

Exploration on Bayesian Group Sequential Designs

Minghan Yang

April 2025

Contents

1	Background Knowledge	2
1.1	Dual Criteria Design	2
1.2	Group Sequential Designs	3
1.3	Futility	3
1.4	Interim analysis	3
1.5	Mathematical Derivation of Bayesian Posterior Parameters	3
2	Methods	4
2.1	Effect and probability thresholds	4
2.2	Interim analyses and operating characteristics	6
3	gsbDesign Package	7
3.1	Key functions in the package	7
4	Reproducing results in Gsponer paper	8
4.1	Example 1: PoC trial on Crohn's disease	8
4.2	Example 2: Phase II single arm trials	11
4.3	Example 3: Phase II trial with count endpoint	14
4.4	Example 4: Phase III trial with Bayesian futility criteria for time-to-event endpoint	16
5	Likelihood Ratios, Unit Information LR, and DOR	19
5.1	Positive and Negative Likelihood Ratios	20
5.2	Unit Information Likelihood Ratios	21
5.3	Diagnostic Odds Ratios	24
5.4	Exploring the impact of the presence of decision value	26
6	Applications to Roychoudhury paper	29
6.1	Promising attempt 1	31
6.2	Not so successful attempts	40

Abstract

This report will focus on exploration of the paper “A practical guide to Bayesian group sequential designs” [Gsponer et al. \(2014\)](#), and look at applications to paper “Beyond p-values: A phase II dual-criterion design with statistical significance and clinical relevance” [Roychoudhury et al. \(2018\)](#).

1 Background Knowledge

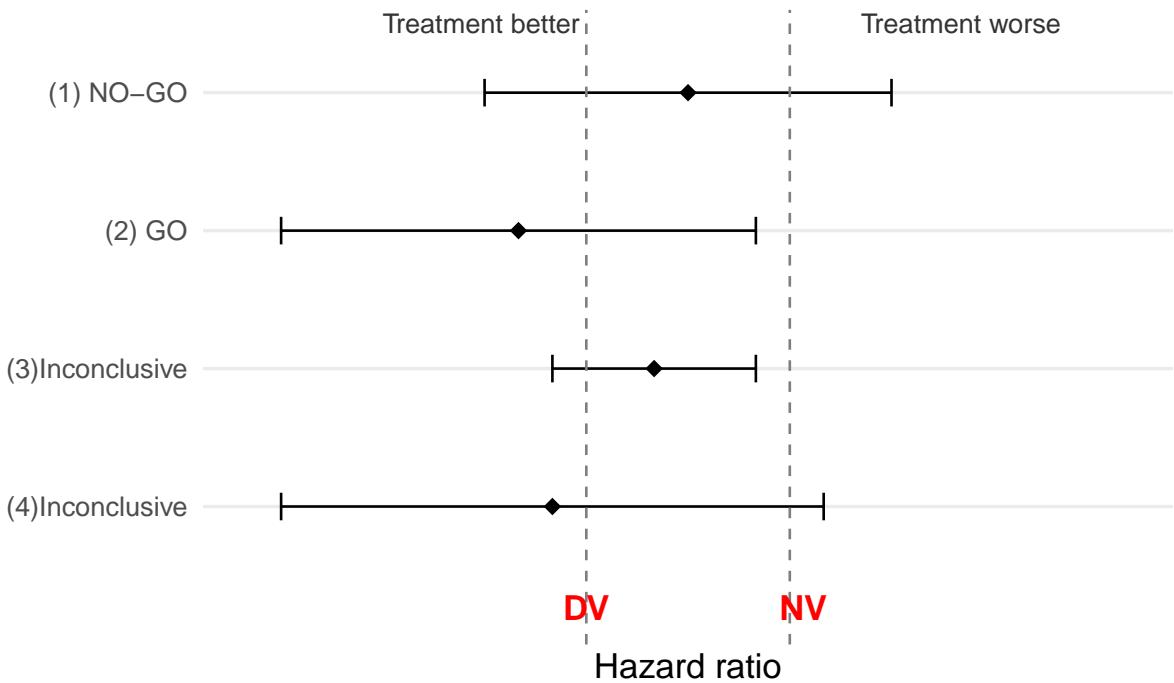
1.1 Dual Criteria Design

By considering both the statistical significance and the clinical relevance criterion, we have both significant and sufficiently large effect estimate.

Required inputs: type I error control (null hypothesis (with null value NV) and type I error α) and a decision value (DV). The DV is the “target difference”. It is the minimal effect estimate needed for trial success (if higher than this value with moderate confidence, then a “GO” decision is made).

The dual-criterion is more demanding, the resulting power of study is less than that of a standard design.

For example, decisions for dual-criterion design when hazard ratio is the primary endpoint:



Sample size calculation in dual criterion design

Given the significance level α , the null value (NV) and the decision value (DV), we can calculate the minimum sample size (for normally distributed data).

$$n_{\min} = \frac{\sigma^2 \times z_{\alpha}^2}{(NV - DV)^2}$$

where σ is the outcome standard deviation, and takes the value 2 under equal randomization for the standard normal approximation to time-to-event data. z_{α} is the $100(1 - \alpha)\%$ quantile of the standard normal distribution. n_{\min} gives the minimum sample size that implies statistical significance if the effect estimate equals the DV. This value is calculated under the situation that both criterion are just satisfied. As illustrated in the below graph, when the effect estimate $\theta = DV$, and the lower bound of the confidence interval just touches the NV so that statistical significance is reached, we have the minimum sample size. Notice that when sample size equals the minimum

sample size, the half-width of the confidence interval $z_\alpha \sqrt{\frac{\sigma^2}{n_{\min}}}$ equals NV-DV, so there will be no “Inconclusive” decisions. When the sample size is larger, the confidence interval becomes narrower, then an “Inconclusive” decision will occur.

1.2 Group Sequential Designs

“Group sequential” means that the data analysis is conducted in interim analyses after every successive group of $2n$ patients. Fixing a maximum number of N groups, a trial is stopped at interim if the (two-sided) p-value is smaller than a pre-specified nominal significance level α , or if N groups of patients have been recruited. The nominal significance level $\hat{\alpha}$ depends on the Type I error rate α and the number of groups N . Standard adjustments for multiple testing are too conservative, since tests are based on accumulating data with a specific dependence structure. The commonly used stopping criteria for success and futility in Bayesian group sequential designs are related to treatment effect size and superior efficacy against placebo. They are often based on the posterior distribution of treatment effect at interim analyses. The Bayesian group sequential methods are designed to provide stopping criteria for related to subsequent decision making. The operational characteristics of interest are probability to stop for success, to stop for futility, or to continue for each interim analysis and the final analysis, as well as corresponding cumulative probabilities. Also, the expected sample size is another characteristic to quantify the gain of group sequential design.

1.3 Futility

Futility refers to stopping a clinical trial early if interim results suggest that the treatment is unlikely to demonstrate a meaningful benefit by the end of the study. This helps to conserve resources, reduce unnecessary exposure to ineffective treatments, and allow quicker redirection of efforts to more promising therapies.

1.4 Interim analysis

[to be added when we explore interim analysis]

1.5 Mathematical Derivation of Bayesian Posterior Parameters

We assume a normal prior distribution for the treatment effect δ : $\delta \sim N(\delta_0, \sigma_0^2)$, and a normal likelihood from observed data D_i : $D_i \sim N(\delta, V_i)$, where δ_0 and σ_0^2 are the prior mean and variance, and V_i is the variance of the observed treatment effect.

Using Bayes’ theorem: $P(\delta | D_i) \propto P(D_i | \delta)P(\delta)$.

Expanding both normal densities in the exponent:

$$\begin{aligned} P(\delta | D_i) &\propto \exp \left(-\frac{1}{2} \left[\frac{(\delta - \delta_0)^2}{\sigma_0^2} + \frac{(D_i - \delta)^2}{V_i} \right] \right) \\ &\propto \exp \left(-\frac{1}{2} \left[\left(\frac{1}{\sigma_0^2} + \frac{1}{V_i} \right) \delta^2 - 2 \left(\frac{\delta_0}{\sigma_0^2} + \frac{D_i}{V_i} \right) \delta \right] \right) \end{aligned}$$

Since a normal distribution has the general form:

$$\exp\left(-\frac{1}{2}\frac{(\delta - \mu_{\text{posterior}})^2}{\sigma_{\text{posterior}}^2}\right) = \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_{\text{posterior}}^2}\delta^2 - 2\frac{\mu_{\text{posterior}}}{\sigma_{\text{posterior}}^2}\delta + C\right]\right),$$

comparing terms, the posterior mean and variance are:

$$\mu_{\text{posterior}} = \frac{\frac{\delta_0}{\sigma_0^2} + \frac{D_i}{V_i}}{\frac{1}{\sigma_0^2} + \frac{1}{V_i}}, \quad \sigma_{\text{posterior}}^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{V_i}}.$$

This formula shows that the inverse variance (precision) of the posterior is the sum of the prior and data precisions. The posterior variance always smaller than either the prior or the likelihood variance.

2 Methods

2.1 Effect and probability thresholds

Traditionally, for treatment effect δ , null hypothesis is $\delta \leq 0$, and alternative is $\delta > 0$.

Given type I error rate α , power $1 - \beta$, treatment effect size δ_1 (minimal clinically relevant effect), in a standard two-group normal case with known standard deviation, the null will be rejected if observed average treatment effect $\hat{\delta}$ exceeds threshold $D_T = z_{1-\alpha} \cdot \frac{\delta_1}{z_{1-\alpha} + z_{1-\beta}}$.

Deduce this D_T : The null hypothesis is rejected when $\hat{\delta}$ exceeds a threshold D_T , so:

$P(\hat{\delta} > D_T | \delta = 0) = \alpha$. Standardizing: $P\left(Z > \frac{D_T}{\sigma/\sqrt{n}} | \delta = 0\right) = \alpha$, which implies: $\frac{D_T}{\sigma/\sqrt{n}} = z_{1-\alpha}$. Thus, $D_T = z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ (1).

For the power condition (type II error control), $P(\hat{\delta} > D_T | \delta = \delta_1) = 1 - \beta$, which gives: $\frac{D_T - \delta_1}{\sigma/\sqrt{n}} = -z_{1-\beta}$. Rearranging, $D_T = \delta_1 - z_{1-\beta} \cdot \frac{\sigma}{\sqrt{n}}$ (2).

Equating (1) and (2), we get $\frac{\sigma}{\sqrt{n}} = \frac{\delta_1}{z_{1-\alpha} + z_{1-\beta}}$.

Substituting into the expression for D_T : $D_T = z_{1-\alpha} \cdot \frac{\delta_1}{z_{1-\alpha} + z_{1-\beta}}$. \square

This threshold is typically smaller than δ_1 , so the null can be rejected for effect sizes that are not seen as clinically relevant.

To avoid situations where a statistically significant result is obtained with a small treatment effect, the double criterion is required to measure the success for clinical trials.

From Bayesian perspective, we base on posterior probabilities, and the success criteria are

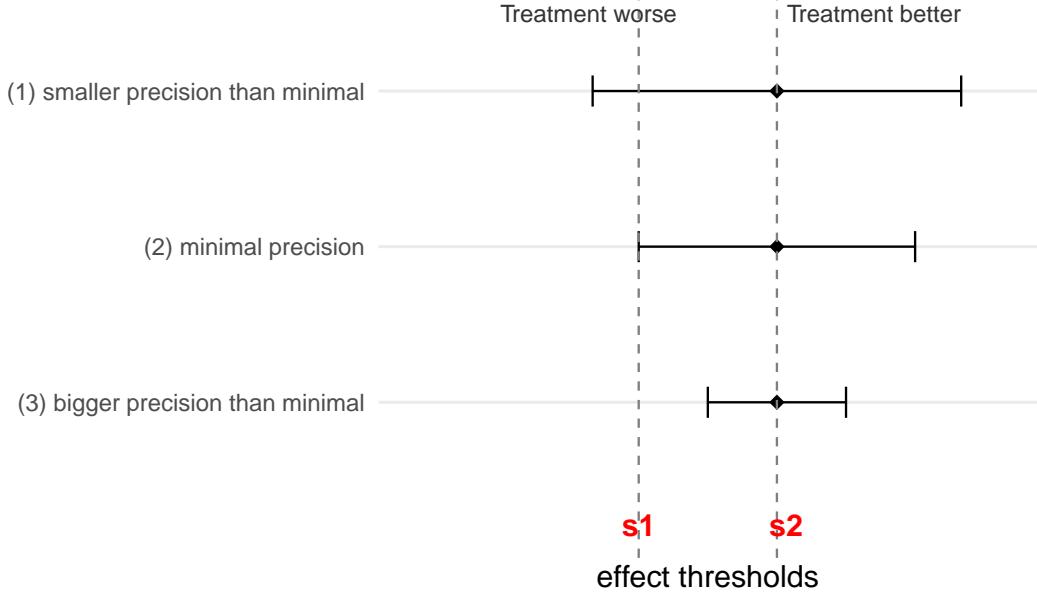
1. $\mathbb{P}(\delta > s_1 | \text{Data}) > p_1$,
2. $\mathbb{P}(\delta > s_2 | \text{Data}) > p_2$,

where s_1, s_2 are some effect thresholds, and p_1, p_2 are some probability thresholds.

Suppose posterior is normally distributed $\delta \sim N(\mathbb{E}(\delta | \text{Data}), \text{Sd}(\delta | \text{Data}))$, then with minimal precision satisfying the double criteria, we have

$$\mathbb{E}(\delta | \text{Data}) = \frac{s_1 z_{1-p_2} - s_2 z_{1-p_1}}{z_{1-p_2} - z_{1-p_1}}, \quad \text{Sd}(\delta | \text{Data}) = \frac{s_2 - s_1}{z_{1-p_2} - z_{1-p_1}}.$$

Derivation: The minimal precision means the maximal variance / widest confidence interval when the two criteria are just met. For larger precision, it may happen that one of the two criteria not satisfied. As shown in the below figure, the confidence interval in case (1) is too wide that the first criterion is not met. When at minimal precision (case (2)), two criteria are just met.



With the mean $\mathbb{E}(\delta | \text{Data})$ and standard deviation $\text{Sd}(\delta | \text{Data})$, when the two criteria are just met:

$$\begin{aligned}\mathbb{P}(\delta > s_1) &= 1 - \Phi\left(\frac{s_1 - \mathbb{E}(\delta | \text{Data})}{\text{Sd}(\delta | \text{Data})}\right) = p_1, \\ \mathbb{P}(\delta > s_2) &= 1 - \Phi\left(\frac{s_2 - \mathbb{E}(\delta | \text{Data})}{\text{Sd}(\delta | \text{Data})}\right) = p_2.\end{aligned}$$

Because cdf is increasing:

$$\begin{aligned}\frac{s_1 - \mathbb{E}(\delta | \text{Data})}{\text{Sd}(\delta | \text{Data})} &= z_{1-p_1}, \\ \frac{s_2 - \mathbb{E}(\delta | \text{Data})}{\text{Sd}(\delta | \text{Data})} &= z_{1-p_2},\end{aligned}$$

Solving for expectation and standard deviation,

$$\mathbb{E}(\delta | \text{Data}) = \frac{s_1 z_{1-p_2} - s_2 z_{1-p_1}}{z_{1-p_2} - z_{1-p_1}},$$

$$\text{Sd}(\delta | \text{Data}) = \frac{s_2 - s_1}{z_{1-p_2} - z_{1-p_1}}. \quad \square$$

Commonly, we use $s_1 = 0$, $s_2 = \delta^*$, $p_1 = 1 - \alpha$, and $p_2 = 0.5$, for the ease of interpretation and communication with clinicians. The δ^* is the required minimal observed difference for a trial to succeed. Then the success criteria are

1. $\mathbb{P}(\delta > 0 | \text{Data}) > 1 - \alpha,$
2. $\mathbb{P}(\delta > \delta^* | \text{Data}) > 0.5.$

This is similar to the double criterion in the frequentist approach when prior information is vague. Thus, the prior information for δ or for both treatment arms is crucial to Bayesian design.

We also define the criteria for futility:

1. $\mathbb{P}(\delta < f_1 | \text{Data}) > q_1,$
2. $\mathbb{P}(\delta < f_2 | \text{Data}) > q_2.$

This is useful in the presence of interim analyses, because it could help to decide early stopping for futility. It is also useful in final analysis to classify a clear success, a clear failure, or an indeterminate outcome. The indeterminate case is undesirable, and happens when neither success criteria nor futility criteria are fulfilled.

2.2 Interim analyses and operating characteristics

Typically, MCMC methods are used for evaluating the posterior distribution. For evaluation of operating characteristics, people use approximate normalised likelihoods and normal priors.

For a two-arm clinical trial with one or more interim analyses, let $Y_{kij} \sim N(\theta_k, \sigma^2)$ be the observation for treatment $k = 1, 2$, in stage $i = 1, \dots, l$, for subject $j = 1, \dots, n_{ki}$, with known standard deviation σ . The treatment effect is $\delta = \theta_2 - \theta_1$. If prior information is $\delta \sim N(\delta_0, \sigma_0^2)$, then the posterior at stage i is

$$\delta | \text{stage } i \sim N\left(\omega_i \delta_0 + (1 - \omega_i) D_i, \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{V_i}}\right)$$

where $\omega_i = 1/(\sigma_0^2/V_i + 1)$, D_i is the aggregated treatment effect at stage i , and V_i is the corresponding variance. Derivation of Bayesain posterior mean and variance could be found in Section 1.5.

The term ω_i represents the weight given to the prior mean in updating the posterior mean:

- If the prior variance σ_0^2 is large (weak prior), then $\omega_i \approx 0$, meaning the posterior is dominated by the observed data.
- If the observed variance V_i is large (noisy data), then $\omega_i \approx 1$, meaning the posterior remains close to the prior.

If priors for two arms are available, say $\theta_1 \sim N(\theta_{10}, \sigma_1^2)$ and $\theta_2 \sim N(\theta_{20}, \sigma_2^2)$, then the posterior at stage i is

$$\delta | \text{stage } i \sim N\left(\{\omega_{2i}\theta_{20} + (1 - \omega_{2i})Y_{2i}\} - \{\omega_{1i}\theta_{10} + (1 - \omega_{1i})Y_{1i}\}, \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{V_{1i}}} + \frac{1}{\frac{1}{\sigma_2^2} + \frac{1}{V_{2i}}}\right),$$

where Y_{ki} is the aggregated mean in arm k in stage i , V_{ki} is the corresponding variance, and $\omega_{ki} = 1/(\sigma_k^2/V_{ki} + 1)$.

3 gsbDesign Package

3.1 Key functions in the package

gsbDesign, gsbSimulation, gsb

```
library(gsbDesign)
design <- gsbDesign(nr.stages = 2,
patients = c(10,20),
sigma=88,
criteria.success = c(0,0.95,50,0.5),
criteria.futility = c(40,0.9),
prior.control = c(49,20),
prior.treatment = c(49,0))
design

##
## *** Trial Design ***
##
## number of stages: 2
##
## prior mean control:    49
## prior patients control:   20
## prior mean treatment:   49
## prior patients treatment: 0
##
## patients:
##           control treatment
## stage 1      10        20
## stage 2      10        20
##
## sigma control: 88 ; sigma treatment: 88
##
## criteria:
##       type stage number value probability
##   success     1      1     0      0.95
##   success     1      2     50     0.50
##   success     2      1     0      0.95
##   success     2      2     50     0.50
##   futility    1      1     40     0.90
##   futility    2      1     40     0.90

# simulation scenarios
simulation1 <- gsbSimulation(
  truth=list(49,c(0,40,50,60,70)),
```

```

grid.type="sliced",
method = "both",
type.update="per arm",
nr.sim=10000,
warnings.sensitivity = 500,
seed=1)

```

If `type.update = "treatment effect"`, the Bayesian update from prior to posterior is calculated on treatment effect delta. If `type.update = "per arm"`, the update is calculated separately in the placebo and the treatment arm. In this case it is possible to enter prior information in only one arm. For `type.update = "per arm"` only a simulation method is implemented.

If `type.update = "per arm"` and `grid.type = "table"` the argument has to be specified as list containing a vector of true control values and true treatment values.

If `type.update = "per arm"` and `grid.type = "sliced"` the argument has to be specified as list containing a vector of true control values and a vector of true deltas (= treatment - control).

4 Reproducing results in Gsponer paper

In this section we look at the four examples in the paper, to summarize and reproduce the results. In the next section, we are going to apply these methods to the cases in the paper by Roychoudhury et al. (2018).

4.1 Example 1: PoC trial on Crohn's disease

```

# design
library(gsbDesign)
design <- gsbDesign(nr.stages = 2,
patients = c(10,20),
sigma=88,
criteria.success = c(0,0.95,50,0.5),
criteria.futility = c(40,0.9),
prior.control = c(49,20),
prior.treatment = c(49,0))
design

##
## *** Trial Design ***
##
## number of stages: 2
##
## prior mean control:    49
## prior patients control:  20
## prior mean treatment:   49

```

```

## prior patients treatment: 0
##
## patients:
##           control treatment
## stage 1      10        20
## stage 2      10        20
##
## sigma control: 88 ; sigma treatment: 88
##
## criteria:
##           type stage number value probability
## success     1      1      0      0.95
## success     1      2     50      0.50
## success     2      1      0      0.95
## success     2      2     50      0.50
## futility    1      1     40      0.90
## futility    2      1     40      0.90

# simulation scenarios
simulation <- gsbSimulation(
  truth=list(49,c(0,40,50,60,70)),
  grid.type="sliced",
  method = "both",
  type.update="per arm",
  nr.sim=10000,
  warnings.sensitivity = 500,
  seed=1)

#simulation1

# operating characteristics
result <- gsb(design=design,simulation=simulation)
#tables and graphics
tab(result, what="cumulative all")

##           control treatment delta stage1.suc stage1.fut stage1.ind stage2.suc
## 1          49        49     0     0.013     0.627     0.360     0.014
## 2          49        89     40     0.324     0.066     0.609     0.412
## 3          49        99     50     0.502     0.026     0.471     0.628
## 4          49       109     60     0.678     0.008     0.314     0.812
## 5          49       119     70     0.816     0.003     0.181     0.924
##           stage2.fut stage2.ind
## 1          0.842     0.145
## 2          0.104     0.483

```

```

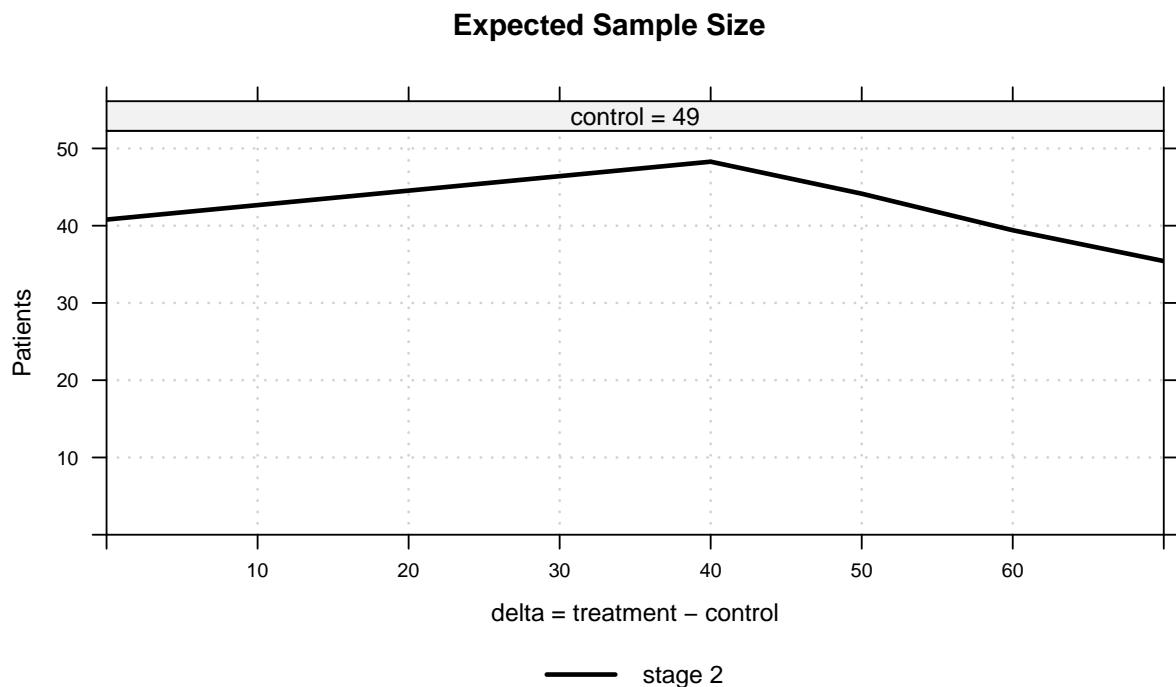
## 3      0.036      0.335
## 4      0.011      0.177
## 5      0.004      0.072

tab(result, what="sample size")

##   control treatment delta stage1 stage2
## 1      49        49     0    30 40.791
## 2      49        89    40    30 48.282
## 3      49        99    50    30 44.142
## 4      49       109    60    30 39.408
## 5      49       119    70    30 35.421

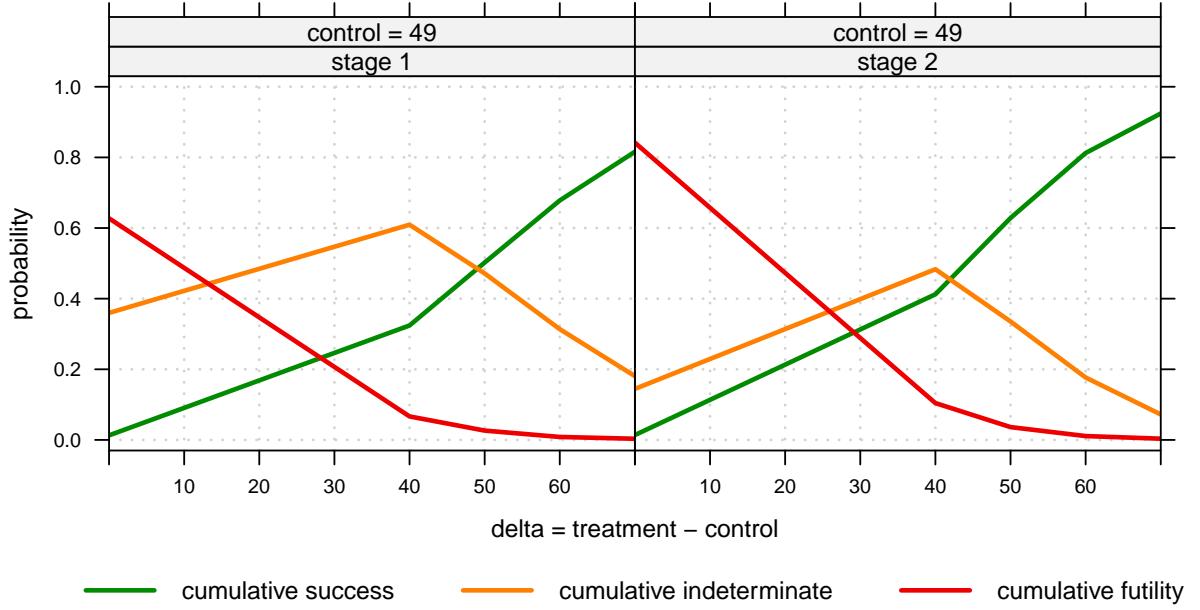
plot(result, what="sample size", sliced=TRUE)

```



```
plot(result, what="cumulative all", sliced=TRUE)
```

Operating Characteristics



4.2 Example 2: Phase II single arm trials

Simon's paper describes how we obtain the values in Table II. The response probability $p_0=0.4$ Simon (1989). In Simon (1989), they searched for optimal design (first stage sample size and total sample size) for a two-stage design.

Two relevant response rates: π_C^*, π_E^* . π_C^* is the assumed historical control rate, and if true underlying response rate is larger than π_E^* , then it is likely that the drug is sufficiently promising. These correspond to p_0, p_1 in Simon's paper. p_0 is similar to the statistical significance criteria threshold, and p_1 is similar to the clinical relevance critaria threshold. In Gsponer et al. (2014), the operating characteristics are obtained by package Gerber and Gsponer (2016), by reformulating single-arm design as two-arm design, and using normal approximation for the endpoint. For r responders in n patients, define y as log-odds of observed response rate. Its distribution is approximately

$$y \sim N(\text{logit}(\pi_E), \sigma_E^2/n),$$

where π_E is the true response rate, and $\sigma_E^2 = 1/\pi_E + 1/(1-\pi_E)$.

Derivation of approximated variance: σ_E^2/n

Logit function $g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$ has derivative $g'(\pi) = \frac{1}{\pi(1-\pi)}$.

Applying the Delta Method: $\text{Var}(\text{logit}(\pi)) \approx \left(\frac{1}{\pi(1-\pi)}\right)^2 \text{Var}(\pi)$.

Since $\text{Var}(\pi) \approx \frac{\pi(1-\pi)}{n}$, we get $\text{Var}(\text{logit}(\pi)) \approx \frac{1}{\pi(1-\pi)} \cdot \frac{\pi(1-\pi)}{n} = \frac{1}{n\pi(1-\pi)}$. \square

A virtual control arm is needed for this single-arm design to mimic a two-arm design. This is done by setting sample size in control group to 0, and by providing an extremely informative prior. The prior is chosen to be centered at π_C^* with an extremely small variance.

```

logit <- function(p) { log(p/(1-p)) }
expit <- function(x) { exp(x)/(1+exp(x)) }
p0 = 0.4
p1 = 0.6
suc = logit(23.5/46) - logit(p0)
fut = logit(7.5/16) - logit(p0)
design <- gsbDesign(nr.stages=2,
patients=cbind( c(0,0), c(16,30) ),
sigma=2,
criteria.success=rbind( c(NA,NA), c(suc,0.5) ),
criteria.futility=rbind( c(fut,0.5), c(NA,NA) ),
prior.control=c( logit(p0),1000),
prior.treatment=c( logit(p0),0.001))

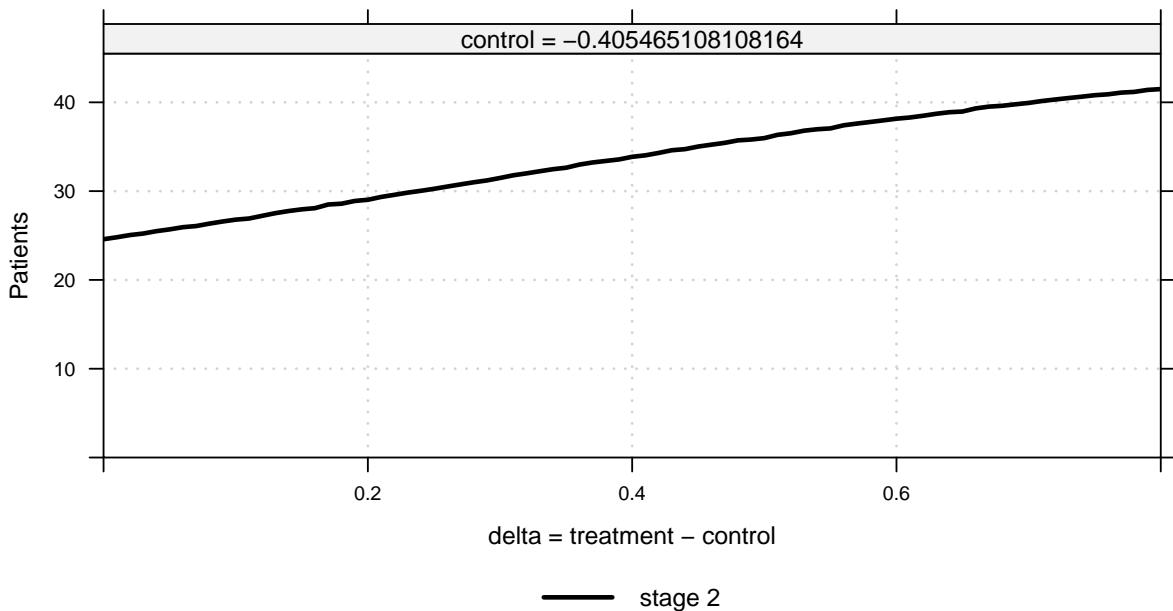
simulation <- gsbSimulation(truth=list(logit(p0),seq(0,0.8,0.01)),
grid.type = "sliced",
type.update = "per arm",
nr.sim = 100000,
warnings.sensitivity = 2000,
seed = 1)
result <- gsb(design,simulation)
result

##
## *** Group Sequential Bayesian Design ***
##
## Analysis N1      N2
##      Prior 1000  0.001
##          1    0 16.000
##          2    0 30.000
##
## sigma treatment: 2  sigma control: 2
##
## access the operating characteristics via the data.frame "OC" in the output of "gsb()"
## or the functions "tab()" and "plot()".
##
## names in output list:
## [1] "OC"           "design"        "simulation"    "delta.grid"    "warnings"
## [6] "system.time"

#tab(result, what="cumulative all")
#tab(result, what="sample size")
plot(result, what="sample size", sliced=TRUE)

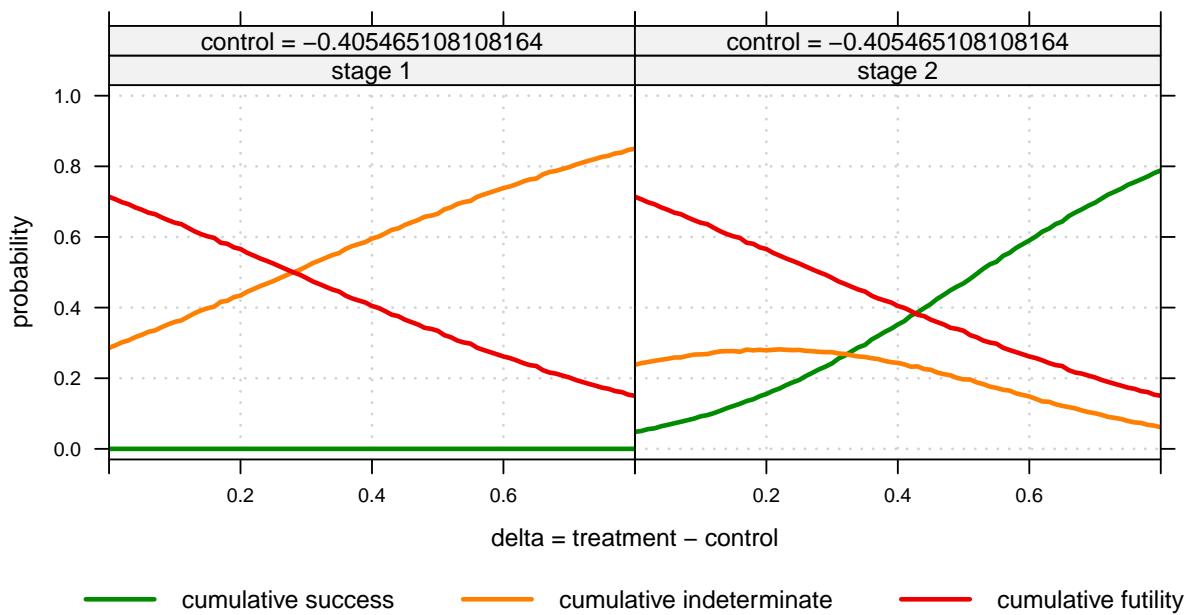
```

Expected Sample Size



```
plot(result, what="cumulative all", sliced=TRUE)
```

Operating Characteristics

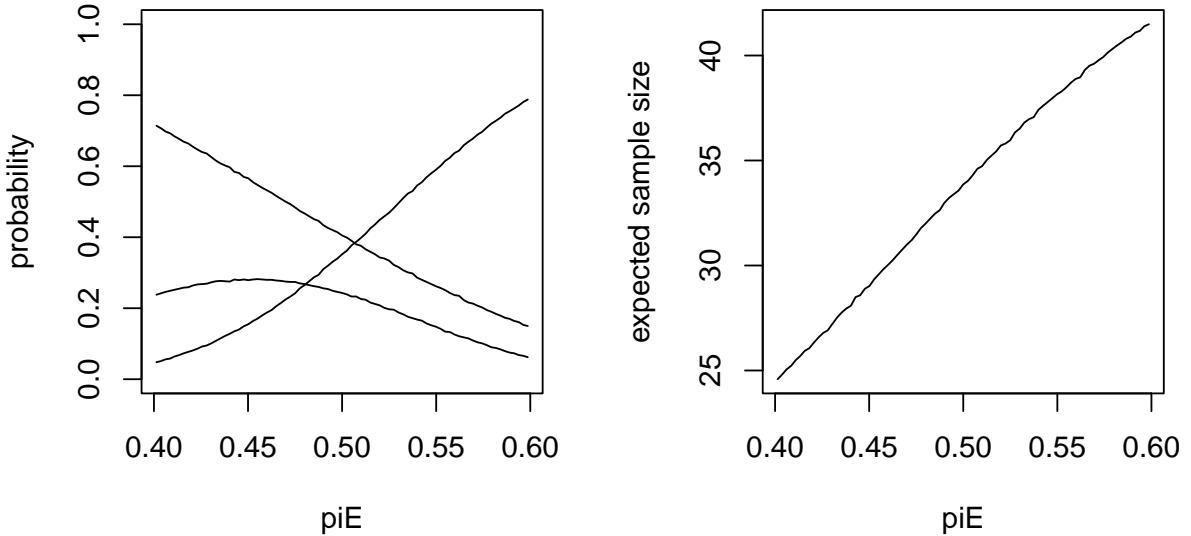


```
par(mfrow=c(1, 2))
y1 = tab(result, what='futility')[,4]
y2 = tab(result, what='success')[,5]
y3 = 1 - y1 - y2
piE = expit( -0.4+seq(0,0.8,0.01) ) # transform back to original scale
plot(piE,y1,type='l',ylim=c(0,1),ylab='probability')
lines(piE,y2)
```

```

lines(piE,y3)
y3 = tab(result,what='sample size')[,5]
plot(piE,y3,type='l',ylab='expected sample size')

```



4.3 Example 3: Phase II trial with count endpoint

```

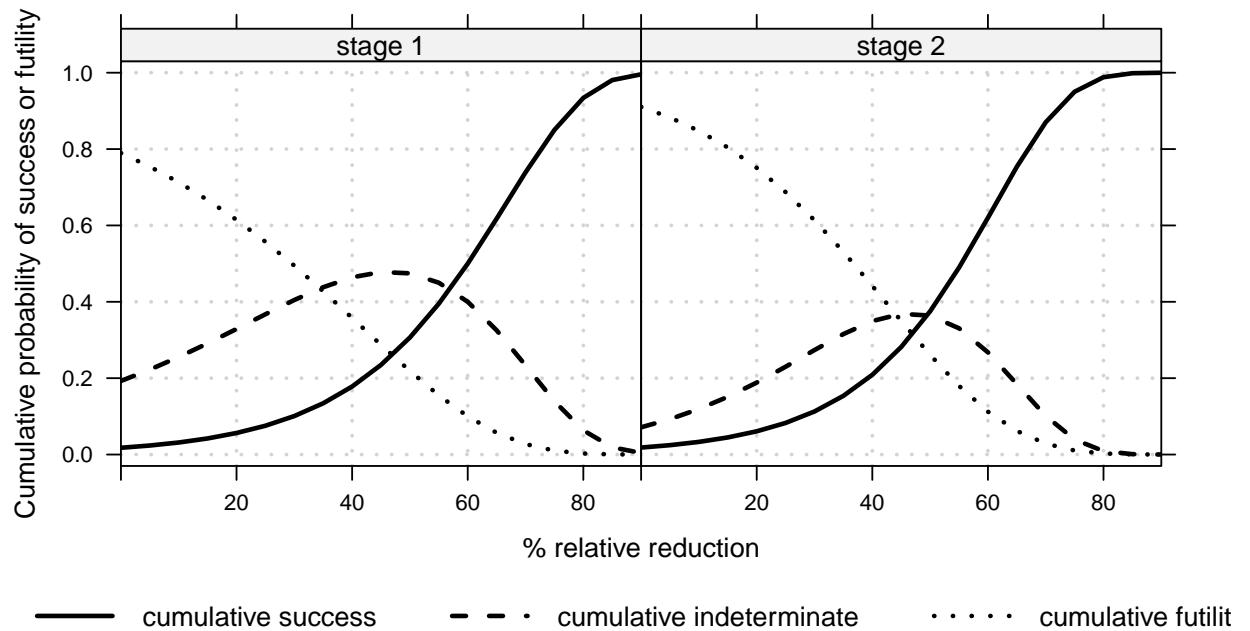
### Case 1 non-informative prior
kappa=2 ; rate1=log(14.8) ; redscen=seq(0,0.9,0.05); OCout2=NULL
for(i in seq(along=redscen)){
  red=redscen[i]
  rate2=log(14.8*(1-red))
  delta=rate1-rate2
  sigma1=sqrt(kappa+(1/rate1)) ; sigma2=sqrt(kappa+(1/rate2))
  design1 <- gsbDesign(nr.stages=2,
  patients=25,
  sigma=c(sigma2,sigma1),
  criteria.success=c(0,0.9, log(2.5), 0.5),
  criteria.futility=c(log(1.42),0.5,log(2.5),0.9),
  prior.difference="non-informative")
  simulation1 <- gsbSimulation(truth=c(-1,delta,2), type.update="treatment effect",
  method= "numerical integration")
  x1 <- gsb(design=design1, simulation=simulation1)
  sub=x1$OC[, "delta"] != -1
  xout=x1$OC[sub,]
  xout$delta=red*100
}

```

```

 0Cout2=rbind(0Cout2,xout)
}
x1$0C=0Cout2
p1=plot(x1,what="cumulative all")
p1$ylab="Cumulative probability of success or futility"; p1$xlab="% relative reduction\n"
p1$main=NULL
p1$panel.args.common$col <- 1 ; p1$panel.args.common$lty <- 1:3
p1$legend$bottom$args$key$lines$col <- 1 ; p1$legend$bottom$args$key$lines$lty <- 1:3
plot(p1)

```



```

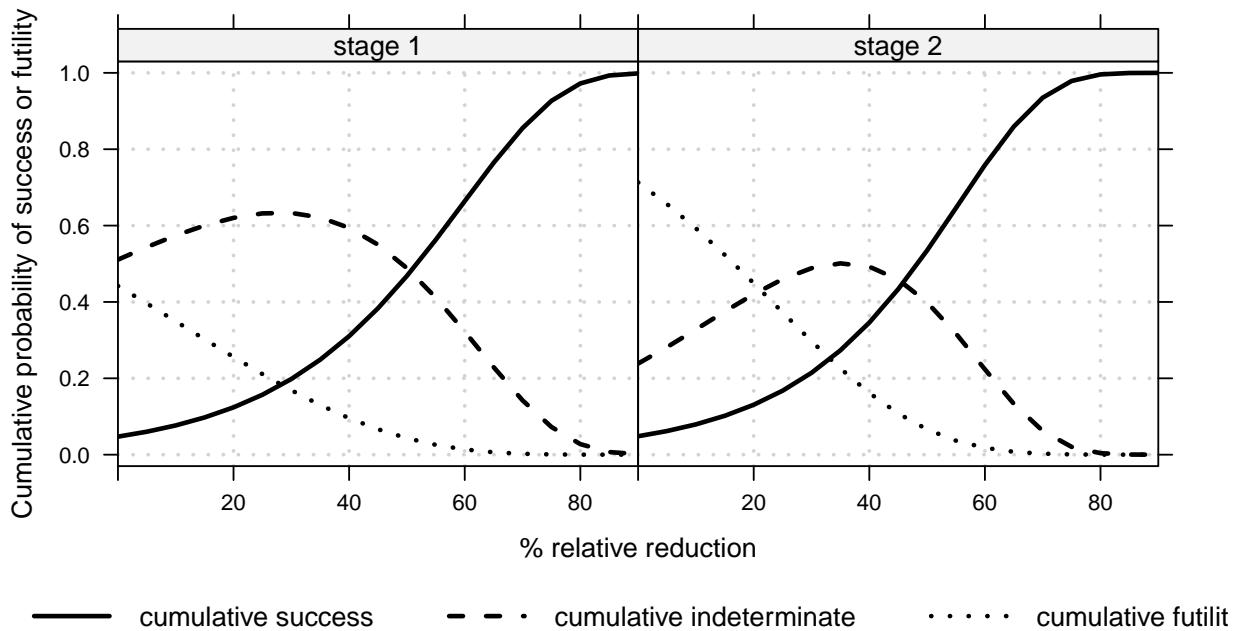
#### Case 2 -informative prior
kappa=2 ; rate1=log(14.8); redscen=seq(0,0.9,0.05) ; 0Cout=NULL
for(i in seq(along=redscen)){
  red=redscen[i]
  rate2=log(14.8*(1-red))
  delta=rate1-rate2
  delta.prior=rate1-log(14.8*(1-0.75))
  sigma1=sqrt(kappa+(1/rate1)) ; sigma2=sqrt(kappa+(1/rate2))
  design1 <- gsbDesign(nr.stages=2,
  patients=25,sigma=c(sigma2,sigma1),
  criteria.success=c(0,0.9, log(2.5), 0.5),
  criteria.futility=c(log(1.42),0.5,log(2.5),0.9),
  prior.difference=c(delta.prior,10,10))
  simulation1 <- gsbSimulation(truth=c(-1,delta,2),
  type.update="treatment effect",

```

```

method= "numerical integration")
x1 <- gsb(design=design1, simulation=simulation1)
sub=x1$OC[, "delta"] !=-1
xout=x1$OC[sub,]
xout$delta=red*100
OCout=rbind(OCout,xout)
}
x1$OC=OCout
p1=plot(x1,what="cumulative all")
p1$xlab="% relative reduction\n" ; p1$ylab="Cumulative probability of success or futility"
p1$main=NULL
p1$panel.args.common$col <- 1 ; p1$panel.args.common$lty <- 1:3
p1$legend$bottom$args$key$lines$col <- 1 ; p1$legend$bottom$args$key$lines$lty <- 1:3
plot(p1)

```



4.4 Example 4: Phase III trial with Bayesian futility criteria for time-to-event endpoint

```

succ <- rbind(c(0,0.99995),
  c(0,0.9995),
  c(log(1.075), 0.975))
fut <- rbind(c(0,0.99995),
  c(0,0.9995),
  NA)
## additional rules for stopping for futility (alternative design)

```

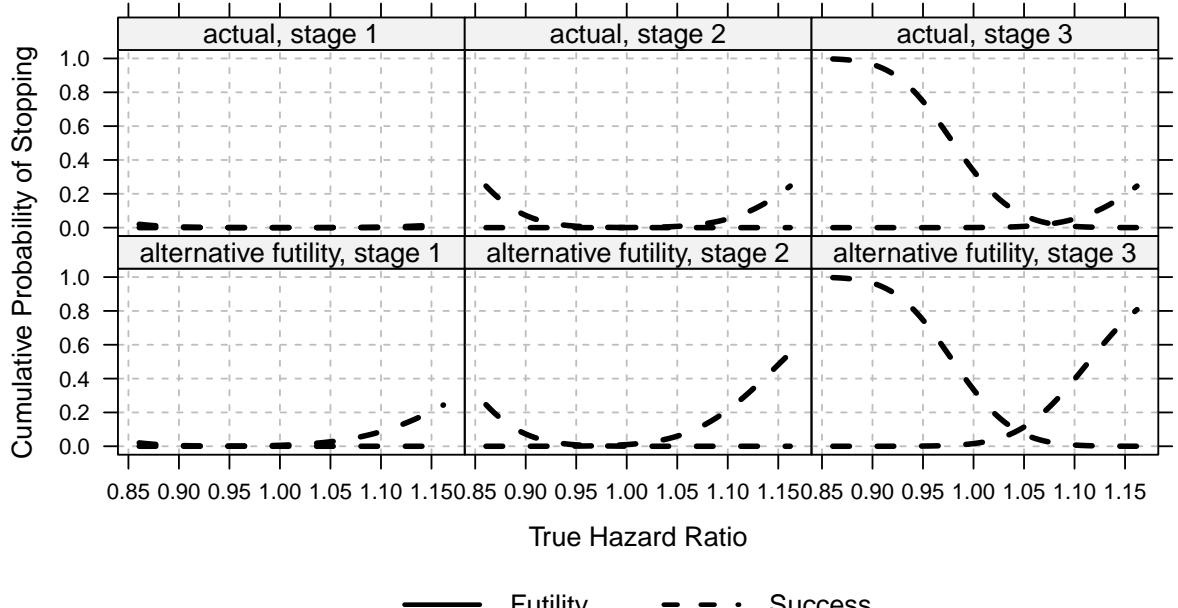
```

fut1 <- rbind(c(log(1.075),0.95),
  c(log(1.075),0.9),
  c(log(1.075),0.8))

## exchange success and futility criteria as smaller values of the endpoint are better here
design1 <- gsbDesign(nr.stages = 3,
  patients = 300,
  criteria.success = fut,
  criteria.futility = succ,
  sigma = 1,
  prior.difference = "non-informative")
design2 <- gsbDesign(nr.stages = 3,
  patients = 300,
  criteria.success = fut1,
  criteria.futility = succ,
  sigma = 1,
  prior.difference = "non-informative")
truth <- seq(-0.15,0.15,by=0.005)
sim <- gsbSimulation(truth = truth,
  grid.type = "manually",
  type.update = "treatment effect",
  method = "numerical integration")

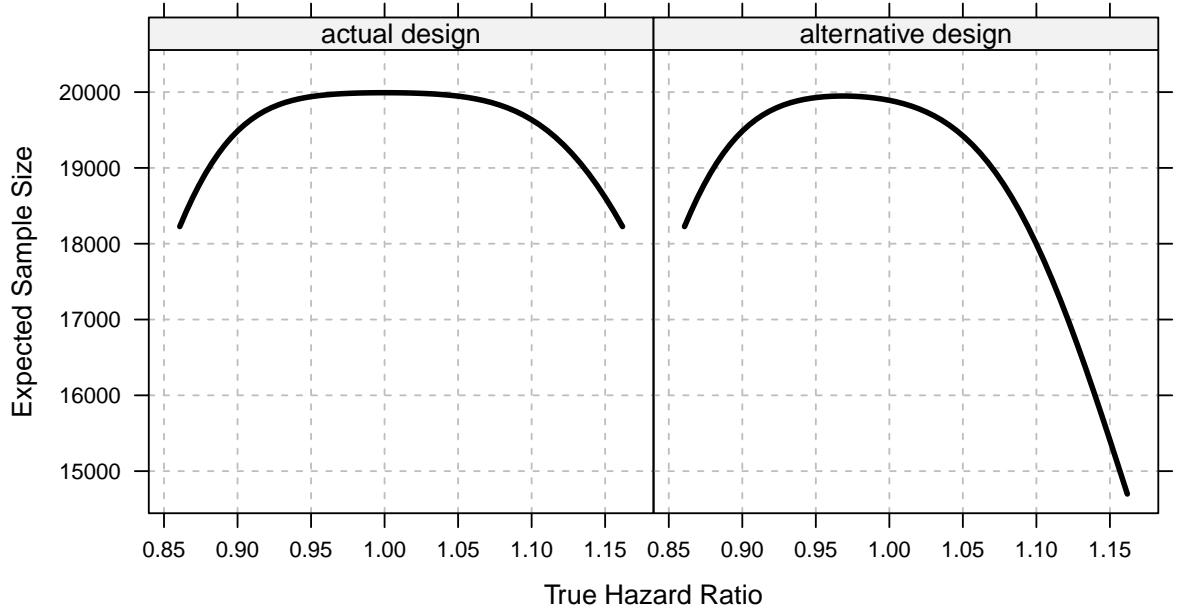
## Calculate the operating characteristics
x1 <- gsb(design = design1, simulation = sim)
x2 <- gsb(design = design2, simulation = sim)
## plot of outcome probabilities (on original HR scale (not log-scale))
dep <- subset(x1$OC,
  x1$OC$type %in% c("cumulative futility", "cumulative success"))
dep$stage <- paste("actual", dep$stage, sep=", ")
dep2 <- subset(x2$OC,
  x2$OC$type %in% c("cumulative futility", "cumulative success"))
dep2$stage <- paste("alternative futility", dep2$stage, sep=", ")
dep0 <- rbind(dep, dep2)
p1 <- xyplot(value~exp(delta)|stage, groups=type, data=dep0, col =1, lty=c(2,2,1),
  layout = c(3,2), as.table = TRUE, scales = list(alternating = 1),
  ylab = "Cumulative Probability of Stopping",
  ylim = c(-0.05,1.05), xlab = "True Hazard Ratio\n",
  panel = function(...){
    panel.grid(h=-1,v=-1, col="grey", lwd=1, lty=2)
    panel.xyplot(... , type="l", lwd=3)},
  key = list(columns = 2, space = "bottom", lines = list(lty=1:2, lwd = 2.5),
  text = list(c("Futility", "Success" ), col = 1), border = FALSE))
plot(p1)

```



```
## plot for expected sample size (on original HR scale (not log-scale))
xx1 <- subset(x1$OC, x1$OC$type %in% c("sample size"))
xx2 <- subset(x2$OC, x2$OC$type %in% c("sample size"))
xx <- rbind(cbind(xx1, design = "actual design"),
            cbind(xx2, design = "alternative design"))
xx <- subset(xx, xx$stage %in% c("stage 3"))
xx$value <- xx$value*1/0.09

p2 <- xyplot(value~exp(delta)|design, data=xx,
              scales = list(alternating = 1),
              panel = function(...){
                panel.grid(h=-1,v=-1, col="grey", lwd=1, lty=2)
                panel.xyplot(... , type="l", lwd=3, col=1)
              }, ylim = c(1300*1/0.09, 1850*1/0.09),
              ylab = "Expected Sample Size", xlab = "True Hazard Ratio")
plot(p2)
```



5 Likelihood Ratios, Unit Information LR, and DOR

Likelihood ratios (LR) quantify the change in disease odds after having observed a positive or negative test result. LRs depend on power and Type I error:

$$LR^+ = \frac{\text{Power}}{\text{Type I Error}}, \quad LR^- = \frac{1 - \text{Power}}{1 - \text{Type I Error}}.$$

In Example 1, we have the following table of operating characteristics of the two-stage design for the PoC trial in Crohn's disease. The probability of success when true $\delta = 0$ (no difference between placebo and treatment) is the Type I error. The probabilities of success when true $\delta = 40, 50, 60, 70$ are the powers in each case. It would be interesting to look at how the likelihood ratios and relevant statistics change with the effect size δ .

Operating Characteristics

δ	Success	Futile	Indeterminate	Expected N
0	0.012	0.847	0.141	40.673
40	0.411	0.108	0.481	48.305
50	0.627	0.037	0.336	44.240
60	0.808	0.011	0.181	39.488
70	0.925	0.002	0.072	35.278

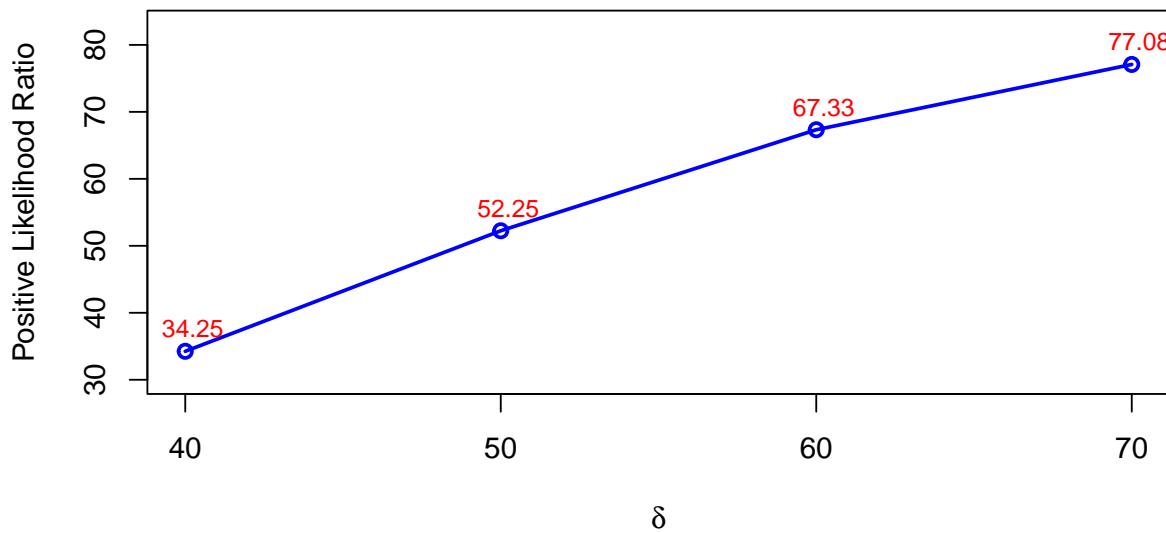
5.1 Positive and Negative Likelihood Ratios

```

library(latex2exp)
delta <- c(40,50,60,70) # true effect sizes
Exp_N <- table[,5]
TypeIError <- table[1,2]
power <- table[2:5,2]
LR_positive <- power/TypeIError
LR_negative <- (1-power)/(1-TypeIError)

plot(delta, LR_positive, type="o", col="blue", lwd=2, xaxt="n", ylim = c(30,83),
      xlab=TeX(r"($\delta$)"), ylab="Positive Likelihood Ratio")
axis(1, at=c(40, 50, 60, 70))
# Add horizontal text labels on the y-axis
text(delta-1.3, LR_positive+3, labels=round(LR_positive,2), pos=4, cex=0.8, col="red")

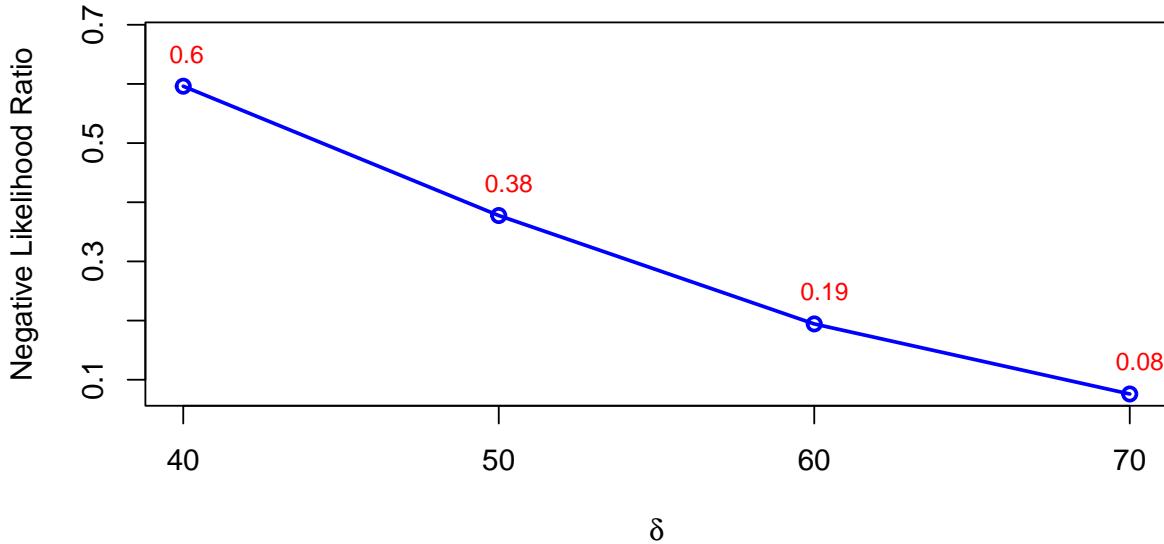
```



```

plot(delta, LR_negative, type="o", col="blue", lwd=2, xaxt="n", ylim = c(0.08,0.68),
      xlab=TeX(r"($\delta$)"), ylab="Negative Likelihood Ratio")
axis(1, at=c(40, 50, 60, 70))
# Add horizontal text labels on the y-axis
text(delta-1, LR_negative+0.05, labels=round(LR_negative,2), pos=4, cex=0.8, col="red")

```



5.2 Unit Information Likelihood Ratios

When we receive multiple independent pieces of evidence E_1, E_2, \dots, E_n , the total likelihood ratio is multiplicative:

$$LR_{\text{total}} = LR_1 \times LR_2 \times \dots \times LR_n$$

Taking the n^{th} root, the Unit Information Likelihood Ratio (LR_{UI}) is:

$$LR_{\text{UI}} = (LR_{\text{total}})^{1/n}.$$

This tells us how much each piece of independent observation contributes on average to the total likelihood ratio.

We know

$$\text{Posterior odds} = \begin{cases} LR^+ = \{LR_{\text{UI}}^+\}^{\mathbb{E}(N)} \\ LR^- = \{LR_{\text{UI}}^-\}^{\mathbb{E}(N)} \end{cases} \cdot \text{Prior odds.}$$

This enables Bayesian updating with likelihood ratios, as likelihood ratios have a multiplicative nature. Then, we can also obtain unit information likelihood ratios ($LR_{\text{UI}} = LR^{1/\mathbb{E}(N)}$), which is the likelihood ratio per unit given the expected sample size. This tells us the information one observation contributes.

For each effect size δ , we can compute two bounds for their UI LRs based on the expected sample size $\mathbb{E}(N)$ for that particular value of δ used to calculate the power and for $\delta = 0$.

```
LR_pos_UI <- as.data.frame(t(matrix(
  c(max(LR_positive[1]^(1/Exp_N[1:2])), min(LR_positive[1]^(1/Exp_N[1:2])),
    max(LR_positive[2]^(1/Exp_N[c(1,3)])), min(LR_positive[2]^(1/Exp_N[c(1,3)])),
    max(LR_positive[3]^(1/Exp_N[c(1,4)])), min(LR_positive[3]^(1/Exp_N[c(1,4)]))),
```

```

max(LR_positive[4]^(1/Exp_N[c(1,5)])), min(LR_positive[4]^(1/Exp_N[c(1,5)]))),
nrow = 2, ncol = 4,
dimnames = list(c("upper bound", "lower bound"),
                c("delta=40", "delta=50", "delta=60", "delta=70"))
))

LR_neg_UI <- as.data.frame(t(matrix(
  c(max(LR_negative[1]^(1/Exp_N[1:2])), min(LR_negative[1]^(1/Exp_N[1:2])),
    max(LR_negative[2]^(1/Exp_N[c(1,3)])), min(LR_negative[2]^(1/Exp_N[c(1,3)])),
    max(LR_negative[3]^(1/Exp_N[c(1,4)])), min(LR_negative[3]^(1/Exp_N[c(1,4)])),
    max(LR_negative[4]^(1/Exp_N[c(1,5)])), min(LR_negative[4]^(1/Exp_N[c(1,5)]))),
  nrow = 2, ncol = 4,
  dimnames = list(c("upper bound", "lower bound"),
                c("delta=40", "delta=50", "delta=60", "delta=70")))
)))

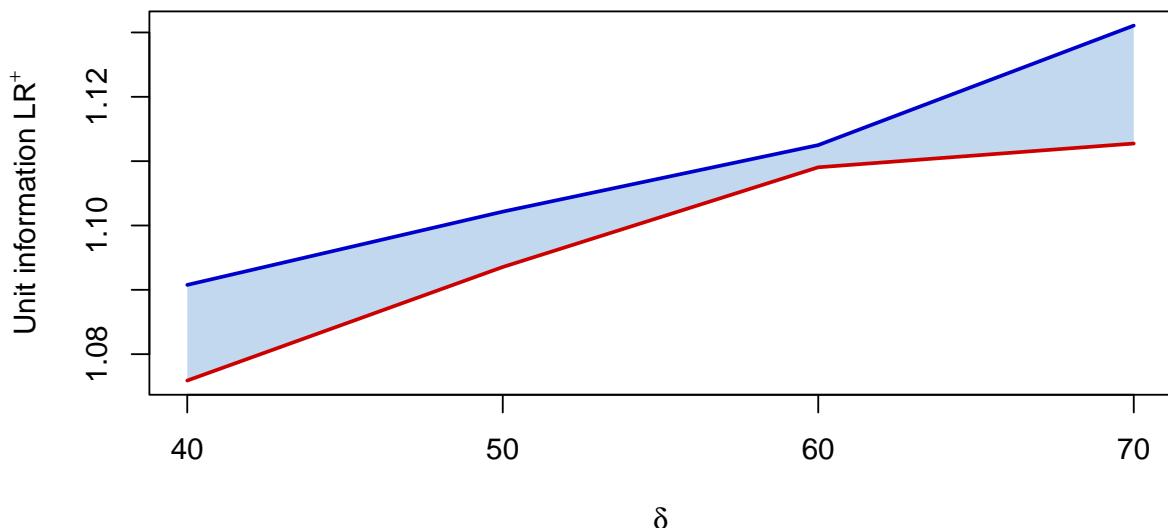
# Plot UI LR+
plot(delta, LR_pos_UI$`upper bound`, type = "n", xaxt="n",
      ylim = range(LR_pos_UI$`lower bound`,LR_pos_UI$`upper bound`),
      xlab = TeX(r"(\delta)"), ylab = TeX(r"(Unit information LR^{+})"),
      main = TeX(r"(UI LR^{+} bounds)"))
axis(1, at=c(40, 50, 60, 70))

# Shade the area between bounds
polygon(c(delta, rev(delta)),
         c(LR_pos_UI$`upper bound`, rev(LR_pos_UI$`lower bound`)),
         col = rgb(0.2, 0.5, 0.8, 0.3), border = NA)

# Add lines for upper and lower bounds
lines(delta, LR_pos_UI$`upper bound`, col = "blue3", lwd = 2)
lines(delta, LR_pos_UI$`lower bound`, col = "red3", lwd = 2)

```

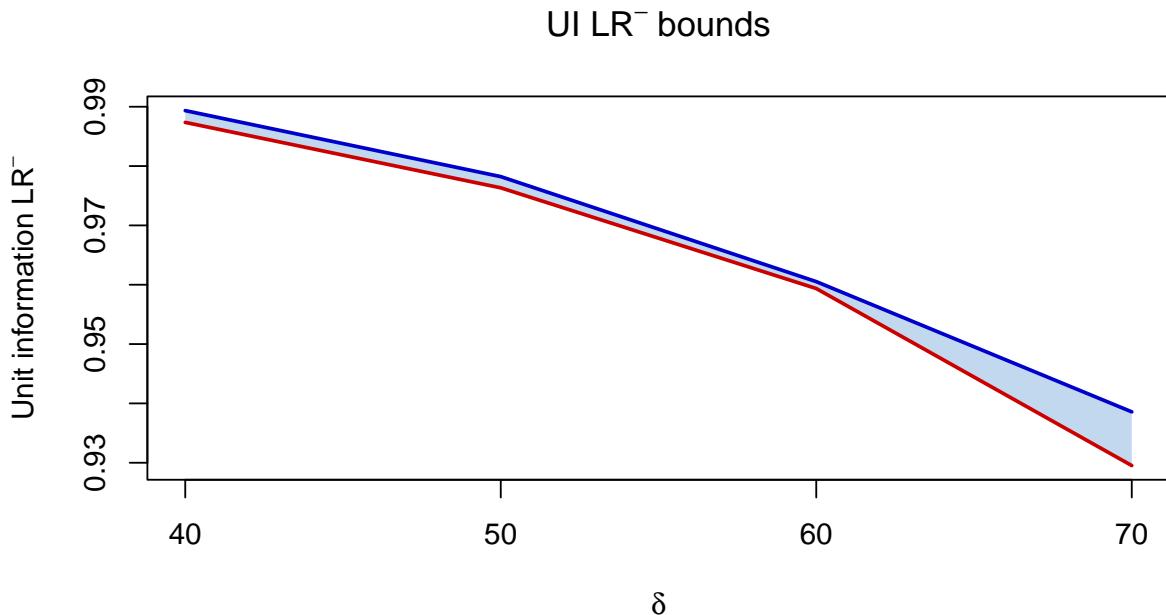
UI LR^+ bounds



```
# Plot UI LR-
plot(delta, LR_neg_UI$`upper bound`, type = "n", xaxt="n",
      ylim = range(LR_neg_UI$`lower bound`,LR_neg_UI$`upper bound`),
      xlab = TeX(r"($\delta$)"), ylab = TeX(r"(Unit information LR$^{-}$)"),
      main = TeX(r"(UI LR$^{-}$ bounds)"))
axis(1, at=c(40, 50, 60, 70))

# Shade the area between bounds
polygon(c(delta, rev(delta)),
         c(LR_neg_UI$`upper bound`, rev(LR_neg_UI$`lower bound`)),
         col = rgb(0.2, 0.5, 0.8, 0.3), border = NA)

# Add lines for upper and lower bounds
lines(delta, LR_neg_UI$`upper bound`, col = "blue3", lwd = 2)
lines(delta, LR_neg_UI$`lower bound`, col = "red3", lwd = 2)
```

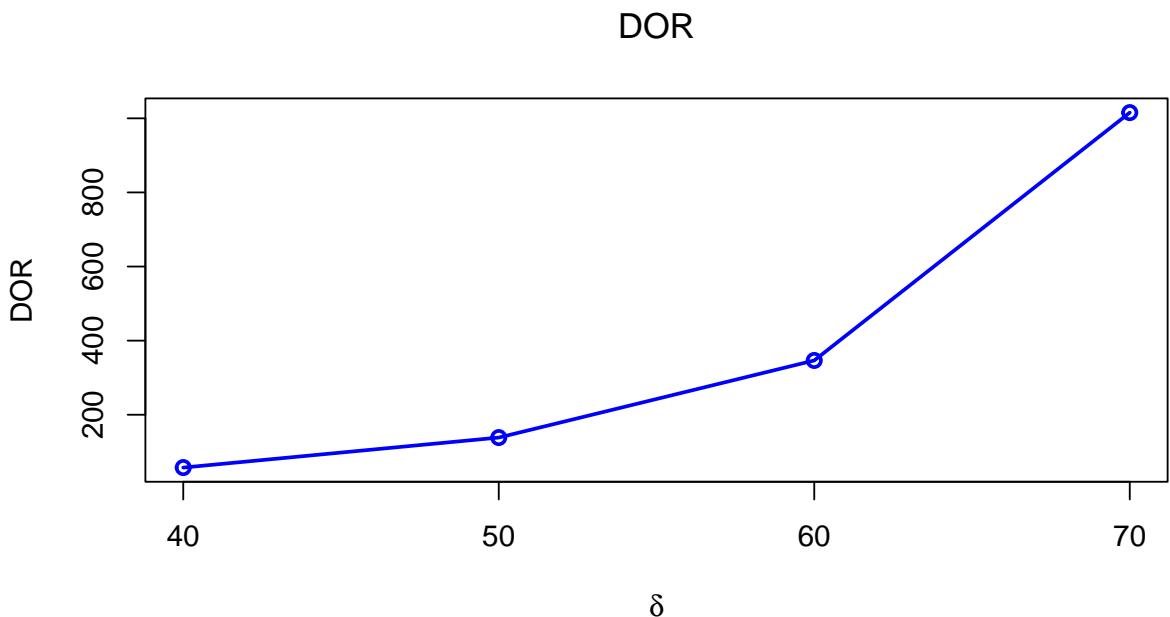


5.3 Diagnostic Odds Ratios

Diagnostic odds ratios (DOR) tells us how much information there is essentially. The unit information DOR implies the amount of information provided by one observation.

```
DOR <- LR_positive/LR_negative
A <- as.matrix(LR_pos_UI)
B <- as.matrix(LR_neg_UI)
DOR_UI <- t(sapply(1:nrow(LR_pos_UI), function(i) as.vector(outer(A[i, ], B[i, ], "/"))))
DOR_UI <- as.data.frame(cbind(DOR_UI, upper_bound=apply(DOR_UI, 1, max),
                               lower_bound=apply(DOR_UI, 1, min)))

# Plot DOR
plot(delta, DOR, type = "o", xaxt="n", lty=1, lwd=2,
      ylim = range(DOR),
      xlab = TeX(r"(\delta)"), ylab = "DOR", col="blue",
      main = TeX(r"(DOR)"))
axis(1, at=c(40, 50, 60, 70))
```



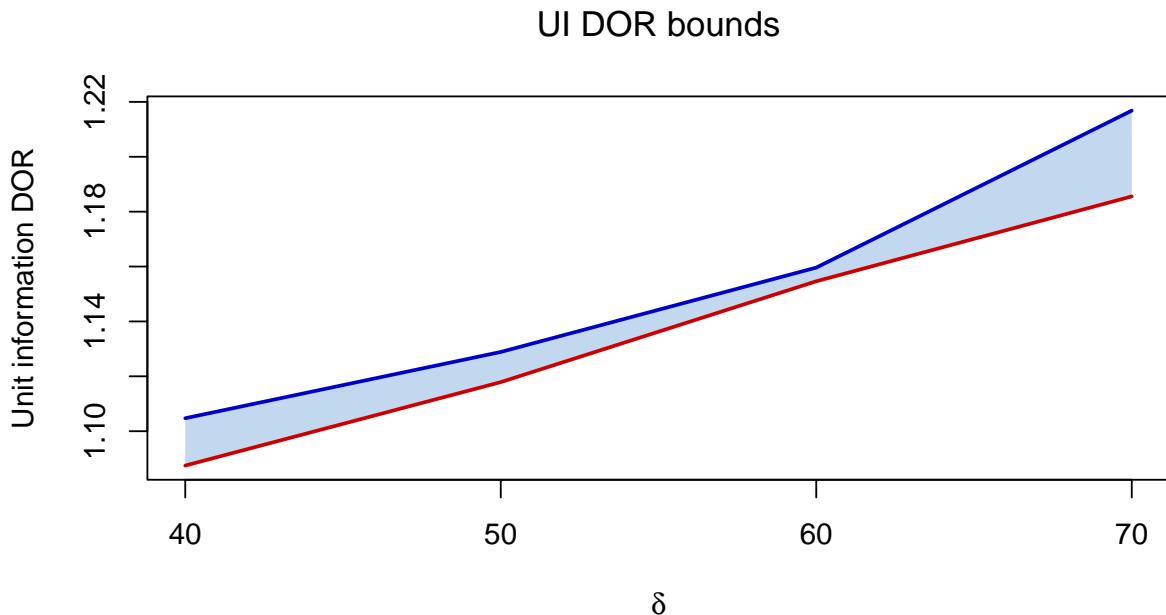
```

# Plot DOR UI
plot(delta, DOR_UI$upper_bound, type = "n", xaxt="n",
      ylim = range(DOR_UI$lower_bound,DOR_UI$upper_bound),
      xlab = TeX(r"($\delta$)", ylab = TeX(r"(Unit information DOR)"),
      main = TeX(r"(UI DOR bounds)"))
axis(1, at=c(40, 50, 60, 70))

# Shade the area between bounds
polygon(c(delta, rev(delta)),
         c(DOR_UI$upper_bound, rev(DOR_UI$lower_bound)),
         col = rgb(0.2, 0.5, 0.8, 0.3), border = NA)

# Add lines for upper and lower bounds
lines(delta, DOR_UI$upper_bound, col = "blue3", lwd = 2)
lines(delta, DOR_UI$lower_bound, col = "red3", lwd = 2)

```



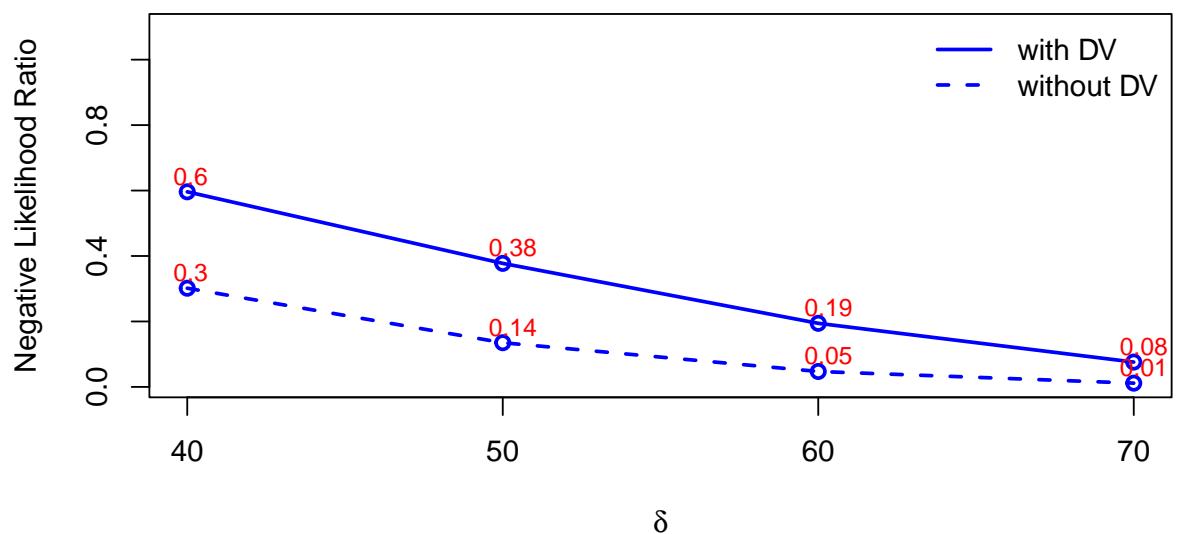
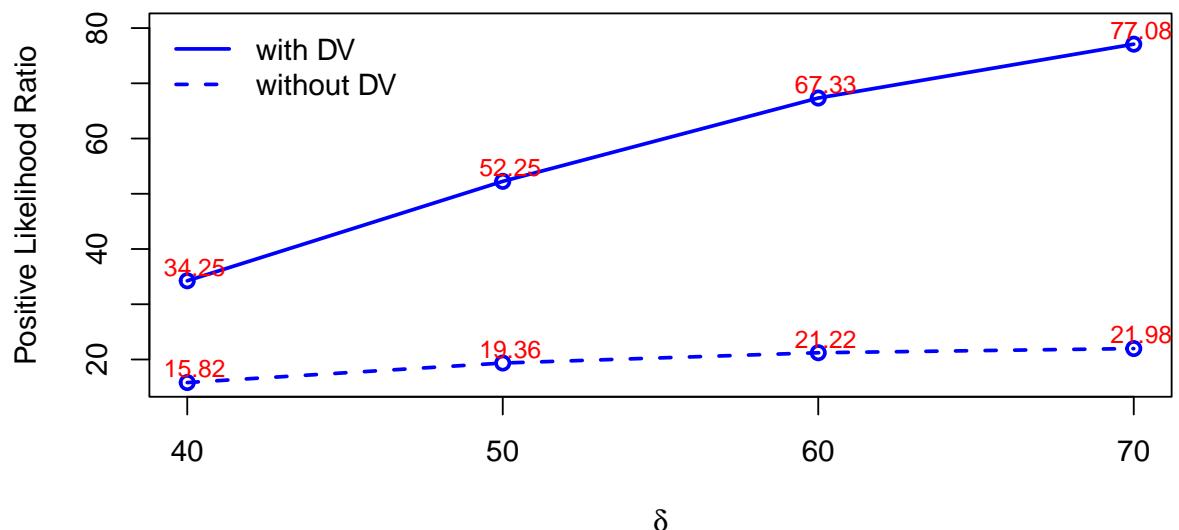
5.4 Exploring the impact of the presence of decision value

Here we remove the DV criterion in Example 1 and see how the results change. Below table contains the operating characteristics and expected sample size.

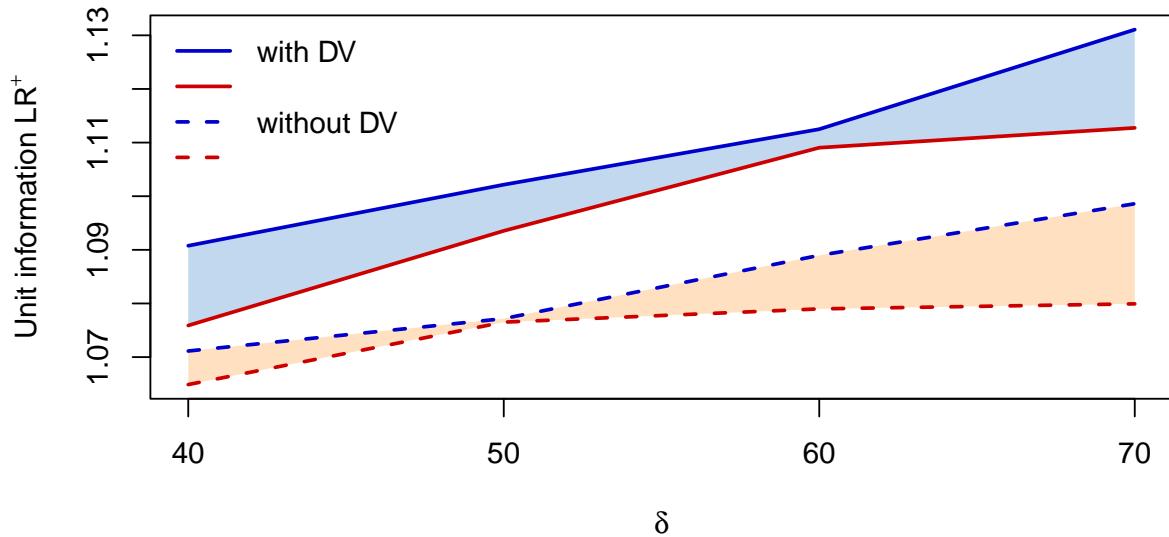
Probability overall

δ	Success	Futile	Indeterminate	Expected N
0	0.045	0.844	0.112	40.185
40	0.712	0.106	0.182	43.928
50	0.871	0.037	0.092	39.857
60	0.955	0.011	0.034	35.842
70	0.989	0.002	0.009	32.861

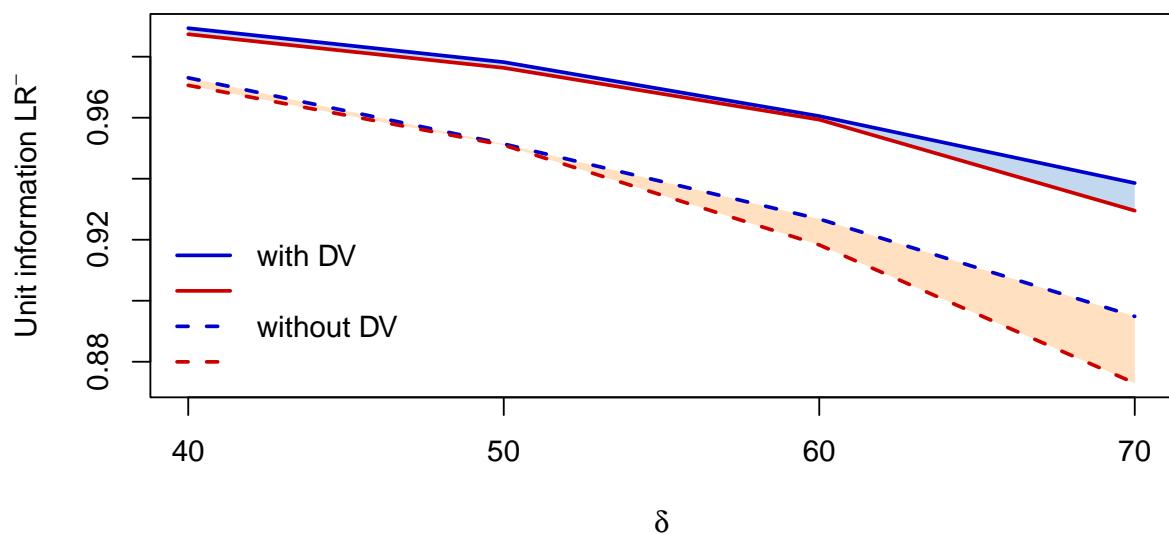
Again we look at the LR^+ , LR^- , unit information LR_{UI}^+ , LR_{UI}^- , DOR and DOR_{UI} . Generally, LR^+ , LR^- , LR_{UI}^+ , and LR_{UI}^- are smaller than that under the presence of DV criterion, while the DOR_{UI} s are generally larger, which could be seen from the below plots.

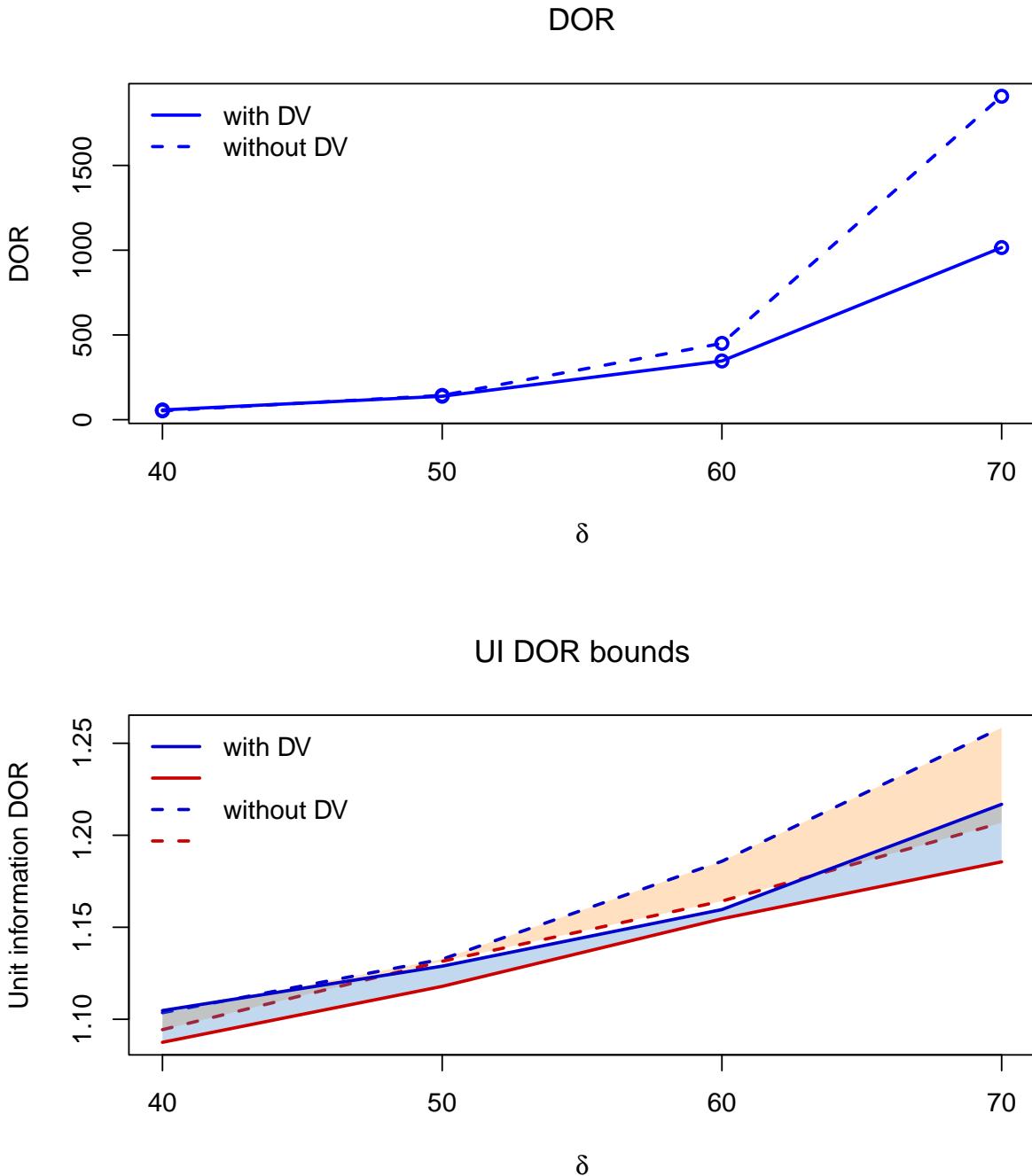


UI LR⁺ bounds



UI LR⁻ bounds





6 Applications to Roychoudhury paper

In this part, we look at the application of the methods in [Gsponer et al. \(2014\)](#) to the single-arm PoC design with binary data in [Roychoudhury et al. \(2018\)](#).

For binary data, the normal approximation is known to be more appropriate on the logit scale than on the proportion scale. Let r be number of responders, and n be the number of treated patients. Defining y as the corresponding log-odds of the observed response rate, its distribution

is approximately given by

$$y \sim N(\text{logit}(\pi), \sigma^2/n)$$

where π is the true response rate and $\sigma^2 = 1/\pi + 1/(1-\pi)$.

In [Roychoudhury et al. \(2018\)](#) paper, the primary endpoint is objective response rate(ORR). The prior is set to be $\text{ORR} \sim \text{Beta}(0.0811, 1)$. Because of the absence of a comparator (in single arm trials), NV is set to 7.5%. A minimum improvement of 10% is considered necessary for further development, so the DV is set to be $10\% + 7.5\% = 17.5\%$. Notice in this example, $\text{NV} < \text{DV}$.

The dual-criterion is:

1. Bayesian statistical significance: $\mathbb{P}(\text{ORR} \geq 7.5\% \mid \text{data}) \geq 0.95$
2. Clinical relevance: Posterior median $\geq 17.5\%$

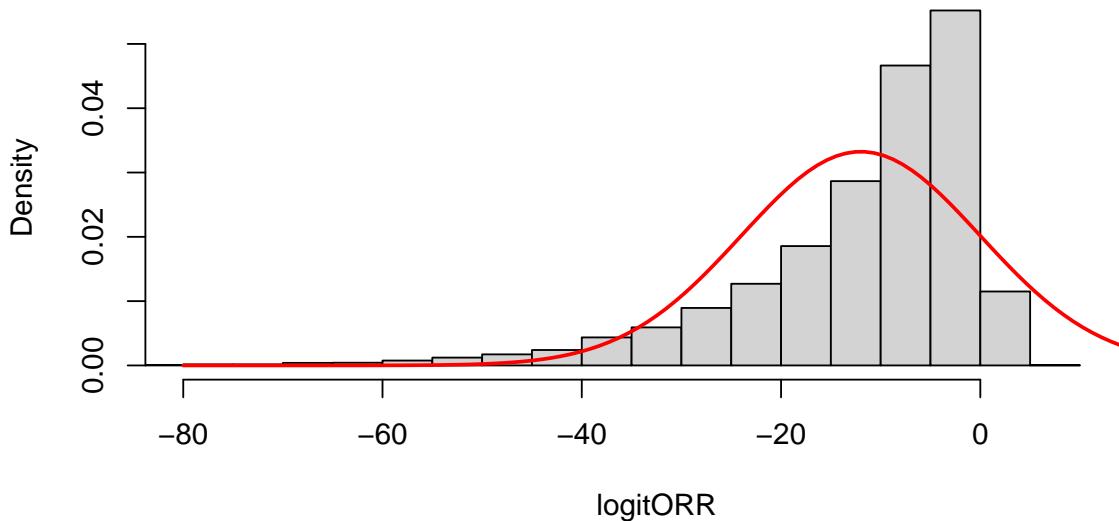
Null hypothesis: there is no effect of the drug, i.e. $\text{ORR}=7.5\%$

If we observe a number of successes (r) and failures ($n - r$), then the posterior is $\text{Beta}(a + r, b + n - r)$.

One key difference between [Gsponer et al. \(2014\)](#) and [Roychoudhury et al. \(2018\)](#) is that [Gsponer et al. \(2014\)](#) is using a logit transformation for normal approximation, whereas [Roychoudhury et al. \(2018\)](#) is using a Beta prior. So here we need to translate [Roychoudhury et al. \(2018\)](#)'s problem into logit scale:

The prior is approximately $\text{logit}(\text{ORR}) \sim N(-12, 12^2)$. This can be visualized below in the histogram of $\text{logit}(\text{ORR})$ where $\text{ORR} \sim \text{Beta}(0.0811, 1)$, and the red curve is the density of $N(-12, 12^2)$.

Normal approximation to logitORR



The translated dual-criterion for success are:

1. Bayesian statistical significance: $\mathbb{P}(\text{logit}(\text{ORR}) \geq \text{logit}(0.075) \mid \text{data}) \geq 0.95$

2. Clinical relevance: $\mathbb{P}(\text{logit(ORR)} \geq \text{logit}(0.175) | \text{data}) \geq 0.5$

The translated dual-criterion for failure are:

1. Bayesian statistical significance: $\mathbb{P}(\text{logit(ORR)} \leq \text{logit}(0.075) | \text{data}) \geq 0.05$
2. Clinical relevance: $\mathbb{P}(\text{logit(ORR)} \leq \text{logit}(0.175) | \text{data}) \geq 0.5$

6.1 Promising attempt 1

This attempt keeps the two-stages design by introducing a virtual control arm, into which no patients are recruited. It uses $N(-12,12)$ for treatment effect, and $N(\text{logit}(0.075), 1/1000)$ as prior for control. It used delta (the changes in logit ORRs), similar to the Gsponer's example 2. This attempt also defined the success and futility criteria using the required responders from [Roychoudhury et al. \(2018\)](#) Table 4. Using above settings, the result gives 0 inconclusive values, which matches the Roychoudhury's results (Table 4 in the paper).

Success criterion: $\mathbb{P}(\delta > \text{logit}(5/25) - \text{logit}(0.075)) \geq 0.5$

Futility criterion: $\mathbb{P}(\delta < \text{logit}(5/25) - \text{logit}(0.075)) \geq 0.5$.

```
logit <- function(p) { log(p/(1-p)) }
expit <- function(x) { exp(x)/(1+exp(x)) }
p0 = 0.075
p1 = 0.175

prior_treatment_precision = 1/12^2

suc = logit(5/25) - logit(p0)
fut = logit(5/25) - logit(p0)
design <- gsbDesign(nr.stages=2,
patients=cbind( c(0,0), c(0,25) ),
sigma=2,
criteria.success=rbind( c(NA,NA), c(suc,0.5) ),
criteria.futility=rbind( c(NA,NA),c(fut,0.5) ),
prior.control=c( logit(p0),1000),
prior.treatment=c(-12,prior_treatment_precision))

simulation <- gsbSimulation(
truth=list(logit(0.075), # true control
           c(logit(0.075)-logit(0.075), logit(0.125)-logit(0.075),
             logit(0.175)-logit(0.075), # true treatment
             logit(0.225)-logit(0.075), logit(0.275)-logit(0.075))),
grid.type = "sliced",
type.update = "per arm",
nr.sim = 100000,
warnings.sensitivity = 2000,
```

```

seed = 1)
result <- gsb(design,simulation)
result

##
## *** Group Sequential Bayesian Design ***
##
## Analysis N1 N2
## Prior 1000 0.00694
## 1 0 0.00000
## 2 0 25.00000
##
## sigma treatment: 2 sigma control: 2
##
## access the operating characteristics via the data.frame "OC" in the output of "gsb()"
## or the functions "tab()" and "plot()".
##
## names in output list:
## [1] "OC"          "design"       "simulation"   "delta.grid"   "warnings"
## [6] "system.time"

#tab(result, what="cumulative all")
#tab(result, what="sample size")
#plot(result, what="sample size", sliced=TRUE)
tab(result, what="cumulative all") # control and treatment columns in logit scale

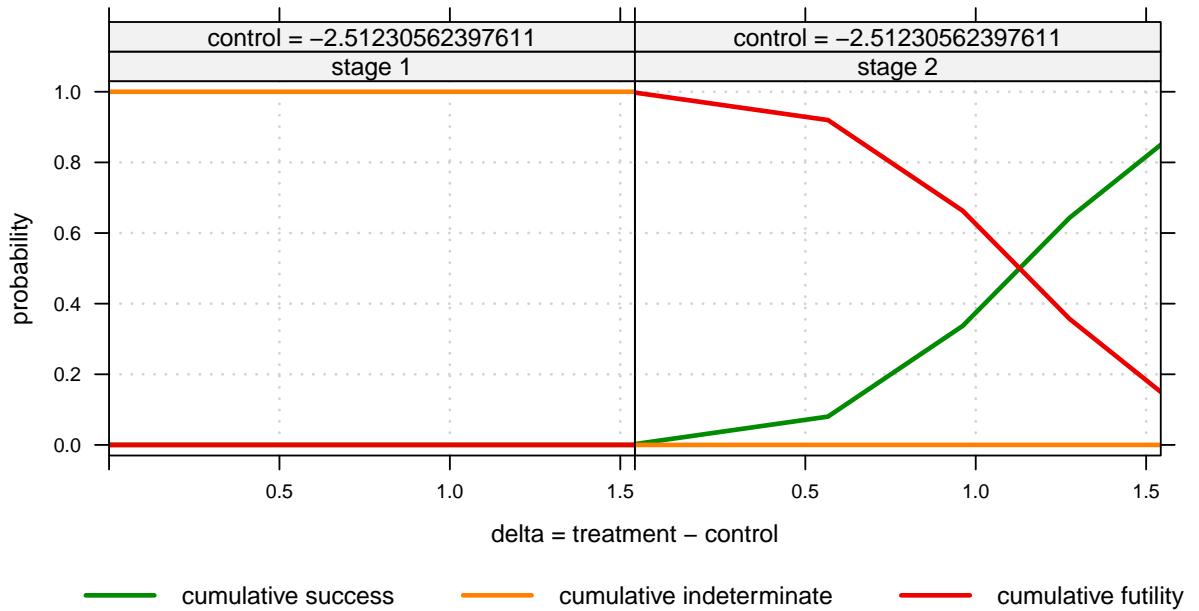
## control treatment delta stage1.suc stage1.fut stage1.ind stage2.suc
## 1 -2.512 -2.512 0.000 0 0 1 0.002
## 2 -2.512 -1.946 0.566 0 0 1 0.080
## 3 -2.512 -1.551 0.962 0 0 1 0.337
## 4 -2.512 -1.237 1.276 0 0 1 0.643
## 5 -2.512 -0.969 1.543 0 0 1 0.850

## stage2.fut stage2.ind
## 1 0.998 0
## 2 0.920 0
## 3 0.663 0
## 4 0.357 0
## 5 0.150 0

plot(result, what="cumulative all", sliced=TRUE)

```

Operating Characteristics

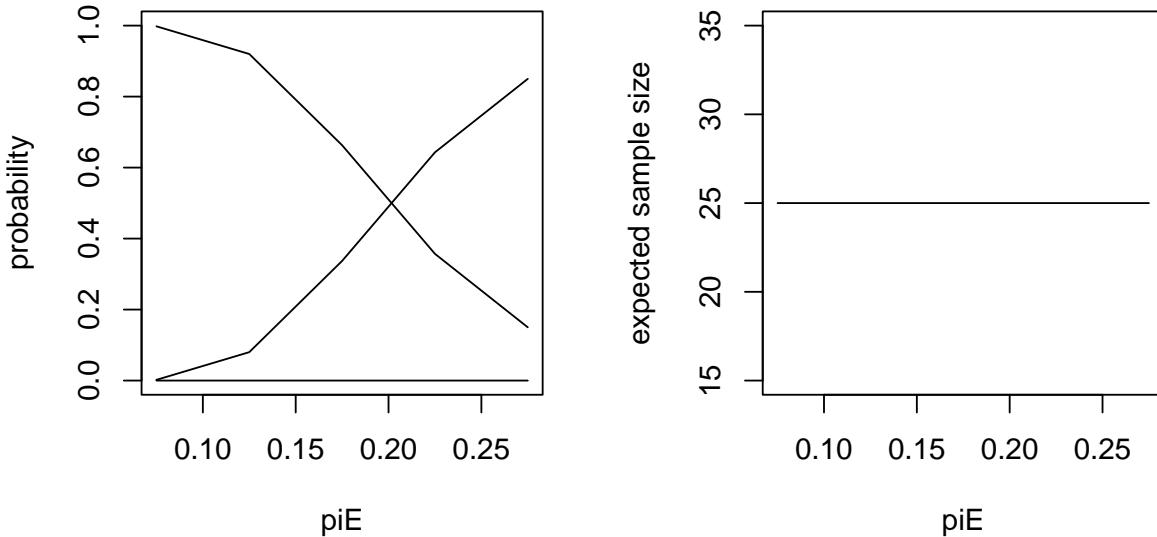


```

par(mfrow=c(1,2))

y1 = tab(result,what='futility')[,5]
y2 = tab(result,what='success')[,5]
y3 = 1 - y1 - y2
piE = expit(c(logit(0.075), logit(0.125),logit(0.175), # true treatment
               logit(0.225), logit(0.275)))
plot(piE,y1,type='l',ylim=c(0,1),ylab='probability') # futility
lines(piE,y2) # success
lines(piE,y3)
y3 = tab(result,what='sample size')[,5]
plot(piE,y3,type='l',ylab='expected sample size')

```



This attempt turns the previous two-stage design to a one-stage design.

```

logit <- function(p) { log(p/(1-p)) }
expit <- function(x) { exp(x)/(1+exp(x)) }
p0 = 0.075
p1 = 0.175
suc = logit(5/25) - logit(p0)
fut = logit(5/25) - logit(p0)
design <- gsbDesign(nr.stages=1,
patients=c(0,25),
sigma=2,
criteria.success= c(suc,0.5),
criteria.futility= c(fut,0.5),
prior.control=c( logit(p0),1000),
prior.treatment=c(-12,1/144))

simulation <- gsbSimulation(
truth=list(logit(0.075), # true control
           c(logit(0.075)-logit(0.075), logit(0.125)-logit(0.075),
             logit(0.175)-logit(0.075), # true treatment
             logit(0.225)-logit(0.075), logit(0.275)-logit(0.075))),
grid.type = "sliced",
type.update = "per arm",
nr.sim = 100000,
warnings.sensitivity = 2000,
seed = 1)

```

```

result <- gsb(design,simulation)
result

##
## *** Group Sequential Bayesian Design ***
##
## Analysis N1      N2
##      Prior 1000  0.00694
##           1     0 25.00000
##
## sigma treatment: 2  sigma control: 2
##
## access the operating characteristics via the data.frame "OC" in the output of "gsb()"
## or the functions "tab()" and "plot()".
##
## names in output list:
## [1] "OC"          "design"       "simulation"   "delta.grid"   "warnings"
## [6] "system.time"

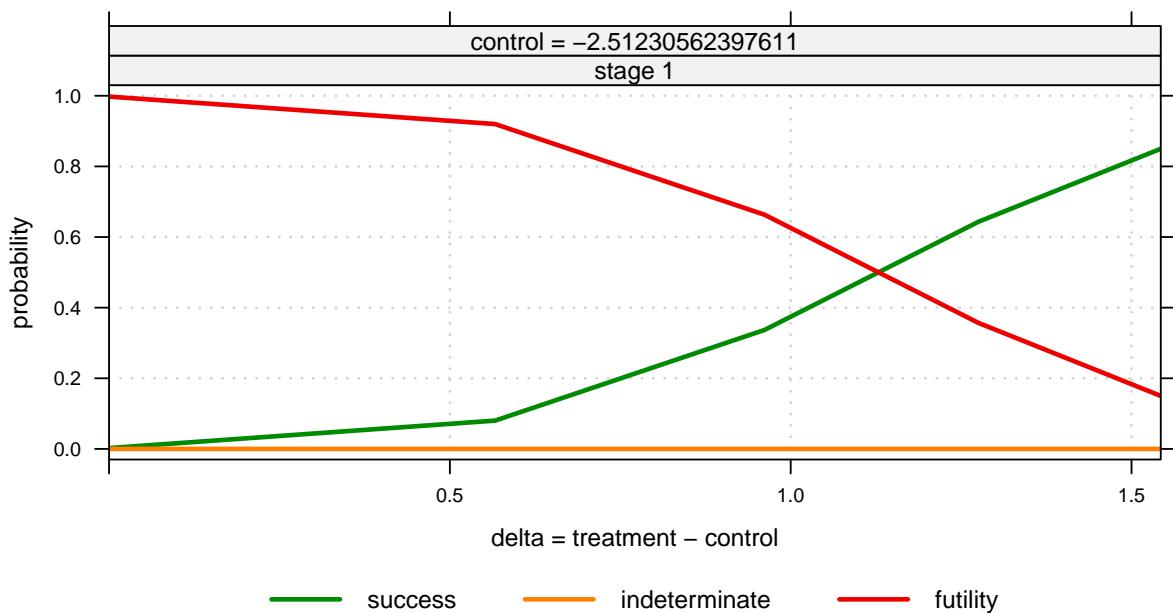
#tab(result, what="cumulative all")
#tab(result, what="sample size")
#plot(result, what="sample size", sliced=TRUE)
tab(result, what="cumulative all") # control and treatment columns in logit scale

## control treatment delta stage1.suc stage1.fut stage1.ind
## 1 -2.512    -2.512 0.000      0.002      0.998      0
## 2 -2.512    -1.946 0.566      0.080      0.920      0
## 3 -2.512    -1.551 0.962      0.337      0.663      0
## 4 -2.512    -1.237 1.276      0.643      0.357      0
## 5 -2.512    -0.969 1.543      0.850      0.150      0

plot(result, what="cumulative all", sliced=TRUE)

```

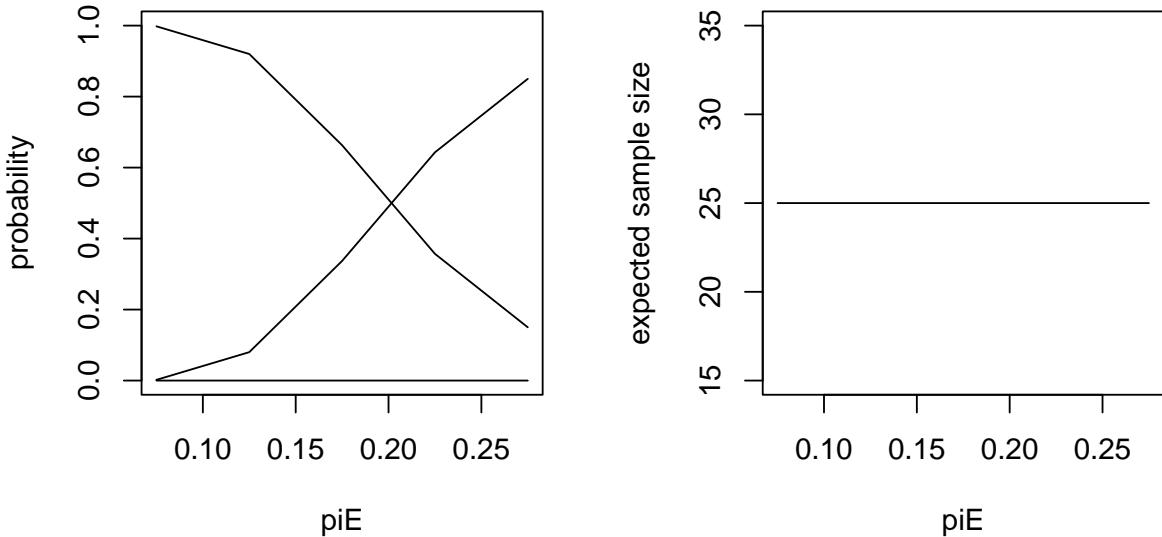
Operating Characteristics



```

par(mfrow=c(1,2))
y1 = tab(result,what='futility')[,4]
y2 = tab(result,what='success')[,4]
y3 = 1 - y1 - y2
piE = expit(c(logit(0.075), logit(0.125),logit(0.175), # true treatment
               logit(0.225), logit(0.275)))
plot(piE,y1,type='l',ylim=c(0,1),ylab='probability')
lines(piE,y2)
lines(piE,y3)
y3 = tab(result,what='sample size')[,4]
plot(piE,y3,type='l',ylab='expected sample size')

```



This attempt look at Roychoudhury's second case in Table 4 (sample size = 36). The results is quite far away from Roychoudhury's values (especially in GO case).

```

logit <- function(p) { log(p/(1-p)) }
expit <- function(x) { exp(x)/(1+exp(x)) }

p0 = 0.075
p1 = 0.175
suc = logit(7/25) - logit(p0)
fut = logit(5/25) - logit(p0)
design <- gsbDesign(nr.stages=1,
patients=c(0,36),
sigma=2,
criteria.success= c(suc,0.5),
criteria.futility= c(fut,0.5),
prior.control=c( logit(p0),1000),
prior.treatment=c(-12,1/144))

simulation <- gsbSimulation(
truth=list(logit(0.075), # true control
c(logit(0.075)-logit(0.075), logit(0.125)-logit(0.075),
logit(0.175)-logit(0.075), # true treatment
logit(0.225)-logit(0.075), logit(0.275)-logit(0.075))),
grid.type = "sliced",
type.update = "per arm",
nr.sim = 100000,
warnings.sensitivity = 2000,

```

```

seed = 1)
result <- gsb(design,simulation)
result

##
## *** Group Sequential Bayesian Design ***
##
## Analysis N1      N2
##     Prior 1000  0.00694
##           1      0 36.00000
##
## sigma treatment: 2  sigma control: 2
##
## access the operating characteristics via the data.frame "OC" in the output of "gsb()"
## or the functions "tab()" and "plot()".
##
## names in output list:
## [1] "OC"          "design"       "simulation"   "delta.grid"   "warnings"
## [6] "system.time"

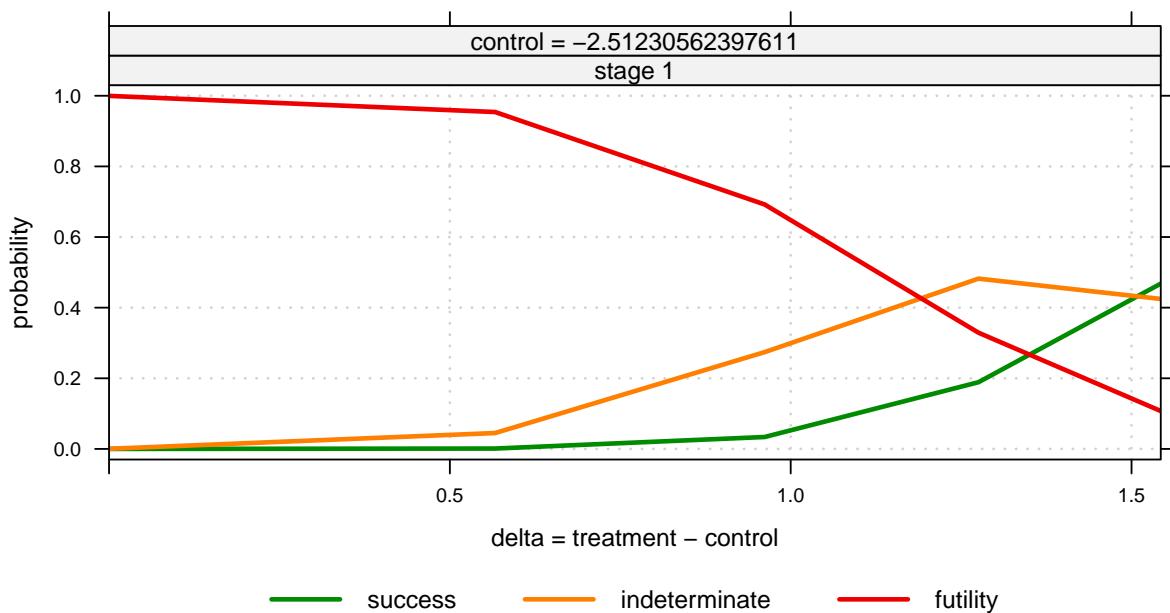
#tab(result, what="cumulative all")
#tab(result, what="sample size")
#plot(result, what="sample size", sliced=TRUE)
tab(result, what="cumulative all") # control and treatment columns in logit scale

##   control treatment delta stage1.suc stage1.fut stage1.ind
## 1 -2.512    -2.512  0.000     0.000    1.000     0.000
## 2 -2.512    -1.946  0.566     0.001    0.954     0.045
## 3 -2.512    -1.551  0.962     0.034    0.692     0.274
## 4 -2.512    -1.237  1.276     0.189    0.329     0.482
## 5 -2.512    -0.969  1.543     0.468    0.107     0.425

plot(result, what="cumulative all", sliced=TRUE)

```

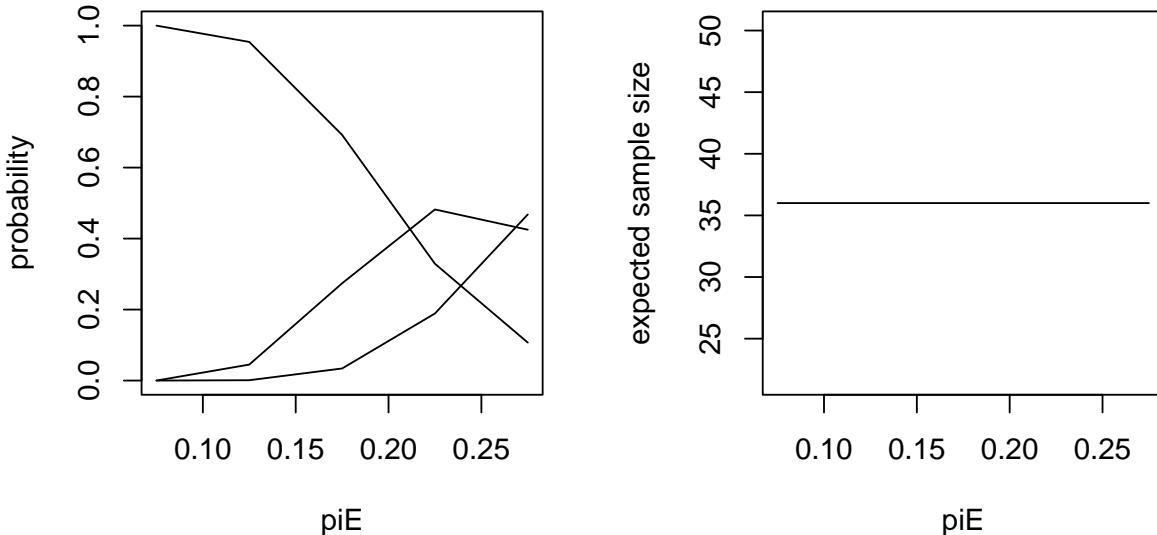
Operating Characteristics



```

par(mfrow=c(1,2))
y1 = tab(result,what='futility')[,4]
y2 = tab(result,what='success')[,4]
y3 = 1 - y1 - y2
piE = expit(c(logit(0.075), logit(0.125),logit(0.175), # true treatment
               logit(0.225), logit(0.275)))
plot(piE,y1,type='l',ylim=c(0,1),ylab='probability')
lines(piE,y2)
lines(piE,y3)
y3 = tab(result,what='sample size')[,4]
plot(piE,y3,type='l',ylab='expected sample size')

```



6.2 Not so successful attempts

I made some attempts but met some problems too:

This attempt uses $N(\text{logit}(0.075), 1/1000)$ as prior for control, and $N(-12, 12^2)$ as prior for treatment. It uses two futility criteria.

```
library(gsbDesign)
logit <- function(p) { log(p/(1-p)) }
expit <- function(x) { exp(x)/(1+exp(x)) }

# single-arm design
design <- gsbDesign(nr.stages=1,
patients=c(0,25),
sigma=2,
criteria.success=c(logit(0.075), 0.95, logit(0.175), 0.5),
criteria.futility=c(logit(0.075), 0.05, logit(0.175), 0.5),
prior.control=c(logit(0.075),1000), # very small variance in control prior
prior.treatment=c(-12, 1/12^2)) # large variance in treatment prior

simulation <- gsbSimulation(
truth=list(logit(0.075), # true control
          c(logit(0.075), logit(0.125),logit(0.175), # true treatment
            logit(0.225), logit(0.275))),
grid.type = "table",
type.update = "per arm", # only simulation method is implemented
nr.sim = 100000,
```

```

warnings.sensitivity = 2000,
seed = 1)

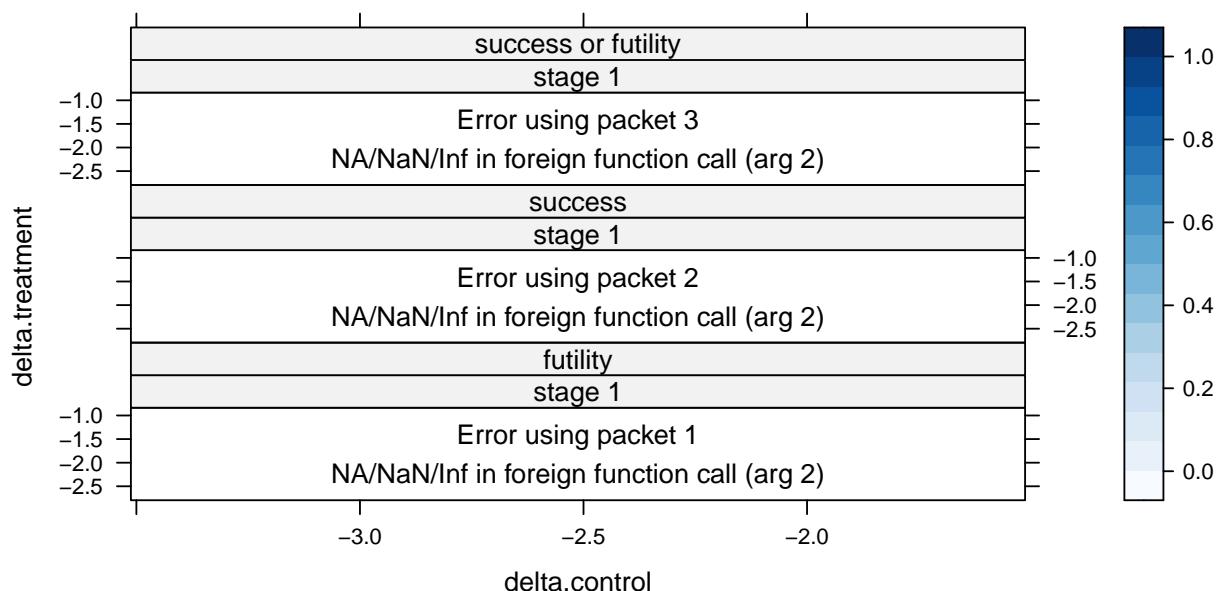
simulation <- gsbSimulation(
  truth=list(logit(0.075), # true control
             c(logit(0.075), logit(0.125), logit(0.175), # true treatment
               logit(0.225), logit(0.275))),
  grid.type = "table",
  type.update = "per arm", # only simulation method is implemented
  nr.sim = 100000,
  warnings.sensitivity = 2000,
  seed = 1)
result <- gsb(design,simulation)
result

## 
## *** Group Sequential Bayesian Design ***
## 
## Analysis   N1      N2
##      Prior 1000  0.00694
##           1     0 25.00000
##
## sigma treatment: 2  sigma control: 2
## 
## access the operating characteristics via the data.frame "OC" in the output of "gsb()"
## or the functions "tab()" and "plot()".
## 
## names in output list:
## [1] "OC"          "design"       "simulation"    "delta.grid"    "warnings"
## [6] "system.time"

#tab(result, what="cumulative all")
#tab(result, what="sample size")
#plot(result, what="sample size", sliced = FALSE)
plot(result, what="cumulative all", sliced=FALSE)

```

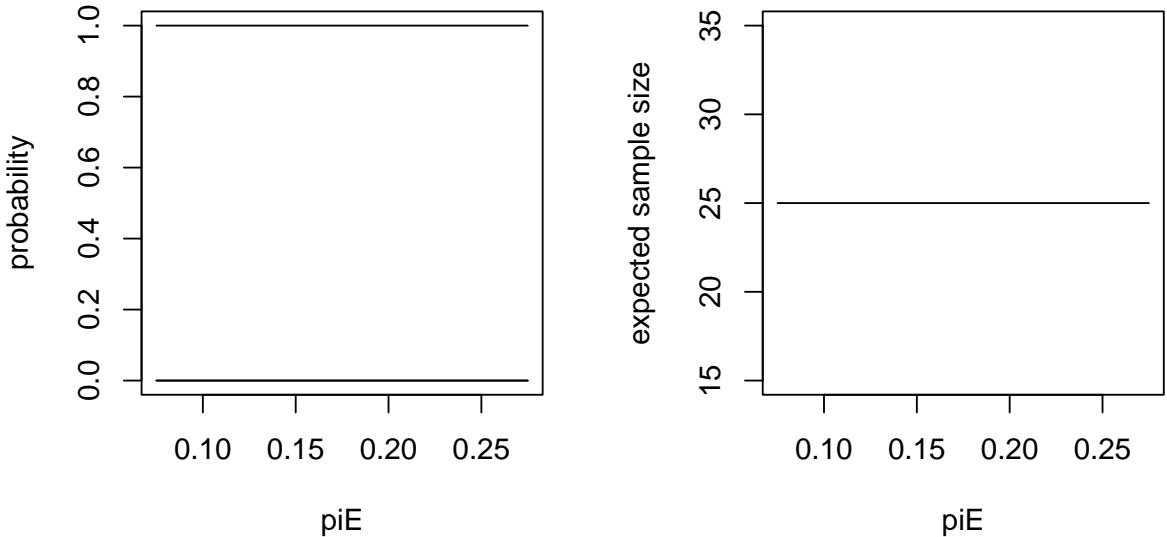
Operating Characteristics



```

par(mfrow=c(1,2))
y1 = tab(result,what='futility')[,4]
y2 = tab(result,what='success')[,4]
y3 = 1 - y1 - y2
piE = expit(c(logit(0.075), logit(0.125),logit(0.175), # true treatment
              logit(0.225), logit(0.275)))
plot(piE,y1,type='l',ylim=c(0,1),ylab='probability')
lines(piE,y2)
lines(piE,y3)
y3 = tab(result,what='sample size')[,4]
plot(piE,y3,type='l',ylab='expected sample size')

```



This attempt uses $N(-12, 1/1000)$ as prior for control, and $N(-12, 12^2)$ as prior for treatment. It only have one criteria for futility(corresponding to 1 stage in total), as the examples in Gsponer's paper. The result returns indeterminate situations.

```

library(gsbDesign)
logit <- function(p) { log(p/(1-p)) }
expit <- function(x) { exp(x)/(1+exp(x)) }

# single-arm design
design <- gsbDesign(nr.stages=1,
patients=c(0,25),
sigma=2,
criteria.success=c(logit(0.075), 0.95, logit(0.175), 0.5),
criteria.futility=c(logit(0.075), 0.05),
prior.control=c(-12,1000), # very small variance in control prior
prior.treatment=c(-12, 1/12^2)) # large variance in treatment prior

simulation <- gsbSimulation(
truth=list(-12, # true control
c(logit(0.075), logit(0.125),logit(0.175), # true treatment
logit(0.225), logit(0.275))),
grid.type = "sliced",
type.update = "per arm", # only simulation method is implemented
nr.sim = 100000,
warnings.sensitivity = 2000,
seed = 1)

```

```

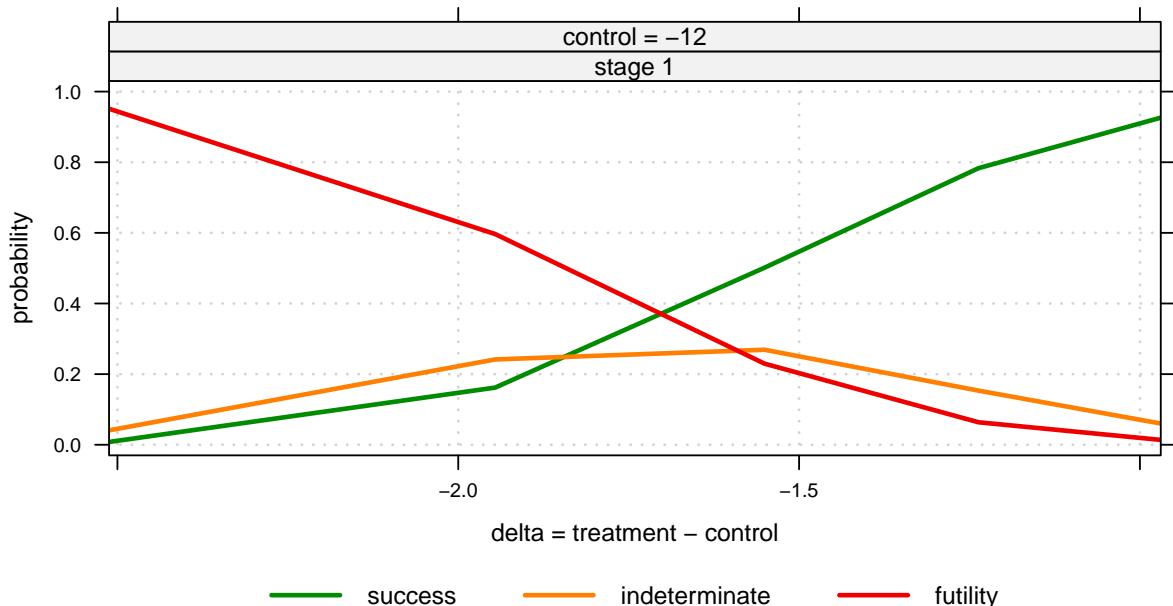
result <- gsb(design,simulation)
result

##
## *** Group Sequential Bayesian Design ***
##
## Analysis N1      N2
## Prior 1000  0.00694
##       1     0 25.00000
##
## sigma treatment: 2  sigma control: 2
##
## access the operating characteristics via the data.frame "OC" in the output of "gsb()"
## or the functions "tab()" and "plot()".
##
## names in output list:
## [1] "OC"          "design"       "simulation"   "delta.grid"   "warnings"
## [6] "system.time"

#tab(result, what="cumulative all")
#tab(result, what="sample size")
#plot(result, what="sample size", sliced = T)
plot(result, what="cumulative all", sliced=T)

```

Operating Characteristics



```

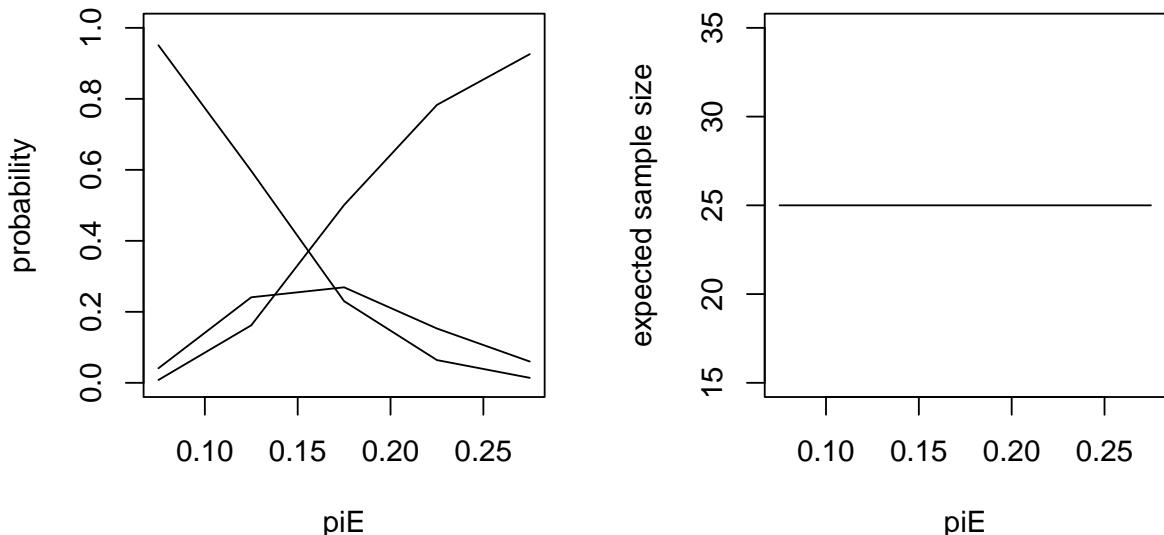
par(mfrow=c(1,2))
y1 = tab(result,what='futility')[,4]

```

```

y2 = tab(result,what='success')[,4]
y3 = 1 - y1 - y2
piE = expit(c(logit(0.075), logit(0.125),logit(0.175), # true treatment
              logit(0.225), logit(0.275)))
plot(piE,y1,type='l',ylim=c(0,1),ylab='probability')
lines(piE,y2)
lines(piE,y3)
y3 = tab(result,what='sample size')[,4]
plot(piE,y3,type='l',ylab='expected sample size')

```



Below code finds the minimal sample size with the help of gsbDesign package.[Not finished]

Appendix

References

- Gerber, F. and Gsponer, T. (2016). gsbdesign: An r package for evaluating the operating characteristics of a group sequential bayesian design. *Journal of Statistical Software*, 69(11):1–23.
- Gsponer, T., Gerber, F., Bornkamp, B., Ohlssen, D., Vandemeulebroecke, M., and Schmidli, H. (2014). A practical guide to Bayesian group sequential designs. *Pharmaceutical statistics*, 13(1):71–80.
- Roychoudhury, S., Scheuer, N., and Neuenschwander, B. (2018). Beyond p-values: A phase II dual-criterion design with statistical significance and clinical relevance. *Clinical trials (London, England)*, 15(5):452–461.

Simon, R. (1989). Optimal two-stage designs for phase ii clinical trials. *Controlled Clinical Trials*, 10(1):1–10.

Computational details

```
## 2025-11-21 15:06:53.84666 Europe/Zurich
## R version 4.4.0 (2024-04-24)
## Platform: aarch64-apple-darwin20
## Running under: macOS 15.6.1
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: Europe/Zurich
## tzcode source: internal
##
## attached base packages:
## [1] grid      stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
## [1] gsbDesign_1.0-3 lattice_0.22-7   gsDesign_3.6.7   latex2exp_0.9.6
## [5] kableExtra_1.4.0 gridExtra_2.3   cowplot_1.1.3   ggplot2_3.5.2
## [9] knitr_1.50
##
## loaded via a namespace (and not attached):
## [1] gtable_0.3.6     highr_0.11      dplyr_1.1.4     compiler_4.4.0
## [5] tidyselect_1.2.1  xml2_1.3.8     stringr_1.5.1   tidyr_1.3.1
## [9] systemfonts_1.2.2 scales_1.3.0    fastmap_1.2.0   R6_2.6.1
## [13] generics_0.1.3   tibble_3.2.1    r2rtf_1.1.4    munsell_0.5.1
## [17] svglite_2.1.3    pillar_1.10.2   rlang_1.1.6    stringi_1.8.7
## [21] xfun_0.52       viridisLite_0.4.2 cli_3.6.5    withr_3.0.2
## [25] magrittr_2.0.3   digest_0.6.37   xtable_1.8-4   rstudioapi_0.17.1
## [29] lifecycle_1.0.4  vctrs_0.6.5     evaluate_1.0.3  glue_1.8.0
## [33] farver_2.1.2    gt_1.0.0       colorspace_2.1-1 purrr_1.0.4
## [37] rmarkdown_2.29   tools_4.4.0    pkgconfig_2.0.3 htmltools_0.5.8.1
```