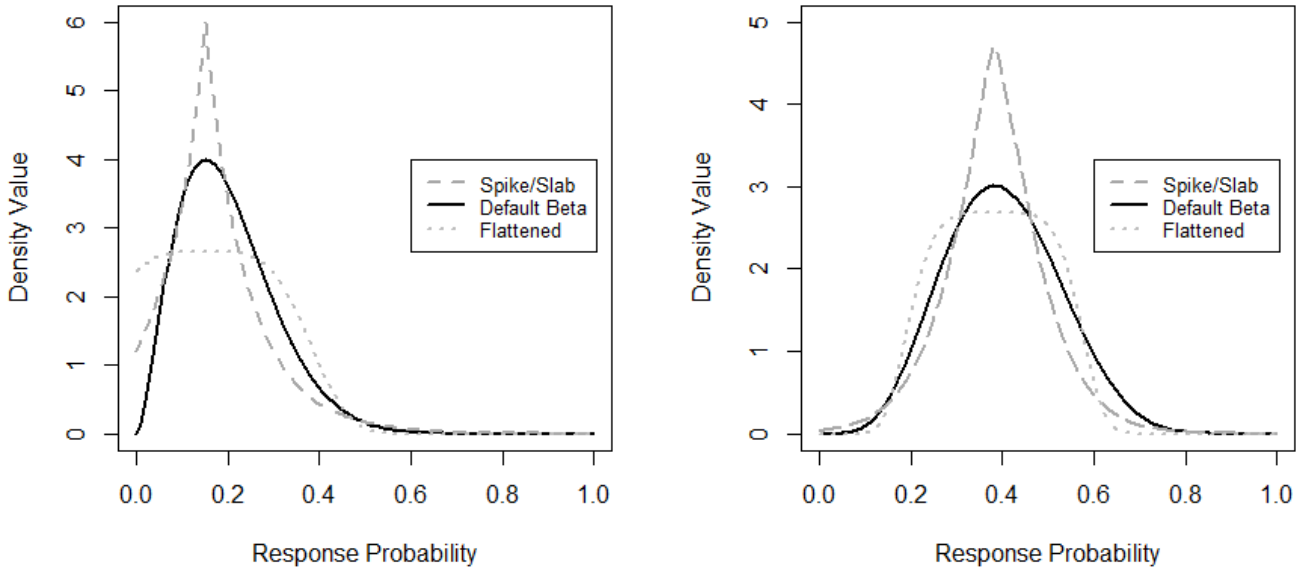

Figure 6: (a) Alternate skeptical priors (b) Alternate enthuastic priors

Table 3: Sequential design properties using alternative monitoring priors

		0.2	0.3	0.4
Default	EFF	0.094	0.693	0.981
	FUT	0.820	0.193	0.013
	INC	0.086	0.114	0.006
	SS	38.8 + 7.0 = 45.9	40.9 + 6.8 = 47.7	24.0 + 7.9 = 31.9
	PM	(I) 0.204 (F) 0.206	(I) 0.320 (F) 0.314	(I) 0.402 (F) 0.398
Spike/Slab	EFF	0.066	0.632	0.978
	FUT	0.738	0.127	0.006
	INC	0.195	0.241	0.016
	SS	46.5 + 6.1 = 52.6	47.7 + 5.8 = 53.5	26.8 + 7.8 = 34.6
	PM	(I) 0.196 (F) 0.199	(I) 0.322 (F) 0.317	(I) 0.400 (F) 0.397
Flattened	EFF	0.142	0.731	0.982
	FUT	0.820	0.22	0.016
	INC	0.039	0.048	0.002
	SS	34.6 + 7.5 = 42.2	34.2 + 7.5 = 41.7	20.6 + 7.9 = 28.5
	PM	(I) 0.211 (F) 0.212	(I) 0.319 (F) 0.315	(I) 0.390 (F) 0.391
SS Skeptical	EFF	0.066	0.620	0.975
	FUT	0.827	0.189	0.011
	INC	0.107	0.190	0.014
	SS	40.3 + 6.8 = 47.1	45.1 + 6.2 = 51.3	26.9 + 7.8 = 34.7
	PM	(I) 0.196 (F) 0.200	(I) 0.316 (F) 0.312	(I) 0.401 (F) 0.400

Use of a spike-slab skeptical prior drastically lowers the probability of terminating enrollment due to efficacy, from 9.4% in the default case to 6.6%. This is because there is less probability mass in the immediate area greater than the null value of 0.2. When using a flattened skeptical prior the probability of terminating enrollment due to efficacy increases to 14.2% since there is more probability mass in the immediate area greater than the null value of 0.2.

4.1 Parallel Two-Group Superiority Trial /w Continuous Binary Endpoint

Interesting because prior is on risk difference [-1,1] while also being non-informative on control group. Will need numerical integration to evalutate posteriors.

4.2 Three-Arm, Placebo Controlled Non-Inferiority Trial w/ Continuous Endpoint

$$P \to \beta_0 \text{ (placebo)}$$
$$C \to \beta_0 + \beta_1 \text{ (control)}$$
$$A \to \beta_0 + \beta_1 + \beta_2 \text{ (active)}$$
$$H_0 : \beta_2 - \delta\beta_1 \leq 0$$

Parameters of interest (β_1, β_2), nuisance parameters (β_0, σ^2).

Need priors $\pi(\beta_0), \pi(\beta_1), \pi(\beta_2|\beta_1)$.

Will use MCMC to evaluate posteriors.

5 Discussion – (MATT/EVAN)

Q: Why not reverse engineer priors to have exact Type 1 error properties?

A: This would basically be a frequentist method, in that the design would have to be adhered to exactly (including number and timing of data monitoring). Philosophically, designing a Bayesian study that requires rigid monitoring rules loses the advantages of Bayes from the likelihood principle.

SUPPLEMENTARY MATERIAL

6 Beta Priors

Beta priors for θ will be used to provide closed-form expressions of the posterior distributions via Beta-Binomial conjugacy (the posterior distribution $p(\theta|\mathbf{D})$ will be Beta distributed). The Beta distribution has two shape parameters. These parameters can be determined uniquely by specifying the desired mean and variance of the distribution. The variance for the skeptical and enthuastic priors is then uniquely determined through by the choice of threshold δ. In particular, let $\pi_S(\theta) \sim \mathcal{B}(\alpha, \beta)$ be Beta distributed with shape parameters (α, β). There is a single choice of (α, β) such that:

$$\theta_0 = E(\pi_S) = \int_\Theta \pi_S(\theta)d\theta = \frac{\alpha}{\alpha+\beta} \text{ and } \delta = \int_{\Theta_0} \pi_S(\theta)d\theta = \int_0^{\theta_0} \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}d\theta$$

where $B(\alpha, \beta)$ is the Beta function.

Alternatively, the variance could be determined by specifying a desired quantile of the prior distribution which would then be reflected in δ. Then there is a single choice of (α, β) such that

$$\theta_0 = E(\pi_S) = \int_\Theta \pi_S(\theta)d\theta = \frac{\alpha}{\alpha+\beta} \text{ and } \lambda = \int_{\theta_A}^1 \pi_S(\theta)d\theta = \int_{\theta_A}^1 \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}d\theta,$$

in which case $\delta = \int_{\Theta_0} \pi_S(\theta)d\theta$ is a deterministic quantity.

7 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*