Web Appendix A: Bayesian Hypothesis Testing

Consider the hypotheses $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ where $\Theta_0 \bigcup \Theta_1 = \Theta$ and $\Theta_0 \bigcap \Theta_1 = \emptyset$. Formal Bayesian hypothesis testing requires the specification of prior probabilities on the hypotheses (e.g., $p(H_i)$ for $i = 0, 1$) and prior distributions for (θ, η) specified over the parameter space defined with respect to each of the hypotheses (e.g. $\pi\left(\theta, \eta \middle| H_i\right)$ for $i = 0, 1$).

The posterior probability of hypothesis H_i is given by

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

where $p(\boldsymbol{D}|H_i) = \int_{\Theta_i} p(\boldsymbol{D}|\theta)\pi(\theta|H_i)d\theta$ is the marginal likelihood associated with hypothesis H_i. In practice, most Bayesian methods for clinical trials perform hypothesis testing based on the posterior probability of the *event defining* H_i. For this approach, one simply needs to specify a prior $\pi(\theta)$ representing belief about θ overall and compute the posterior distribution. The posterior probability that $\theta \in \Theta_i$ is given by

$$P(\theta \in \Theta_i|\boldsymbol{D}) = \frac{\int_{\Theta_i} p(\boldsymbol{D}|\theta)\pi(\theta|\theta \in \Theta_i)d\theta \cdot P(\theta \in \Theta_i)}{\sum_{j=0,1} \int_{\Theta_j} p(\boldsymbol{D}|\theta)\pi(\theta|\theta \in \Theta_j)d\theta \cdot P(\theta \in \Theta_j)} \tag{2}$$

where $P(\theta \in \Theta_i) = \int_{\Theta_i} \pi(\theta)d\theta$. We can readily see that the $P(\theta \in \Theta_i|\boldsymbol{D})$ is equal to $p(H_i|\boldsymbol{D})$ if one takes $p(H_i) = P(\theta \in \Theta_i)$ and $\pi\left(\theta \middle| H_i\right) = \pi\left(\theta \middle| \theta \in \Theta_i\right)$ for $i = 0, 1$. If in fact $\pi(\theta)$ does represent belief about θ, these choices are perhaps the most intuitive and thus we should have no reservation referring to $P(\theta \in \Theta_i|\boldsymbol{D})$ as the probability that hypothesis H_i is true.

Web Appendix B: Monitoring Prior Parameterization

Normal Distribution $\mathcal{N}_p(\tilde{\mu}, q)$Suppose $\theta \sim \mathcal{N}(\mu, \sigma)$ is a normal random variable that satisfies mode$(\theta) = \tilde{\mu}$ and $P(\theta \leqslant q) = p$. The values for the mean and standard deviation are $\mu = \tilde{\mu}$ and $\sigma = \frac{q-\mu}{\Phi^{-1}(p)}$, where Φ denotes the cumulative distribution function for a standard normal distribution and Φ^{-1} denotes its quantile function. Therefore we can denote the distribution with the desired mode and tail probability constraint as $\theta \sim \mathcal{N}_p(\tilde{\mu}, q)$, which is well-defined for values (μ, q, p) that satisfy $\frac{q-\mu}{p-0.5} > 0$. Since the normal distribution is

completely specified by (μ, σ), quantities such as $P(\theta \leqslant \tilde{q})$ are also specified for any $\tilde{q} \in \mathbb{R}$. In particular, if $\theta \sim \mathcal{N}_p(\tilde{\mu}, q)$ then $P(\theta \leqslant \frac{q+\mu}{2}) = \Phi(\frac{\Phi^{-1}(p)}{2})$. Furthermore, $P(\theta \in (\mu, \frac{q+\mu}{2})) = |p - \Phi(\frac{\Phi^{-1}(p)}{2})|$.

Generalized Normal Distributions $\mathcal{GN}_p(\tilde{\mu}, q, \gamma)$Let $F_{\mu,\alpha,\beta}$ denote the cumulative distribution function of $\mathcal{GN}(\mu, \alpha, \beta)$. The CDF of a generalized normal random variable $\theta \sim \mathcal{GN}(\mu, \alpha, \beta)$ can be expressed as (Griffin, 2018)

$$P(\theta \leqslant q|\mu, \alpha, \beta) = \frac{1}{2} + \frac{\text{sign}(q-\mu)}{2} \int_0^{|q-\mu|^\beta} \frac{w^{1/\beta - 1}}{\alpha\Gamma(1/\beta)} \exp\left\{-\left(\frac{1}{\alpha}\right)^\beta w\right\} dw \tag{3}$$

Define $\theta \sim \mathcal{GN}_p(\tilde{\mu}, q, \gamma)$ as the generalized normal distribution $\mathcal{GN}(\mu, \alpha, \beta)$ that satisfies $\text{mode}(\theta) = \tilde{\mu}$, $P(\theta \leqslant q) = p$, and $P(\theta \in (q, \frac{q+\mu}{2})) = \gamma \cdot |p - \Phi(\frac{\Phi^{-1}(p)}{2})|$. The mode is equal to $\mu = \tilde{\mu}$, and α and β are determined to minimize the function $(F_{\mu,\alpha,\beta}(q) - p)^2 + (F_{\mu,\alpha,\beta}(\frac{q+\mu}{2}) - \gamma \cdot |p - \Phi(\frac{\Phi^{-1}(p)}{2})|)^2$ with box-constrained optimization (Byrd et al., 1995).

Using this notation to define the distributions in main article Figure 1 yields the default enthusiastic prior as $\mathcal{GN}_{p=0.025}(\tilde{\mu} = \theta_1, q = \theta_0, \gamma = 1)$, the concentrated enthusiastic prior as $\mathcal{GN}_{p=0.025}(\tilde{\mu} = \theta_1, q = \theta_0, \gamma = 0.75)$, the flattened enthusiastic prior as $\mathcal{GN}_{p=0.025}(\tilde{\mu} = \theta_1, q = \theta_0, \gamma = 1.5)$, and the locally non-informative prior as $\mathcal{GN}_{p=0.025}(\tilde{\mu} = \frac{\theta_0+\theta_1}{2}, q = \frac{3\theta_0-\theta_1}{2}, \gamma = 1.5)$.

Truncated Generalized Normal Distribution $\mathcal{GN}_{p,\Theta}(\tilde{\mu}, q, \gamma)$The density for a generalized normal distribution truncated to the interval domain $\Theta = (\theta_{min}, \theta_{max})$, denoted by $\mathcal{GN}_\Theta(\mu, \alpha, \beta)$, is $f(\theta) = c \cdot \exp\left\{-\frac{|\theta-\mu|}{\alpha}^\beta\right\} I(\theta \in \Theta)$ where $c = \frac{\beta}{2\alpha\Gamma(1/\beta)}(F_{\mu,\alpha,\beta}(\theta_{max}) - F_{\mu,\alpha,\beta}(\theta_{min}))^{-1}$. Define $\theta \sim \mathcal{GN}_{p,\Theta}(\tilde{\mu}, q, \gamma)$ as the truncated generalized normal distribution $\mathcal{GN}_\Theta(\mu, \alpha, \beta)$ that satisfies $\text{mode}(\theta) = \tilde{\mu}$, $P(\theta \leqslant q) = p$, and $P(\theta \in (q, \frac{q+\mu}{2})) = \gamma|p - \Phi(\frac{\Phi^{-1}(p)}{2})|$.

Using this notation to define the distribution in main article Figure 2 yields the joint distribution $\pi(\theta, \eta) = \pi(\theta) \times \pi(\eta|\theta)$, as $\pi(\theta) \sim \mathcal{GN}_{p=0.975,\Theta=[-1,1]}(\tilde{\mu} = \theta_0, q = \theta_1, \gamma = 1)$ and $\pi(\eta|\theta) \sim \mathcal{GN}_{p=0.975,\Theta=[\max(-\theta,0),\min(1,1+\theta)]}(\tilde{\mu} = \eta_0, q = \eta_1, \gamma = 1)$.

Web Appendix C: Type 1 Error Rate Depending on Enrollment Schemes

Recall main article Figure 5 showed Type 1 error properties for the single-arm design. Figure 1 shows results from a design that has a longer follow-up period. The interim sample sizes are the same for each monitoring frequency, however, the final sample sizes under the longer follow-up designs are much larger (over 20 patients in follow-up for monitoring frequencies of 8 or fewer, compared to approximately 6 patients in the shorter follow-up designs). The final probability of efficacy criteria being satisfied is generally slightly lower in the longer follow-up design, which is what we would expect since the larger final sample size contains more data consistent with a null result.

[Figure 1 about here.]

Web Appendix D: Robustness of Parameterizations of Monitoring Priors

Recall the analyses done in main article Section 3.1 used a concentrated skeptical prior and default enthusiastic prior. In this section we show the four possible designs using the combinations of skeptical and enthusiastic prior given in main article Figure 1.

Figures 2 and 3 show what happens when the enthusiastic prior shifts from default to flattened, with the skeptical prior remaining fixed. Note that in the region between θ_0 and $\frac{\theta_0+\theta_1}{2}$ as the enthusiastic prior shifts from default to flattened, (a) the probability of stopping early for futility increases (b) the probability of inconclusive findings decreases and (c) the intermediate and final sample sizes decrease. This is because the enthusiastic prior gives more mass in for this region of θ. The flattened enthusiastic prior was used in Section **??** to enhance the ability of futility monitoring to reduce the sample size.

Contrasting 2 and 3, we see that the probability of stopping early for efficacy is much higher at θ_0 when the default skeptical prior is used rather than the concentrated skeptical prior. This is because the default skeptical prior has less mass around $\theta = \theta_0$, therefore it

is easier to convince the skeptic that $\theta > \theta_0$ under the null result $\theta = \theta_0$. The concentrated skeptical prior was used in main article Section 3.1 to limit this probability and provide better Type 1 error control.

The choice of skeptical and enthusiastic prior affects the analysis, and their specification (e.g. default, skeptical, enthusiastic) should be made with these properties in mind.

[Figure 2 about here.]

[Figure 3 about here.]

References

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* **16,** 1190–1208.

Griffin, M. (2018). Working with the exponential power distribution using gnorm.

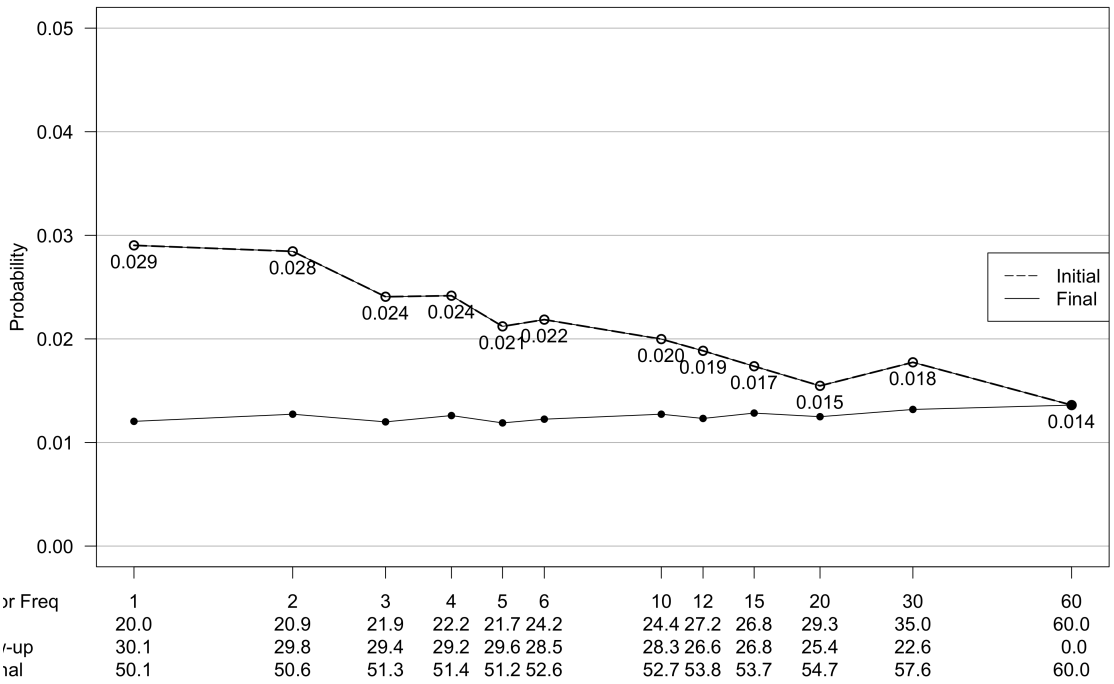


Figure 1. Single arm design from main article Section 3.1 with a longer follow-up period. Probability of efficacy criteria being satisfied when $\theta = \theta_0$. SS; sample size. Monitor Freq; monitoring frequency.

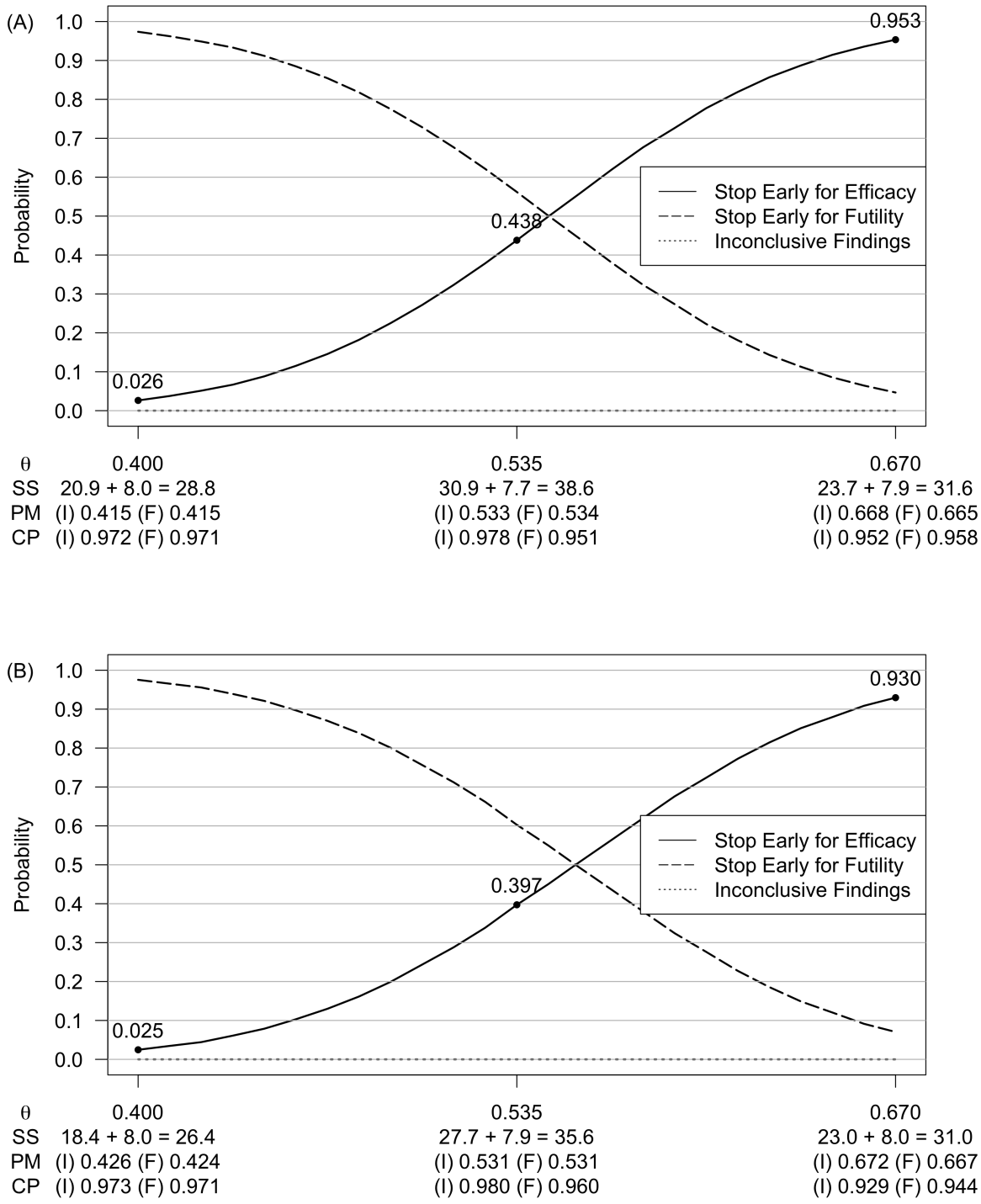
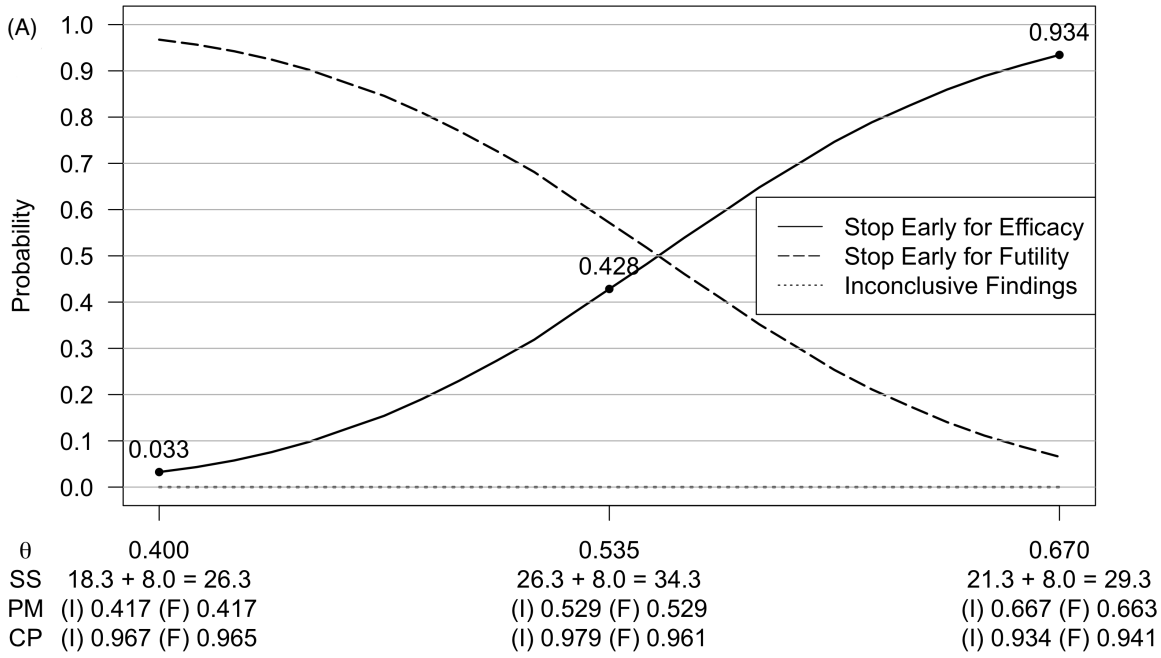


Figure 2. Modification of enthusiastic prior parameterization from main article Section 3.1. A, default enthusiastic prior (main article Figure 1(c)). B, flattened enthusiastic prior (main article Figure 1(d)). Both designs use concentrated skeptical prior (main article Figure 1(b)).



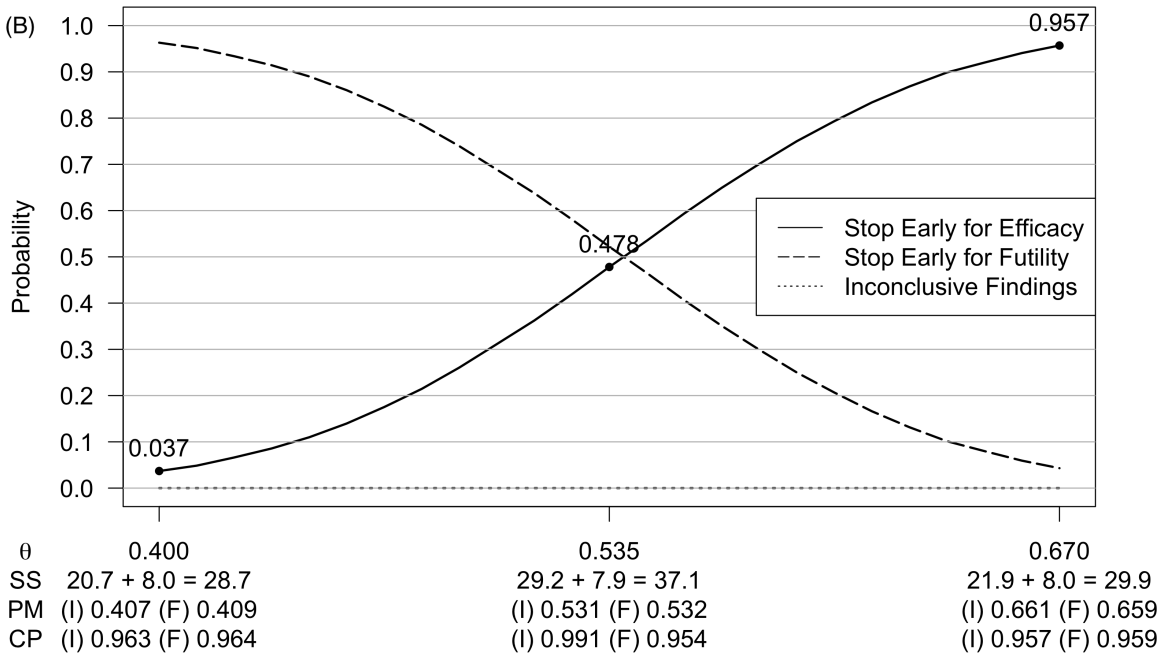


Figure 3. Modification of enthusiastic prior parameterization in main article Section 3.1. A, default enthusiastic prior (main article Figure 1(c)). B, flattened enthusiastic prior (main article Figure 1(d)). Both designs use default skeptical prior (main article Figure 1(a)).