Clinical trial analysis

André Moser, CTU Bern, Department of Clinical Research, University of Bern

20/11/2023

Table of contents

1 Introduction

This web book serves as a documentation source of some analysis approaches I used during my work at CTU Bern. Learning is an active process. I learn while writing my thoughts down and experiment with statistical methodology. My notes contain errors, because I am a human being. Any notification of errors is highly appreciated.

1.1 General notation and abbreviations

- iid: Independent and identically distributed.
- pdf: Probability density function. Most often denoted as $f(\cdot)$. For bivariate pdf we use the notation $f_2(\cdot)$.
- cdf: Cumulative distribution function. Most often denoted as $F(\cdot)$. For bivariate pdf we use the notation $F_2(\cdot)$.
- N_2: Bivariate cdf of the Gaussian distribution.
- ϕ: pdf of the standard Gaussian distribution.
- Φ: cdf of the standard Gaussian distribution.
- Φ^{-1}: Quantile function of the standard Gaussian distribution function.

1.2 'Power' vocabulary

In their supplement Kunzmann et al. [1] provide a literature review of the terminology used in articles. We provide here a summary of this terminology:

- **Frequentist power**: Probability of rejection given that the alternative hypothesis is true.

- **Average power**: Prior averaged probability of rejection. Often also called 'probability of success', 'assurance', 'Bayesian predictive power'.
- **Prior adjusted power**: Joint probability of rejection and that the treatment effect is effective.

1.3 Some 'Bayesian' concepts

- **Prior**: A random variable Θ with pdf $f(\theta)$ representing that the uncertainty of a parameter of interest.
- **Design prior**: Prior used before data collection as data generating mode [2].
- **Analysis prior**: Prior used for Bayesian analysis of the collected data [2].
- **Prior predictive distribution**: Situation *before* a sample was taken. Let θ be a realisation of a random variable Θ with pdf $f(\theta)$. Then for a future observation $\tilde{X}$ the pdf is

$$f(\tilde{x}) = \int_{\Theta} f(\tilde{x}, \theta) d\theta = \int_{\Theta} \underbrace{f(\tilde{x}|\theta)}_{likelihood} \underbrace{f(\theta)}_{prior} d\theta.$$

- **Posterior predictive distribution**: Situation *after* a sample was taken. Let θ be a realisation of a random variable Θ with pdf $f(\theta)$. Then for a future observation $\tilde{X}$ and observed X (since X is independent $\tilde{X}$) the pdf is

$$f(\tilde{x}|x) = \int_{\Theta} f(\tilde{x}|\theta, x) f(\theta|x) d\theta = \int_{\Theta} \underbrace{f(\tilde{x}|\theta)}_{likelihood} \underbrace{f(\theta|x)}_{posterior} d\theta.$$

- **Improper prior**: A prior with $\int_{\Theta} f(\theta) = \infty$.
- **Jeffrey's prior**: For an unknown parameter θ Jeffrey's (scalar) prior is defined as $f(\theta) \propto \sqrt{I(\theta)}$, where $I(\theta)$ is the expected Fisher information of θ [3]. Jeffrey's prior can be improper [3]. Bayesian point estimates using Jeffrey's prior are often very close to maximum likelihood estimators [3].

Example: Jeffrey's prior for the binomial model

The likelihood of the binomial model is

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

and thus

$$L := log(f(x|\theta)) = x \log(\theta) + (n-x) \log(1-\theta).$$

Simple algebra leads to

$$\frac{dL}{d\theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta}, \quad \frac{d^2L}{d\theta^2} = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}.$$

The expected Fisher information is

$$I(\theta) = -E_\theta \left(\frac{d^2L}{d\theta^2} \right) = \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)} \propto \frac{1}{\theta(1-\theta)}.$$

Thus, Jeffrey's prior for the binomial model is

$$f(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}} = \theta^{-0.5}(1-\theta)^{-0.5} = beta(0.5, 0.5).$$

Jeffrey's prior for the binomial model is a proper prior [3].

1.4 Used R libraries

```
# All tidyverse functions
library(tidyverse)
# For frequentist sample size calculation
library(epiR)
# Multivariate Gaussian and t-distributions
library(mvtnorm)
# Beta binomial distribution
library(extraDistr)
```

```r
# Agresti-Caffo confidence intervals for risk difference
library(PropCIs)
```

2 Power and sample size calculations

2.1 Background

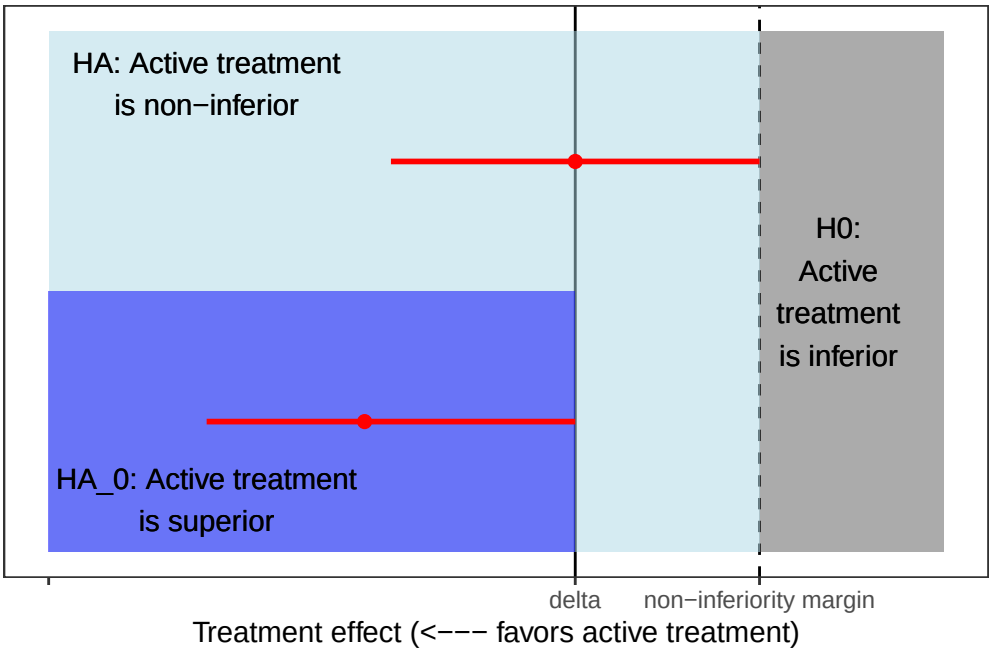
The terminology of 'power' is often imprecisely used [1]. Kunzmann et al. suggest to use the neutral term 'probability to reject'. The classical (frequentist) 'power' is defined as the probability to reject given that the alternative hypothesis is true. Frequentist power calculations do not include uncertainties of the treatment effect, whereas Bayesian and hybrid approaches include such uncertainties in their calculations. In the following we use different approaches for the calculation of the 'probability to reject' (frequentist, Bayesian and hybrid) for different clinical trial designs.

Definition 'hybrid' ([4], Section 6.5.2)

'[...] we have a prior distribution to use in our study design, but that the conclusions of the study will be entirely classical and will not make use of the prior [...]'

2.2 Two-arm non-inferiority setting

In this section we consider a non-inferiority clinical trial setting with a null hypothesis $H_0 : \delta > \delta^*$ and alternative hypothesis $H_a : \delta \leq \delta^*$, where $\delta^* > 0$ is a fixed non-inferiority margin and a treatment effect δ, for example, a continuous difference or a risk difference.



2.2.1 Binomial outcome

Here p_1 and p_0 are event probabilities from an active treatment arm and a control arm, respectively. $\delta = p_1 - p_0$ is the true treatment effect expressed as a risk difference.

In this section we assume that the variances in both groups are known, but might be different.

Working example

We use the SAFE-SSPE trial as a working example [5]. In brief, this non-inferiority randomised placebo-controlled trial compares clinical surveillance versus anticoagulant treatment in low-risk patients with isolated subsegmental pulmonary embolism (SSPE). The primary outcome is the proportion of 3-month recurrence of venous thromboembolism (VTE).

The null hypothesis H_0 is *'clinical surveillance is inferior to anticoagulant treatment'* versus the alternative hypothesis H_a *'clinical surveillance is non-inferior to anticoagulant treatment'*. Thus, $H_0 : p_1 - p_0 > \delta^*$

vs $H_a : p_1 - p_0 \leq \delta^*$, where p_1 is the VTE proportion in the clinical surveillance arm and p_0 is the VTE proportion in the control arm.

The non-inferiority margin was set at 3.5% and it was assumed that the proportion of VTE in both groups was 1%.

2.2.1.1 Frequentist approach

Let $Y_i = (Y_{i,1}, Y_{i,2}, \cdots, Y_{i,n_i})^\top$ be a sample of size n_i from a binomial distribution $Y_i \sim Binomial(p_i, n_i)$, $i \in \{0, 1\}$, where p_i, $i \in \{0, 1\}$, are the true event proportions. Denote the estimated event proportions as $\overline{p}_i = \frac{1}{n_i} \sum_{k \leq n_i} Y_{i,k}$, $i \in 0, 1$, and the estimated risk difference as $D = \overline{p}_1 - \overline{p}_0$. Then D is asymptotically Gaussian distributed $D \sim N\left(\delta, \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}\right)$, where $\sigma_i^2 = p_i(1 - p_i)$, $i \in \{0, 1\}$. For notational purposes we denote the variance of the treatment effect as $\sigma_{treat}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}$.

We are interested whether the upper $(1 - \alpha)$%-confidence limit of D is smaller than the non-inferiority margin δ^*, that is,

$$D + z_{1-\alpha} \sqrt{\frac{n_0 \sigma_1^2 + n_1 \sigma_0^2}{n_1 n_0}} \leq \delta^* \quad \Rightarrow \quad D \leq -z_{1-\alpha} \sqrt{\frac{n_0 \sigma_1^2 + n_1 \sigma_0^2}{n_1 n_0}} + \delta^*,$$

where $z_{1-\alpha} = \Phi^{-1}(1 - \alpha)$. Note that $D_{suc}^{\delta^*} := -z_{1-\alpha} \sqrt{\frac{n_0 \sigma_1^2 + n_1 \sigma_0^2}{n_1 n_0}} + \delta^*$ is the **required risk difference** for a 'successful' rejection of the null hypothesis. Then

$$\begin{aligned} P_\delta(D \leq D_{suc}^{\delta^*}) &= \Phi\left(-z_{1-\alpha} - \sqrt{\frac{n_1 n_0}{n_0 \sigma_1^2 + n_1 \sigma_0^2}}(\delta - \delta^*)\right) \\ &= \Phi\left(-z_{1-\alpha} - \frac{(\delta - \delta^*)}{\sigma_{treat}}\right), \end{aligned}$$

since under regularity conditions,

$$Z = \frac{D - \delta^*}{\sigma_{treat}} \to N(0, 1), \quad \min(n_1, n_0) \to \infty.$$

The conditional probability $P_\delta(D \leq D^{\delta^*}_{suc})$ is the **probability to reject given** δ. For $\delta > \delta^*$ this is the 'type-I error' and for $\delta \leq \delta^*$ the frequentist 'power'.

For a δ_A it holds that

$$-z_{1-\alpha} - \sqrt{\frac{n_1 n_0}{n_0\sigma_1^2 + n_1\sigma_0^2}}(\delta_A - \delta^*) = \Phi^{-1}(1-\beta) = z_{1-\beta}$$

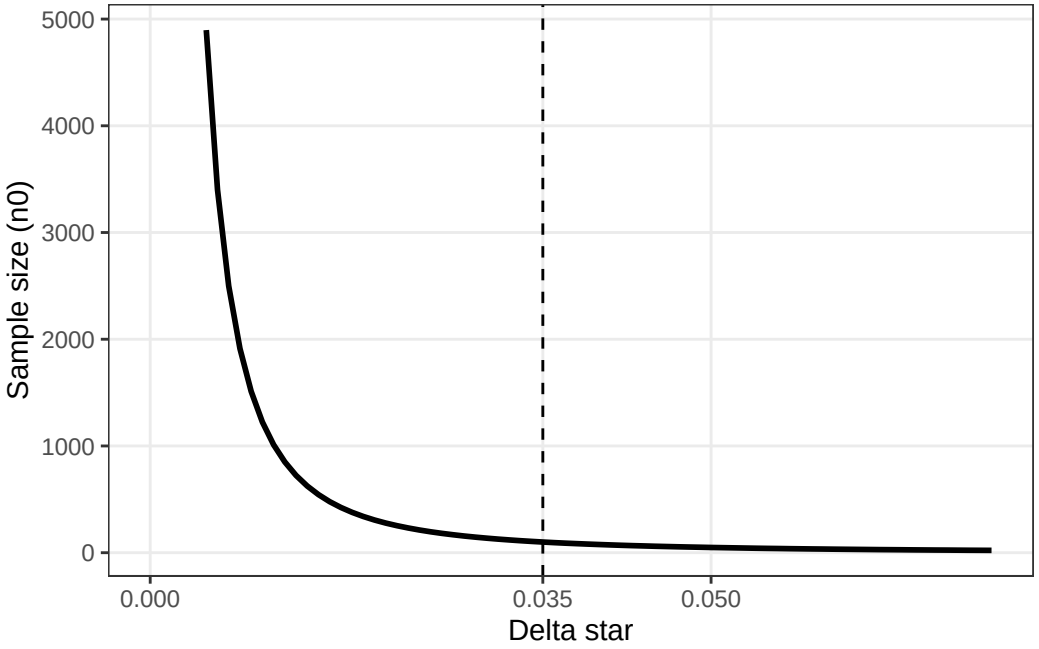
and so the sample size can then be calculated as

$$\frac{(z_{1-\beta} + z_{1-\alpha})^2}{(\delta_A - \delta^*)^2} = \frac{an_0^2}{n_0\sigma_1^2 + an_0\sigma_0^2},$$

where $a = n_1/n_0$ is an allocation ratio, such that

$$n_0 = (z_{1-\beta} + z_{1-\alpha})^2 \frac{\sigma_1^2 + a\sigma_0^2}{a(\delta_A - \delta^*)^2}, \quad n_1 = an_0.$$

δ_0, δ_A and δ^* are assumed as fixed and known constants in a frequentist approach. Their choices are of high importance, because all trial conclusions are based on those choices and affect the sample size calculation. The plot below shows how the sample size increase as δ^* approaches δ.



Working example (continued)

We calculate the required sample size for the SAFE-SSPE trial using a frequentist approach with the following parameters:

- $p_1 = 0.01$, $p_0 = 0.01$, $\delta^* = 0.035$, $1 - \beta = 0.8$, $\alpha = 0.05$, $a = 1/1$

```
library(epiR)

alpha <- 0.05
beta <- 0.2
p_0 <- 0.01
p_1 <- 0.01
delta <- p_1 - p_0
delta_star <- 0.035
sd_0 <- sqrt(p_0 * (1 - p_0))
sd_1 <- sqrt(p_1 * (1 - p_1))
a <- 1/1

epi.ssninfb(treat = p_1, control = p_0, delta = delta_star, power = 1 -
    beta, r = a, alpha = alpha, n = NA)
```

```
$n.total
[1] 200

$n.treat
[1] 100

$n.control
[1] 100

$delta
```

```
[1] 0.035
```

```
$power
[1] 0.8
```

```
n_0 <- ((qnorm(1 - beta) + qnorm(1 - alpha))^2) * ((sd_1^2 +
    a * sd_0^2)/(a * (delta - delta_star)^2))
n_0
```

```
[1] 99.93031
```

```
n_1 <- n_0 * a
n_1
```

```
[1] 99.93031
```

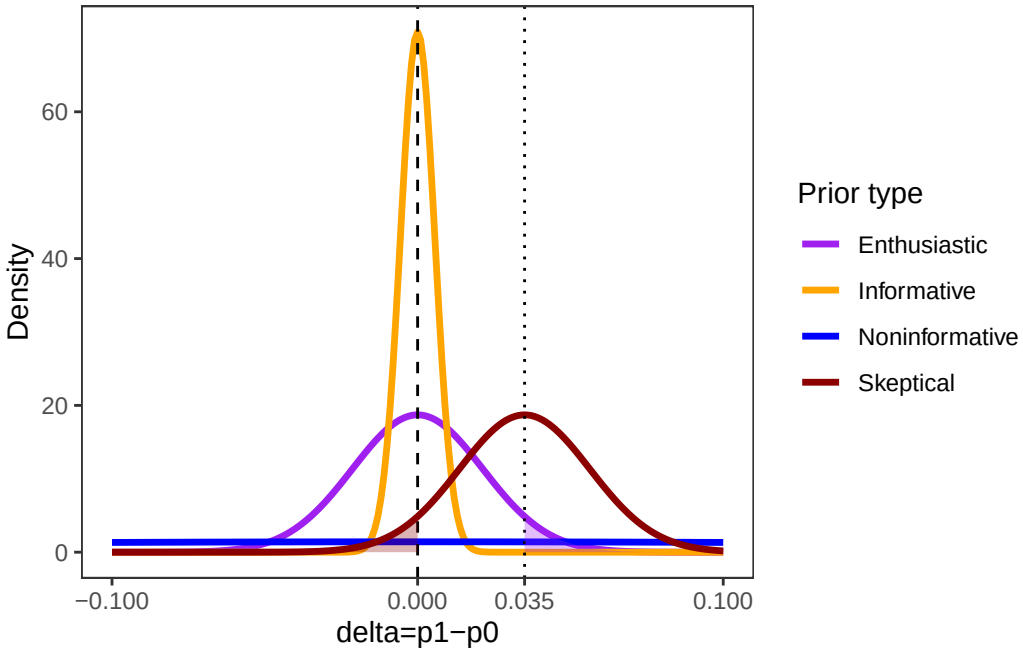
Under the specified parameters a sample size of 200 patients (100 per arm) is needed to reject the null hypothesis of inferiority. This is more or less the sample size mentioned in the study protocol of the SAFE-SSPE trial but without dropouts and adjustments for rare events.

2.2.1.2 Hybrid approach: Prior on the risk difference

Suppose that the true treatment effect δ is a realization from a random variable Δ with $f(\delta)$. In this subsection we assume that the design prior comes from a Gaussian distribution function so that $\Delta \sim N\left(d, \frac{\sigma_1^2+\sigma_0^2}{m}\right)$. Note that this prior can be thought as a realisation from m Gaussian 'prior observations' with variance $\sigma_1^2 + \sigma_0^2$. Again for notational purposes we denote the variance of the design prior of the treatment effect as $\sigma^2_{prior} = \frac{\sigma_1^2+\sigma_0^2}{m}$.

In the following we will use the following design priors:

- **Enthusiastic prior** (favors non-inferiority): $d = 0$, $m = 6.6$, $P(\Delta > \delta^*) = 0.05$. This prior is centered on the treatment effect such that there is a low probability (here 5%) of inferiority.
- **Skeptical prior** (favors inferiority): $d = \delta^*$, $m = 6.6$, $P(\Delta > 0) = 0.05$. This prior is centered on the non-inferiority margin such that there is a low probability (here 5%) of superiority.
- **Informative prior** (clinical expert knowledge): $d = 0$ with $m = 25$.
- **Noninformative prior**: $d = 0$ with $m = 0.5$.



Let

$$AP := \int_{\Delta} P_{\delta}(D \leq D_{suc}^{\delta^*}) f(\delta) d\delta,$$

be the **'average power'** [6] (also called 'assurance' [7], 'probability of success' [1], [4] or Bayesian predictive power [8]. Kunzmann et al. provide in their supplemental section a literature review of the used terminology in publications from clinical trials [1]).

In the hybrid approach we are interested in trial conclusions from a frequentist point of view, thus we are interested in

$$D \leq -z_{1-\alpha} \sqrt{\frac{n_0 \sigma_1^2 + n_1 \sigma_0^2}{n_1 n_0}} + \delta^*.$$

Then, for a $\tilde{D}$ from future observations,

$$
\begin{aligned}
AP &= \int_{-\infty}^{-\infty} P_\delta(D \leq D_{suc}^{d,\delta^*}) f(\delta) d\delta \\
&= \int_{-\infty}^{-\infty} \left\{ \int_{-\infty}^{-z_{1-\alpha}\sqrt{\frac{n_0\sigma_1^2+n_1\sigma_0^2}{n_1 n_0}}+\delta^*} f(\tilde{\delta}|\delta) d\tilde{\delta} \right\} f(\delta) d\delta \\
&= \int_{-\infty}^{-z_{1-\alpha}\sqrt{\frac{n_0\sigma_1^2+n_1\sigma_0^2}{n_1 n_0}}+\delta^*} \left\{ \int_{-\infty}^{-\infty} f(\tilde{\delta}|\delta) f(\delta) d\delta \right\} d\tilde{\delta} \\
&= \int_{-\infty}^{-z_{1-\alpha}\sqrt{\frac{n_0\sigma_1^2+n_1\sigma_0^2}{n_1 n_0}}+\delta^*} f(\tilde{\delta}) d\tilde{\delta} \\
&= \Phi\left(\frac{1}{\sigma_{prior}} \left[-z_{1-\alpha}\sigma_{treat} - (d - \delta^*) \right] \right),
\end{aligned}
$$

since the predictive distribution of $\tilde{D}$ is Gaussian distributed $\tilde{D} \sim N\left(d, \sigma_{treat}^2 + \sigma_{prior}^2\right)$ under a design prior $\Delta \sim N\left(d, \sigma_{prior}^2\right)$ ([9], [6]). Note that as $m \to \infty$, then $AP \to \Phi\left(-z_{1-\alpha} - \frac{(d-\delta^*)}{\sigma_{treat}}\right)$, that is, the frequentist power at d.

Working example (continued)

We calculate the AP under the assumed prior distributions and parameters for the SAFE-SSPE trial.

```
# Enthusiastic prior
m <- 6.6
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- data.frame(type = "Enthusiastic", n_0, n_1, AP = round(AP,
```

```
    2))

# Skeptical prior
m <- 6.6
prior_mean <- delta_star
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- rbind(data_output, data.frame(type = "Skeptical",
    n_0, n_1, AP = round(AP, 2)))

# Informative prior
m <- 25
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- rbind(data_output, data.frame(type = "Informative",
    n_0, n_1, AP = round(AP, 2)))

# Noninformative prior
m <- 0.5
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
```

```r
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- rbind(data_output, data.frame(type = "Noninformative",
    n_0, n_1, AP = round(AP, 2)))

data.frame(data_output %>%
    arrange(type))
```

```
           type      n_0      n_1   AP
1   Enthusiastic 99.93031 99.93031 0.59
2    Informative 99.93031 99.93031 0.66
3 Noninformative 99.93031 99.93031 0.52
4      Skeptical 99.93031 99.93031 0.34
```

Under an 'enthusiastic prior' we get an average power of 59%. For a 'skeptical prior' the average power decreases to 34%. These values are lower than the frequentist power of 80%, which holds under the assumption that the specified treatment effect (the alternative hypothesis) is true.

Rufibach et al. give a closed a formula for the distribution of $P_\Delta(D \leq D_{suc}^{\delta^*})$, where $\Delta \sim N(d, \sigma_{prior}^2)$, and discuss the shape under different prior choices [8]. Kunzmann et al. call this the 'random probability to reject' (RPR), $RPR := P_\Delta(D \leq D_{suc}^{\delta^*})$.

For $0 < y < 1$, the random variable RPR has a pdf

$$f(y) = \frac{\sigma_{treat}}{\sigma_{prior}} \phi\left(-z_{1-\alpha}\frac{\sigma_{treat}}{\sigma_{prior}} - \frac{(d-\delta^*)}{\sigma_{prior}} - \frac{\sigma_{treat}}{\sigma_{prior}}\Phi^{-1}(y)\right)\left[\phi\left(\Phi^{-1}(y)\right)\right]^{-1},$$

see [8].

Special case

For the case that $n_0 = n_1 = n$ and $\sigma_1^2 = \sigma_0^2 = \sigma^2$, then $\sigma_{treat}^2 = \frac{2\sigma^2}{n}$ and $\sigma_{prior}^2 = \frac{2\sigma^2}{m}$, and the formula above reduces to

$$f(y) = \sqrt{\frac{m}{n}}\phi\left(\sqrt{\frac{m}{n}}\left[-z_{1-\alpha} - \sqrt{\frac{n}{2\sigma^2}}(d-\delta^*) - \Phi^{-1}(y)\right]\right)\left[\phi\left(\Phi^{-1}(y)\right)\right]^{-1}, \quad 0 < y < 1.$$

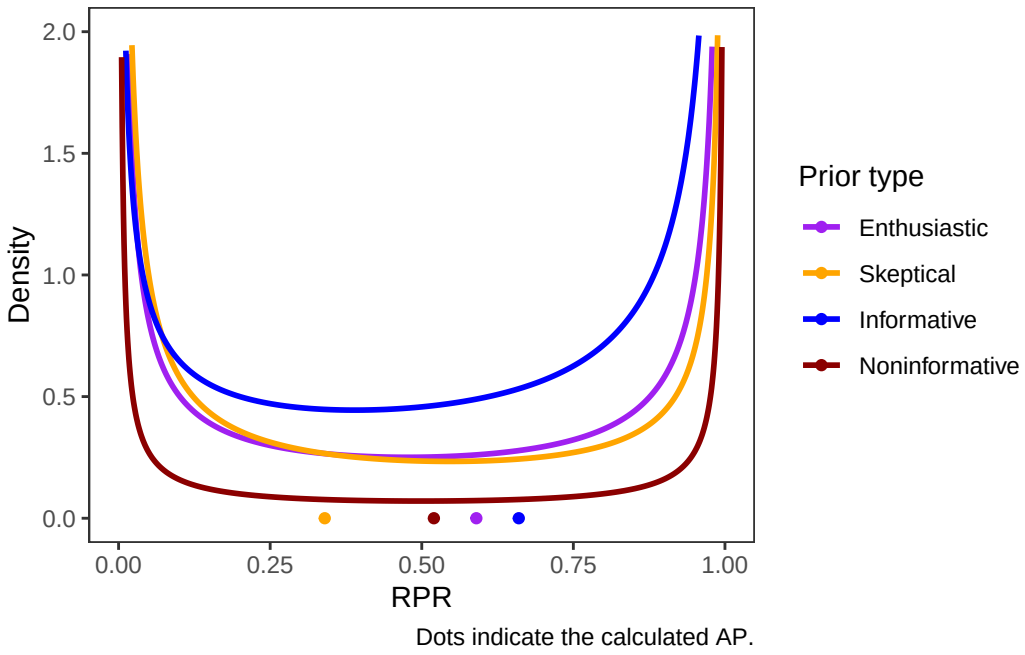
Working example (continued)

We derive the probability densities for the different assumed design priors for the SAFE-SSPE study.

```
x <- seq(0.001, 0.999, 0.001)

# Enthusiastic prior
m <- 6.6
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

## Rufibach 2016: formula (4)
y <- sqrt(m * n_0 * sd_1^2 + m * n_1 * sd_0^2)/(sqrt(n_1 * n_0 *
    (sd_0^2 + sd_1^2))) * dnorm(-sqrt(m * n_0 * sd_1^2 + m *
    n_1 * sd_0^2)/(sqrt(n_1 * n_0 * (sd_0^2 + sd_1^2))) * qnorm(1 -
    alpha) - sqrt(m)/sqrt((sd_0^2 + sd_1^2)) * (prior_mean -
    delta_star) - sqrt(m * n_0 * sd_1^2 + m * n_1 * sd_0^2)/(sqrt(n_1 *
    n_0 * (sd_0^2 + sd_1^2))) * qnorm(x)) * (dnorm(qnorm(x)))^(-1)
```

```r
data_power <- data.frame(x, y, type = "Enthusiastic")

# Skeptical prior
m <- 6.6
prior_mean <- delta_star
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

y <- sigma_treat/sigma_prior * dnorm(-qnorm(1 - alpha) * sigma_treat/sigma_prior -
    (1/sigma_prior) * (prior_mean - delta_star) - (sigma_treat/sigma_prior) *
    qnorm(x)) * (dnorm(qnorm(x)))^(-1)

data_power <- rbind(data_power, data.frame(x, y, type = "Skeptical"))

# Informative prior
m <- 25
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

y <- sigma_treat/sigma_prior * dnorm(-qnorm(1 - alpha) * sigma_treat/sigma_prior -
    (1/sigma_prior) * (prior_mean - delta_star) - (sigma_treat/sigma_prior) *
    qnorm(x)) * (dnorm(qnorm(x)))^(-1)

data_power <- rbind(data_power, data.frame(x, y, type = "Informative"))

# Noninformative prior
m <- 0.5
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
```

```r
sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

y <- sigma_treat/sigma_prior * dnorm(-qnorm(1 - alpha) * sigma_treat/sigma_prior -
    (1/sigma_prior) * (prior_mean - delta_star) - (sigma_treat/sigma_prior) *
    qnorm(x)) * (dnorm(qnorm(x)))^(-1)

data_power <- rbind(data_power, data.frame(x, y, type = "Noninformative"))

data_output2 <- data_output %>%
    select(type, x = AP) %>%
    mutate(y = 0)

data_power$type <- factor(data_power$type, levels = c("Enthusiastic",
    "Skeptical", "Informative", "Noninformative"))

ggplot(data_power, aes(x, y, colour = type)) + geom_line(linewidth = 1) +
    geom_point(data = data_output2, aes(x = x, y = y, colour = type)) +
    scale_colour_manual("Prior type", values = c("purple", "orange",
        "blue", "darkred")) + theme_bw() + theme(panel.grid = element_blank()) +
    ylab("Density") + xlab("RPR") + ylim(c(0, 2)) +
    labs(caption = "Dots indicate the calculated AP.")
```



Dots indicate the calculated AP.

The cdf of RPR is

$$P(RPR \leq y) = 1 - \Phi\left(\frac{\sigma_{treat}}{\sigma_{prior}}\left[-z_{1-\alpha} - \frac{(d-\delta^*)}{\sigma_{treat}} - \Phi^{-1}(y)\right]\right), \quad 0 < y < 1.$$

Special case

For the case that $n_0 = n_1 = n$ and $\sigma_1^2 = \sigma_0^2 = \sigma^2$, then $\sigma_{treat}^2 = \frac{2\sigma^2}{n}$ and $\sigma_{prior}^2 = \frac{2\sigma^2}{m}$, and the formula above reduces to

$$P(RPR \leq y) = 1 - \Phi\left(\frac{\sigma_{treat}}{\sigma_{prior}}\left[-z_{1-\alpha} - \frac{(d-\delta^*)}{\sigma_{treat}} - \Phi^{-1}(y)\right]\right), \quad 0 < y < 1.$$

Note that if n_0 and n_1 are the planned treatment arm sample sizes from a frequentist power calculation as described in Section 2.1.1 above, then, if $d = \delta_A$,

$$-z_{1-\alpha} - \frac{(d-\delta^*)}{\sigma_{treat}} = \Phi^{-1}(y) = \Phi^{-1}(1-\beta) \quad \Rightarrow \quad P(RPR \leq y) = 0.5, \quad y = 1-\beta.$$

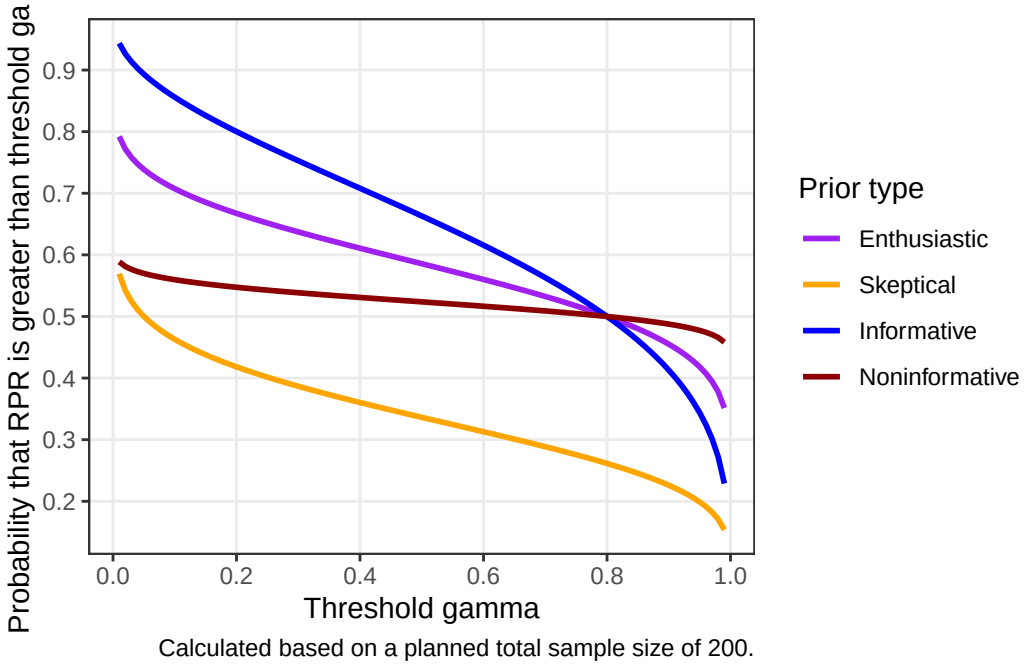
Working example (continued)

With the formula above we can calculate the probability $P(RPR > \gamma)$, $0 < \gamma < 1$, for the SAFE-SSPE study and the above specified design priors.

```
# gamma range
gamma <- seq(0.01, 0.99, 0.01)

# Centered on p1-p0
prior_mean <- 0
m_range <- c(0.5, 6.6, 25)
n_0_range <- n_0
data_rpr <- expand.grid(gamma, m_range, n_0_range)
names(data_rpr) <- c("gamma", "m_range", "n_0_range")
data_rpr$n_1_range <- a * data_rpr$n_0_range
data_rpr$sigma_treat <- sqrt(sd_0^2/data_rpr$n_0_range + sd_1^2/data_rpr$n_1_range)
data_rpr$sigma_prior <- sqrt(sd_0^2/data_rpr$m_range + sd_1^2/data_rpr$m_range)

data_rpr$p_rpr <- pnorm(-qnorm(1 - alpha) * data_rpr$sigma_treat/data_rpr$sigma_prior -
    (1/data_rpr$sigma_prior) * (prior_mean - delta_star) -
      (data_rpr$sigma_treat/data_rpr$sigma_prior) *
    qnorm(data_rpr$gamma))

data_rpr$type <- NA
data_rpr$type[data_rpr$m_range == 0.5] <- "Noninformative"
```

```r
data_rpr$type[data_rpr$m_range == 6.6] <- "Enthusiastic"
data_rpr$type[data_rpr$m_range == 25] <- "Informative"

# Centered on noninferiority margin
prior_mean <- delta_star
m_range <- c(6.6)
n_0_range <- n_0
data_rpr_skep <- expand.grid(gamma, m_range, n_0_range)
names(data_rpr_skep) <- c("gamma", "m_range", "n_0_range")
data_rpr_skep$n_1_range <- a * data_rpr_skep$n_0_range
data_rpr_skep$sigma_treat <- sqrt(sd_0^2/data_rpr_skep$n_0_range +
    sd_1^2/data_rpr_skep$n_1_range)
data_rpr_skep$sigma_prior <- sqrt(sd_0^2/data_rpr_skep$m_range +
    sd_1^2/data_rpr_skep$m_range)

data_rpr_skep$p_rpr <- pnorm(-qnorm(1 - alpha) *
                               data_rpr_skep$sigma_treat/data_rpr_skep$sigma_prior -
    (1/data_rpr_skep$sigma_prior) * (prior_mean - delta_star) -
    (data_rpr_skep$sigma_treat/data_rpr_skep$sigma_prior) * qnorm(data_rpr_skep$gamma))

data_rpr_skep$type <- "Skeptical"

data_rpr <- rbind(data_rpr, data_rpr_skep)

data_rpr$type <- factor(data_rpr$type, levels = c("Enthusiastic",
    "Skeptical", "Informative", "Noninformative"))

ggplot(data_rpr, aes(x = gamma, y = p_rpr, colour = type)) +
    geom_line(linewidth = 1) + theme_bw() + theme(panel.grid.minor = element_blank()) +
    scale_colour_manual("Prior type", values = c("purple", "orange",
```

```
    "blue", "darkred")) + scale_y_continuous(breaks = seq(0,
1, 0.1)) + scale_x_continuous(breaks = seq(0, 1, 0.2)) +
labs(caption = str_glue("Calculated based on a planned total sample size of ",
    round(n_0_range + a * n_0_range, 0), ".")) + xlab("Threshold gamma") +
ylab("Probability that RPR is greater than threshold gamma")
```

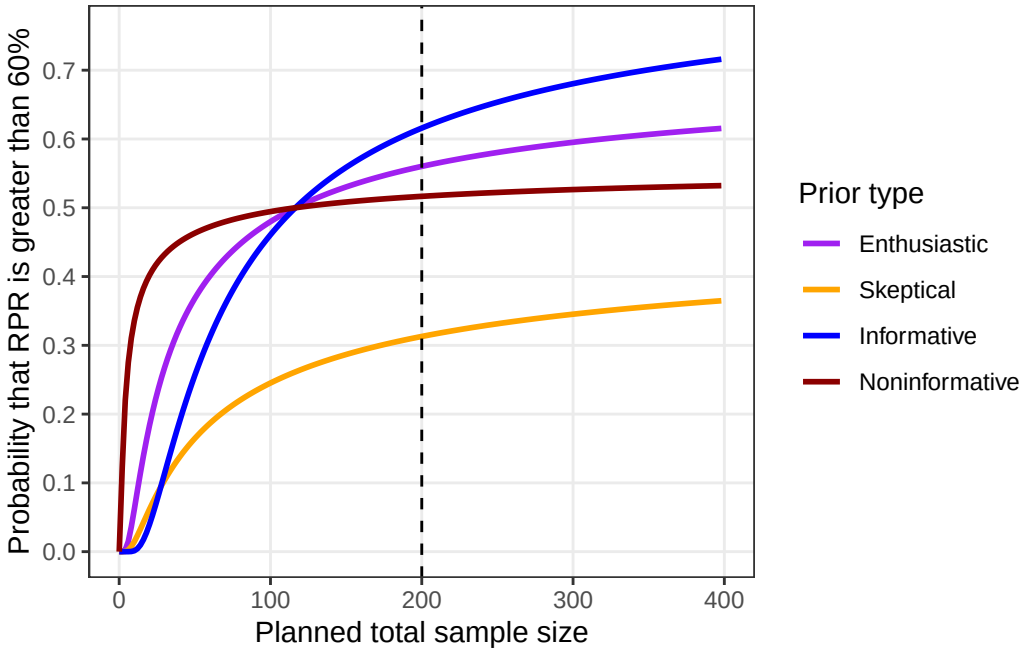


For a specific threshold, say $\gamma = 0.6$, and varying 'prior sample sizes' m we obtain

```
gamma <- 0.6

# Centered on p1-p0
prior_mean <- 0

m_range <- c(0.5, 6.6, 25)
n_0_range <- seq(0.01, 300, by = 1)

data_rpr <- expand.grid(m_range, n_0_range)
```

```r
names(data_rpr) <- c("m_range", "n_0_range")
data_rpr$n_1_range <- a * data_rpr$n_0_range
data_rpr$sigma_treat <- sqrt(sd_0^2/data_rpr$n_0_range + sd_1^2/data_rpr$n_1_range)
data_rpr$sigma_prior <- sqrt(sd_0^2/data_rpr$m_range + sd_1^2/data_rpr$m_range)

data_rpr$p_rpr <- pnorm(data_rpr$sigma_treat/data_rpr$sigma_prior *
    (-qnorm(1 - alpha) - (1/data_rpr$sigma_treat) * (prior_mean -
        delta_star) - qnorm(gamma)))

data_rpr$type <- NA
data_rpr$type[data_rpr$m_range == 0.5] <- "Noninformative"
data_rpr$type[data_rpr$m_range == 6.6] <- "Enthusiastic"
data_rpr$type[data_rpr$m_range == 25] <- "Informative"


# Centered on noninferiority margin
prior_mean <- delta_star
m_range <- c(6.6)
n_0_range <- seq(0.01, 300, by = 1)

data_rpr_skep <- expand.grid(m_range, n_0_range)
names(data_rpr_skep) <- c("m_range", "n_0_range")
data_rpr_skep$n_1_range <- a * data_rpr_skep$n_0_range
data_rpr_skep$sigma_treat <- sqrt(sd_0^2/data_rpr_skep$n_0_range +
    sd_1^2/data_rpr_skep$n_1_range)
data_rpr_skep$sigma_prior <- sqrt(sd_0^2/data_rpr_skep$m_range +
    sd_1^2/data_rpr_skep$m_range)

data_rpr_skep$p_rpr <- pnorm(data_rpr_skep$sigma_treat/data_rpr_skep$sigma_prior *
    (-qnorm(1 - alpha) - (1/data_rpr_skep$sigma_treat) * (prior_mean -
```

```
        delta_star) - qnorm(gamma)))

data_rpr_skep$type <- "Skeptical"

data_rpr <- rbind(data_rpr, data_rpr_skep)

data_rpr$type <- factor(data_rpr$type, levels = c("Enthusiastic",
    "Skeptical", "Informative", "Noninformative"))

ggplot(data_rpr, aes(x = n_1_range + n_0_range, y = p_rpr, colour = type)) +
    geom_line(linewidth = 1) + theme_bw() + theme(panel.grid.minor = element_blank()) +
    scale_colour_manual("Prior type", values = c("purple", "orange",
        "blue", "darkred")) + xlab("Planned total sample size") +
    ylab(str_glue("Probability that RPR is greater than ", gamma *
        100, "%")) + scale_x_continuous(breaks = c(0, 100, 200,
    300, 400), limits = c(0, 400)) + geom_vline(xintercept = 200,
    linetype = "dashed") + scale_y_continuous(breaks = seq(0,
    1, 0.1))
```



Note that the average power AP integrates over the whole Δ range. This might include also 'non-favorable' regions. To see that one can decompose AP as follows (see [1], [6]):

$$AP = \overbrace{P(D \leq D_{suc}^{\delta^*}, \Delta > \delta^*)}^{(1)} + \overbrace{P(D \leq D_{suc}^{\delta^*}, 0 < \Delta \leq \delta^*)}^{(2)} + \overbrace{P(D \leq D_{suc}^{\delta^*}, \Delta \leq 0)}^{(3)}$$

where

- (1) Probability of Type-I error,
- (2) 'Non-inferior region, but treatment effect not relevant',
- (3) 'Non-inferior region, treatment effect relevant'.

Spiegelhalter et al. highlight that $AP \approx P(D \leq D_{suc}^{\delta^*}, \Delta \leq 0)$ because the type-I error is often small and one has strong believe for the alternative hypothesis in designing a clinical trial [4]. [1] and [7] discuss the practical relevance of the AP decomposition. For example, pharmaceutical companies might favour (1)+(2)+(3) taking into account shortterm risk, wheras regulators are interested in (3) or (2)+(3), that is non-inferior outcomes (2)+(3) with relevant treatment effects (3).

In the non-inferiority setting we are interested in (2)+(3), that is $\Delta \leq \delta^*$:

$$P(D \leq D_{suc}^{\delta^*}, \Delta \leq \delta^*) = P(D \leq D_{suc}^{\delta^*} | \Delta \leq \delta^*) P(\Delta \leq \delta^*) = \underbrace{E\left[P_{\Delta \leq \delta^*}(D \leq D_{suc}^{\delta^*})\right]}_{EP} P(\Delta \leq \delta^*),$$

Kunzmann et al. denote EP the **'expected power'** [1]. Note that

$$\underbrace{P(D \leq D_{suc}^{\delta^*}, \Delta \leq \delta^*)}_{PAP} = \underbrace{E\left[P_{\Delta \leq \delta^*}(D \leq D_{suc}^{\delta^*})\right]}_{EP} \underbrace{P(\Delta \leq \delta^*)}_{constant}.$$

Spiegelhalter calls $P(D \leq D_{suc}^{\delta^*}, \Delta \leq \delta^*)$ the **'prior adjusted power'** (PAP).

The AP decomposition can be visualised using the posterior predictive distribution. Let $f(\tilde{\delta})$ be the pdf of the posterior predictive distribution then joint distribution of $\tilde{\delta}$ and δ is

$$f_2(\tilde{\delta}, \delta) = N_2\left(\begin{pmatrix} d \\ d \end{pmatrix}, \begin{pmatrix} \sigma_{treat}^2 + \sigma_{prior}^2 & \sigma_{prior}^2 \\ \sigma_{prior}^2 & \sigma_{prior}^2 \end{pmatrix}\right),$$

see formula (2.11) in [6].

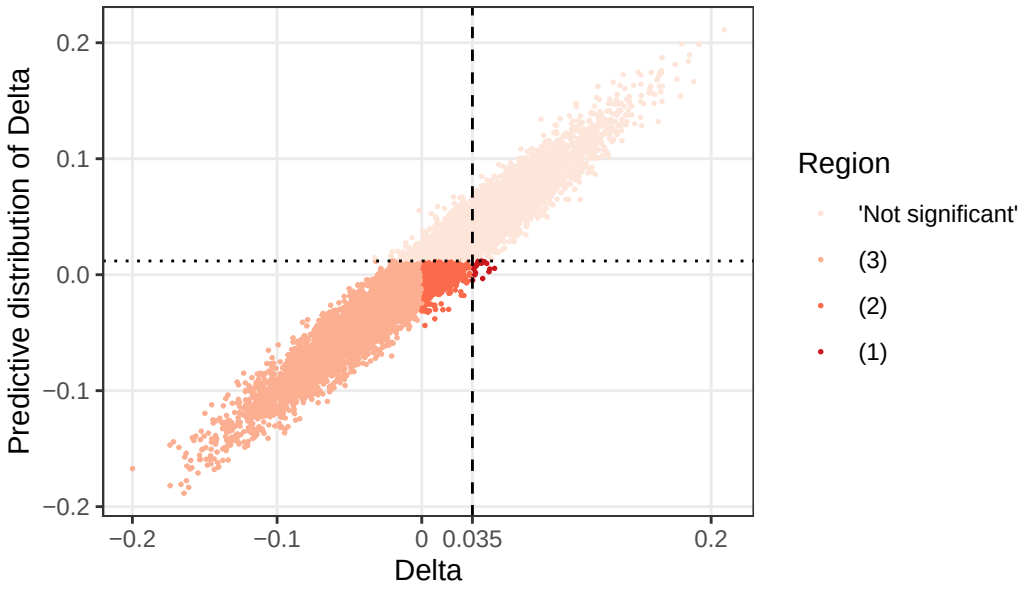
Working example (continued)

For the SAFE-SSPE study the joint distribution $f_2(\tilde{\delta}, \delta)$ under an enthusiastic prior can be plotted as:

```
library(mvtnorm)
library(tidyverse)
library(ggplot2)

delta_star <- 0.035
prior_mean <- 0
m <- 6.6

sigma_sim <- matrix(c((sd_0^2/n_0 + sd_1^2/n_1 + (sd_1^2 +
```

```
    sd_0^2)/m), (sd_1^2 + sd_0^2)/m, (sd_1^2 + sd_0^2)/m,
    (sd_1^2 + sd_0^2)/m), nrow = 2, ncol = 2, byrow = T)

set.seed(1)
data_sim <- data.frame(rmvnorm(n = 10000, mean = c(prior_mean,
    prior_mean), sigma_sim))
names(data_sim) <- c("delta_pred", "delta")

data_sim$region <- ifelse(data_sim$delta <= 0 & data_sim$delta_pred <=
    -qnorm(1 - alpha) * sqrt(sd_1^2/n_1 + sd_0^2/n_0) -
        (prior_mean - delta_star), 1, 0)
data_sim$region <- ifelse(data_sim$delta > 0 & data_sim$delta <=
    delta_star & data_sim$delta_pred <= -qnorm(1 -
    alpha) * sqrt(sd_1^2/n_1 + sd_0^2/n_0) - (prior_mean -
    delta_star), 2, data_sim$region)
data_sim$region <- ifelse(data_sim$delta > delta_star &
    data_sim$delta_pred <= -qnorm(1 - alpha) * sqrt(sd_1^2/n_1 +
        sd_0^2/n_0) - (prior_mean - delta_star), 3,
    data_sim$region)

data_sim$region <- factor(data_sim$region, levels = 0:3,
    labels = c("'Not significant'", "(3)", "(2)", "(1)"))

ggplot(data_sim, aes(x = delta, y = delta_pred, colour = factor(region))) +
    geom_point(size = 0.3) + theme_bw() + theme(panel.grid.minor = element_blank()) +
    geom_vline(xintercept = 0.035, linetype = "dashed") +
    geom_hline(yintercept = -qnorm(1 - alpha) * sqrt(sd_1^2/n_1 +
        sd_0^2/n_0) - (prior_mean - delta_star), linetype = "dotted") +
    ylab("Predictive distribution of Delta") + xlab("Delta") +
    scale_color_brewer("Region", palette = "Reds") +
```

```
    scale_x_continuous(breaks = c(-0.2, -0.1, 0, 0.035,
        0.2), labels = c(-0.2, -0.1, 0, 0.035, 0.2)) +
    labs(caption = "Dashed line: Noninferiority margin.\nDotted line: Upper CI from predicted c
```



Dashed line: Noninferiority margin.
Dotted line: Upper CI from predicted delta <= non−inferiority margin.

```
data_output <- data.frame(data_sim %>%
    group_by(region) %>%
    summarise(prop = n()/nrow(data_sim)))

data_output
```

```
             region   prop
1 'Not significant' 0.4148
2               (3) 0.4909
3               (2) 0.0920
4               (1) 0.0023
```

From the above values the average power can be calculated as (1) 0.0023 + (2) 0.092 + (3) 0.4909, which is equal to 0.5852. This is the value we reported above (59%).

The results of EP, PAP and AP are shown in the table below:

```
delta_star <- 0.035
prior_mean <- 0
m <- 6.6

set.seed(1)

sd_prior <- sqrt((sd_1^2 + sd_0^2))/sqrt(m)
draws <- rnorm(10000, mean = prior_mean, sd = sd_prior)

power_classic <- pnorm(-qnorm(1 - alpha) - sqrt(n_0 * n_1)/sqrt((n_0 *
    sd_1^2 + n_1 * sd_0^2)) * (draws - delta_star))

data_ep <- data.frame(ep = mean(power_classic[draws <= delta_star]),
    pap = mean(power_classic[draws <= delta_star]) * pnorm(delta_star,
        mean = prior_mean, sd = sd_prior), ap = mean(power_classic),
    const = pnorm(delta_star, mean = prior_mean, sd = sd_prior),
    type = "Skeptical")

set.seed(1)
prior_mean <- delta_star
m <- 6.6
draws <- rnorm(10000, mean = prior_mean, sd = sd_prior)

power_classic <- pnorm(-qnorm(1 - alpha) - sqrt(n_0 * n_1)/sqrt((n_0 *
    sd_1^2 + n_1 * sd_0^2)) * (draws - delta_star))

data_ep <- rbind(data_ep, data.frame(ep = mean(power_classic[draws <=
    delta_star]), pap = mean(power_classic[draws <= delta_star]) *
```

```
    pnorm(delta_star, mean = prior_mean, sd = sd_prior), ap = mean(power_classic),
    const = pnorm(delta_star, mean = prior_mean, sd = sd_prior),
    type = "Enthusiastic"))


prior_mean <- 0
m <- 25
set.seed(1)
draws <- rnorm(10000, mean = prior_mean, sd = sd_prior)

power_classic <- pnorm(-qnorm(1 - alpha) - sqrt(n_0 * n_1)/sqrt((n_0 *
    sd_1^2 + n_1 * sd_0^2)) * (draws - delta_star))

data_ep <- rbind(data_ep, data.frame(ep = mean(power_classic[draws <=
    delta_star]), pap = mean(power_classic[draws <= delta_star]) *
    pnorm(delta_star, mean = prior_mean, sd = sd_prior), ap = mean(power_classic),
    const = pnorm(delta_star, mean = prior_mean, sd = sd_prior),
    type = "Informative"))

prior_mean <- 0
m <- 0.5
set.seed(1)
draws <- rnorm(1e+05, mean = prior_mean, sd = sd_tilde/sqrt(0.5))

power_classic <- pnorm(-qnorm(1 - alpha) - sqrt(n_0 * n_1)/sqrt((n_0 *
    sd_1^2 + n_1 * sd_0^2)) * (draws - delta_star))

data_ep <- rbind(data_ep, data.frame(ep = mean(power_classic[draws <=
    delta_star]), pap = mean(power_classic[draws <= delta_star]) *
    pnorm(delta_star, mean = prior_mean, sd = sd_prior), ap = mean(power_classic),
```

```
    const = pnorm(delta_star, mean = prior_mean, sd = sd_prior),
    type = "Noninformative"))


data_ep$ep <- round(data_ep$ep, 4)
data_ep$ap <- round(data_ep$ap, 4)
data_ep$pap <- round(data_ep$pap, 4)
data_ep$const <- round(data_ep$const, 4)

data_ep
```

```
      ep    pap     ap  const           type
1 0.7930 0.5857 0.5872 0.7386      Skeptical
2 0.6740 0.3370 0.3437 0.5000   Enthusiastic
3 0.7930 0.5857 0.5872 0.7386    Informative
4 0.9186 0.6785 0.5233 0.7386 Noninformative
```

2.2.1.3 Bayesian approach: Prior on the risk difference

In this section we use a proper Bayesian approach for the calculation of the probability to reject the null hypothesis $H_0 : \delta > \delta^*$, where $\delta = p_1 - p_0$ ([4], [10]). Suppose that $\Delta \sim N(d, \sigma^2_{prior})$ is an analysis prior for the treatment effect δ. Remember that $D = \overline{p}_1 - \overline{p}_0$ is the estimated (observed) treatment effect with $D \sim N(\delta, \sigma^2_{treat})$. A 'trial success' can then be defined as

$$P(\Delta \leq \delta^* | D) > 1 - \epsilon,$$

where ϵ is small, say $\epsilon = 0.05$ ([4], [10]). Using the above specified prior and likelihood of the data, the posterior distribution is given as

$$\Delta|D \sim N\left(\frac{\frac{d}{\sigma^2_{prior}} + \frac{D}{\sigma^2_{treat}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{treat}}}, \frac{1}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{treat}}}\right) = N\left(\frac{\sigma^2_{treat}d + \sigma^2_{prior}D}{\sigma^2_{treat} + \sigma^2_{prior}}, \frac{\sigma^2_{treat}\sigma^2_{prior}}{\sigma^2_{treat} + \sigma^2_{prior}}\right),$$

see for example [9].

Special case

If $n_0 = n_1 = n$ and $\sigma_1^2 = \sigma_0^2 = \sigma^2$, then $\sigma^2_{treat} = \frac{2\sigma^2}{n}$ and $\sigma^2_{prior} = \frac{2\sigma^2}{m}$, and the formula above reduces to

$$\Delta|D \sim N\left(\frac{m\delta + nD}{n+m}, \frac{2\sigma^2}{n+m}\right).$$

The trial is successful if

$$\frac{\sigma^2_{treat}d + \sigma^2_{prior}D}{\sigma^2_{treat} + \sigma^2_{prior}} + z_{1-\epsilon}\frac{\sigma_{treat}\sigma_{prior}}{\sqrt{\sigma^2_{treat} + \sigma^2_{prior}}} \leq \delta^*,$$

where the left side of the above equation is the upper $(1-\epsilon)$-credible interval. A simple algebraic step gives

$$D \leq -z_{1-\epsilon}\frac{\sigma_{treat}}{\sigma_{prior}}\sqrt{\sigma^2_{treat} + \sigma^2_{prior}} + \delta^*\left(1 + \frac{\sigma^2_{treat}}{\sigma^2_{prior}}\right) - \frac{\sigma^2_{treat}}{\sigma^2_{prior}}d.$$

Here, $D^{d,\delta^*}_{suc} = -z_{1-\epsilon}\frac{\sigma_{treat}}{\sigma_{prior}}\sqrt{\sigma^2_{treat} + \sigma^2_{prior}} + \delta^*\left(1 + \frac{\sigma^2_{treat}}{\sigma^2_{prior}}\right) - \frac{\sigma^2_{treat}}{\sigma^2_{prior}}d$ is the **required risk difference** for a successful rejection of the null hypothesis taking into account the uncertainty of δ.

Since $D \sim N(\delta, \sigma^2_{treat})$,

$$P(D \leq D^{d,\delta^*}_{suc}|\delta) = \Phi\left(-z_{1-\epsilon}\sqrt{1 + \frac{\sigma^2_{treat}}{\sigma^2_{prior}}} - \frac{1}{\sigma_{treat}}\left[\delta - \delta^*\left\{1 + \frac{\sigma^2_{treat}}{\sigma^2_{prior}}\right\}\right] - \frac{\sigma_{treat}}{\sigma^2_{prior}}d\right).$$

This is the **conditional Bayesian probability to reject** (or Bayesian power) under an (assumed) *known* true treatment effect δ.

Special case

If $n_0 = n_1 = n$ and $\sigma_1^2 = \sigma_0^2 = \sigma^2$, then $\sigma_{treat}^2 = \frac{2\sigma^2}{n}$ and $\sigma_{prior}^2 = \frac{2\sigma^2}{m}$, and the formula above reduces to

$$P(D \leq D_{suc}^{d,\delta^*}|\delta) = \Phi\left(-z_{1-\epsilon}\sqrt{\frac{n+m}{n}} - \frac{m}{\sqrt{2n\sigma^2}}d - \sqrt{\frac{n}{2n\sigma^2}}\left\{\delta - \delta^*\left(\frac{n+m}{n}\right)\right\}\right).$$

Working example (continued)

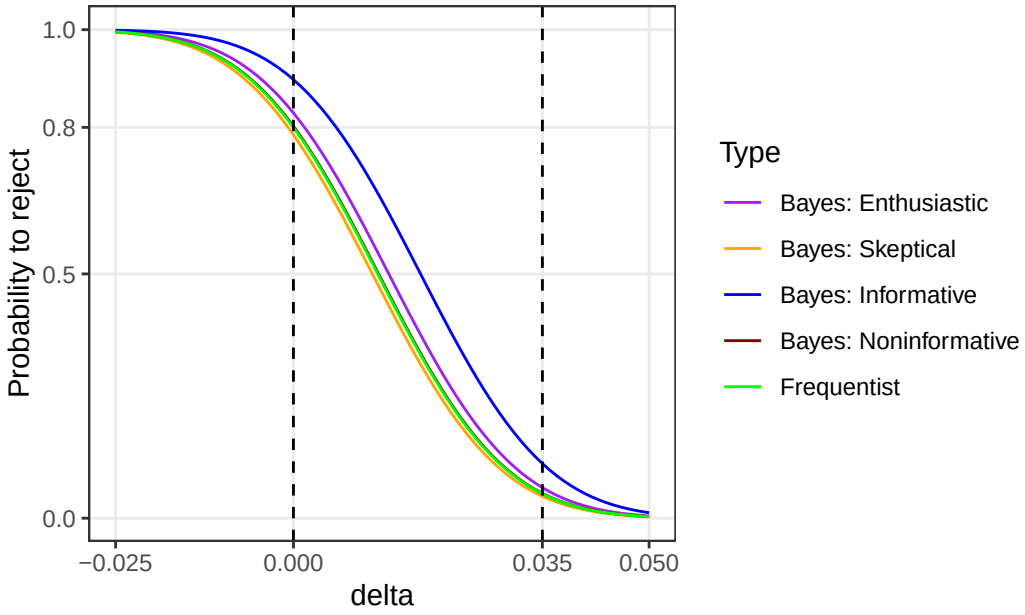
We derive the conditional Bayesian probability to reject for the SAFE-SSPE study for different δ values and analysis priors.

```
# Delta range
delta <- seq(-0.025, 0.05, 0.001)

# Epsilon
epsilon <- 0.05

# Enthusiastic prior
prior_mean <- 0
m <- 6.6

ap_bayes <- pnorm(-qnorm(1 - epsilon) * (sqrt(1 + m * (n_0 *
    sd_1^2 + n_1 * sd_0^2)/(n_1 * n_0 * (sd_1^2 + sd_0^2)))) -
    (m * sqrt(n_0 * sd_1^2 + n_1 * sd_0^2))/(sqrt(n_1 * n_0) *
```

```
        (sd_1^2 + sd_0^2)) * prior_mean - sqrt((n_1 * n_0)/((n_0 *
    sd_1^2 + n_1 * sd_0^2))) * (delta - delta_star * (1 + (m *
    (n_0 * sd_1^2 + n_1 * sd_0^2))/(n_1 * n_0 * (sd_1^2 + sd_0^2)))))

data_power <- data.frame(delta, y = ap_bayes, type = "Bayes: Enthusiastic")

# Skeptical prior
prior_mean <- 0.035
m <- 6.6

ap_bayes <- pnorm(-qnorm(1 - epsilon) * (sqrt(1 + m * (n_0 *
    sd_1^2 + n_1 * sd_0^2)/(n_1 * n_0 * (sd_1^2 + sd_0^2)))) -
    (m * sqrt(n_0 * sd_1^2 + n_1 * sd_0^2))/(sqrt(n_1 * n_0) *
        (sd_1^2 + sd_0^2)) * prior_mean - sqrt((n_1 * n_0)/((n_0 *
    sd_1^2 + n_1 * sd_0^2))) * (delta - delta_star * (1 + (m *
    (n_0 * sd_1^2 + n_1 * sd_0^2))/(n_1 * n_0 * (sd_1^2 + sd_0^2)))))

data_power <- rbind(data_power, data.frame(delta, y = ap_bayes,
    type = "Bayes: Skeptical"))

# Informative prior
prior_mean <- 0
m <- 25

ap_bayes <- pnorm(-qnorm(1 - epsilon) * (sqrt(1 + m * (n_0 *
    sd_1^2 + n_1 * sd_0^2)/(n_1 * n_0 * (sd_1^2 + sd_0^2)))) -
    (m * sqrt(n_0 * sd_1^2 + n_1 * sd_0^2))/(sqrt(n_1 * n_0) *
        (sd_1^2 + sd_0^2)) * prior_mean - sqrt((n_1 * n_0)/((n_0 *
    sd_1^2 + n_1 * sd_0^2))) * (delta - delta_star * (1 + (m *
    (n_0 * sd_1^2 + n_1 * sd_0^2))/(n_1 * n_0 * (sd_1^2 + sd_0^2)))))
```

```
data_power <- rbind(data_power, data.frame(delta, y = ap_bayes,
    type = "Bayes: Informative"))

# Noninformative prior
prior_mean <- 0
m <- 0.5

ap_bayes <- pnorm(-qnorm(1 - epsilon) * (sqrt(1 + m * (n_0 *
    sd_1^2 + n_1 * sd_0^2)/(n_1 * n_0 * (sd_1^2 + sd_0^2)))) -
    (m * sqrt(n_0 * sd_1^2 + n_1 * sd_0^2))/(sqrt(n_1 * n_0) *
        (sd_1^2 + sd_0^2)) * prior_mean - sqrt((n_1 * n_0)/((n_0 *
    sd_1^2 + n_1 * sd_0^2))) * (delta - delta_star * (1 + (m *
    (n_0 * sd_1^2 + n_1 * sd_0^2))/(n_1 * n_0 * (sd_1^2 + sd_0^2)))))

data_power <- rbind(data_power, data.frame(delta, y = ap_bayes,
    type = "Bayes: Noninformative"))

# Frequentist
y_freq <- pnorm(-(delta - delta_star) * sqrt(n_1 * n_0)/sqrt(n_0 *
    sd_1^2 + n_1 * sd_0^2) - qnorm(1 - alpha))
data_power <- rbind(data_power, data.frame(delta, y = y_freq,
    type = "Frequentist"))

data_power$type <- factor(data_power$type, levels = c("Bayes: Enthusiastic",
    "Bayes: Skeptical", "Bayes: Informative", "Bayes: Noninformative",
    "Frequentist"))


ggplot(data_power, aes(x = delta, y = y, colour = type)) + geom_line() +
```

```
    theme_bw() + theme(panel.grid.minor = element_blank()) +
    scale_colour_manual("Type", values = c("purple", "orange",
        "blue", "darkred", "green")) + ylab("Probability to reject") +
    ylab("Probability to reject") + scale_x_continuous(breaks = c(-0.025,
    0, 0.035, 0.05)) + scale_y_continuous(breaks = c(0, 0.5,
    0.8, 1)) + geom_vline(xintercept = 0, linetype = "dashed") +
    geom_vline(xintercept = 0.035, linetype = "dashed") +
    labs(caption=str_glue("For treatment arm sample sizes: n0=",
    round(n_0, 0), ", n1=", round(n_1, 0)))
```




```
data_power$y <- round(data_power$y, 2)
data_power %>%
    filter(delta %in% c("0", "0.035"))
```

```
  delta    y                   type
1 0.000 0.83   Bayes: Enthusiastic
2 0.035 0.06   Bayes: Enthusiastic
```

```
3  0.000 0.78      Bayes: Skeptical
4  0.035 0.04      Bayes: Skeptical
5  0.000 0.90    Bayes: Informative
6  0.035 0.11    Bayes: Informative
7  0.000 0.80 Bayes: Noninformative
8  0.035 0.05 Bayes: Noninformative
9  0.000 0.80           Frequentist
10 0.035 0.05           Frequentist
```

The Bayesian probability to reject $H_0 : \delta > \delta^*$ given $\delta = p_1 - p_0 = 0$ and an informative prior is 90%. This is higher than the frequentist probability to reject of 80% given $\delta = p_1 - p_0 = 0$. In contrast, the probability to reject $H_0 : \delta > \delta^*$ for $\delta = 0.035$ under an informative prior is 11%. This is higher than the frequentist 'type-I error' of 5%. The probability to reject under the null hypothesis is higher for the Bayesian approach under an informative prior because the prior has substantial believe in non-inferiority.

Suppose $\tilde{D}$ is calculated from a future sample given we have already observed D. The posterior predictive distribution of $\tilde{D}$ is Gaussian distributed $\tilde{D} \sim N(d, \sigma^2_{treat} + \sigma^2_{prior})$ under an analysis prior $\Delta \sim N(d, \sigma^2_{prior})$. The **Bayesian average (marginal) probability to reject** (BAP) can be calcu-

lated as

$$
\begin{aligned}
BAP &= \int_{-\infty}^{-\infty} P(D \leq D_{suc}^{d,\delta^*}|\delta) f(\delta) d\delta \\
&= \int_{-\infty}^{-\infty} \left\{ \int_{-\infty}^{-z_{1-\epsilon}\frac{\sigma_{treat}}{\sigma_{prior}}\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}+\delta^*\left(1+\frac{\sigma_{treat}^2}{\sigma_{prior}^2}\right)-\frac{\sigma_{treat}^2}{\sigma_{prior}^2}d} f(\tilde{\delta}|\delta) d\tilde{\delta} \right\} f(\delta) d\delta \\
&= \int_{-\infty}^{-z_{1-\epsilon}\frac{\sigma_{treat}}{\sigma_{prior}}\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}+\delta^*\left(1+\frac{\sigma_{treat}^2}{\sigma_{prior}^2}\right)-\frac{\sigma_{treat}^2}{\sigma_{prior}^2}d} \left\{ \int_{-\infty}^{-\infty} f(\tilde{\delta}|\delta) f(\delta) d\delta \right\} d\tilde{\delta} \\
&= \int_{-\infty}^{-z_{1-\epsilon}\frac{\sigma_{treat}}{\sigma_{prior}}\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}+\delta^*\left(1+\frac{\sigma_{treat}^2}{\sigma_{prior}^2}\right)-\frac{\sigma_{treat}^2}{\sigma_{prior}^2}d} f(\tilde{\delta}) d\tilde{\delta} \\
&= \Phi\left(-z_{1-\epsilon}\frac{\sigma_{treat}}{\sigma_{prior}}\frac{\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}}{\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}} + \frac{\delta^*}{\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}}\left\{1+\frac{\sigma_{treat}^2}{\sigma_{prior}^2}\right\} \right. \\
&\qquad \left. -\frac{d}{\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}}\left\{1+\frac{\sigma_{treat}^2}{\sigma_{prior}^2}\right\} \right) \\
&= \Phi\left(-z_{1-\epsilon}\frac{\sigma_{treat}}{\sigma_{prior}} - \frac{\sqrt{\sigma_{treat}^2+\sigma_{prior}^2}}{\sigma_{prior}^2}(d-\delta^*) \right).
\end{aligned}
$$

Special case

If $n_0 = n_1 = n$ and $\sigma_1^2 = \sigma_0^2 = \sigma^2$, then $\sigma_{treat}^2 = \frac{2\sigma^2}{n}$ and $\sigma_{prior}^2 = \frac{2\sigma^2}{m}$, and the formula above reduces to

$$
BAP = -z_{1-\epsilon}\sqrt{\frac{m}{n}} - \sqrt{\frac{(m+n)m}{2n\sigma^2}}(d-\delta^*)
$$

Working example (continued)

We derive the Bayesian average probability to reject for the SAFE-SSPE study.

```r
### Hybrid AP

sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))
digit_round <- 4

# Enthusiastic prior
m <- 6.6
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- data.frame(type = "Enthusiastic", n_0, n_1, AP = round(AP,
    digit_round))

# Skeptical prior
m <- 6.6
prior_mean <- delta_star
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- rbind(data_output, data.frame(type = "Skeptical",
    n_0, n_1, AP = round(AP, digit_round)))

# Informative prior
m <- 25
prior_mean <- 0
```

```r
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- rbind(data_output, data.frame(type = "Informative",
    n_0, n_1, AP = round(AP, digit_round)))

# Noninformative prior
m <- 0.5
prior_mean <- 0
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)

AP <- pnorm(1/sigma_prior * (-qnorm(1 - alpha) * sigma_treat -
    (prior_mean - delta_star)))

data_output <- rbind(data_output, data.frame(type = "Noninformative",
    n_0, n_1, AP = round(AP, digit_round)))

data_output <- data_output %>%
    select(type, n_0, n_1, AP)

### Bayesian AP

sigma_treat <- sqrt((sd_0^2/n_0 + sd_1^2/n_1))

# Enthusiastic prior
prior_mean <- 0
m <- 6.6
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)
```

```r
ap_bayes <- pnorm(-qnorm(1 - alpha) * sigma_treat/sigma_prior -
    sqrt(sigma_treat^2 + sigma_prior^2)/sigma_prior^2 * (prior_mean -
        delta_star))

data_output2 <- data.frame(type = "Enthusiastic", n_0, n_1, AP_bayes = round(ap_bayes,
    digit_round))

# Skeptical prior
prior_mean <- delta_star
m <- 6.6
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)

ap_bayes <- pnorm(-qnorm(1 - alpha) * sigma_treat/sigma_prior -
    sqrt(sigma_treat^2 + sigma_prior^2)/sigma_prior^2 * (prior_mean -
        delta_star))

data_output2 <- rbind(data_output2, data.frame(type = "Skeptical",
    n_0, n_1, AP_bayes = round(ap_bayes, digit_round)))

# Informative prior
prior_mean <- 0
m <- 25
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)

ap_bayes <- pnorm(-qnorm(1 - alpha) * sigma_treat/sigma_prior -
    sqrt(sigma_treat^2 + sigma_prior^2)/sigma_prior^2 * (prior_mean -
        delta_star))

data_output2 <- rbind(data_output2, data.frame(type = "Informative",
```

```
    n_0, n_1, AP_bayes = round(ap_bayes, digit_round)))

# Noninformative prior
prior_mean <- 0
m <- 0.5
sigma_prior <- sqrt((sd_0^2 + sd_1^2))/sqrt(m)

ap_bayes <- pnorm(-qnorm(1 - alpha) * sigma_treat/sigma_prior -
    sqrt(sigma_treat^2 + sigma_prior^2)/sigma_prior^2 * (prior_mean -
        delta_star))

data_output2 <- rbind(data_output2, data.frame(type = "Noninformative",
    n_0, n_1, AP_bayes = round(ap_bayes, digit_round)))

data_output <- left_join(data_output, data_output2 %>%
    select(type, AP_bayes), by = "type")

data_output %>%
    select(type, n_0, n_1, AP_hybrid = AP, AP_bayes)
```

```
            type      n_0      n_1 AP_hybrid AP_bayes
1   Enthusiastic 99.93031 99.93031    0.5856   0.5937
2      Skeptical 99.93031 99.93031    0.3363   0.3363
3    Informative 99.93031 99.93031    0.6631   0.7149
4 Noninformative 99.93031 99.93031    0.5237   0.5239
```

The average Bayesian power under an informative prior is 72% compared to 66% from a hybrid approach.

2.2.1.4 Hybrid approach: Prior on event probabilities

In the former subsections we assumed that the prior was directly specified on the treatment effect δ. In this subsection design priors are specified on the event probabilities p_i, $i = \{0, 1\}$. Suppose that these design priors are beta distributed $\pi_i \sim Beta(a_i, b_i)$, $i = \{0, 1\}$. It is well known (see for example [9]) that the prior predictive distribution of the number of events R_i, $i = \{0, 1\}$, from a future binomial sample of size n_i, $i = \{0, 1\}$, is beta-binomial distributed $R_i \sim betabinom(r_i|n_i, a_i, b_i)$ with

$$E(R_i) = n_i \frac{a_i}{a_i + b_i}, \quad Var(R_i) = \frac{n_i a_i b_i (n_i + a_i + b_i)}{(a_i + b_i)^2 (a_i + b_i + 1)}.$$

Working example (continued)

For the SAFE-SSPE trial we specify the following design beta priors for the expected event probabilites $p_1 = p_0 = 0.01$.

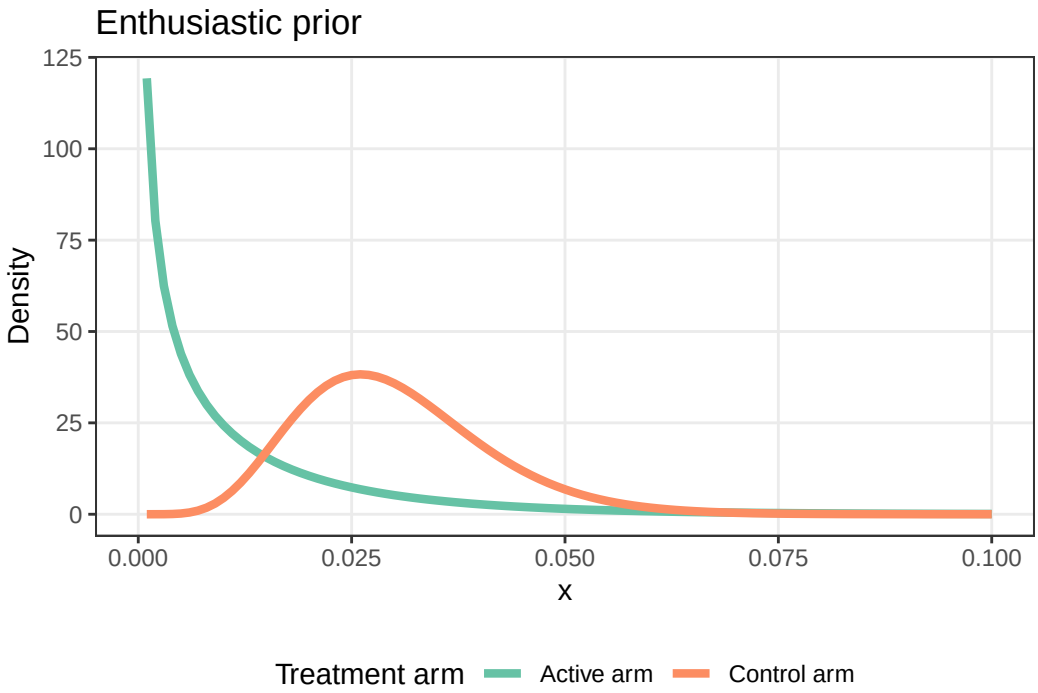
Enthusiastic prior: The beta prior for the active treatment arm is centered at 0.01 (the expected event proportion) with $P(\pi_1 > 0.05) = 0.025$, that is, the probability that the prior is greater than the safety margin is 2.5%. The mean of the beta prior for the control arm is centered at 0.03 (the safety margin) with $P(\pi_0 > 0.05) = 0.05$.

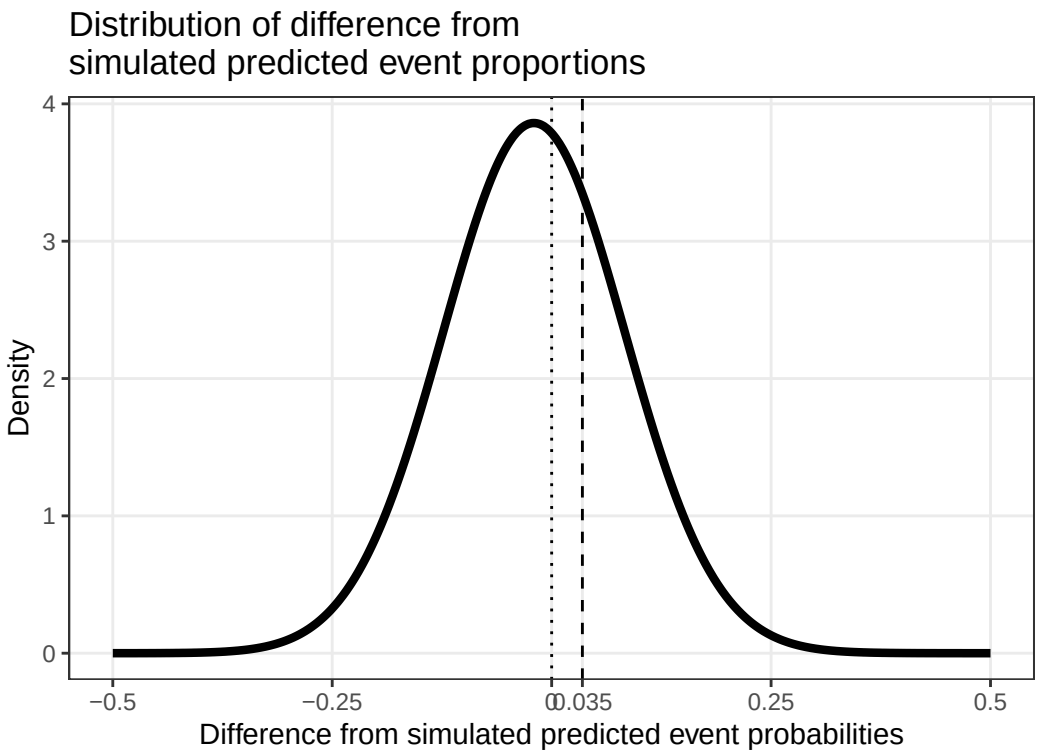
- Parameters of beta prior for active treatment arm: $a_1 = 0.5$, $b_1 = 49.5$.
- Parameters of beta prior for control treatment arm: $a_0 = 7.2$, $b_0 = 232.8$.

```
library(extraDistr)

x <- seq(0.001, 0.999, 0.001)

## Enthusiastic prior
a_1 <- 0.5
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 7.2
b_0 <- (1 - 0.03)/0.03 * a_0
```

```r
prior_0 <- dbeta(x, a_0, b_0)
prior_1 <- dbeta(x, a_1, b_1)

prior_plot <- data.frame(x, prior_0, prior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0,
    prior_1))
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name,
    "_0") == T, "Control arm", "Active arm")

fig1 <- ggplot(prior_plot_long, aes(x = x, y = value,
    colour = group)) + geom_line(linewidth = 1.5) +
    theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "horizontal") +
    scale_x_continuous(limits = c(0, 0.1)) + ylab("Density") +
    labs(caption = str_glue("Assumed event probabilities: Control arm (p_0=",
        p_0, "), ", "Active arm (p_1=", p_1, ")")) +
    scale_color_brewer("Treatment arm", palette = "Set2") +
    scale_linetype("Type") + ggtitle("Enthusiastic prior")

diff_prior <- data.frame(x = rbbinom(1e+06, size = ceiling(n_1),
    a_1, b_1)/ceiling(n_1) - rbbinom(1e+06, size = ceiling(n_0),
    a_0, b_0)/ceiling(n_0))

fig2 <- ggplot(diff_prior, aes(x), group = x) + geom_density(linewidth = 1.5,
    bw = 0.1) + theme_bw() + theme(panel.grid.minor = element_blank()) +
    xlab("Difference from simulated predicted event probabilities") +
    ggtitle("Distribution of difference from\nsimulated predicted event proportions") +
    ylab("Density") + geom_vline(xintercept = delta_star,
    linetype = "dashed") + geom_vline(xintercept = 0,
```

```
    linetype = "dotted") + scale_x_continuous(breaks = c(-0.5,
    -0.25, 0, 0.035, 0.25, 0.5), labels = c(-0.5, -0.25,
    0, 0.035, 0.25, 0.5), limits = c(-0.5, 0.5))

cowplot::plot_grid(fig1, fig2, ncol = 1)
```





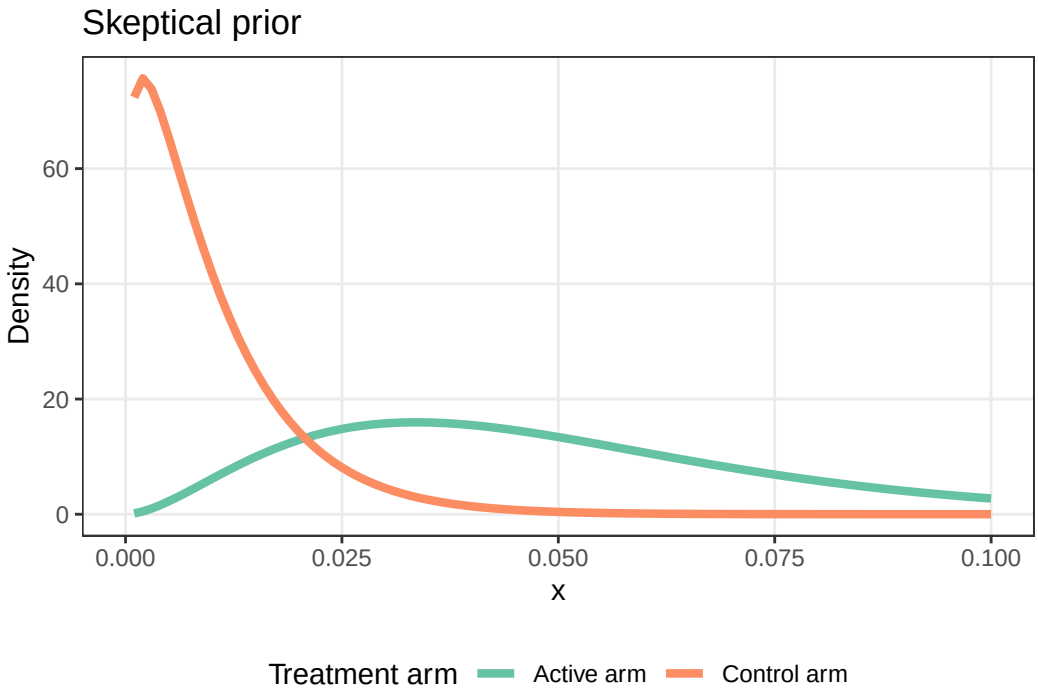
Skeptical prior: The beta prior for the active treatment arm is centered at 0.05 (the safety margin) with

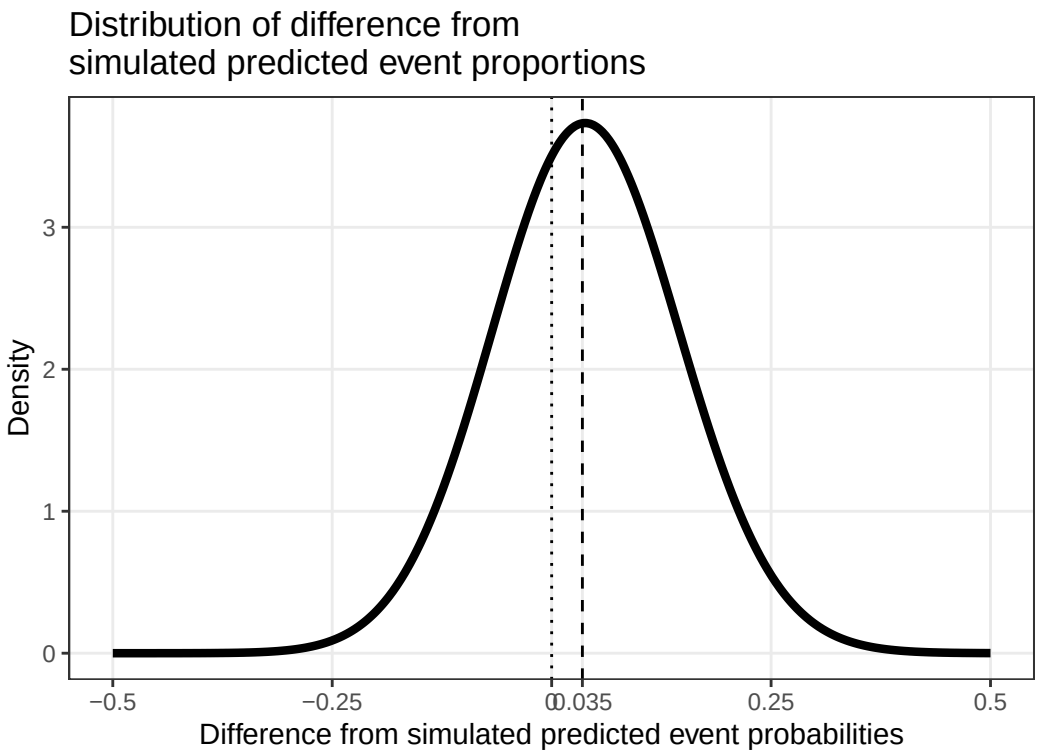
$P(\pi_1 \leq 0.01) = 0.025$, that is the probability that the prior is smaller or equal than the expected event proportion is 2.5%. The mean of the beta prior for the control arm is is centered at 0.01 (the expected event proportion) with $P(\pi_0 > 0.05) = 0.025$.

- Parameters of beta prior for active treatment arm: $a_1 = 2.84$, $b_1 = 53.96$.
- Parameters of beta prior for control treatment arm: $a_i = 1.24$, $b_i = 122.76$.

```
x <- seq(0.001, 0.999, 0.001)

# Skeptical prior
a_1 <- 2.84
b_1 <- (1 - 0.05)/0.05 * a_1
a_0 <- 1.24
b_0 <- (1 - 0.01)/0.01 * a_0

prior_0 <- dbeta(x, a_0, b_0)
prior_1 <- dbeta(x, a_1, b_1)

prior_plot <- data.frame(x, prior_0, prior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0,
    prior_1))
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name,
    "_0") == T, "Control arm", "Active arm")

fig1 <- ggplot(prior_plot_long, aes(x = x, y = value,
    colour = group)) + geom_line(linewidth = 1.5) +
    theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "horizontal") +
    scale_x_continuous(limits = c(0, 0.1)) + ylab("Density") +
    labs(caption = str_glue("Assumed event probabilities: Control arm (p_0=",
        p_0, "), ", "Active arm (p_1=", p_1, ")")) +
```

```
    scale_color_brewer("Treatment arm", palette = "Set2") +
    scale_linetype("Type") + ggtitle("Skeptical prior")

diff_prior <- data.frame(x = rbbinom(1e+06, size = ceiling(n_1),
    a_1, b_1)/ceiling(n_1) - rbbinom(1e+06, size = ceiling(n_0),
    a_0, b_0)/ceiling(n_0))

fig2 <- ggplot(diff_prior, aes(x), group = x) + geom_density(linewidth = 1.5,
    bw = 0.1) + theme_bw() + theme(panel.grid.minor = element_blank()) +
    xlab("Difference from simulated predicted event probabilities") +
    ggtitle("Distribution of difference from\nsimulated predicted event proportions") +
    ylab("Density") + geom_vline(xintercept = delta_star,
    linetype = "dashed") + geom_vline(xintercept = 0,
    linetype = "dotted") + scale_x_continuous(breaks = c(-0.5,
    -0.25, 0, 0.035, 0.25, 0.5), labels = c(-0.5, -0.25,
    0, 0.035, 0.25, 0.5), limits = c(-0.5, 0.5))

cowplot::plot_grid(fig1, fig2, ncol = 1)
```





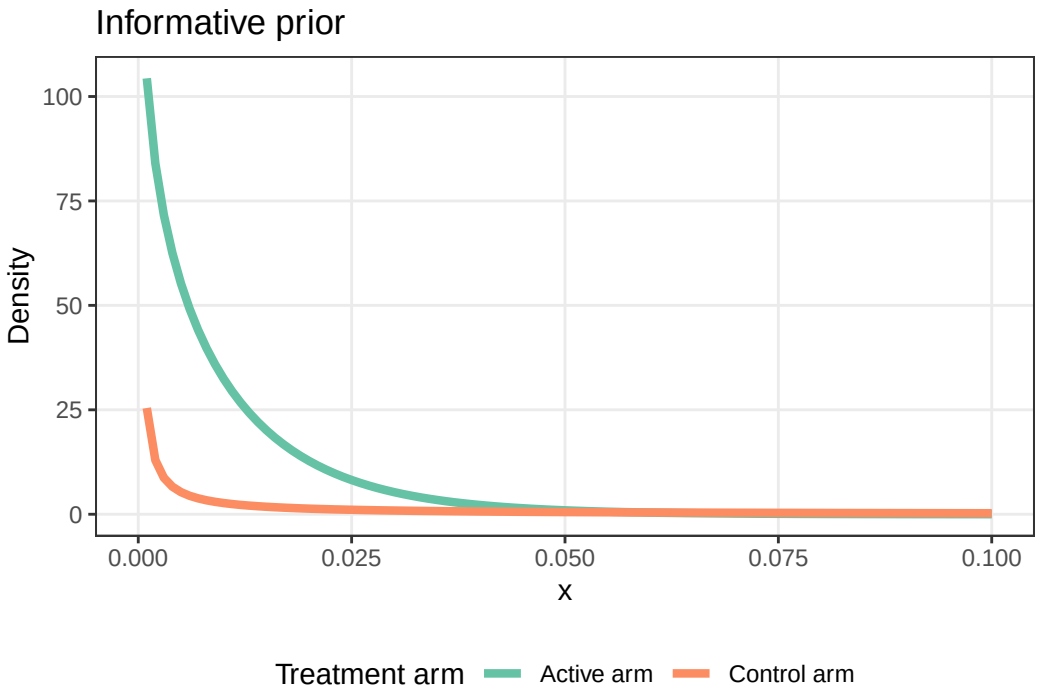
Informative prior: The beta prior for the active treatment arm and control arm is centered at 0.01 (the

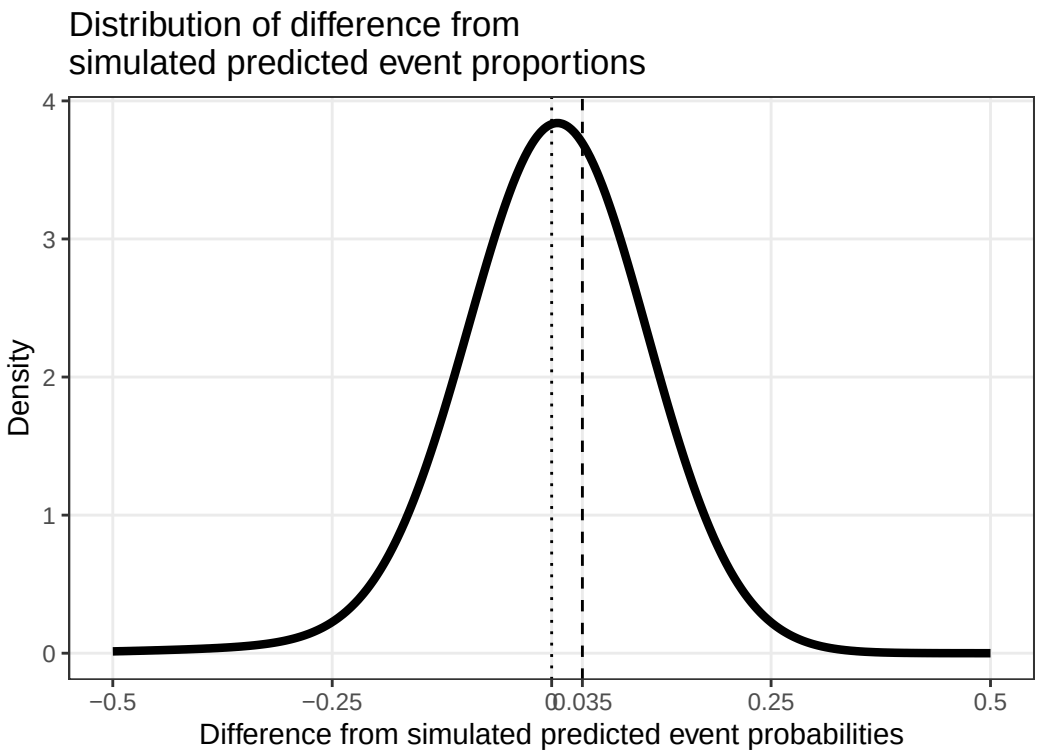
expected event proportion) with $P(\pi_1 \leq 0.05) = 0.01$ and $P(\pi_1 \leq 0.05) = 0.05$.

- Parameters of beta prior for active treatment arm: $a_1 = 0.8$, $b_1 = 79.2$.
- Parameters of beta prior for control treatment arm: $a_0 = 0.03$, $b_0 = 2.97$.

```
x <- seq(0.001, 0.999, 0.001)

## Informative prior
a_1 <- 0.8
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 0.03
b_0 <- (1 - p_0)/p_0 * a_0

prior_0 <- dbeta(x, a_0, b_0)
prior_1 <- dbeta(x, a_1, b_1)

prior_plot <- data.frame(x, prior_0, prior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0,
    prior_1))
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name,
    "_0") == T, "Control arm", "Active arm")

fig1 <- ggplot(prior_plot_long, aes(x = x, y = value,
    colour = group)) + geom_line(linewidth = 1.5) +
    theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "horizontal") +
    scale_x_continuous(limits = c(0, 0.1)) + ylab("Density") +
    labs(caption = str_glue("Assumed event probabilities: Control arm (p_0=",
        p_0, "), ", "Active arm (p_1=", p_1, ")")) +
    scale_color_brewer("Treatment arm", palette = "Set2") +
    scale_linetype("Type") + ggtitle("Informative prior")
```

```r
diff_prior <- data.frame(x = rbbinom(1e+06, size = ceiling(n_1),
    a_1, b_1)/ceiling(n_1) - rbbinom(1e+06, size = ceiling(n_0),
    a_0, b_0)/ceiling(n_0))

fig2 <- ggplot(diff_prior, aes(x), group = x) + geom_density(linewidth = 1.5,
    bw = 0.1) + theme_bw() + theme(panel.grid.minor = element_blank()) +
    xlab("Difference from simulated predicted event probabilities") +
    ggtitle("Distribution of difference from\nsimulated predicted event proportions") +
    ylab("Density") + geom_vline(xintercept = delta_star,
    linetype = "dashed") + geom_vline(xintercept = 0,
    linetype = "dotted") + scale_x_continuous(breaks = c(-0.5,
    -0.25, 0, 0.035, 0.25, 0.5), labels = c(-0.5, -0.25,
    0, 0.035, 0.25, 0.5), limits = c(-0.5, 0.5))

cowplot::plot_grid(fig1, fig2, ncol = 1)
```





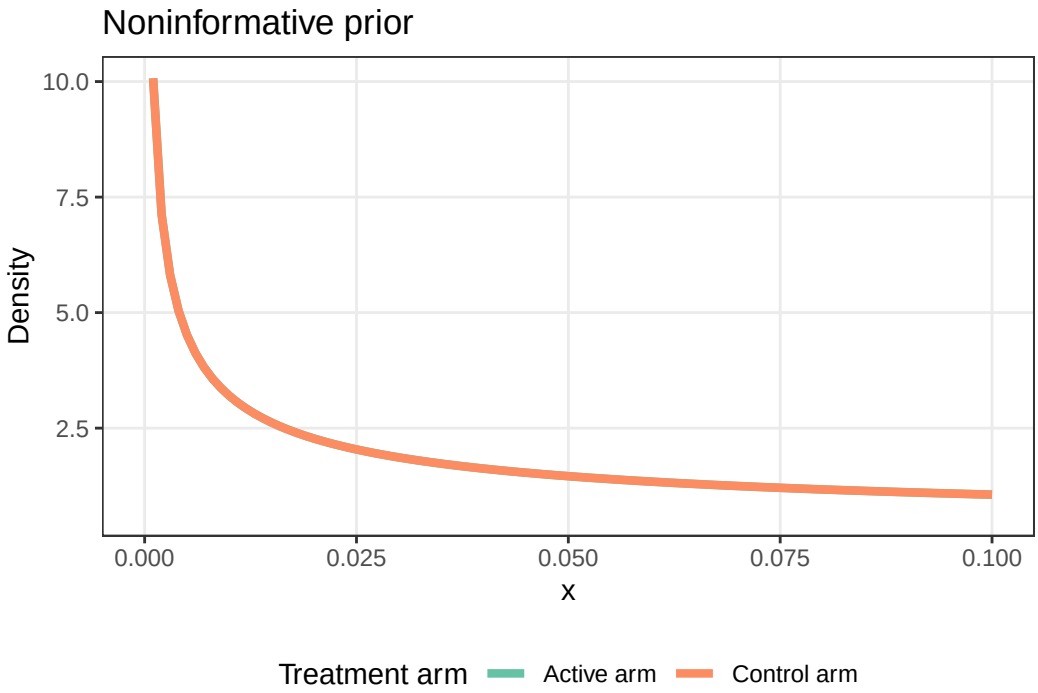
Noninformative prior: Agresti and Min suggest to use a 'diffuse' prior and recommend to use Jeffrey's

prior [11]. Jeffrey's prior for a binomial likelihood is $beta(0.5, 0.5)$-distributed.

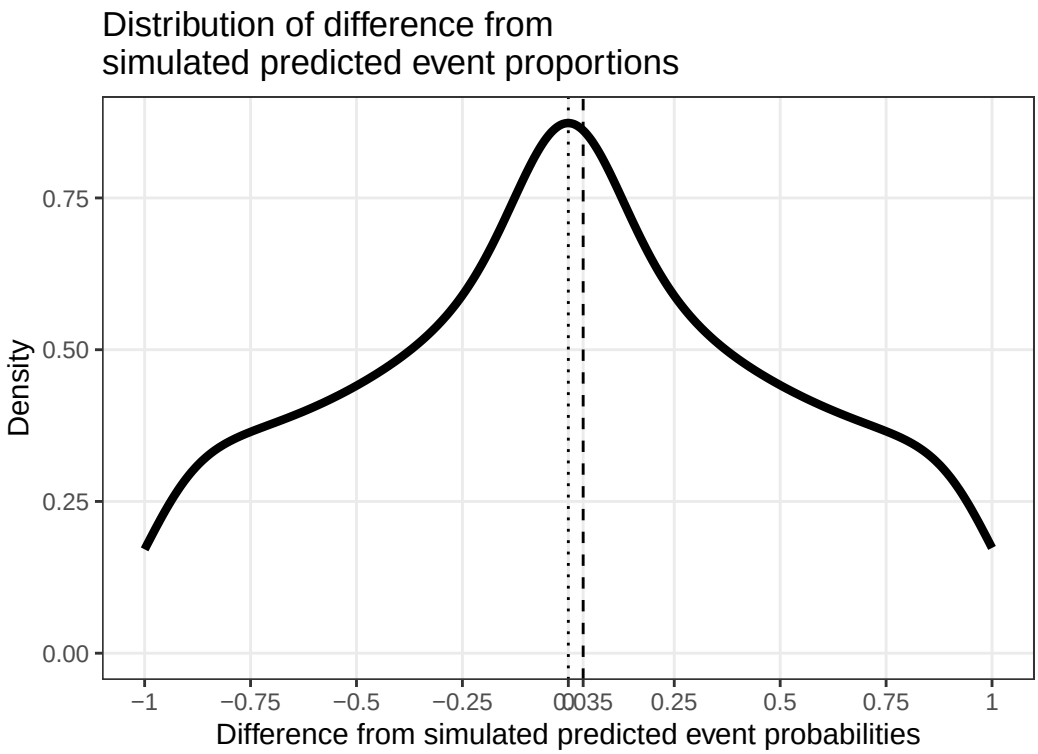
- Parameters of beta prior for active treatment arm: $a_1 = 0.5$, $b_1 = 0.5$.
- Parameters of beta prior for control treatment arm: $a_0 = 0.5$, $b_0 = 0.5$.

```
## Noninformative prior
a_0 <- 0.5
b_0 <- 0.5
a_1 <- 0.5
b_1 <- 0.5

prior_0 <- dbeta(x, a_0, b_0)
prior_1 <- dbeta(x, a_1, b_1)

prior_plot <- data.frame(x, prior_0, prior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0,
    prior_1))
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name,
    "_0") == T, "Control arm", "Active arm")

fig1 <- ggplot(prior_plot_long, aes(x = x, y = value,
    colour = group)) + geom_line(linewidth = 1.5) +
    theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "horizontal") +
    scale_x_continuous(limits = c(0, 0.1)) + ylab("Density") +
    labs(caption = str_glue("Assumed event probabilities: Control arm (p_0=",
        p_0, "), ", "Active arm (p_1=", p_1, ")")) +
    scale_color_brewer("Treatment arm", palette = "Set2") +
    scale_linetype("Type") + ggtitle("Noninformative prior")

diff_prior <- data.frame(x = rbbinom(1e+06, size = ceiling(n_1),
```

```
    a_1, b_1)/ceiling(n_1) - rbbinom(1e+06, size = ceiling(n_0),
    a_0, b_0)/ceiling(n_0))

fig2 <- ggplot(diff_prior, aes(x), group = x) + geom_density(linewidth = 1.5,
    bw = 0.1) + theme_bw() + theme(panel.grid.minor = element_blank()) +
    xlab("Difference from simulated predicted event probabilities") +
    ggtitle("Distribution of difference from\nsimulated predicted event proportions") +
    ylab("Density") + geom_vline(xintercept = delta_star,
    linetype = "dashed") + geom_vline(xintercept = 0,
    linetype = "dotted") + scale_x_continuous(breaks = c(-1, -0.75, -0.5,
    -0.25, 0, 0.035, 0.25, 0.5, 0.75, 1), labels = c(-1, -0.75, -0.5, -0.25,
    0, 0.035, 0.25, 0.5, 0.75, 1), limits = c(-1, 1))

cowplot::plot_grid(fig1, fig2, ncol = 1)
```



Assumed event probabilities: Control arm (p_0=0.01), Active arm (p_1=0.01)



One can calculate the average power as follows:

2.2.1.4.1 Approach 1: Joint prior predictive probability

The joint probability approach has been described in [6] and includes the following steps:

- Calculate the joint pdf for all pairs $r_i \in \{0, 1, \cdots, n_i\}$, $i \in \{0, 1\}$.
- The sum of the joint predictive probabilites over r_i pairs where the $(1 - \alpha)$-upper confidence limit of the risk difference is smaller than the non-inferiority margin is the average power.

Here we use Agresti-Caffo confidence intervals for the risk difference [12].

```
x <- c()

library(extraDistr)
library(PropCIs)

res_ap <- c()

## Enthusiastic prior

a_1 <- 0.5
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 7.2
b_0 <- (1 - 0.03)/0.03 * a_0

r_0 <- 0:(ceiling(n_0)-1)
r_1 <- 0:(ceiling(n_1)-1)

data_points <- expand.grid(r_1 = r_1, r_0 = r_0)

data_points$pdf_r_1 <- dbbinom(data_points$r_1, size = ceiling(n_1), a_1,
```

```
    b_1)
data_points$pdf_r_0 <- dbbinom(data_points$r_0, size = ceiling(n_0), a_0,
    b_0)
data_points$joint_pdf <- data_points$pdf_r_1 * data_points$pdf_r_0

data_points$sig <- NA

for (i in 1:nrow(data_points)) {
    data_points$sig[i] <- ifelse(wald2ci(data_points$r_1[i], n_1, data_points$r_0[i],
        n_0, conf.level = 0.95, adjust = "AC")$conf.int[2] <= delta_star,
        1, 0)
}

res_ap <- data.frame(type = "Enthusiastic",
                             ap = sum(data_points$joint_pdf[data_points$sig == 1]))

## Skeptical prior

a_1 <- 2.84
b_1 <- (1 - 0.05)/0.05 * a_1
a_0 <- 1.24
b_0 <- (1 - 0.01)/0.01 * a_0

r_0 <- 0:(ceiling(n_0)-1)
r_1 <- 0:(ceiling(n_1)-1)

data_points <- expand.grid(r_1 = r_1, r_0 = r_0)

data_points$pdf_r_1 <- dbbinom(data_points$r_1, size = ceiling(n_1), a_1,
    b_1)
```

```r
data_points$pdf_r_0 <- dbbinom(data_points$r_0, size = ceiling(n_0), a_0,
    b_0)
data_points$joint_pdf <- data_points$pdf_r_1 * data_points$pdf_r_0

data_points$sig <- NA

for (i in 1:nrow(data_points)) {
    data_points$sig[i] <- ifelse(wald2ci(data_points$r_1[i], n_1, data_points$r_0[i],
        n_0, conf.level = 0.95, adjust = "AC")$conf.int[2] <= delta_star,
        1, 0)
}

res_ap <- rbind(res_ap, data.frame(type = "Skeptical",
                           ap = sum(data_points$joint_pdf[data_points$sig == 1])))

## Informative prior

a_1 <- 0.8
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 0.03
b_0 <- (1 - p_0)/p_0 * a_0

r_0 <- 0:(ceiling(n_0)-1)
r_1 <- 0:(ceiling(n_1)-1)

data_points <- expand.grid(r_1 = r_1, r_0 = r_0)

data_points$pdf_r_1 <- dbbinom(data_points$r_1, size = ceiling(n_1), a_1,
    b_1)
data_points$pdf_r_0 <- dbbinom(data_points$r_0, size = ceiling(n_0), a_0,
```

```
    b_0)
data_points$joint_pdf <- data_points$pdf_r_1 * data_points$pdf_r_0

data_points$sig <- NA

for (i in 1:nrow(data_points)) {
    data_points$sig[i] <- ifelse(wald2ci(data_points$r_1[i], n_1, data_points$r_0[i],
        n_0, conf.level = 0.95, adjust = "AC")$conf.int[2] <= delta_star,
        1, 0)
}

res_ap <- rbind(res_ap, data.frame(type = "Informative",
                           ap = sum(data_points$joint_pdf[data_points$sig == 1])))

## Noninformative prior

a_1 <- 0.5
b_1 <- 0.5
a_0 <- 0.5
b_0 <- 0.5

r_0 <- 0:(ceiling(n_0)-1)
r_1 <- 0:(ceiling(n_1)-1)

data_points <- expand.grid(r_1 = r_1, r_0 = r_0)

data_points$pdf_r_1 <- dbbinom(data_points$r_1, size = ceiling(n_1), a_1,
    b_1)
data_points$pdf_r_0 <- dbbinom(data_points$r_0, size = ceiling(n_0), a_0,
    b_0)
```

```r
data_points$joint_pdf <- data_points$pdf_r_1 * data_points$pdf_r_0

data_points$sig <- NA

for (i in 1:nrow(data_points)) {
    data_points$sig[i] <- ifelse(wald2ci(data_points$r_1[i], n_1, data_points$r_0[i],
        n_0, conf.level = 0.95, adjust = "AC")$conf.int[2] <= delta_star,
        1, 0)
}

res_ap <- rbind(res_ap, data.frame(type = "Noninformative prior",
                           ap = sum(data_points$joint_pdf[data_points$sig == 1])))
res_ap$ap <- round(res_ap$ap, 2)

res_ap
```

```
                 type   ap
1         Enthusiastic 0.76
2            Skeptical 0.09
3          Informative 0.55
4 Noninformative prior 0.40
```

2.2.1.4.2 Approach 2: Sampling from prior predictive distribution

The second approach samples q event outcomes from the predictive distribution and calculates for each draw whether the $(1-\alpha)$-upper confidence limit of the risk difference is smaller than the non-inferiority margin. The sum of positive outcomes divided by the number of samples q gives the average power. Also here we use Agresti-Caffo confidence intervals for the risk difference [12].

```r
# Number of draws
q <- 10000

res_ap <- c()

## Enthusiastic prior

a_1 <- 0.5
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 7.2
b_0 <- (1 - 0.03)/0.03 * a_0

set.seed(1)
r_1 <- rbbinom(q, size = ceiling(n_1), a_1, b_1)
r_0 <- rbbinom(q, size = ceiling(n_0), a_0, b_0)

res <- c()

for (i in 1:length(r_1)) {
    res <- c(res, ifelse(wald2ci(r_1[i], n_1, r_0[i], n_0, conf.level = 0.95,
        adjust = "AC")$conf.int[2] <= delta_star, 1, 0))
}

res_ap <- data.frame(type = "Enthusiastic", ap = sum(res)/q)

## Skeptical prior

a_1 <- 2.84
b_1 <- (1 - 0.05)/0.05 * a_1
a_0 <- 1.24
```

```r
b_0 <- (1 - p_0)/p_0 * a_0

set.seed(1)
r_1 <- rbbinom(q, size = ceiling(n_1), a_1, b_1)
r_0 <- rbbinom(q, size = ceiling(n_0), a_0, b_0)

res <- c()

for (i in 1:length(r_1)) {
    res <- c(res, ifelse(wald2ci(r_1[i], n_1, r_0[i], n_0, conf.level = 0.95,
        adjust = "AC")$conf.int[2] <= delta_star, 1, 0))
}

res_ap <- rbind(res_ap, data.frame(type = "Skeptical", ap = sum(res)/q))

## Informative prior

a_1 <- 0.8
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 0.03
b_0 <- (1 - p_0)/p_0 * a_0

set.seed(1)
r_1 <- rbbinom(q, size = ceiling(n_1), a_1, b_1)
r_0 <- rbbinom(q, size = ceiling(n_0), a_0, b_0)

res <- c()

for (i in 1:length(r_1)) {
    res <- c(res, ifelse(wald2ci(r_1[i], n_1, r_0[i], n_0, conf.level = 0.95,
```

```
        adjust = "AC")$conf.int[2] <= delta_star, 1, 0))
}

res_ap <- rbind(res_ap, data.frame(type = "Informative", ap = sum(res)/q))

## Noninformative prior

a_0 <- 0.5
b_0 <- 0.5
a_1 <- 0.5
b_1 <- 0.5

set.seed(1)
r_1 <- rbbinom(q, size = ceiling(n_1), a_1, b_1)
r_0 <- rbbinom(q, size = ceiling(n_0), a_0, b_0)

res <- c()

for (i in 1:length(r_1)) {
    res <- c(res, ifelse(wald2ci(r_1[i], n_1, r_0[i], n_0, conf.level = 0.95,
        adjust = "AC")$conf.int[2] <= delta_star, 1, 0))
}

res_ap <- rbind(res_ap, data.frame(type = "Noninformative", ap = sum(res)/q))

res_ap$ap <- round(res_ap$ap, 2)
res_ap
```

```
          type   ap
1  Enthusiastic 0.76
```

```
2      Skeptical 0.09
3    Informative 0.55
4 Noninformative 0.45
```

2.2.1.5 Bayesian approach: Prior on event probabilities

In the former subsection we assumed that the prior was directly specified on the treatment effect. In this subsection we assume that that design prior is specified on the event probabilities p_i, $i = \{0, 1\}$. Suppose that the priors on the event probability are beta distributed $\pi_i \sim Beta(a_i, b_i)$ and that we have observed a binomial distributed random sample of size n_i and number of events r_i. Then the posterior distribution is

$$\pi_i | r_i \sim Beta\left(a_i + r_i, b_i + n_i - r_i\right) = \frac{\pi_i^{a_i + r_i - 1}(1 - \pi_i)^{b_i + n_i - r_i - 1}}{B(a_i + r_i, b_i + n_i - r_i)}$$

with $B(\cdot)$ the beta function and

$$\mu_{i,r_i} := E(\pi_i | r_i) = \frac{a_i + r_i}{a_i + b_i + n_i}, \quad \sigma^2_{i,r_i} := Var(\pi_i | r_i) = \frac{(a_i + r_i)\left(b_i + n_i - r_i\right)}{(a_i + b_i + n_i)^2 (a_i + b_i + n_i + 1)}.$$

Working example (continued)

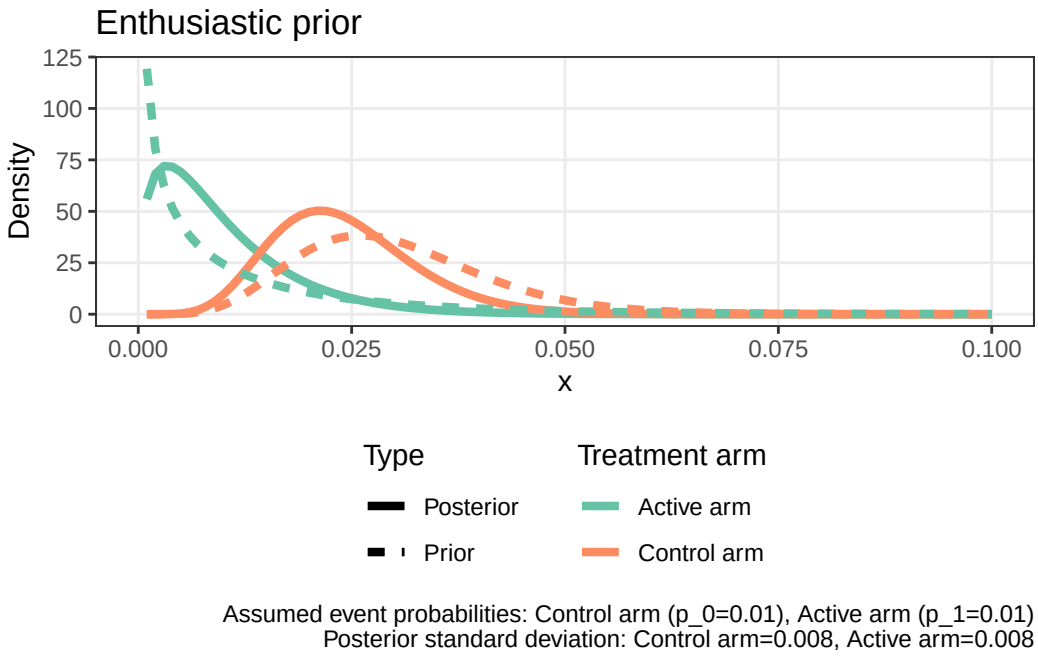
We use the same prior specifications as from the former subsection for the analysis priors.

Enthusiastic prior: The beta prior for the active treatment arm is centered at 0.01 (the expected event proportion) with $P(\pi_1 > 0.05) = 0.025$, that is, the probability that the prior is greater than the safety margin is 2.5%. The mean of the beta prior for the control arm is centered at 0.03 (the safety margin) with $P(\pi_0 > 0.05) = 0.05$.

- Parameters of beta prior for active treatment arm: $a_1 = 0.5$, $b_1 = 49.5$.
- Parameters of beta prior for control treatment arm: $a_0 = 7.2$, $b_0 = 232.8$.

```r
x <- seq(0.001, 0.999, 0.001)


## Enthusiastic prior
```

```r
a_1 <- 0.5
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 7.2
b_0 <- (1 - 0.03)/0.03 * a_0

prior_0 <- dbeta(x, a_0, b_0)
prior_1 <- dbeta(x, a_1, b_1)
posterior_0 <- dbeta(x, a_0 + n_0 * p_0, b_0 + n_0 - n_0 * p_0)
posterior_1 <- dbeta(x, a_1 + n_1 * p_1, b_1 + n_1 - n_1 * p_1)
var_posterior_0 <- ((a_0 + n_0 * p_0) * (b_0 + n_0 - n_0 * p_0))/((a_0 +
    b_0 + n_0)^2 * (a_0 + b_0 + n_0 + 1))
var_posterior_1 <- ((a_1 + n_1 * p_1) * (b_1 + n_1 - n_1 * p_1))/((a_1 +
    b_1 + n_1)^2 * (a_1 + b_1 + n_1 + 1))

prior_plot <- data.frame(x, prior_0, prior_1, posterior_0, posterior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0, prior_1,
    posterior_0, posterior_1))
prior_plot_long$type <- ifelse(str_detect(prior_plot_long$name, "posterior") ==
    T, "Posterior", "Prior")
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name, "_0") ==
    T, "Control arm", "Active arm")

ggplot(prior_plot_long, aes(x = x, y = value, linetype = type, colour = group)) +
    geom_line(linewidth = 1.5) + theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "vertical") + scale_x_continuous(limits = c(
    0.1)) + ylab("Density") + labs(caption = str_glue("Assumed event probabilities: Control arm
    p_0, "), ", "Active arm (p_1=", p_1, ")\nPosterior standard deviation: Control arm=",
    round(sqrt(var_posterior_0), 3), ", ", "Active arm=", round(sqrt(var_posterior_1),
        3))) + scale_color_brewer("Treatment arm", palette = "Set2") +
    scale_linetype("Type") + ggtitle("Enthusiastic prior")
```



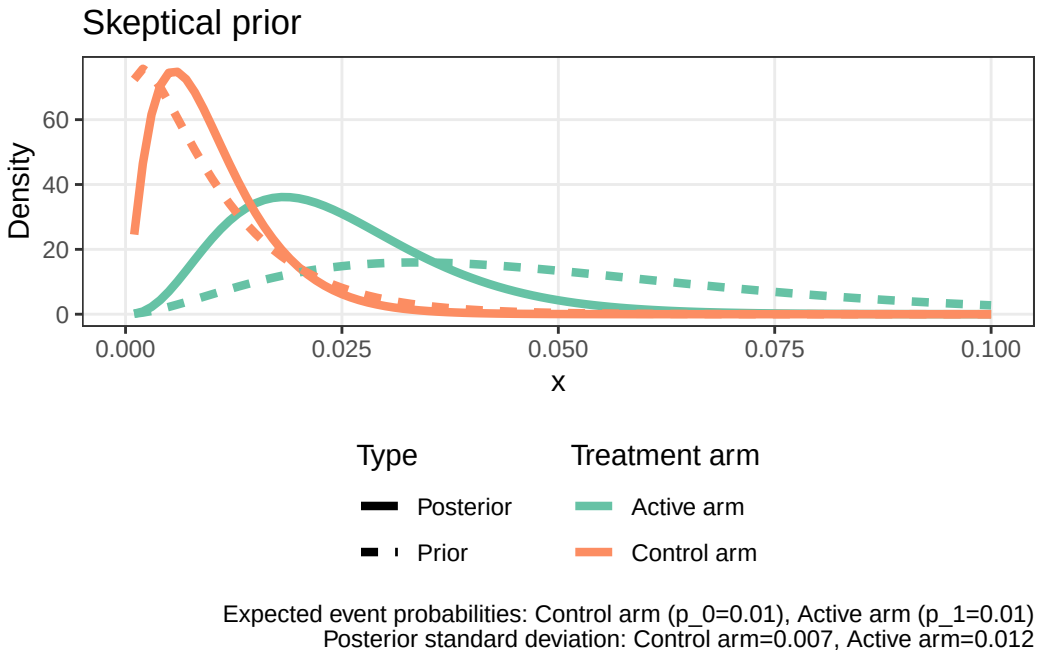
Assumed event probabilities: Control arm (p_0=0.01), Active arm (p_1=0.01)
Posterior standard deviation: Control arm=0.008, Active arm=0.008

Skeptical prior: The beta prior for the active treatment arm is centered at 0.05 (the safety margin) with $P(\pi_1 \leq 0.01) = 0.025$, that is the probability that the prior is smaller or equal than the expected event proportion is 2.5%. The mean of the beta prior for the control arm is is centered at 0.01 (the expected event proportion) with $P(\pi_0 > 0.05) = 0.025$.

- Parameters of beta prior for active treatment arm: $a_1 = 2.84$, $b_1 = 53.96$.
- Parameters of beta prior for control treatment arm: $a_i = 1.24$, $b_i = 122.76$.

```
x <- seq(0.001, 0.999, 0.001)


# Skeptical prior
a_1 <- 2.84
b_1 <- (1 - 0.05)/0.05 * a_1
a_0 <- 1.24
b_0 <- (1 - p_0)/p_0 * a_0


prior_0 <- dbeta(x, a_0, b_0)
```

```r
prior_1 <- dbeta(x, a_1, b_1)
posterior_0 <- dbeta(x, a_0 + n_0 * p_0, b_0 + n_0 - n_0 * p_0)
posterior_1 <- dbeta(x, a_1 + n_1 * p_1, b_1 + n_1 - n_1 * p_1)
var_posterior_0 <- ((a_0 + n_0 * p_0) * (b_0 + n_0 - n_0 * p_0))/((a_0 +
    b_0 + n_0)^2 * (a_0 + b_0 + n_0 + 1))
var_posterior_1 <- ((a_1 + n_1 * p_1) * (b_1 + n_1 - n_1 * p_1))/((a_1 +
    b_1 + n_1)^2 * (a_1 + b_1 + n_1 + 1))

prior_plot <- data.frame(x, prior_0, prior_1, posterior_0, posterior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0, prior_1,
    posterior_0, posterior_1))
prior_plot_long$type <- ifelse(str_detect(prior_plot_long$name, "posterior") ==
    T, "Posterior", "Prior")
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name, "_0") ==
    T, "Control arm", "Active arm")

ggplot(prior_plot_long, aes(x = x, y = value, linetype = type, colour = group)) +
    geom_line(linewidth = 1.5) + theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "vertical") + scale_x_continuous(limits = c(
    0.1)) + ylab("Density") + labs(caption = str_glue("Expected event probabilities: Control ar
    p_0, "), ", "Active arm (p_1=", p_1, ")\nPosterior standard deviation: Control arm=",
    round(sqrt(var_posterior_0), 3), ", ", "Active arm=", round(sqrt(var_posterior_1),
        3))) + scale_color_brewer("Treatment arm", palette = "Set2") +
    scale_linetype("Type") + ggtitle("Skeptical prior")
```



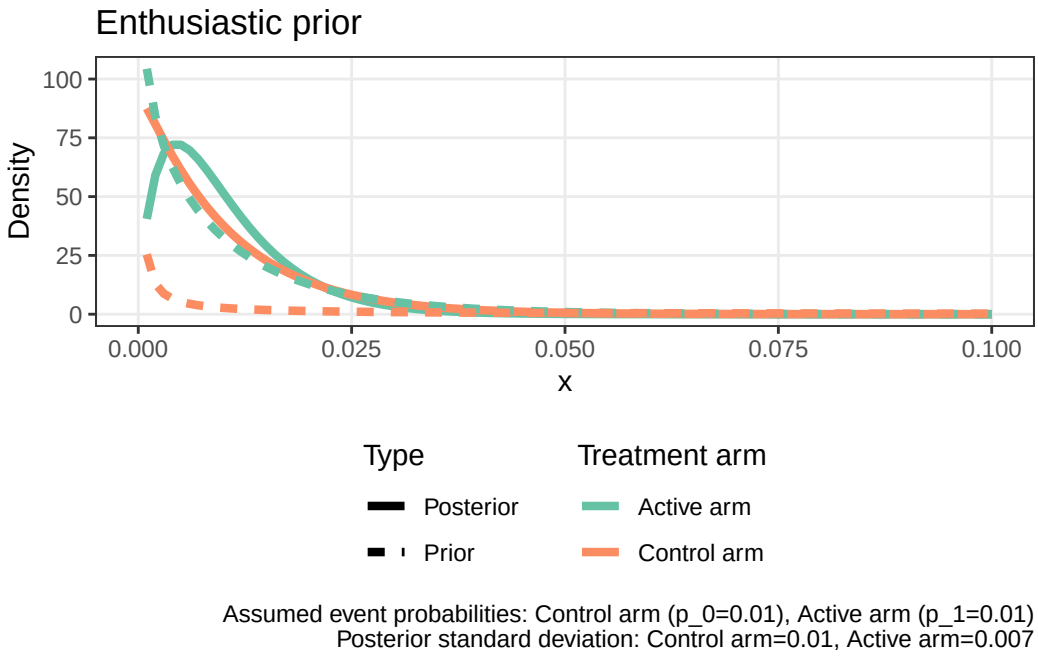
Expected event probabilities: Control arm (p_0=0.01), Active arm (p_1=0.01)
Posterior standard deviation: Control arm=0.007, Active arm=0.012

Informative prior: The beta prior for the active treatment arm and control arm is centered at 0.01 (the expected event proportion) with $P(\pi_1 \leq 0.05) = 0.01$ and $P(\pi_1 \leq 0.05) = 0.05$.

- Parameters of beta prior for active treatment arm: $a_1 = 0.8$, $b_1 = 79.2$.
- Parameters of beta prior for control treatment arm: $a_0 = 0.03$, $b_0 = 2.97$.

```
x <- seq(0.001, 0.999, 0.001)


## Informative prior
a_1 <- 0.8
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 0.03
b_0 <- (1 - p_0)/p_0 * a_0


prior_0 <- dbeta(x, a_0, b_0)
prior_1 <- dbeta(x, a_1, b_1)
posterior_0 <- dbeta(x, a_0 + n_0 * p_0, b_0 + n_0 - n_0 * p_0)
```

```r
posterior_1 <- dbeta(x, a_1 + n_1 * p_1, b_1 + n_1 - n_1 * p_1)
var_posterior_0 <- ((a_0 + n_0 * p_0) * (b_0 + n_0 - n_0 * p_0))/((a_0 +
    b_0 + n_0)^2 * (a_0 + b_0 + n_0 + 1))
var_posterior_1 <- ((a_1 + n_1 * p_1) * (b_1 + n_1 - n_1 * p_1))/((a_1 +
    b_1 + n_1)^2 * (a_1 + b_1 + n_1 + 1))

prior_plot <- data.frame(x, prior_0, prior_1, posterior_0, posterior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0, prior_1,
    posterior_0, posterior_1))
prior_plot_long$type <- ifelse(str_detect(prior_plot_long$name, "posterior") ==
    T, "Posterior", "Prior")
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name, "_0") ==
    T, "Control arm", "Active arm")

ggplot(prior_plot_long, aes(x = x, y = value, linetype = type, colour = group)) +
    geom_line(linewidth = 1.5) + theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "vertical") + scale_x_continuous(limits = c(
    0.1)) + ylab("Density") + labs(caption = str_glue("Assumed event probabilities: Control arm
    p_0, "), ", "Active arm (p_1=", p_1, ")\nPosterior standard deviation: Control arm=",
    round(sqrt(var_posterior_0), 3), ", ", "Active arm=", round(sqrt(var_posterior_1),
        3))) + scale_color_brewer("Treatment arm", palette = "Set2") +
    scale_linetype("Type") + ggtitle("Enthusiastic prior")
```

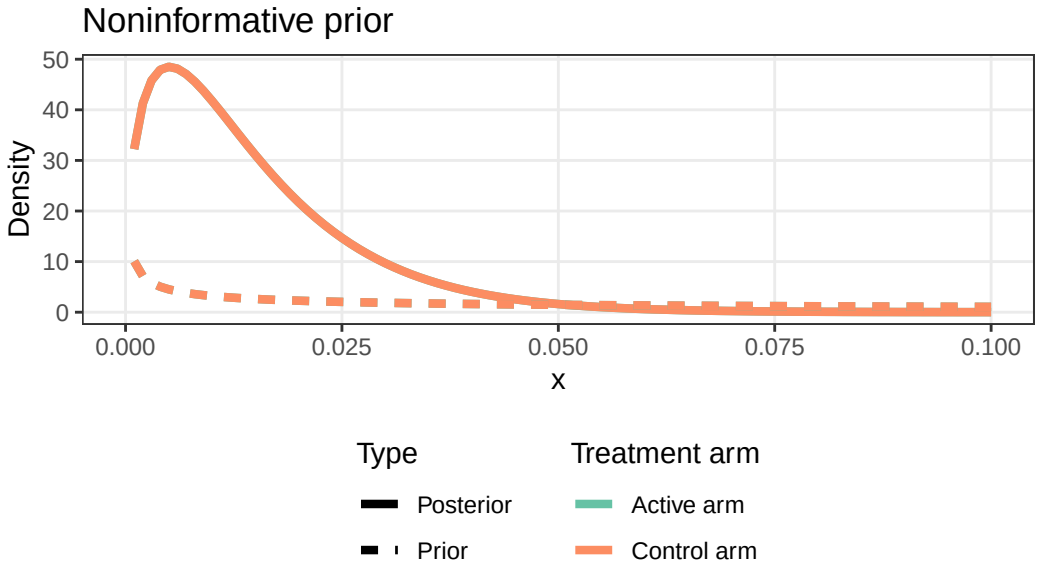


Noninformative prior: Agresti and Min suggest to use a 'diffuse' prior and recommend to use Jeffrey's prior [11]. Jeffrey's prior for a binomial likelihood is $beta(0.5, 0.5)$-distributed.

- Parameters of beta prior for active treatment arm: $a_1 = 0.5$, $b_1 = 0.5$.
- Parameters of beta prior for control treatment arm: $a_0 = 0.5$, $b_0 = 0.5$.

```
## Noninformative prior
a_0 <- 0.5
b_0 <- 0.5
a_1 <- 0.5
b_1 <- 0.5

prior_0 <- dbeta(x, a_0, b_0)
prior_1 <- dbeta(x, a_1, b_1)
posterior_0 <- dbeta(x, a_0 + n_0 * p_0, b_0 + n_0 -
    n_0 * p_0)
posterior_1 <- dbeta(x, a_1 + n_1 * p_1, b_1 + n_1 -
```

```
    n_1 * p_1)
var_posterior_0 <- ((a_0 + n_0 * p_0) * (b_0 + n_0 -
    n_0 * p_0))/((a_0 + b_0 + n_0)^2 * (a_0 + b_0 +
    n_0 + 1))
var_posterior_1 <- ((a_1 + n_1 * p_1) * (b_1 + n_1 -
    n_1 * p_1))/((a_1 + b_1 + n_1)^2 * (a_1 + b_1 +
    n_1 + 1))

prior_plot <- data.frame(x, prior_0, prior_1, posterior_0,
    posterior_1)
prior_plot_long <- pivot_longer(prior_plot, cols = c(prior_0,
    prior_1, posterior_0, posterior_1))
prior_plot_long$type <- ifelse(str_detect(prior_plot_long$name,
    "posterior") == T, "Posterior", "Prior")
prior_plot_long$group <- ifelse(str_detect(prior_plot_long$name,
    "_0") == T, "Control arm", "Active arm")

ggplot(prior_plot_long, aes(x = x, y = value, linetype = type,
    colour = group)) + geom_line(linewidth = 1.5) +
    theme_bw() + theme(panel.grid.minor = element_blank(),
    legend.position = "bottom", legend.direction = "vertical") +
    scale_x_continuous(limits = c(0, 0.1)) + ylab("Density") +
    labs(caption = str_glue("Assumed event probabilities: Control arm (p_0=",
        p_0, "), ", "Active arm (p_1=", p_1, ")\nPosterior standard deviation: Control arm=",
        round(sqrt(var_posterior_0), 3), ", ", "Active arm=",
        round(sqrt(var_posterior_1), 3))) + scale_color_brewer("Treatment arm",
    palette = "Set2") + scale_linetype("Type") + ggtitle("Noninformative prior")
```



Assumed event probabilities: Control arm (p_0=0.01), Active arm (p_1=0.01)
Posterior standard deviation: Control arm=0.012, Active arm=0.012

We denote the joint posterior distribution as

$$f_2(\pi_1, \pi_0 | r_0, r_1, n_0, n_1) = \prod_{i \in \{0,1\}} \frac{\pi_i^{a_i + r_i - 1} (1 - \pi_i)^{b_i + n_i - r_i - 1}}{B(a_i + r_i, b_i + n_i - r_i)}.$$

and thus, by using Fubini's theorem and the explanations of Patricia Altham [13],

$$
\begin{aligned}
P\left(\pi_1 - \pi_0 \leq \delta^* | \, r_0, r_1, n_0, n_1\right) &= \int_0^1 \int_0^{\pi_0+\delta^*} f_2(\pi_1, \pi_0 | r_0, r_1, n_0, n_1) d\pi_1 d\pi_0 \\
&= \int_0^1 \frac{\pi_0^{a_0+r_0-1}(1-\pi_0)^{b_0+n_0-r_0-1}}{B(a_0+r_0, b_0+n_0-r_0)} \left(\int_0^{\pi_0+\delta^*} \frac{\pi_1^{a_1+r_1-1}(1-\pi_1)^{b_1+n_1-r_1-1}}{B(a_1+r_1, b_1+n_1-r_1)} d\pi_1 \right) d\pi_0 \\
&\overset{u=\pi_0-\delta^*}{=} \int_{-\delta^*}^{1-\delta^*} \frac{u^{a_0+r_0-1}(1-u)^{b_0+n_0-r_0-1}}{B(a_0+r_0, b_0+n_0-r_0)} \left(\int_0^{u} \frac{\pi_1^{a_1+r_1-1}(1-\pi_1)^{b_1+n_1-r_1-1}}{B(a_1+r_1, b_1+n_1-r_1)} d\pi_1 \right) du \\
&= \int_{-\delta^*}^{1-\delta^*} \frac{u^{a_0+r_0-1}(1-u)^{b_0+n_0-r_0-1}}{B(a_0+r_0, b_0+n_0-r_0)} \left(\sum_{s=a_1+r_1}^{a_1+b_1+n_1-1} \binom{a_1+b_1+n_1-1}{s} u^s (1-u)^{a_1+b_1+n_1-1-s} \right) du \\
&= \sum_{s=a_1+r_1}^{a_1+b_1+n_1-1} \binom{a_1+b_1+n_1-1}{s} \frac{1}{B(a_0+r_0, b_0+n_0-r_0)} \left(\int_{-\delta^*}^{1-\delta^*} u^{a_0+r_0+s-1}(1-u)^{a_1+b_1+n_1-1+b_0+n_0-r_0-s-1} du \right) \\
&\overset{z=u+\delta^*}{=} \sum_{s=a_1+r_1}^{a_1+b_1+n_1-1} \binom{a_1+b_1+n_1-1}{s} \frac{1}{B(a_0+r_0, b_0+n_0-r_0)} \left(\int_0^{1} z^{a_0+r_0+s-1}(1-z)^{a_1+b_1+n_1-1-s+b_0+n_0-r_0-1} dz \right) \\
&= \sum_{s=a_1+r_1}^{a_1+b_1+n_1-1} \binom{a_1+b_1+n_1-1}{s} \frac{B(a_0+r_0+s, a_1+b_1+n_1-1-s+b_0+n_0-r_0)}{B(a_0+r_0, b_0+n_0-r_0)}
\end{aligned}
$$

is the **Bayesian probability to reject** given the observed r_0, r_1, n_0, n_1. The last term in the above formula is the upper tail of a beta-binomial distribution. Altham [13] showed that this could also be written as a hypergeometric distribution (see also [6], [14]).

Working example (continued)

We calculate the Bayesian probability to reject for the SAFE-SSPE study.

```
## Enthusiastic prior
a_1 <- 0.5
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 7.2
b_0 <- (1 - 0.03)/(0.03) * a_0
```

```r
s <- ceiling(seq(a_1 + n_1 * p_1, b_1 + n_1 - n_1 *
    p_1 - 1))
bap <- sum(dbbinom(s, size = ceiling(a_1 +
    b_1 + n_1 - 1), alpha = a_0 + n_0 * p_0, beta = b_0 +
    n_0 - n_0 * p_0))

res_bap <- data.frame(type="Enthusiastic", bap)

# Skeptical prior
a_1 <- 2.84
b_1 <- (1 - 0.05)/0.05 * a_1
a_0 <- 1.24
b_0 <- (1 - p_0)/p_0 * a_0

s <- ceiling(seq(a_1 + n_1 * p_1, b_1 + n_1 - n_1 *
    p_1 - 1))

bap <- sum(dbbinom(s, size = ceiling(a_1 +
    b_1 + n_1 - 1), alpha = a_0 + n_0 * p_0, beta = b_0 +
    n_0 - n_0 * p_0))

res_bap <- rbind(res_bap, data.frame(type="Skeptical", bap))

# Informative prior
a_1 <- 0.8
b_1 <- (1 - p_1)/p_1 * a_1
a_0 <- 0.03
b_0 <- (1 - p_0)/p_0 * a_0

s <- ceiling(seq(a_1 + n_1 * p_1, b_1 + n_1 - n_1 *
```

```
    p_1 - 1))
bap <- sum(dbbinom(s, size = ceiling(a_1 +
    b_1 + n_1 - 1), alpha = a_0 + n_0 * p_0, beta = b_0 +
    n_0 - n_0 * p_0))

res_bap <- rbind(res_bap, data.frame(type="Informative", bap))



## Noninformative prior
a_0 <- 0.5
b_0 <- 0.5
a_1 <- 0.5
b_1 <- 0.5

s <- ceiling(seq(a_1 + n_1 * p_1, b_1 + n_1 - n_1 *
    p_1 - 1))
bap <- sum(dbbinom(s, size = ceiling(a_1 +
    b_1 + n_1 - 1), alpha = a_0 + n_0 * p_0, beta = b_0 +
    n_0 - n_0 * p_0))

res_bap <- rbind(res_bap, data.frame(type="Noninformative", bap))

res_bap$bap <- round(res_bap$bap, 2)

res_bap
```

```
           type  bap
1   Enthusiastic 0.83
2      Skeptical 0.12
3    Informative 0.42
```

```
4 Noninformative 0.38
```

To get the Bayesian average probability to reject we calculate (IS THIS NECESSARY????)

$$BAP = \sum_{r_0=0}^{n_0-1} \sum_{r_1=0}^{n_1-1} P(\pi_1 - \pi_0 \leq \delta^* | r_0, r_1, n_0, n_1) f_2(r_0, r_1 | n_0, n_1)$$

3 Phase II trials

Phase II trials are often single-arm trials in early phases but can also be multi-arm (non-)randomized controlled trials in later phases [15]. The aim of early phase-II designs is to assess efficacy of a new treatment measured by clinical response in a small population of 30-100 patients. Early phase II trials can be designed as two- or three stage trials using frequentist approaches like Simon's two stage design. In this chapter we use a Bayesian predictive probability approach to analyse an early phase-II trial [16] [17]. The aim is to compare two different decision rules for early stopping the trial: One uses the posterior distribution given the interim data, the second one uses the predictive distribution given the interim data and future events.

3.1 Background

Let p the true toxicity rate and $\pi \sim Beta(a, b)$ be a design prior on p. Our null hypothesis of interest is $H_0 : p \leq p_0$ versus $H_1 : p > p_0$, that is, we compare, say, the clinical response p toa prespecified threshold p_0. Suppose that we plan interim looks $j \geq 1$ after N_j patients were recruited and r_j toxicity events occurred. Under the prior π the posterior distribution is beta distributed $p|r_j \sim Beta(a + r_j, b + N_j - r_j)$. Suppose that we plan to recruit N_{max} patients. Let Y_j be the number of toxicities in future $m_j = N_{max} - N_j$ patients at interim look $j \geq 1$. Under the prior π, Y_j is beta-binomial distributed $Y_j \sim Betabinomal(i|m_j, a + r_j, b + N_j - r_j)$ and $p|r_j, y_j \sim Beta(a + r_j + y_j, b + N_{max} - N_j - r_j - y_i)$.

Working example

Suppose that the goal of a phase-II trial is to evaluate safety of new chemotherapy treatment investigating the discontinuation rate p as primary endpoint. Discontinuation is defined based on toxicity grading from the US National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE).

3.2 Decision rules for stopping the trial

Early stopping using the posterior distribution

The decision to the stop the trial early is based on the decision rule $P(p > p_0|N_j) > \theta_T$, that is, the posterior distribution given the interim data. If the trial is not stopped early we recruit more patients until N_{max} is reached.

Early stopping using the predictive distribution

Let $T_{j,i} := P(p > p_0|N_j, Y_j = i)$, that is, the posterior distribution given the interim data and future events. Note that because Y_j are not observed events $T_{j,i}$ are random variables. The predictive probability for a successful (safe) trial is

$$PP_j = \sum_{i \leq m} P(Y_j = i|r_j) \cdot I(T_{j,i} > \theta_T), \quad j \geq 1.$$

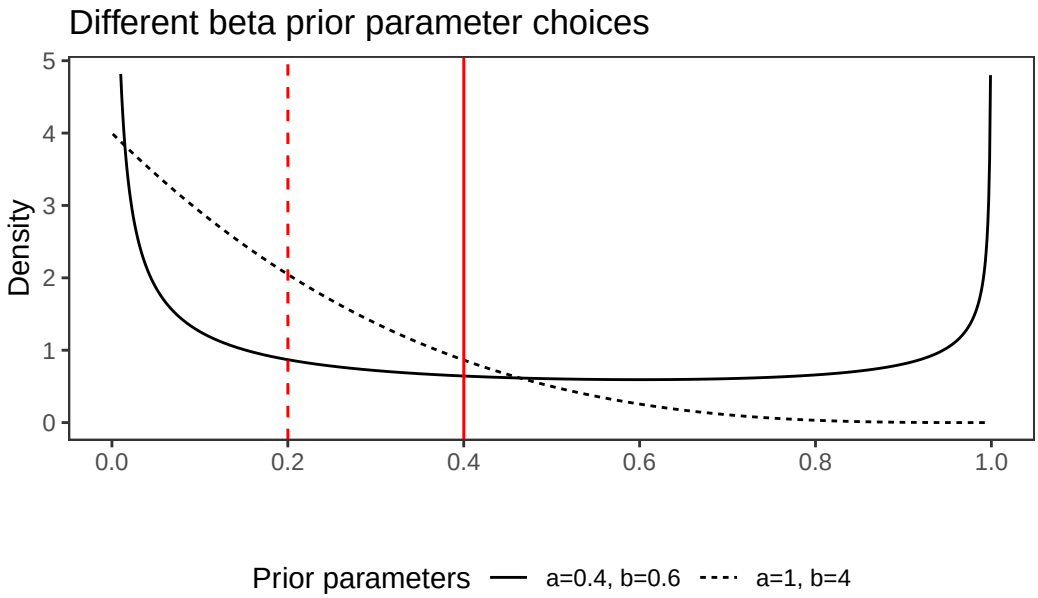
This is the expectation of $T_{j,i}$ given the interim data $j \geq 1$ [16]. If $PP_j > \theta_S$ then we stop the trial because of safety concerns, otherwise we recruit more patients until N_{max} is reached. Usually, θ_S is chosen to be high (>0.8) [17].

Working example

Based on prior evidence p_0 is set to 0.2. This threshold is based on a result from a previous study where the treatment discontinuation was estimated with 0.17 with 95% confidence intervals (0.10-0.27). Because we want to be conservative we set the parameters for a beta prior to $a = 0.4$ and $b = 0.6$. The mean

of the prior is then 0.4. For illustrating purposes we also show a prior with a prior weight of 5 patients which is centered on $p = 0.2$ ($a = 1$, $b = 4$).

```
# Prior information
a_0 <- 0.4
b_0 <- 0.6

x <- seq(0.001,0.999,0.001)

data_prior <- data.frame(x, y=dbeta(x, a_0, b_0), type=1)

# Prior information
a_0 <- 1
b_0 <- 4

data_prior <- bind_rows(data_prior, data.frame(x, y=dbeta(x, a_0, b_0), type=2))

data_prior$type <- factor(data_prior$type, levels=1:2,
                          labels=c("a=0.4, b=0.6", "a=1, b=4"))
ggplot(data_prior %>% filter(y<5), aes(x=x, y=y, linetype=type, group=type))+
  geom_line()+theme_bw()+scale_x_continuous(breaks=seq(0,1,0.2))+
  theme(panel.grid = element_blank(), legend.position = "bottom")+
  scale_linetype("Prior parameters")+geom_vline(xintercept=0.4, colour="red", linetype=1)+
  geom_vline(xintercept=0.2, colour="red", linetype=2)+
  ylab("Density")+xlab("")+ggtitle("Different beta prior parameter choices")+
  labs(caption=c("Red vertical lines indicate mean values of priors."))
```



Red vertical lines indicate mean values of priors.

```r
p_0_max <- 0.2

# Prior information
a_0 <- 0.4
b_0 <- 0.6

n <- 35
r <- 0:n

n_max <- 35
n <- seq(10, n_max, 5)
r <- 0:n_max
m <- n_max-r
x <- seq(0,1,0.001)

theta_T <- 0.75
```

```r
data <- expand.grid(n=n, r=r) %>% arrange(n) %>% filter(n>=r) %>% mutate(i=n_max-n+1)
data <- expandRows(data, count=3)
data <- data %>% group_by(n, r) %>% mutate(m=n_max-n, ind=1, i=cumsum(ind)-1, ind=NULL)

data <- data %>% mutate(cond_postprob=dbbinom(i, size=m, alpha=a_0+r, beta=b_0+n-r),
                        Ti=1-pbeta(p_0_max, a_0+r+i, b_0+n_max-r-i),
                        ind=ifelse(Ti>theta_T, 1, 0), n_max)

data <- data %>% select(n_max, n, r, m, i, cond_postprob, Ti, ind)
```

Suppose we plan a phase-II with a maximum of $N_{max} = 35$ patients. Let us consider the situation where we have reached an initial $N_1 = 10$ with $r_1 = 2$ observed toxicities and we have to decide whether to stop or continue the trial. Specifically, we calculate for each potential future toxicity from m_1 patients and decide whether $T_{i,1} > \theta_T = 0.75$. The column *cond_postprob* in the following table is $p|r_1, y_i$ (because it is conditional on the future outcome $Y_{j,i} = i$) and the column Ti corresponds to $T_{1,i}$.

```r
data %>% filter(n==10, r==2, i<=15)
```

```
# A tibble: 16 x 8
# Groups:   n, r [1]
   n_max     n     r     m     i cond_postprob      Ti   ind
   <dbl> <dbl> <int> <dbl> <dbl>         <dbl>   <dbl> <dbl>
 1    35    10     2    25     0        0.0434  0.00789     0
 2    35    10     2    25     1        0.0799  0.0315      0
 3    35    10     2    25     2        0.103   0.0881      0
 4    35    10     2    25     3        0.114   0.190       0
 5    35    10     2    25     4        0.114   0.334       0
 6    35    10     2    25     5        0.107   0.500       0
 7    35    10     2    25     6        0.0958  0.661       0
 8    35    10     2    25     7        0.0821  0.793       1
 9    35    10     2    25     8        0.0678  0.887       1
```

```
10    35    10     2    25     9       0.0542  0.945       1
11    35    10     2    25    10       0.0419  0.976       1
12    35    10     2    25    11       0.0313  0.990       1
13    35    10     2    25    12       0.0227  0.997       1
14    35    10     2    25    13       0.0159  0.999       1
15    35    10     2    25    14       0.0107  1.00        1
16    35    10     2    25    15       0.00690 1.00        1
```

Based on this information we calculate PP_1 and decide based on the threshold $\theta_S = 0.9$ whether to stop the trial early or to accrual more patients. For the decision rule based on the posterior probability we calculate the decision to stop the trial using the posterior distribution and use the threshold $\theta_T = 0.75$.

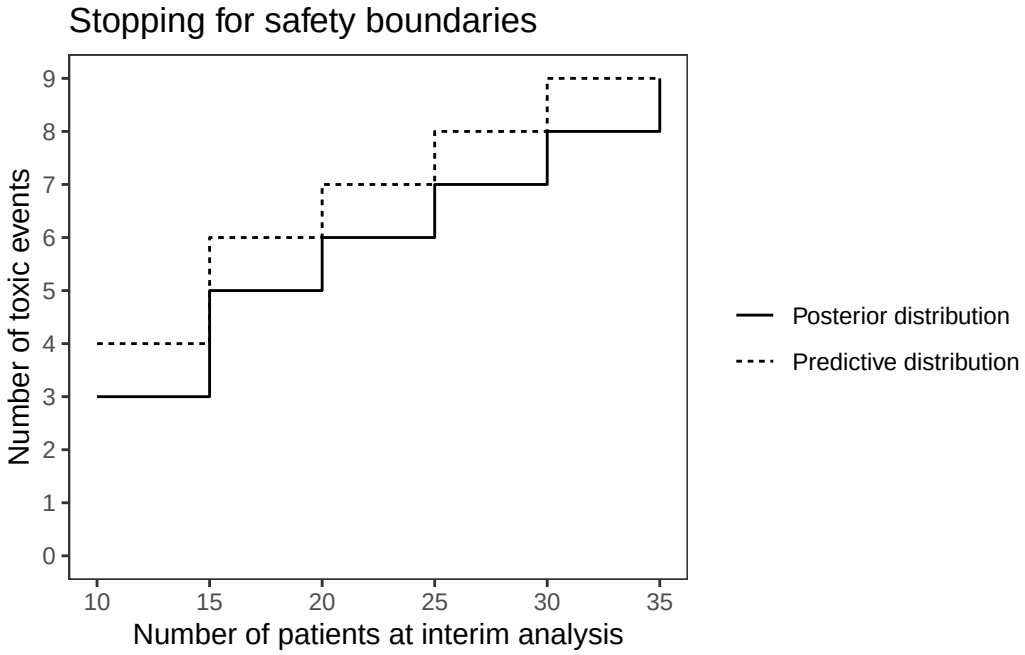
```
threshold_safety <- 0.9

## Early stopping using predictive distribution
data_pp <- data %>% group_by(n, r) %>% summarise(pp=round(sum(cond_postprob*ind),6),
                                                 stop_pp=ifelse(pp>threshold_safety, 1, 0))

## Early stopping using posterior distribution
data_pp <- data_pp %>%
  mutate(stop_postdist=ifelse(1-pbeta(p_0_max, a_0+r, b_0+n-r)>theta_T, 1, 0))

data_pp %>% filter(n==10)
```

```
# A tibble: 11 x 5
# Groups:   n [1]
       n     r      pp stop_pp stop_postdist
   <dbl> <int>   <dbl>   <dbl>         <dbl>
 1    10     0 0.00552       0             0
 2    10     1 0.0889        0             0
 3    10     2 0.343         0             0
```

4	10	3	0.678	0	1
5	10	4	0.903	1	1
6	10	5	0.983	1	1
7	10	6	0.999	1	1
8	10	7	1.00	1	1
9	10	8	1.00	1	1
10	10	9	1	1	1
11	10	10	1	1	1

For each interim look $j \geq 1$ we can then calculate stopping of safety boundaries of the trial. Here we assumed that we recruit 5 additional patients if we do not stop the trial early.

```
data_pp_long <- pivot_longer(data_pp, cols=4:5)

stop_boundaries <- data_pp_long %>% filter(value==1) %>%
  group_by(name, n) %>% summarise(r=min(r), stop=NULL)

stop_boundaries$name <- factor(stop_boundaries$name,
                               levels=c("stop_postdist", "stop_pp"),
                               labels=c("Posterior distribution",
                                        "Predictive distribution"))

ggplot(stop_boundaries, aes(x=n, y=r, linetype=name))+geom_step()+
  scale_y_continuous(limits=c(0, max(stop_boundaries$r)),
                     breaks=0: max(stop_boundaries$r))+
  ylab("Number of toxic events")+xlab("Number of patients at interim analysis")+
  theme_bw()+theme(panel.grid = element_blank())+
  ggtitle("Stopping for safety boundaries")+
  scale_linetype("")
```



3.3 Simulation study

3.3.1 Prior with a=0.4, b=0.6

We set up a simulation study to quantify operations characteristics (type-I error, type-II error, the probability to declare toxicity and expected number of patients and toxicities) with the following parameters

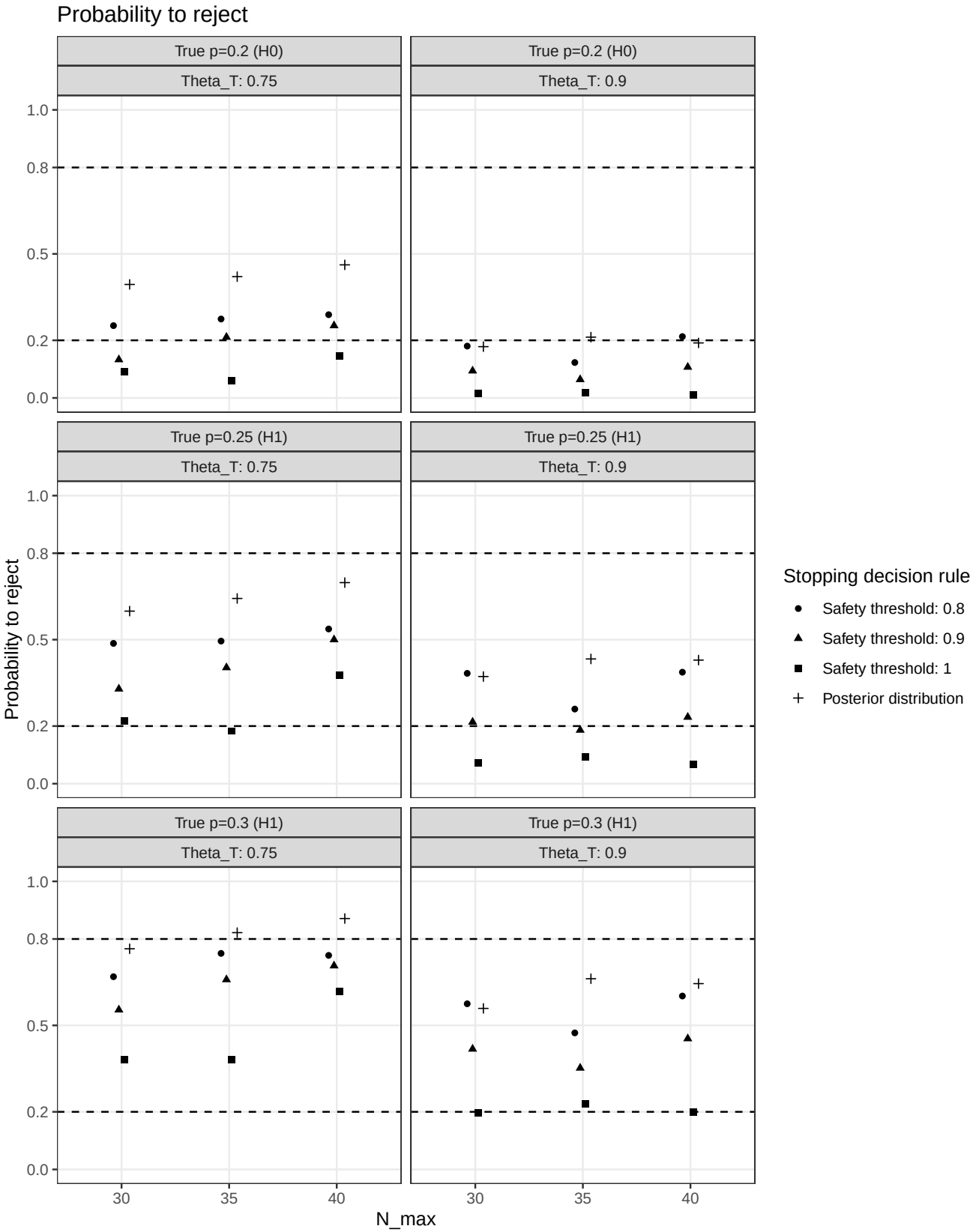
- Beta prior parameters: a=0.4, b=0.6
- True p: 0.2, 0.25, 0.3
- p_0: 0.2
- N_0: 10
- Cohort increase: 5
- N_{max}: 30, 35, 40
- θ_T: 0.75, 0.9
- θ_S: 0.8, 0.9, 1

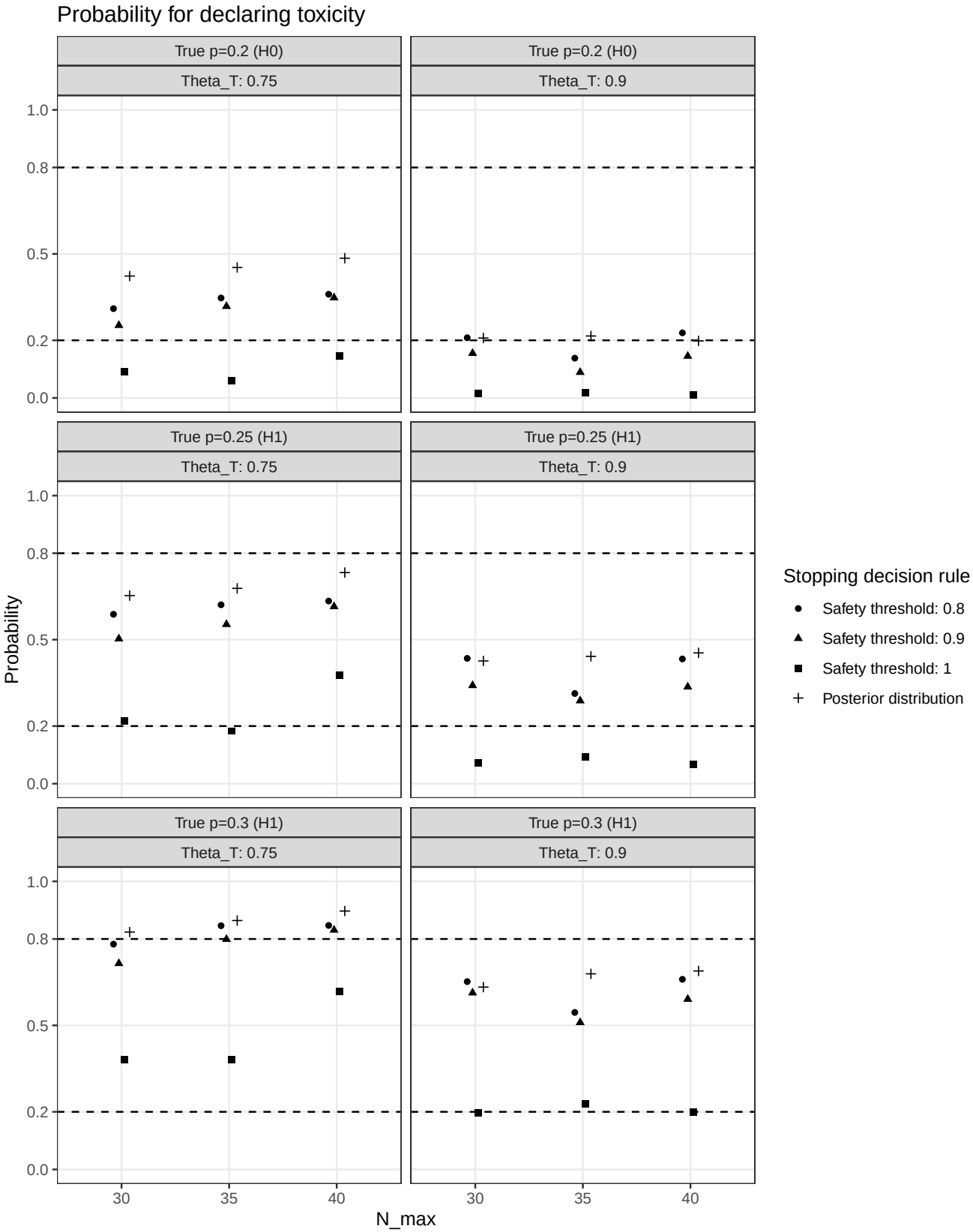
- Number of simulation runs: 1000

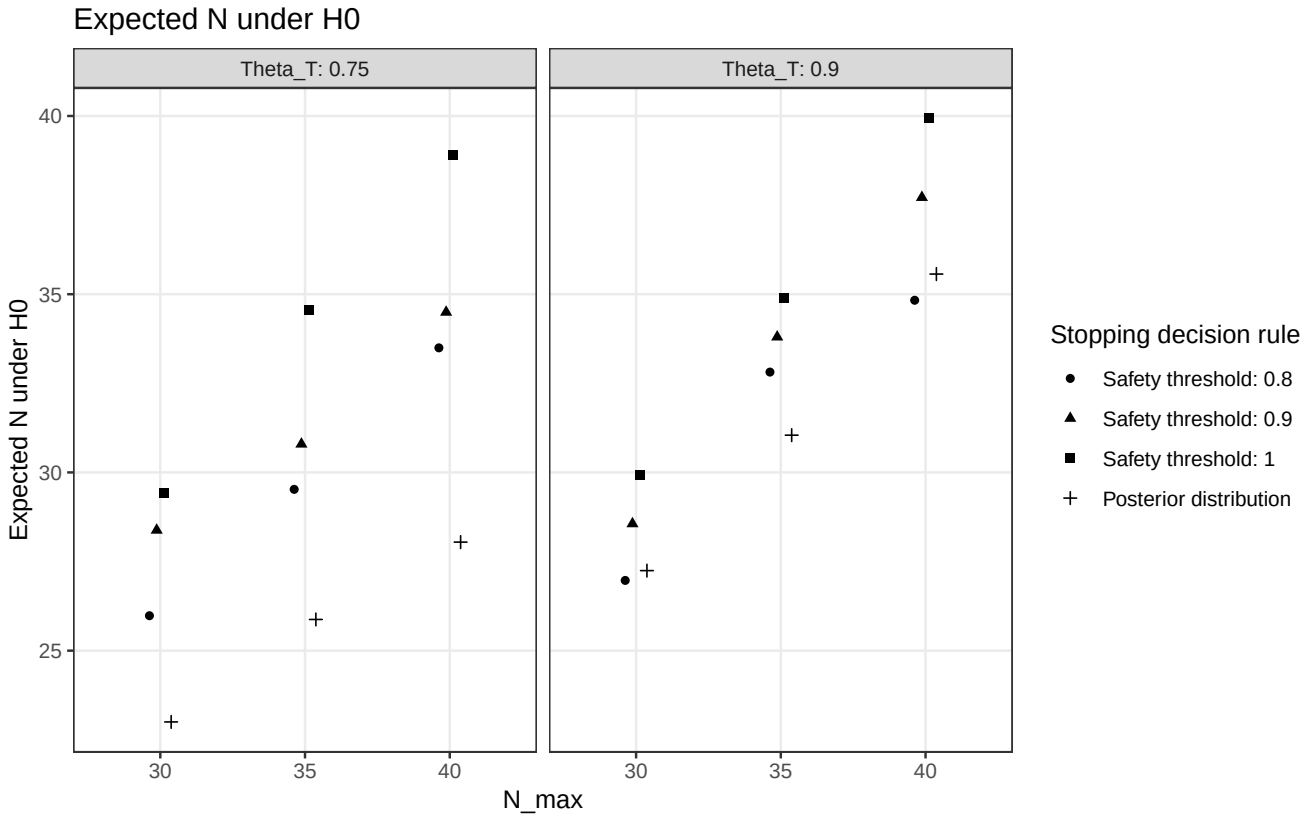
Stopping decision rules:

- Posterior distribution: Stop early if $P(p > p_0|N_j) > \theta_T$,
- Safety threshold: Stop early if $T_{j,i} > \theta_T$ and $PP_j > \theta_S$.

Note that the type-I and type-II errors correspond to the proportion of trials which stopped early. If the maximal number of patients was reached the trial cannot be stopped early.

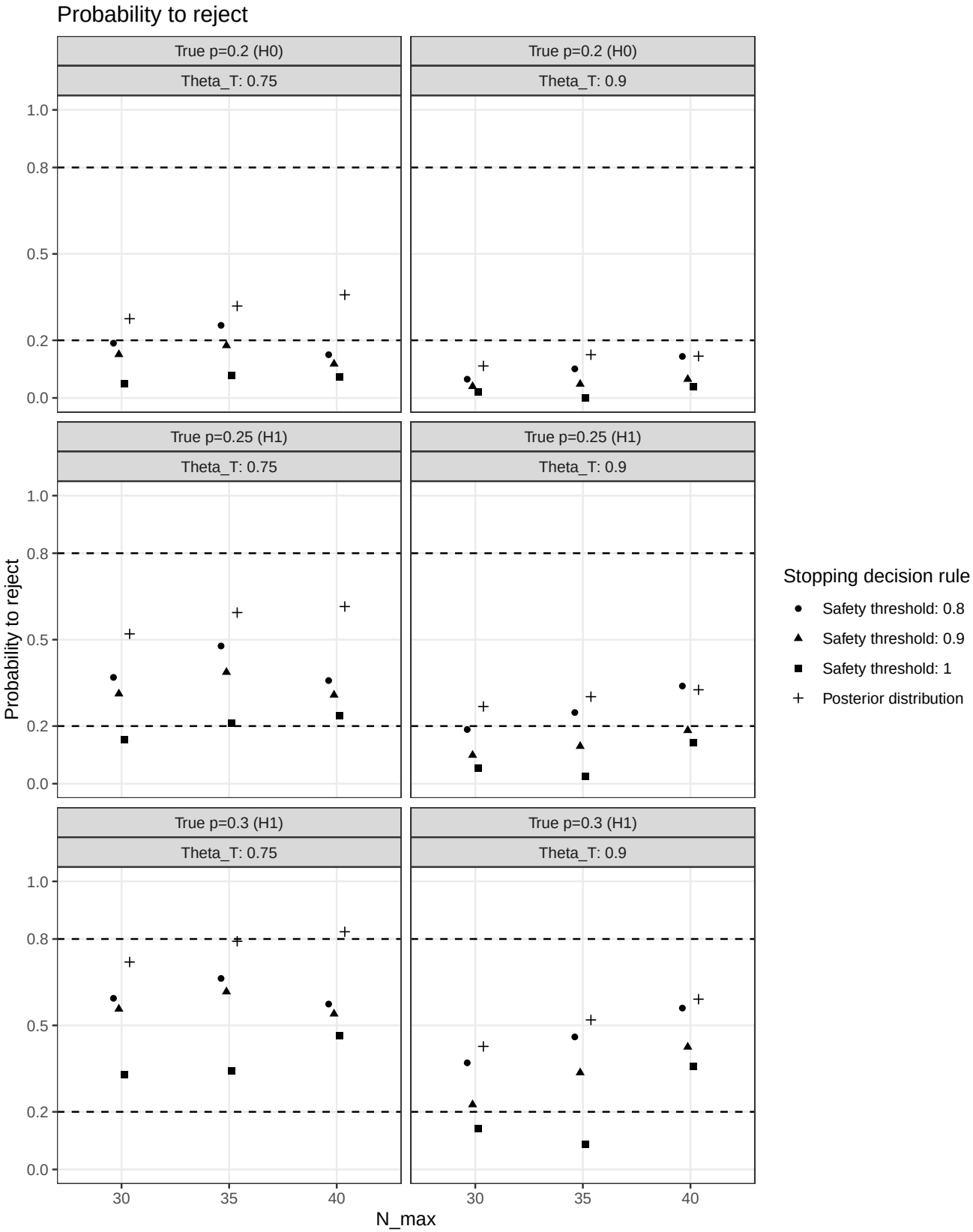

Probability to reject
True p=0.2 (H0)
Theta_T: 0.75
True p=0.2 (H0)
Theta_T: 0.9
True p=0.25 (H1)
Theta_T: 0.75
True p=0.25 (H1)
Theta_T: 0.9
True p=0.3 (H1)
Theta_T: 0.75
True p=0.3 (H1)
Theta_T: 0.9
Probability to reject
1.0
0.8
0.5
0.2
0.0
30
35
40
N_max
Stopping decision rule
Safety threshold: 0.8
Safety threshold: 0.9
Safety threshold: 1
Posterior distribution

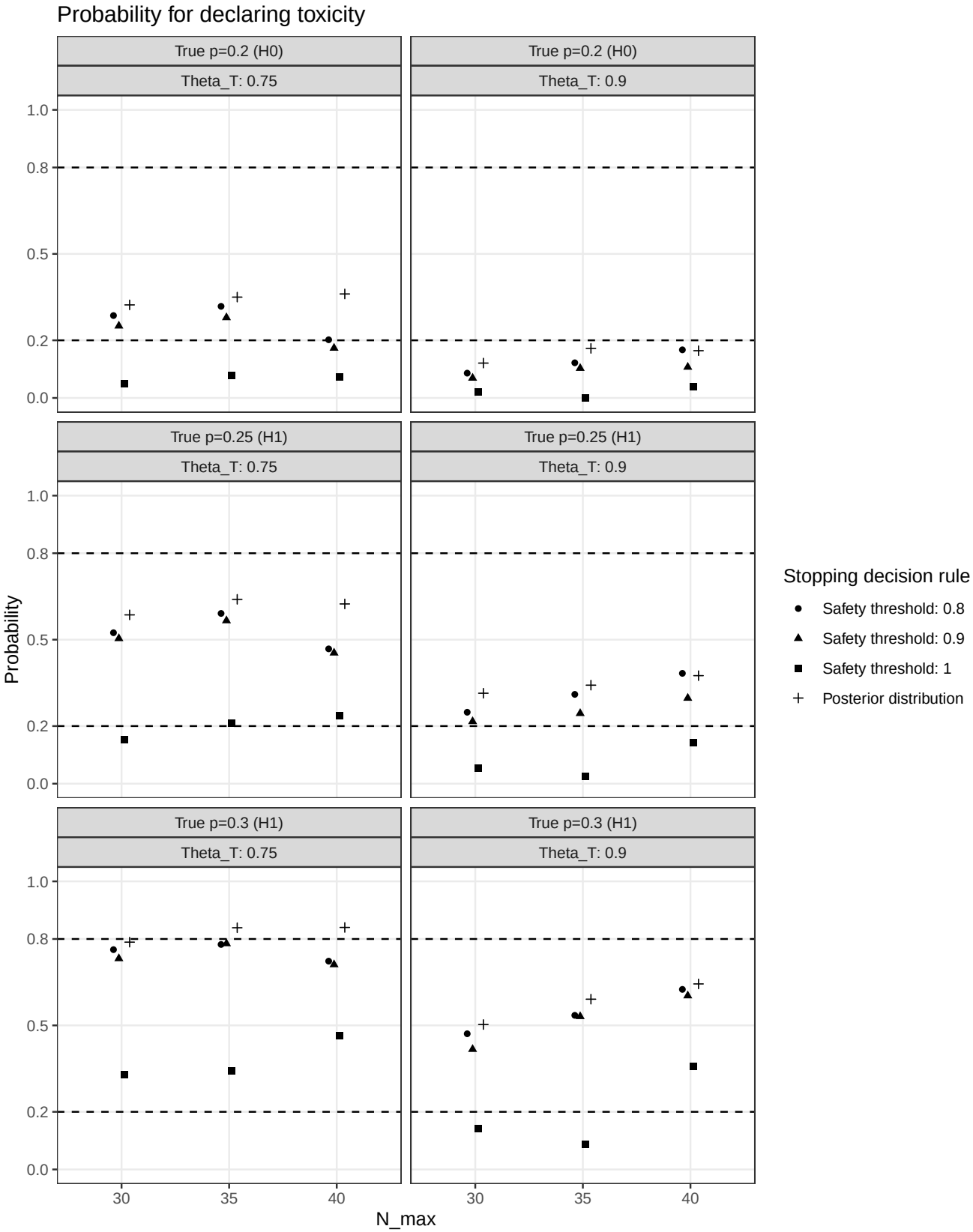

Probability for declaring toxicity
True p=0.2 (H0)
Theta_T: 0.75
True p=0.2 (H0)
Theta_T: 0.9
True p=0.25 (H1)
Theta_T: 0.75
True p=0.25 (H1)
Theta_T: 0.9
True p=0.3 (H1)
Theta_T: 0.75
True p=0.3 (H1)
Theta_T: 0.9
Probability
1.0
0.8
0.5
0.2
0.0
30
35
40
N_max
Stopping decision rule
Safety threshold: 0.8
Safety threshold: 0.9
Safety threshold: 1
Posterior distribution

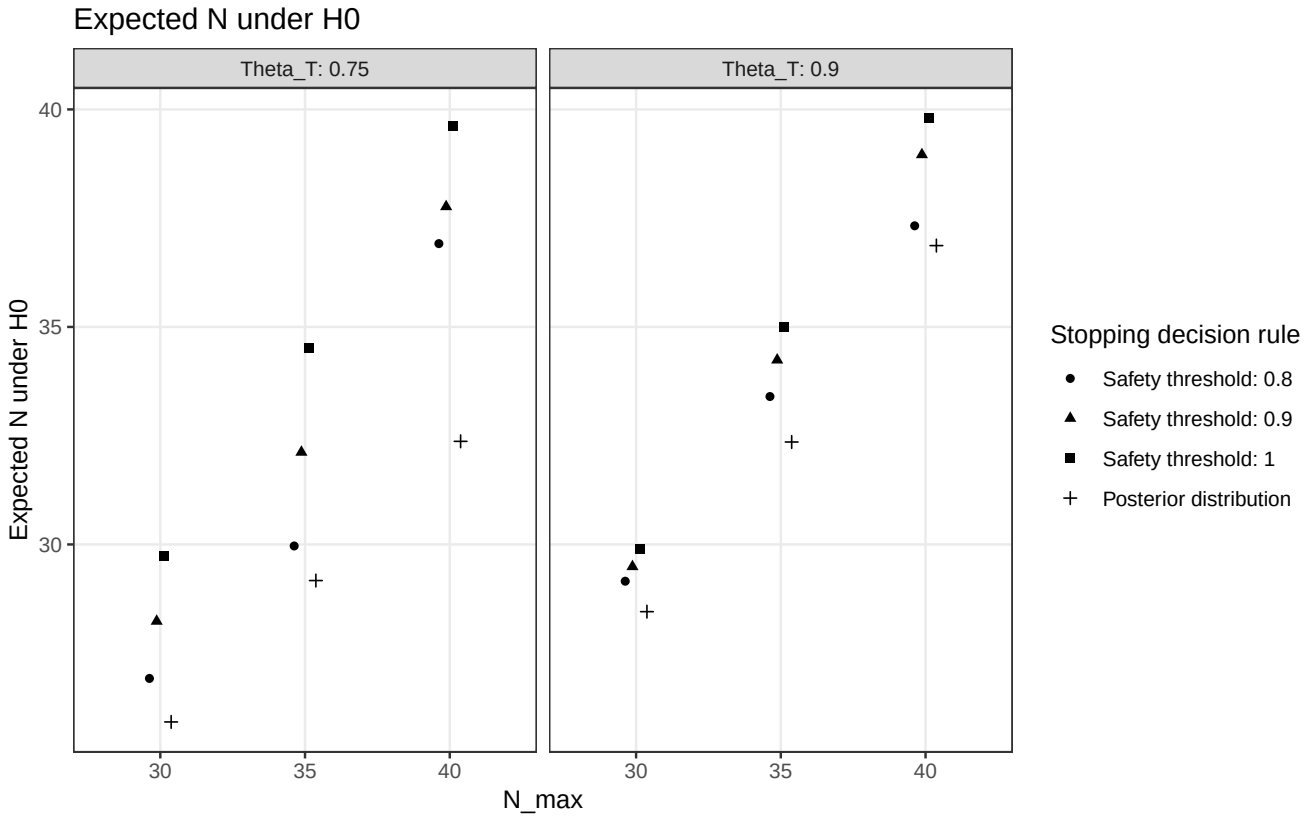


3.3.2 Prior with a=1, b=4

We use the same simulation approach as in the previous subsection but using a a Beta prior with parameters a=1 and b=4.


Probability to reject
True p=0.2 (H0)
Theta_T: 0.75
True p=0.2 (H0)
Theta_T: 0.9
True p=0.25 (H1)
Theta_T: 0.75
True p=0.25 (H1)
Theta_T: 0.9
True p=0.3 (H1)
Theta_T: 0.75
True p=0.3 (H1)
Theta_T: 0.9
Probability to reject
1.0
0.8
0.5
0.2
0.0
30
35
40
N_max
Stopping decision rule
Safety threshold: 0.8
Safety threshold: 0.9
Safety threshold: 1
Posterior distribution


Probability for declaring toxicity
True p=0.2 (H0)
Theta_T: 0.75
True p=0.2 (H0)
Theta_T: 0.9
True p=0.25 (H1)
Theta_T: 0.75
True p=0.25 (H1)
Theta_T: 0.9
True p=0.3 (H1)
Theta_T: 0.75
True p=0.3 (H1)
Theta_T: 0.9
Probability
1.0
0.8
0.5
0.2
0.0
30
35
40
N_max
Stopping decision rule
Safety threshold: 0.8
Safety threshold: 0.9
Safety threshold: 1
Posterior distribution


Expected N under H0
Theta_T: 0.75
Theta_T: 0.9
Expected N under H0
40
35
30
30
35
40
N_max
Stopping decision rule
Safety threshold: 0.8
Safety threshold: 0.9
Safety threshold: 1
Posterior distribution

References

1. Kunzmann K, Grayling MJ, Lee KM, Robertson DS, Rufibach K, Wason JMS. A review of bayesian perspectives on sample size derivation for confirmatory trials. The American Statistician. 2021;75: 424–432. doi:10.1080/00031305.2021.1901782

2. Stefan AM, Gronau QF, Schönbrodt FD, Wagenmakers E-J. A tutorial on bayes factor design analysis using an informed prior. Behavior Research Methods. 2019;51: 1042–1058. doi:10.3758/s13428-018-01189-8

3. Held L, Bové DS. Likelihood and bayesian inference. Springer Berlin Heidelberg; 2020. doi:10.1007/978-3-662-60792-3

4. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Wiley; 2003. doi:10.1002/0470092602

5. Baumgartner C, Klok FA, Carrier M, Limacher A, Moor J, Righini M, et al. Clinical surveillance vs. Anticoagulation for low-risk patiEnts with isolated SubSegmental pulmonary embolism: Protocol for a multicentre randomised placebo-controlled non-inferiority trial (SAFE-SSPE). BMJ Open. 2020;10: e040151. doi:10.1136/bmjopen-2020-040151

6. Grieve AP. Hybrid frequentist/bayesian power and bayesian power in planning clinical trials. Chapman; Hall/CRC; 2022. doi:10.1201/9781003218531

7. O'Hagan A, Stevens JW, Campbell MJ. Assurance in clinical trial design. Pharmaceutical Statistics. 2005;4: 187–201. doi:10.1002/pst.175

8. Rufibach K, Burger HU, Abt M. Bayesian predictive power: Choice of prior and some recommendations for its use as probability of success in drug development. Pharmaceutical Statistics. 2016;15: 438–446. doi:10.1002/pst.1764

9. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Chapman; Hall/CRC; 2013. doi:10.1201/b16018

10. Muirhead RJ, Şoaita AI. On an approach to bayesian sample sizing in clinical trials. 2013; 126–137. doi:10.1214/12-IMSCOLL1007

11. Agresti A, Min Y. Frequentist performance of bayesian confidence intervals for comparing proportions in 2 × 2 contingency tables. Biometrics. 2005;61: 515–523. doi:10.1111/j.1541-0420.2005.031228.x

12. Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. The American Statistician. 2000;54: 280. doi:10.2307/2685779

13. Altham PME. Exact bayesian analysis of a 2 times 2 contingency table, and fisher's "exact" significance test. Journal of the Royal Statistical Society: Series B (Methodological). 1969;31: 261–269. doi:10.1111/j.2517-6161.1969.tb00786.x

14. Grieve AP. Idle thoughts of a "well-calibrated" bayesian in clinical drug development. Pharmaceutical Statistics. 2016;15: 96–108. doi:10.1002/pst.1736

15. Mossop H, Grayling MJ, Gallagher FA, Welsh SJ, Stewart GD, Wason JMS. Advantages of multi-arm non-randomised sequentially allocated cohort designs for phase II oncology trials. British Journal of Cancer. 2022;126: 204–210. doi:10.1038/s41416-021-01613-5

16. Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clinical Trials. 2008;5: 93–106. doi:10.1177/1740774508089279

17. Sambucini V. Efficacy and toxicity monitoring via bayesian predictive probabilities in phase II clinical trials. Statistical Methods & Applications. 2021;30: 637–663. doi:10.1007/s10260-020-00537-3