Bayesian Sequential Monitoring for Pediatric Clinical Trials

Matthew A. Psioda

Department of Biostatistics

The University of North Carolina at Chapel Hill

September 24, 2019

ASA Biopharmaceutical Section

Regulatory-Industry Statistics Workshop

Outline

Why Bayes and Why Sequential Monitoring?

Consistent Interpretation of Evidence

- Evidence in favor of a hypothesis at *every* point in time can be evaluated against a common standard.

- Because Bayesian inference obeys the likelihood principle, it does not depend on the stopping rule for data collection or how many times the data have been analyzed before.

  "*It is entirely appropriate to collect data until a point has been proven or disproven, or until the data collector runs out of time, money, or patience.*" – Edwards et al. (1963)

Intuitiveness of Monitoring Criteria

- Summarizing Spiegelhalter et al. (1993) and others: A Bayesian may monitor data continually and stop collection when any of the following criteria have been met:
  - A sufficiently *skeptical person* is convinced H_1 is true.
  - A sufficiently *enthusiastic person* is convinced H_1 is false or that the benefit of treatment is not likely to be what was expected.
  - The probability of *eventually* proving that H_1 is true is sufficiently low.
  - The resources allocated have been exhausted.

Common Criticisms Levied Against Bayesian Methods

- Bayesian inference depends on the prior. How does one pick the prior?
  - Most Bayesians advocate performing analyses using multiple priors instead of simply picking one with the possible exception being when the prior is informed by concrete, objective information (i.e., data).

    Much work has been done in this area. See Berger and Berliner (1986), Greenhouse and Waserman (1995), Carlin and Sargent (1996), and Spiegelhalter et al. (1993) for examples.

- Repeated analysis of the data with a goal of early stoppage for efficacy results in an inflated type I error rate. In extreme cases sequential monitoring can lead to *sampling to a foregone conclusion*.
  - Sampling to a foregone conclusion is a theoretical concern, not a practical one.

    See Spiegelhalter et al. (1993) for some discussion that traces back to Cornfield (1966).

Motivating Example: Pediatric Ulcerative Colitis Trial

Motivating Example: Pediatric Ulcerative Colitis Trial

- Consider an example application based on the drug infliximab, which is FDA approved for the treatment of several diseases, including ulcerative colitis (UC).
- Prior information $\rightarrow$ UC trials in adults – ACT1 and ACT2 trials.

Study	Placebo	**5mg infliximab**	10mg infliximab
ACT1	121 (37.2%)	**121 (69.4%)**	122 (61.5%)
ACT2	123 (29.3%)	**121 (64.5%)**	120 (69.2%)
Total	244 (33.2%)	**242 (67.0%)**	242 (65.3%)

- Infliximab was subsequently investigated in the T72 pediatric UC trial using the 5mg dose and same response rate primary endpoint.
- A response rate equal to 73.3% (N=60) was observed.

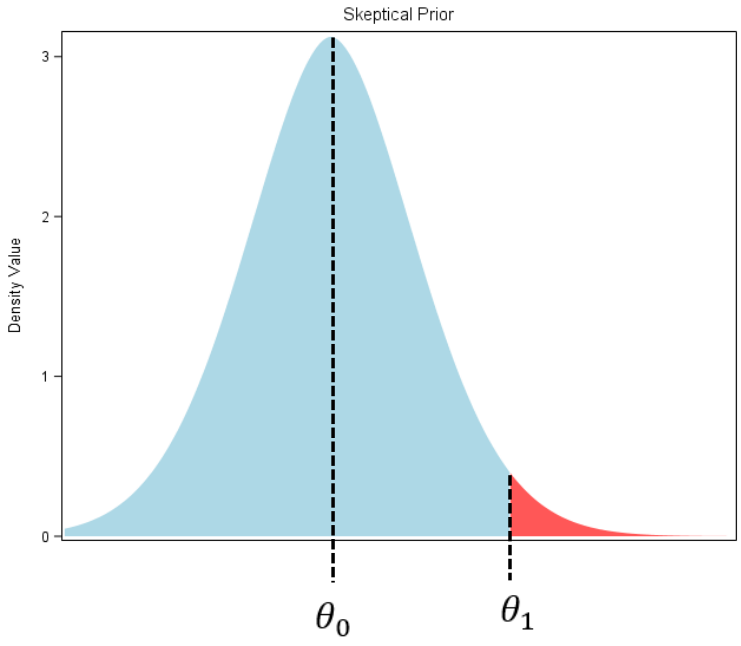
Priors for Sequential Monitoring

Defining Substantial Evidence

- Consider testing the hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ where θ is a treatment effect of interest.
- Suppose an effect $\theta_1 > \theta_0$ is thought to be highly clinically relevant and plausible given data from adult trials.
- Consider a clinical trial with a single analysis and fixed sample size chosen so that there is high probability of proving H_1 when $\theta = \theta_1$.
  - Bayesian decision rule $\rightarrow$ Reject H_0 when $P\left(\theta > \theta_0 \middle| \mathbf{D}\right) \geq 0.975$.
  - Type I error rate ≈ 0.025 if $\theta = \theta_0$ for non-informative prior.
  - Substantial evidence $\leftrightarrow P\left(\theta > \theta_0 \middle| \mathbf{D}\right) \geq 0.975$.

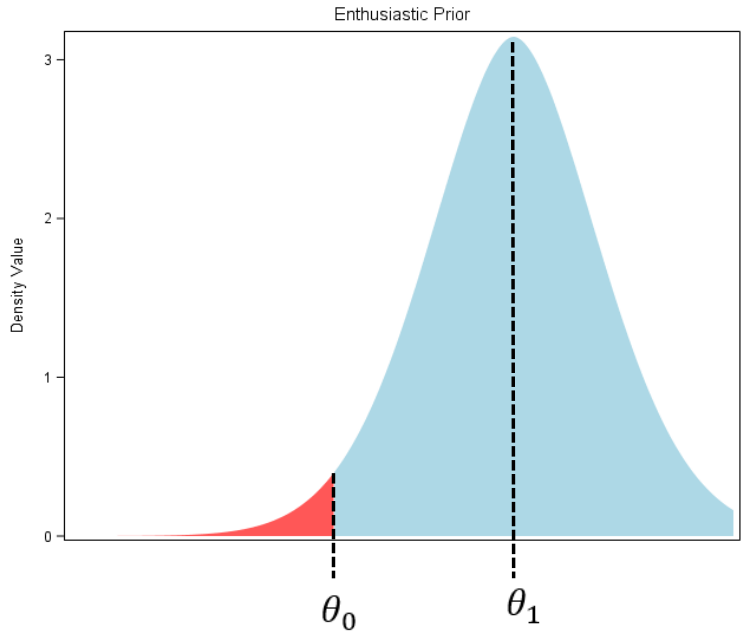
Skeptical/Enthusiastic Priors

- If one were to observe $P\left(\theta > \theta_0 \middle| \mathbf{D}\right) = 0.975$, you might say they would be **all but convinced** H_1 is true.
- Skeptic $\rightarrow$ someone all but convinced that $\theta < \theta_1$ and who believes $\theta = \theta_0$ is most likely
- Skeptical prior $\rightarrow \pi_0^{(s)}(\theta)$



Skeptical/Enthusiastic Priors

- Enthusiastic person $\rightarrow$ someone all but convinced that $\theta > \theta_0$ and who believes $\theta = \theta_1$ is most likely
- Enthusiastic prior $\rightarrow \pi_0^{(e)}(\theta)$



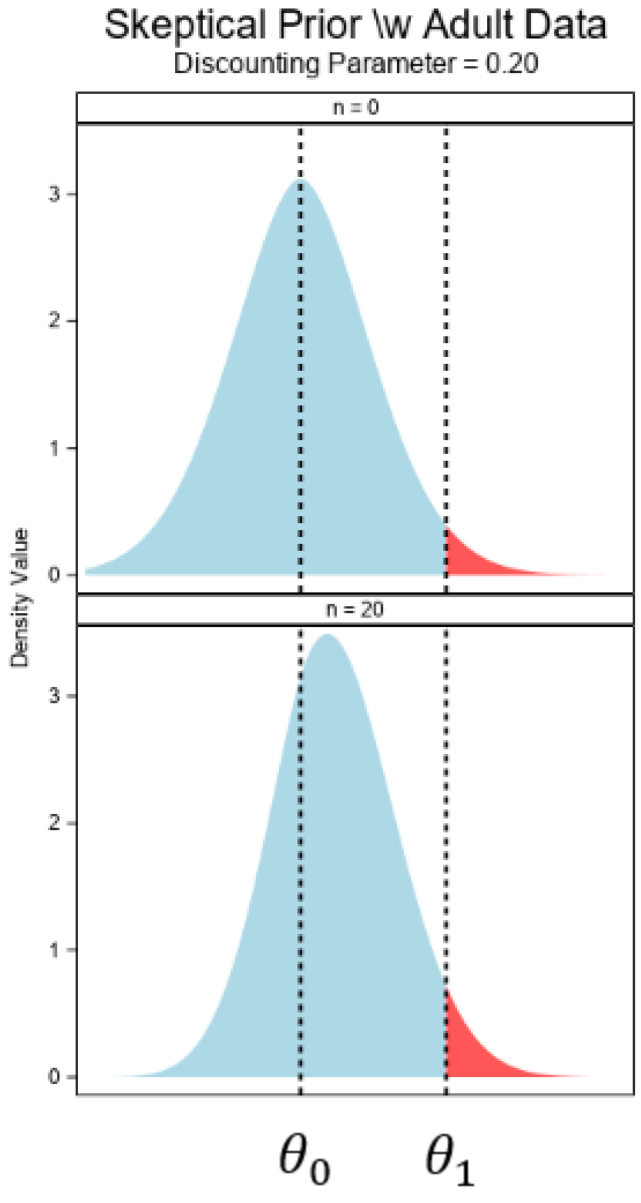
Skeptical Power Prior Incorporating Adult Data

- Let $a_0 \in [0, 1]$ be a scalar that discounts the adult data.
- Adult-data informed skeptical power prior is

$$\pi_{a_0}^{(s)}(\theta) \propto \left[\mathcal{L}\left(\theta \middle| \mathbf{D}_a\right)\right]^{\omega_0} \pi_0^{(s)}(\theta),$$

where

  - $\mathbf{D}_a$ is the adult data,
  - $\omega_0 = a_0 \times \min\left(1, {}^{n_p}/_{n_a}\right)$, and
  - n_p and n_a are the sample sizes for the pediatric trial (currently) and adult trial.



Using Prior Information: Applicability & Compatibility

- Applicability $\rightarrow$ elicited by expert opinion during design $\rightarrow$ a_0 fixed
- Compatibility $\rightarrow$ adult data used in accordance with compatibility with pediatric data $\rightarrow$ a_0 changes as pediatric data accumulate
  - **Bayesian p-value $\rightarrow$ How extreme are the pediatric data with respect to their prior predictive distribution given the adult data? (Nott et al. (2018), Evans and Moshonov (2006))**

A Bayesian p-value – ψ

- $\pi\left(\theta\middle|\mathbf{D}_{\mathsf{a}}\right) \propto \mathcal{L}\left(\theta\middle|\mathbf{D}_{\mathsf{a}}\right)\pi_0\left(\theta\right) \rightarrow$ posterior distribution for θ based on the adult data where $\pi_0\left(\theta\right)$ is some initial, non-informative prior.
- Prior predictive distribution for the marginal likelihood for pediatric data:

$$m_{\mathsf{pred}} = p\left(\mathbf{D}_{\mathsf{pred}}\middle|\mathbf{D}_{\mathsf{a}}\right) = \int_{\theta} p\left(\mathbf{D}_{\mathsf{pred}}\middle|\theta\right)\pi\left(\theta\middle|\mathbf{D}_{\mathsf{a}}\right)\mathrm{d}\theta$$

- Compatibility can be assessed using the prior predictive probability of observing more extreme pediatric data than has been observed:

$$\psi = P\left(m_{\mathsf{pred}} \leq m_{\mathsf{p}}\middle|\mathbf{D}_{\mathsf{a}}\right)$$

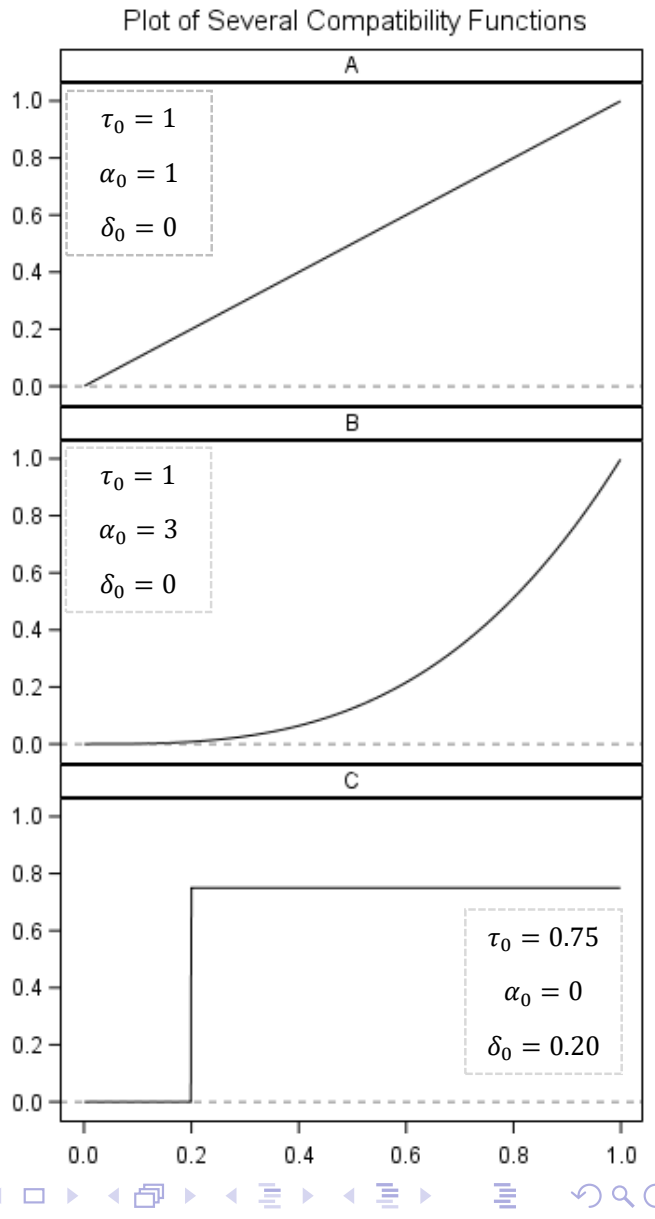
  - $m_{\mathsf{p}} = p\left(\mathbf{D}_{\mathsf{p}}\middle|\mathbf{D}_{\mathsf{a}}\right)$ is the marginal likelihood for observed pediatric data
  - $\mathbf{D}_{\mathsf{p}}$ is the observed pediatric data at some point in time

Customized Borrowing Based on Compatibility

- $\mathbf{D}_p$ highly compatible with $\mathbf{D}_a$, $\psi \approx 1$.
- As $\mathbf{D}_p$ and $\mathbf{D}_a$ become increasingly discrepant, $\psi \to 0$.
- Define $a_0 = f(\psi)$ where $f : [0,1] \to [0,1]$ is a non-decreasing function.
- A flexible family of functions is given by

$$f(\psi) = \tau_0 \cdot \psi^{\alpha_0} \cdot 1[\delta_0 \leq \psi]$$

where $\tau_0 \in [0,1]$, $\alpha_0 \geq 0$, and $\delta_0 \in [0,1]$.



Monitoring Criteria & Enrollment

Monitoring Criteria

- The goal of the trial is to test the hypotheses:

$$H_0 : \theta \leq \theta_0 = 0.40 \text{ versus } H_1 : \theta > \theta_0$$

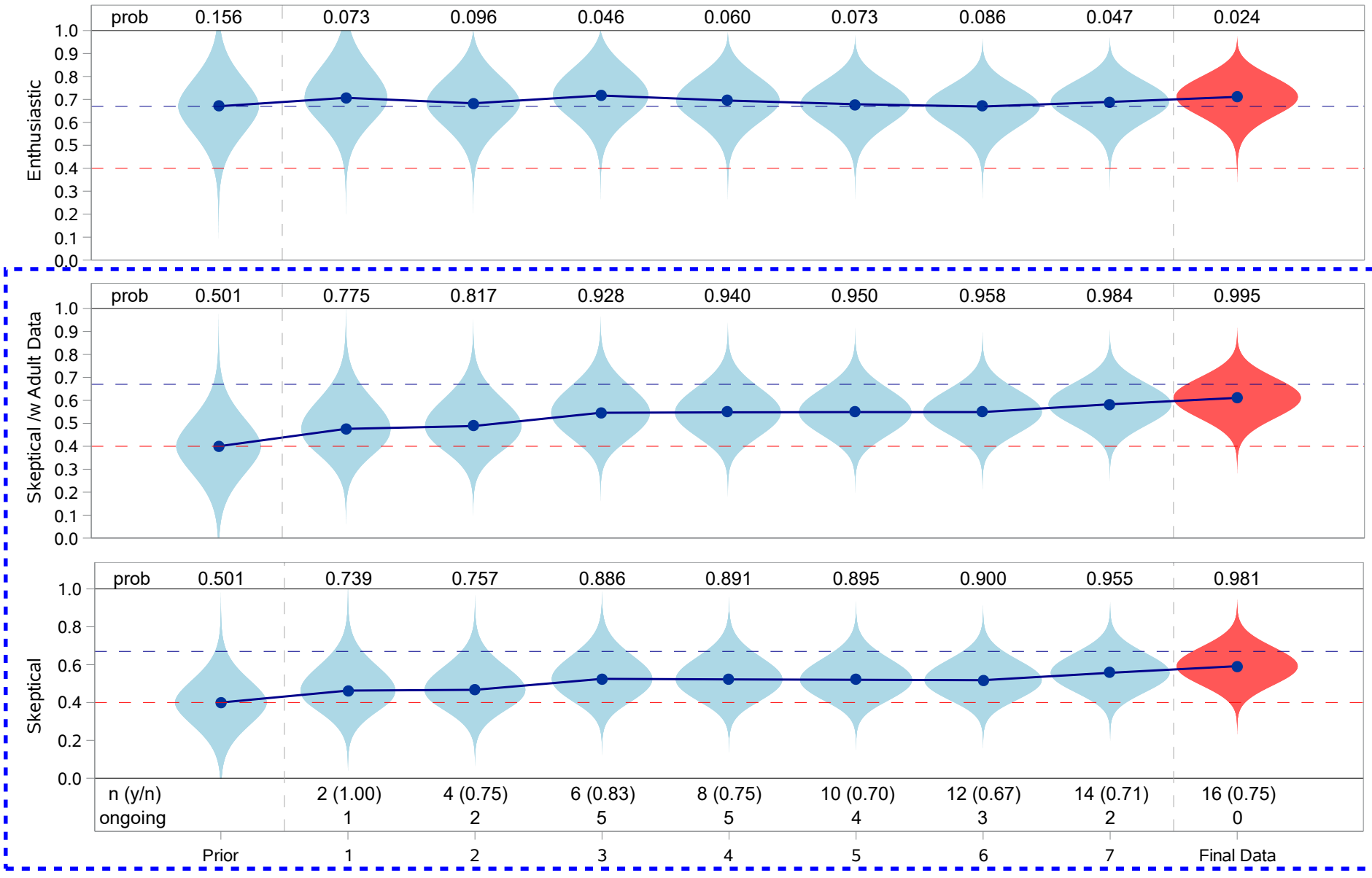
- From adult data – $\theta_1 = 0.67$
- Enrollment stoppage criteria:
  - Efficacy:
    - (1) $P\left(\theta > \theta_0 \middle| \mathbf{D}, \pi_0^{(s)}\right) \geq 0.80$ (persuasive based on skeptical prior)
    - (2) $P\left(\theta > \theta_0 \middle| \mathbf{D}, \pi_{a_0}^{(s)}\right) \geq 0.975$ (convincing based on skeptical power prior)
  - Futility:
    - (1) $P\left(\theta \leq \frac{\theta_0+\theta_1}{2} \middle| \mathbf{D}, \pi_0^{(e)}\right) \geq 0.80$
      (persuasive to enthusiast that efficacy is unacceptably low)
  - Exhausted resources: $N = 80$ patients enrolled

Enrollment

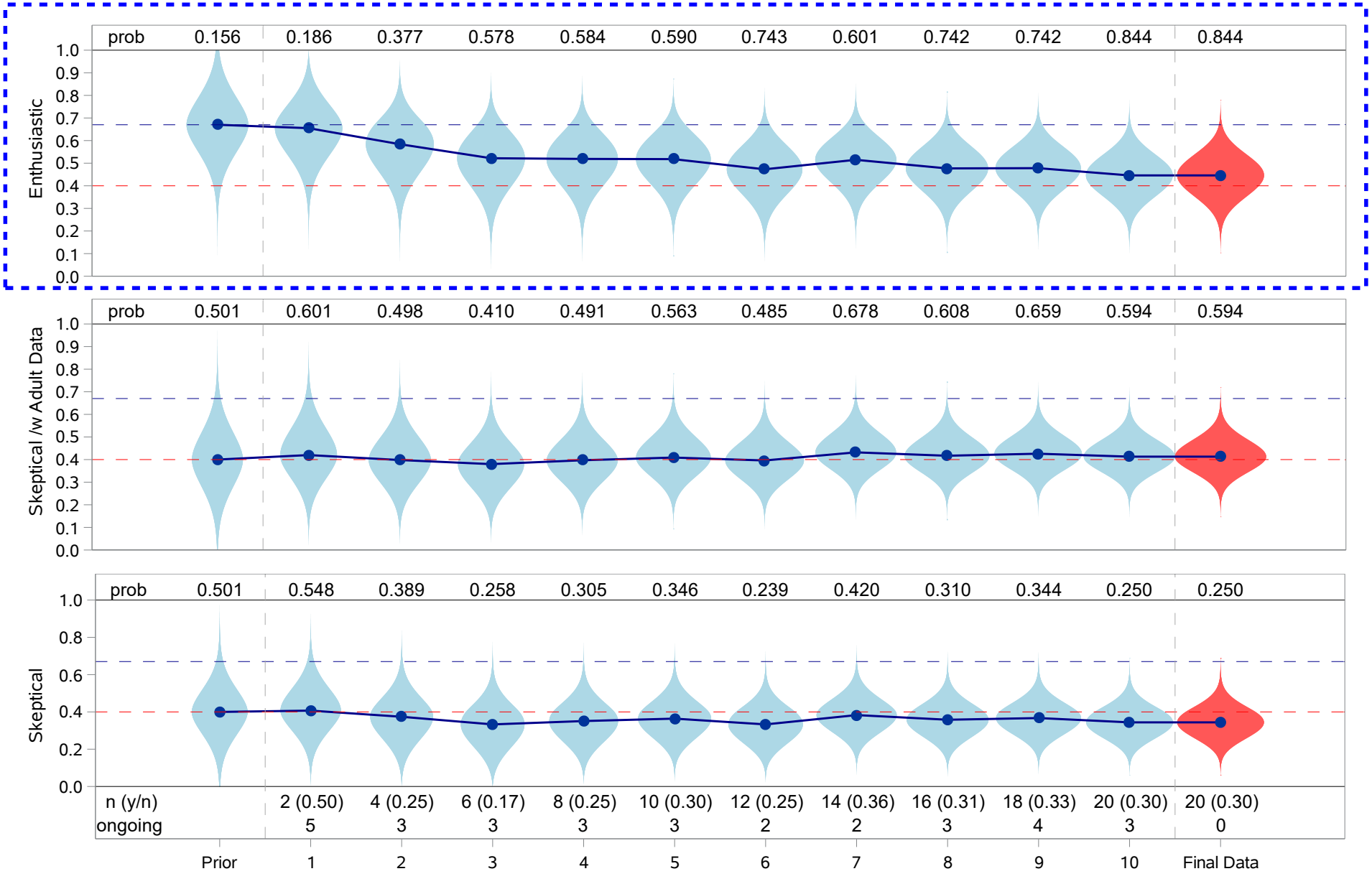
- Inferences are updated after every 2 outcomes are ascertained.
- T72 trial enrollment:
  (1) First patient first visit $\rightarrow$ August 25, 2006.
  (2) Last patient last visit $\rightarrow$ June 24, 2010.
- Based on the 54-week follow-up period, we can infer enrollment took place over approximately 33 months (approximately 1 patient per 0.55 months).

Illustrations of Monitoring Process

Example Data – Stopping for Efficacy – $a_0 = 0.5$

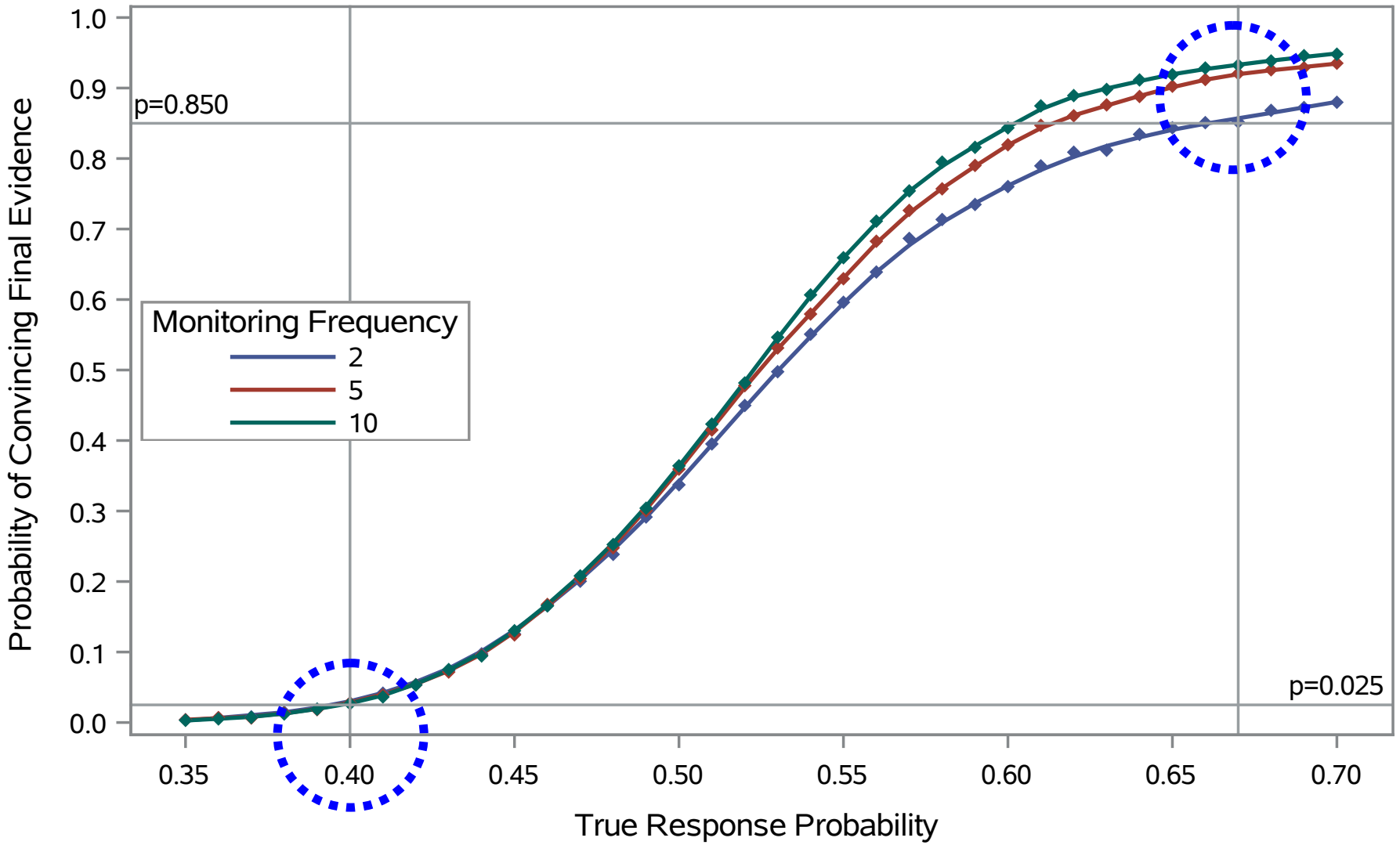


Example Data – Stopping for Futility – $a_0 = 0.5$



Preposterior Analysis (Frequentist Properties)

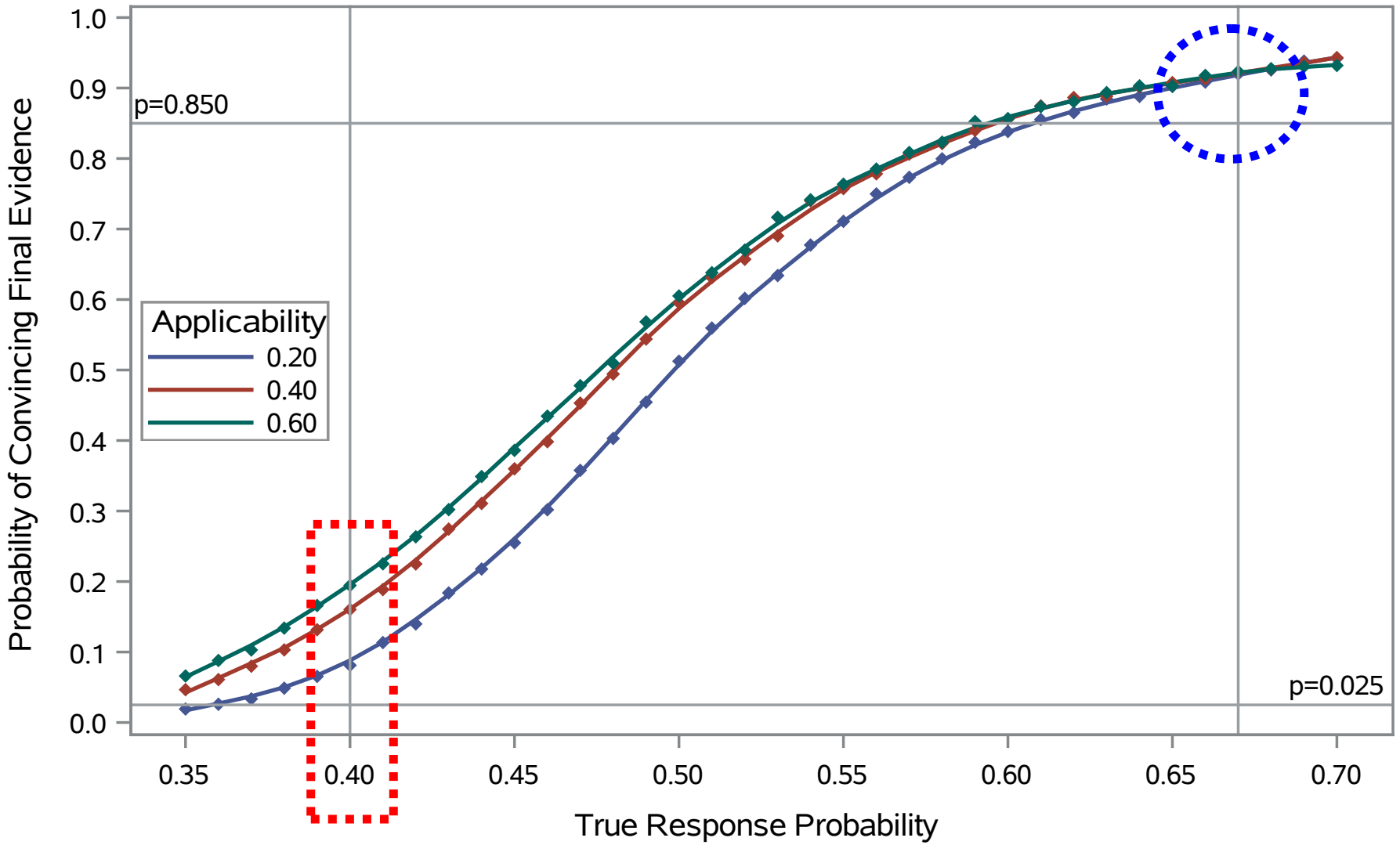
Operating Characteristics – No Use of Adult Data



E[Final N] / E[N]

2	34.7 / 34.5	49.3 / 47.6	37.0 / 33.6	26.7 / 23.1
5	37.1 / 36.9	51.7 / 50.1	40.0 / 36.6	29.3 / 25.7
10	41.4 / 41.3	56.6 / 55.0	43.1 / 39.8	31.6 / 28.0

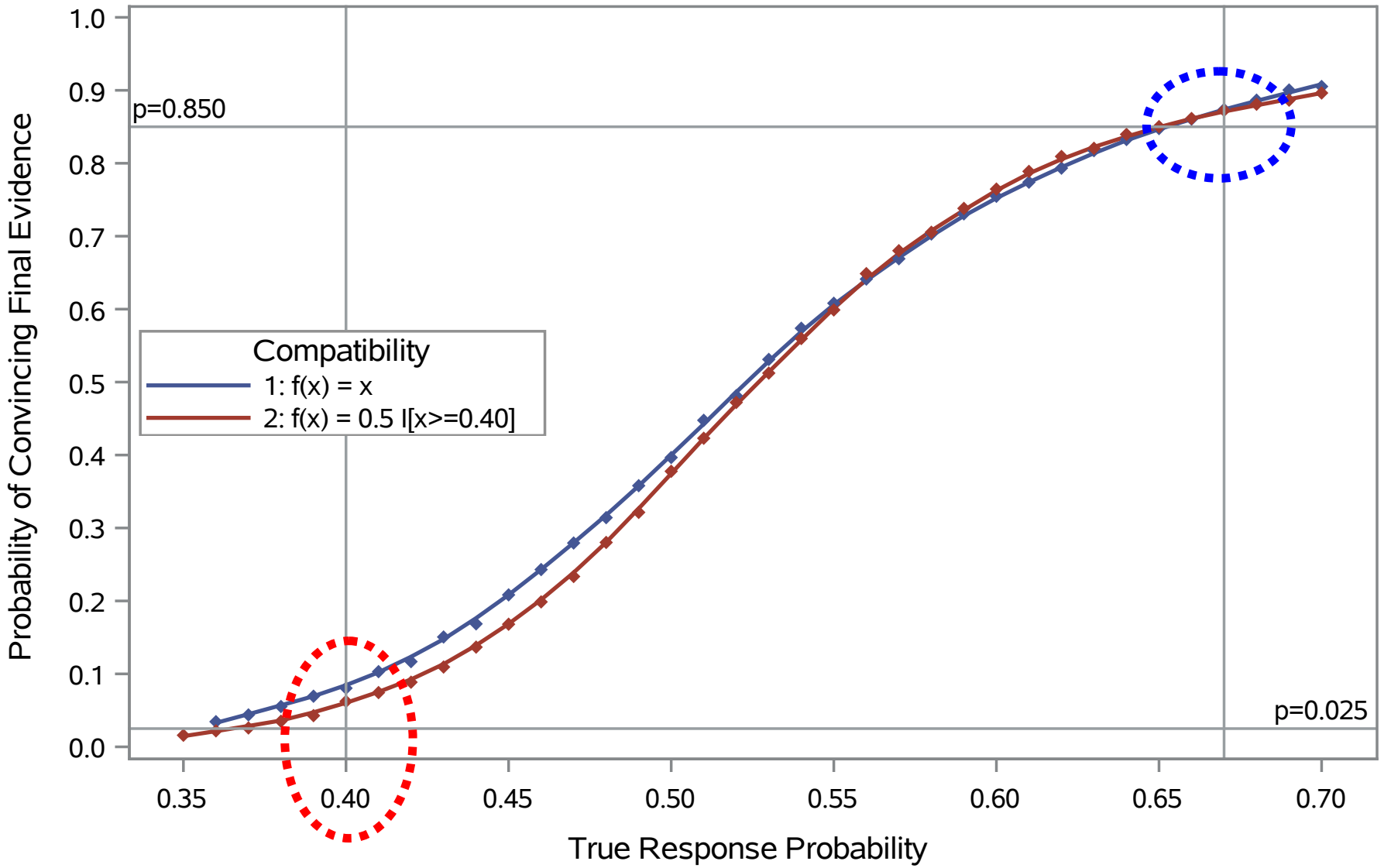
Operating Characteristics - with Adult Data - Applicability



E[Final N] / E[N]

0.20	32.7 / 32.3	39.8 / 37.4	30.5 / 27.0	23.8 / 20.1
0.40	29.4 / 28.5	32.2 / 29.5	25.5 / 22.0	20.7 / 17.0
0.60	26.9 / 25.8	28.2 / 25.4	22.3 / 18.7	18.3 / 14.7

Operating Characteristics - with Adult Data - Compatibility



E[Final N] / E[Interim N]

1	29.7 / 29.1	33.9 / 31.7	23.4 / 20.0	17.6 / 14.0
2	32.2 / 31.8	41.3 / 39.3	28.7 / 25.2	20.5 / 16.9

Closing Remarks

Closing Remarks

- The Bayesian approach provides an intuitive framework for continually updating inferences about a parameter in a sequential trial.
- It is better suited for the incoporation of prior information.
- Good Frequentist properties are desirable and so it is no surprise that reasonable Bayesian approaches exhibit them.
- We should move away from the type I error rate as the *primary* determining factor for whether the prior information *should* be used.
- "Why Bayes" $\rightarrow$ See Berry (1993) and Spiegelhalter et al. (1993)
- Tutorial of Bayesian data monitoring $\rightarrow$ Fayers et al. (1997)

References I

[1] Berger, J. and Berliner, L. M. (1986), "Robust Bayes and Empirical Bayes Analysis with Epsilon-Contaminated Priors," *Ann. Statist.*, 14, 461–486.

[2] Berry, D. A. (1993), "A case for bayesianism in clinical trials," *Statistics in Medicine*, 12, 1377–1393.

[3] Carlin, B. P. and Sargent, D. J. (1996), "Robust Bayesian Approaches for Clinical Trial Monitoring," *Statistics in Medicine*, 15, 1093–1106.

[4] Cornfield, J. (1966), "Sequential Trials, Sequential Analysis and the Likelihood Principle," *The American Statistician*, 20, 18–23.

[5] Edwards, W., Lindman, H., and Savage, L. J. (1963), "Bayesian statistical inference for psychological research." *Psychological Review*, 70, 193–242.

[6] Evans, M. and Moshonov, H. (2006), "Checking for prior-data conflict," *Bayesian Anal.*, 1, 893–914.

References II

[7] Fayers, P. M., Ashby, D., and Parmar, M. K. B. (1997), "Tutorial in Biostatistics: Bayesian Data Monitoring in Clinical Trials," *Statistics in Medicine*, 16, 1413–1430.

[8] Greenhouse, J. B. and Waserman, L. (1995), "Robust bayesian methods for monitoring clinical trials," *Statistics in Medicine*, 14, 1379–1391.

[9] Nott, D. J., Drovandi, C. C., Mengersen, K., and Evans, M. (2018), "Approximation of Bayesian Predictive p -Values with Regression ABC," *Bayesian Anal.*, 13, 59–83.

[10] Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. B. (1993), "Applying Bayesian ideas in drug development and clinical trials," *Statistics in Medicine*, 12, 1501–1511.

THANK YOU!

Questions?