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

August 22, 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

As you introduce ideas that come from or extend other ideas in the literature, cite the relevant literature.

2.1 Monitoring versus Estimation Priors (EVAN – draft on 6/21)

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 (i.e. "nuisance parameters"). Define the sample spaces for the unknown quantities as $\theta \in \Theta$ and $\psi \in \Psi$.

Suppose the hypotheses under consideration are $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$. These hypotheses are adjudicated based on posterior probabilities of θ by evaluating its marginal likelihood

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

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

2.1.2 Compelling level of evidence & prior elicitation

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.

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 consider the impact of interim analyses on subject enrollment, with the potential for early termination of enrollment.

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. A skeptical prior is an informative or subjective prior that gives substantial preference to H_0 such that it

is "all but convinced" that H_0 is true a-priori. This prior $\pi_S(\theta, \psi) \equiv \pi_S$ has the property that $P(\theta \in \Theta_0 | \pi_S) \geq \delta$ (equivalently $P(\theta \in \Theta_1 | \pi_S) < 1 - \delta$). 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 skeptical prior places 95% prior probability that $\theta \in \Theta_0$. An enthuastic prior $\pi_E(\theta, \psi) \equiv \pi_E$ similarly gives preference to H_1 through the property that $P(\theta \in \Theta_1 | \pi_E) \geq \delta$ (equivalently $P(\theta \in \Theta_0 | \pi_E) < 1 - \delta$). For purposes of interpretation, a *skeptical person* is someone whose a-priori opinions of θ are reflected through a skeptical prior and a *enthuastic person* is someone whose a-priori opinions of θ are reflected through an ethuastic prior. The prior distributions discussed are generally "non-informative" over the nuisance parameters.

2.1.3 Sequential monitoring & final inference

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 disilusioned by the data to stop further experimentation" (Spiegelhalter et al. (1994)).

Early termination of the trial 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). It is then reasonable to stop data collection if, upon seeing the data, a *skeptical person* changes their opinion to be "all but convinced" that H_1 is true ($P(\theta \in \Theta_1 | \mathbf{D}, \pi_S) \geq \delta$), or an *enthuastic person* becomes "all but convinced" that H_1 is false ($P(\theta \in \Theta_0 | \mathbf{D}, \pi_E) \geq \delta$).

Final inference on θ is made once enrollment is stopped based on the monitoring priors or at the planned end of the trial. An inference prior $\pi_I(\theta, \psi) \equiv \pi_I$ is often non-informative or objective in the sense that it does not show a-priori preference to H_0 or H_1 ($P(\theta \in \Theta_0|\pi_I) \approx P(\theta \in \Theta_1|\pi_I)$). There are many ways to formulate an inference prior with this property. 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 is impartial to H_0 and H_1, and is a practical choice of π_I is to be determined before the start of data collection. 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 $\omega = p(\mathbf{D}|\pi_S)/(p(\mathbf{D}|\pi_S) + p(\mathbf{D}|\pi_E))$.

The determination of a significant trial result is given by

$$P(\theta \in \Theta_1|\mathbf{D}, \pi_I) \geq \delta.$$

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:

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

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. $\xi < 0.05$).

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.

Consider testing the hypothesis $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Consider a highly clinically relevant treatment effect $\theta_A > \theta_0$. A design property of the trial will be the posterior probability of $\theta \geq \theta_0$ given $\theta = \theta_A$ (similar to frequentist power).

Monitoring priors for this trial will be made using the concepts of a skeptical prior and an enthuastic prior. Recall a skeptic is "all but convinced" that H_0 is true a-priori, therefore $P(\theta \in \Theta_0 | \pi_S) \geq \delta$. An optimist is "all but convinced" that H_1 is true a-priori, therefore $P(\theta \in \Theta_1 | \pi_E) \geq \delta$.

It is intuitive to center the skeptical and enthuastic priors around the quantities θ_0 and θ_A respectively, so that $E(\pi_S) = \theta_0$ and $E(\pi_E) = \theta_A$. For example, suppose it is desirable that the skeptical prior places small probability $\lambda > 0$ that the $\theta \geq \theta_A$, which is the highly clinically relevant treatment effect.

Assume that the outcomes are ascertained after approximately 4 months of follow-up and 2 patients per month on average are enrolled. The study may stop enrollment at a given point in time, but all enrolled patients will be followed for outcome ascertainment.

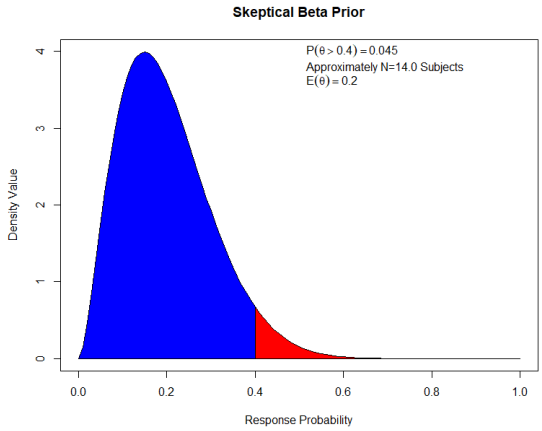
Beta priors for θ will be used to provide closed-form expressions of the posterior distributions via Beta-Binomial conjugacy (the posterior distribution $p(\theta|\mathbf{D})$ will be Beta distributed).

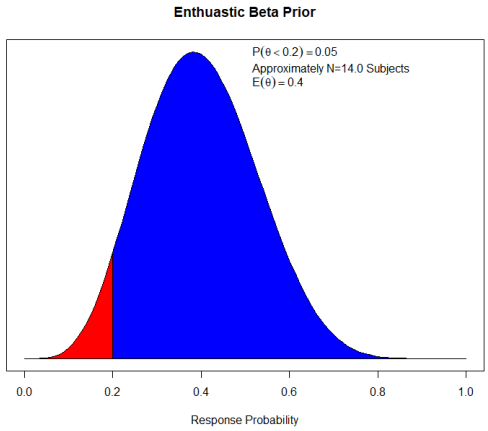
Consider the hypotheses

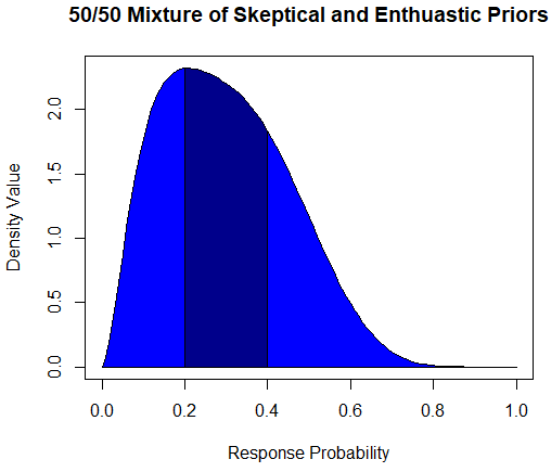
$$H_0 : \theta \leq 0.15$$

$$H_1 : \theta > 0.15$$

The skeptical prior will be Beta distributed, centered around $\theta_0 = 0.15$, and have 2.5% prior probability that $\theta > \theta_A$. Similarly, the enthuastic prior will be centered around $\theta_A = 0.45$ and have 2.5% prior probability that $\theta < \theta_0$.


Skeptical Beta Prior
P(θ > 0.4) = 0.045
Approximately N=14.0 Subjects
E(θ) = 0.2
Density Value
Response Probability


Enthuastic Beta Prior
P(θ < 0.2) = 0.05
Approximately N=14.0 Subjects
E(θ) = 0.4
Response Probability


50/50 Mixture of Skeptical and Enthuastic Priors
Density Value
Response Probability

3.1.2 Sequential monitoring

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

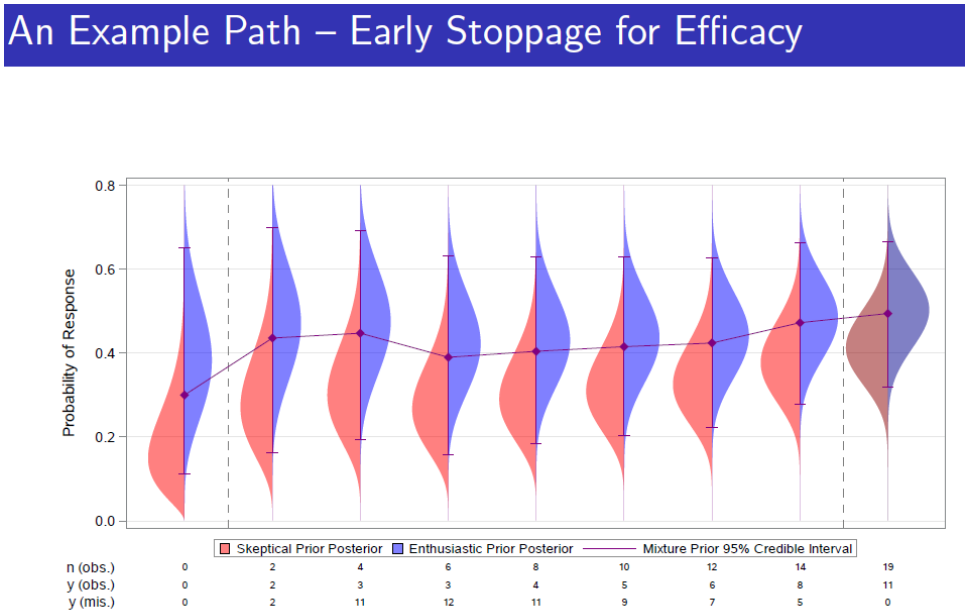
$$\begin{aligned}
\text{Efficacy criteria: } & P(\theta > 0.20|\mathbf{D}, \pi_S) \geq 0.95 \\
\text{Futility criteria: } & P(\theta \leq 0.30|\mathbf{D}, \pi_E) \geq 0.85 \\
\text{Exhausted resources: } & N = 76 \text{ patient outcomes obtained}
\end{aligned}$$

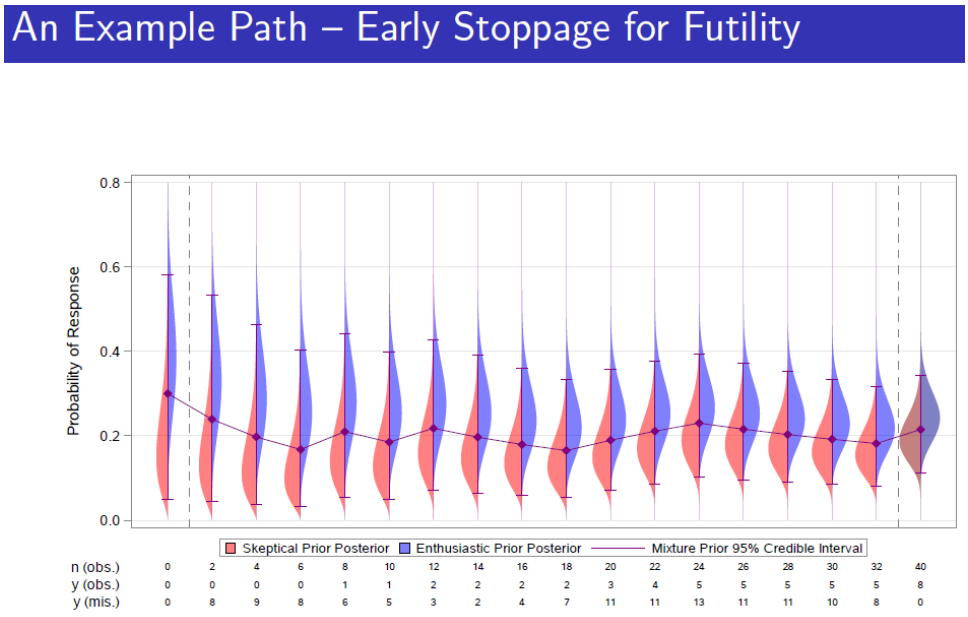
Let y_1 be the number of successes and y_0 be the number of failures.

$$\begin{aligned}
\text{Skeptical prior: } & \pi_S(\theta) \sim \mathcal{B}(\alpha_S, \beta_S) \\
\text{Skeptical posterior: } & p(\theta|\mathbf{D}, \pi_S) \sim \mathcal{B}(\alpha_S + y_1, \beta_S + y_0) \\
\text{Efficacy stopping: } & P(\theta \in \Theta_1|\mathbf{D}, \pi_S) \geq \delta
\end{aligned}$$

$$\begin{aligned}
\text{Enthuastic prior: } & \pi_E(\theta) \sim \mathcal{B}(\alpha_E, \beta_E) \\
\text{Enthuastic posterior: } & p(\theta|\mathbf{D}, \pi_E) \sim \mathcal{B}(\alpha_E + y_1, \beta_E + y_0) \\
\text{Futility stopping: } & P(\theta \in \Theta_0|\mathbf{D}, \pi_E) \geq \delta)
\end{aligned}$$

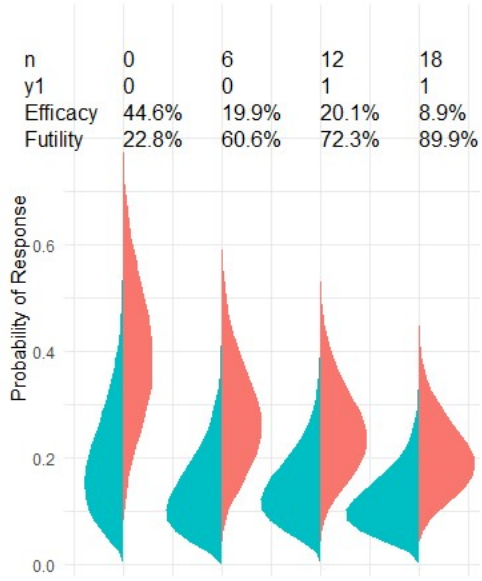
3.1.3 Example paths (violin plots)



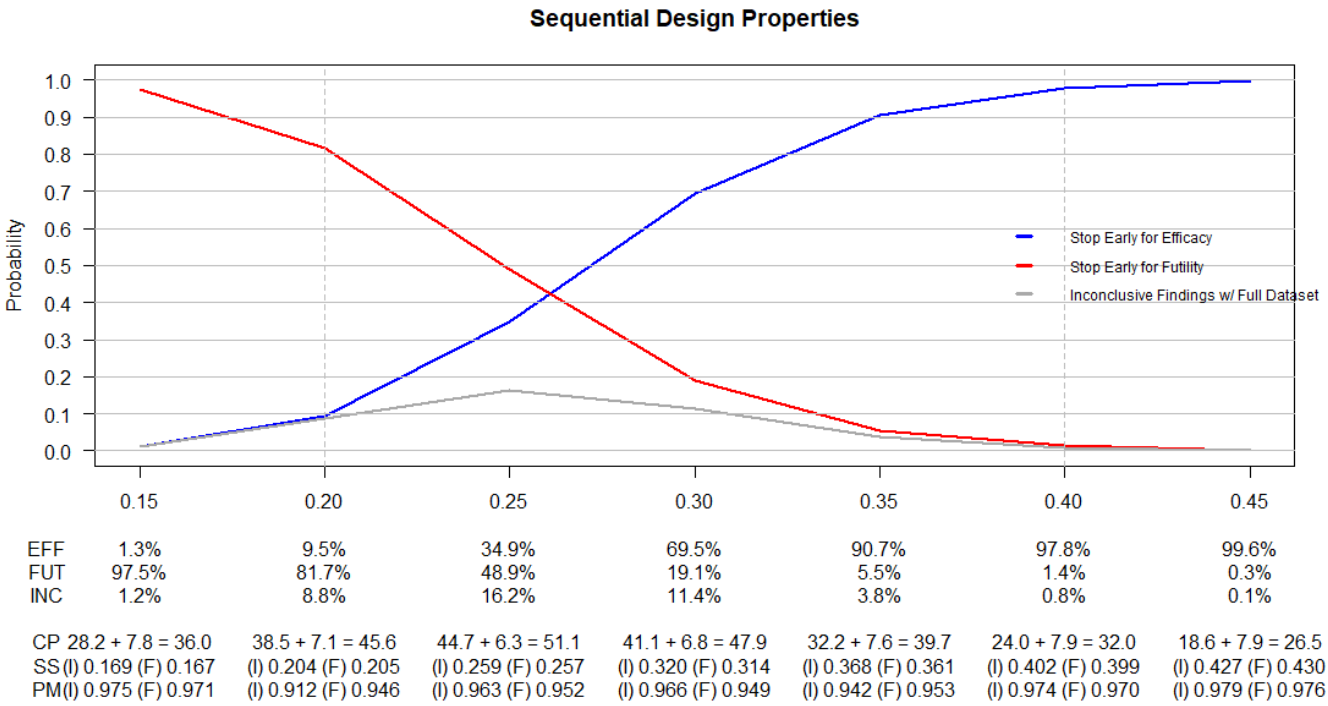


An Example Path - Early Stoppage for Efficacy

An Example Path - Early Stoppage for Futility



3.1.4 Design properties: Results



- EFF is the probability of the trial stopping early for efficacy.
- FUT is the probability of the trial stopping early for futility.
- INC is the probability of reaching the maximum sample size without a conclusive monitoring result.
- CP is the coverage probability using the mixture prior.
- SS is the average sample size at the interim analysis and at the end of follow-up. Note there are approximately 8 subjects undergoing follow-up when enrollment is

terminated.

- PM is the posterior mean using the mixture prior.

3.1.5 Design properties: Slow vs. fast accrual

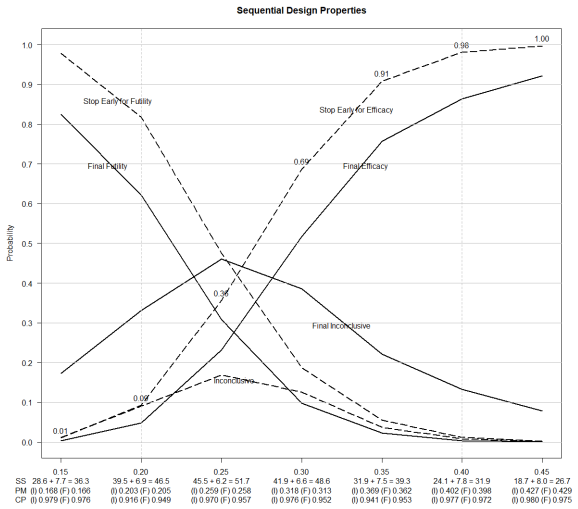
3.1.6 Type 1 error rate by the frequency of data monitoring

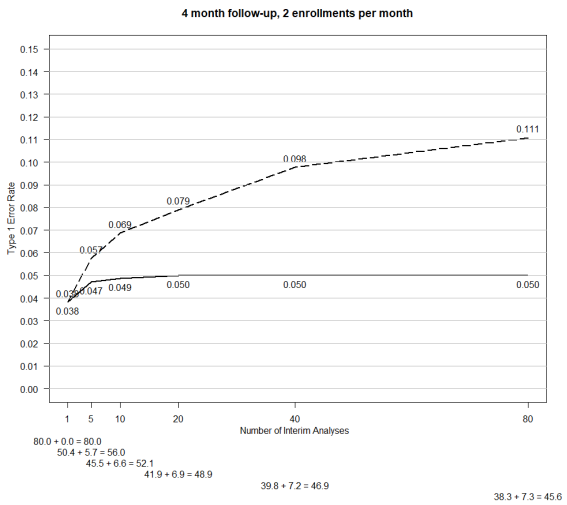
Figures needed:

1. Sequential design properties - slow vs. fast accrual (average sample size, posterior mean, coverage probabilities, distribution of final posterior probability).

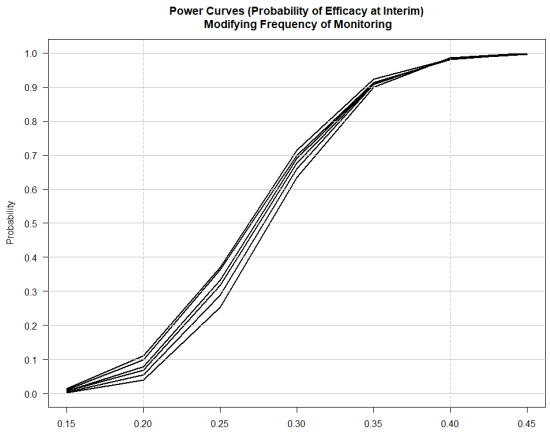
Look at varying the types of priors while keeping the tail areas the same. More spiked are OBF, less spiked are Pocock. Vary decay rate will affect bias not Type I/II error. Not dissimilar to Frequentist: flatter-Pocock, mass at null-OBF.

- Use correct terminology: discuss the probability of stopping early vs. probability we conclude the treatment works.
- The two sides of the discussion: first is what happens during the trial regarding sequential monitoring, such as % of time stopping early vs. trial done to completion and expected sample size. Second is the final determination of efficacy or futility and how that relates to Type 1 Error and power.
- Make the point that regardless of frequency of monitoring there are good Type 1 error rates. Remember the best case for sequential monitoring is slow enrollment relative to outcome ascertainment. Slow enrollment means there is a benefit to ending trial early and reach a conclusion faster. Outcome ascertainment needs to be somewhat fast to ensure a good # of outcomes are generated.
- Consider hyperlinking formulas.





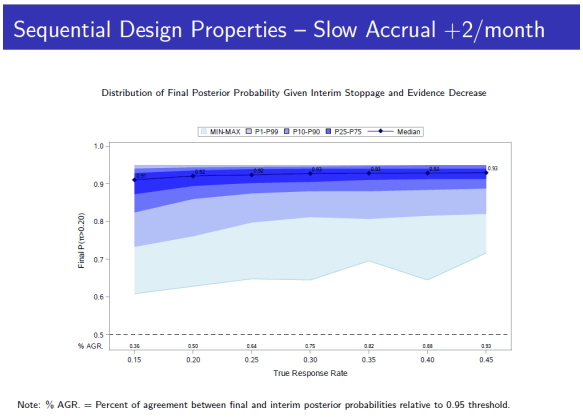
- Want to highlight that the ultimate inference will have no Type 1 error inflation. At this point the ultimate inference for efficacy is still made with skeptical prior.
- Label lines nicely.
- "Only bad thing to do is to stop learning"
- Enrollment rate and % of ongoing data as operating characteristics are interesting ideas, but focus on the plots already created.
- Mat: Scaling on # of subjects in each interim analysis rather than # of interim analyses (e.g. flip X axis). Make labels go vertical or diagonal.



Looking at spike/slab and flat priors...

- Test the hypothesis that power curves will be identical and what will change is the expected sample sizes.

Include details for how Bayesian monitoring has good frequentist properties even with frequent interim analyses.



- Plot is conditional on evidence going down. For example, at $\theta = 0.45$ there is a 93% chance of agreement, which means there is evidence decrease only 7% of the time. At $\theta = 0.15$ there is evidence decrease 64% of the time. Note that minimum is very sensitive to the number of simulations and should be removed from the plot.

- Note that there is functionally a minimum sample size already that is the # of subjects needed to overwhelm the skeptical prior. Actually compute the # (perhaps after 5 consecutive successes).

- There is a drop in the final power which is undesirable, but the drop in type 1 error is desirable. These move in tandem.

- Move discussion away from being scared of evidence decrease. There is an attenuation with respect to dichotomous threshold, but not overall. In this sense % agreement may be misleading. Look at the % of change in the evidence rather than the % change in the agreement based on a dichotomous agreement. The methods section may need to be modified to reduce focus on dichotmous decision. However, don't abandon hypothesis testing, but look at it holistically. Offer different perspectives on attenuation in the discussion. After all, 0.93 posterior probability isn't that different from 0.95.

3.2 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.

3.3 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.

4 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.

Meeting Notes 7/19/19

- For slides 22/23, it should say nmis not ymis.
- Violin plots are particular examples.

- Run replicates of trial to get frequentist properties.
- Poisson process enrollment, λ=rate parameter, $1/\lambda$=monthly enrollment. Consider the outcome ascertainment length in months.
- Interim/final results are consistent if quick/high enrollment relative to outcome ascertainment (think of the extreme example of instant ascertainment). Iterim/final results could be inconsistent if slow/low enrollment relative to outcome ascertainment.
- The efficacy criteria (proving the null is not true (showing skeptical prior based centered H_0 is now convinced)) is usually not changed in simulations.
- The futility criteria takes two forms: use futility prior based on intermediate value (between H_0 and H_1) if the intermediate value has clinical significant, otherwise use probability of success (POS). Pick one method to show in the paper, and show the other way is equivalent in the supplement.
- *Major point: Regardless of the frequency of monitoring there are good type 1 error rates.*
- Recall the best case for sequential monitoring is slow enrollment relative to outcome ascertainment (outcome ascertainment is quick to provide consistency with interim and final results, enrollment is slow so there is substantial pragmatic benefit to ending trial early if possible).
- Create panel graph, with Slow, Medium, and High enrollment. Slow=1 subject per month, instantanous outcome ascertainment. Medium=2 subjects per month,

1 month for outcome ascertainment. High=2 subjects per month, 2 months for outcome ascertainment. Show interim and final probability of rejecting H_0. Show how the type 1 error changes in each scenario. There is no gap in Slow case, and gap is highest in High case.

- Main ideas

    1. Show how the Bayesian monitoring can be done with examples based on violin plot idea. These examples will have variable enrollment rates to show agreement of interim/final.
    2. Show frequentist properties of sequential monitoring Bayesian trial when enrollment is slow relative to outcome ascertainment.
    3. Show how the frequentist Type 1 error rate inflates as a function of the # of times the data are monitored.
    4. Explore alternative distributions for skeptical and enthuastic priors and discuss relationships to OBF and Pocock α-spending. The alternative distributions could be mixtures of Betas.

SUPPLEMENTARY MATERIAL

5 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:

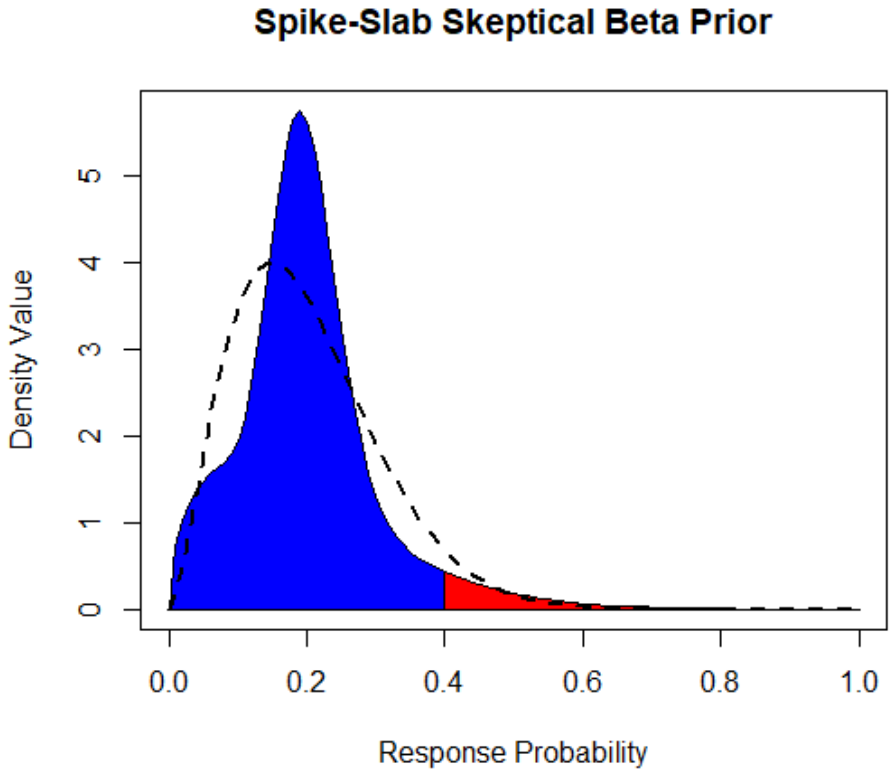
$$\theta_0 = E(\pi_S) = \int_\Theta \pi_S(\theta)d\theta = \frac{\alpha}{\alpha+\beta} \text{ and } \delta = \int_{\Theta_0} \pi_S(\theta)d\theta = \int_0^{\theta_0} \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}d\theta$$

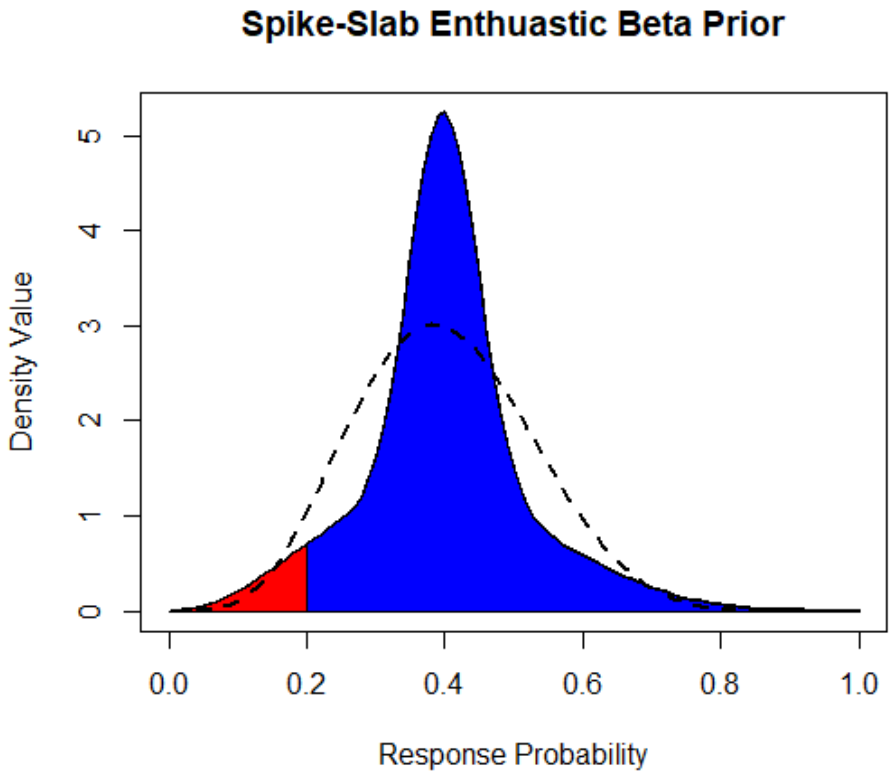
where $B(\alpha, \beta)$ is the Beta function.

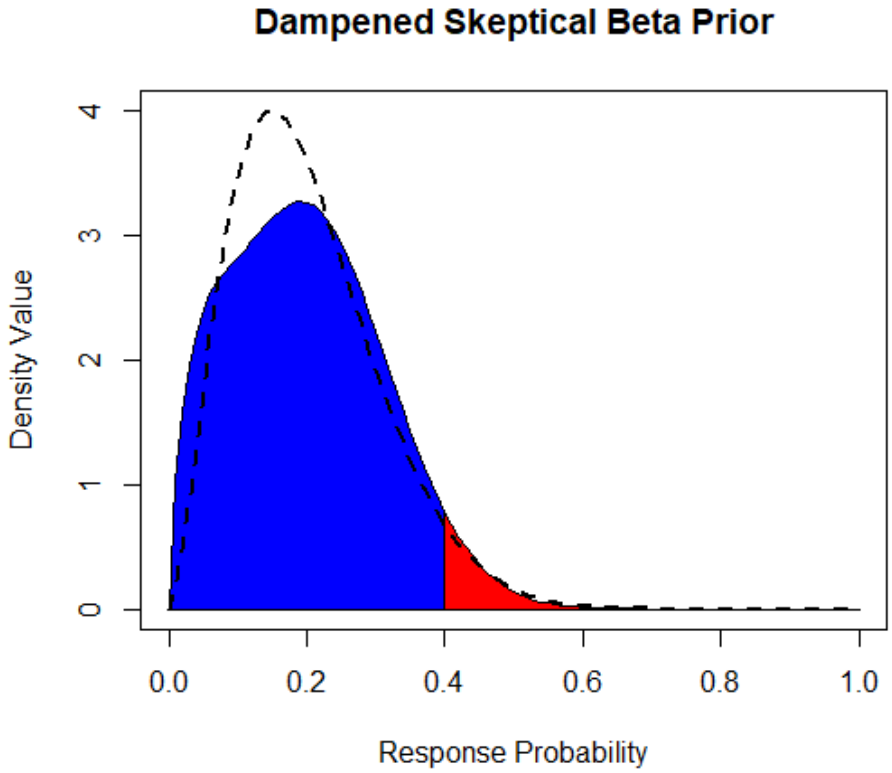
Alternatively, the variance could be determined by specifying a desired quantile of the prior distribution which would then be reflected in δ. Then there is a single choice of (α, β) such that

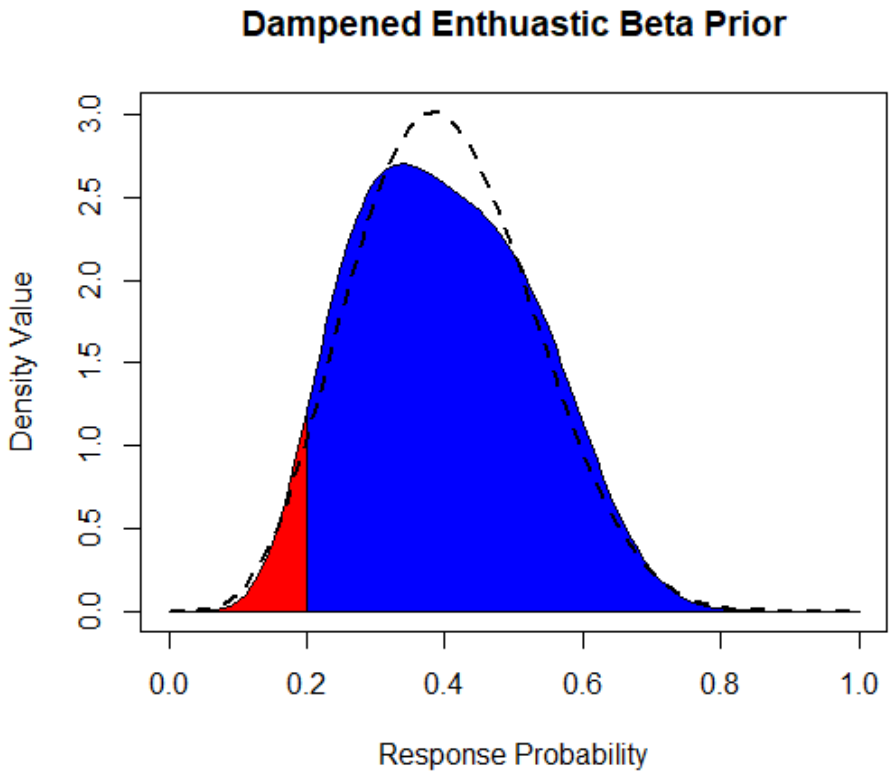
$$\theta_0 = E(\pi_S) = \int_\Theta \pi_S(\theta)d\theta = \frac{\alpha}{\alpha+\beta} \text{ and } \lambda = \int_{\theta_A}^1 \pi_S(\theta)d\theta = \int_{\theta_A}^1 \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}d\theta,$$

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


Spike-Slab Skeptical Beta Prior
Density Value
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Response Probability


Spike-Slab Enthuastic Beta Prior
Density Value
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Response Probability


Dampened Skeptical Beta Prior
Density Value
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Response Probability



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%*
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*

Hyman, D. M., Puzanov, I., Subbiah, V., Faris, J. E., Chau, I., Blay, J.-Y., Wolf, J., Raje, N. S., Diamond, E. L., Hollebecque, A., Gervais, R., Elez-Fernandez, M. E., Italiano, A., Hofheinz, R.-D., Hidalgo, M., Chan, E., Schuler, M., Lasserre, S. F., Makrutzki, M., Sirzen, F., Veronese, M. L., Tabernero, J. & Baselga, J. (2015), 'Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations', *New England Journal of Medicine* **373**(8), 726–736.
URL: *http://www.nejm.org/doi/10.1056/NEJMoa1502309*

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*