A novel group-sequential phase II design for clinical trials with binary endpoints based on Bayesian evidence values

Riko Kelter

Department of Mathematics, University of Siegen

Introduction

- In clinical research, the initial efficacy assessment of a new agent is typically considered in a phase IIA *'proof-of-concept'* study which investigates the response rate of patients to the agent under consideration (Matthews, 2006)

- In clinical research, the initial efficacy assessment of a new agent is typically considered in a phase IIA *'proof-of-concept'* study which investigates the response rate of patients to the agent under consideration (Matthews, 2006)
- A meta-analysis in for phase II studies in *Nature Reviews drug discovery* for 2007-2010 concluded: **"Phase II success rates are currently at 18%, lower than at any other phase of drug development."** (Arrowsmith, 2011)

- From 2015-2016, a total of 9985 clinical and regulatory phase transitions were recorded and analyzed from 7455 development programs, across 1103 companies in the Biomedtracker database (Biotechnology Innovation Organization, 2015):

  "Phase II clinical programs continue to experience the lowest success rate of the four development phases, with only 30.7% of developmental candidates advancing to Phase III."

- From 2015-2016, a total of 9985 clinical and regulatory phase transitions were recorded and analyzed from 7455 development programs, across 1103 companies in the Biomedtracker database (Biotechnology Innovation Organization, 2015):

  > ***"Phase II clinical programs continue to experience the lowest success rate of the four development phases, with only 30.7% of developmental candidates advancing to Phase III."***

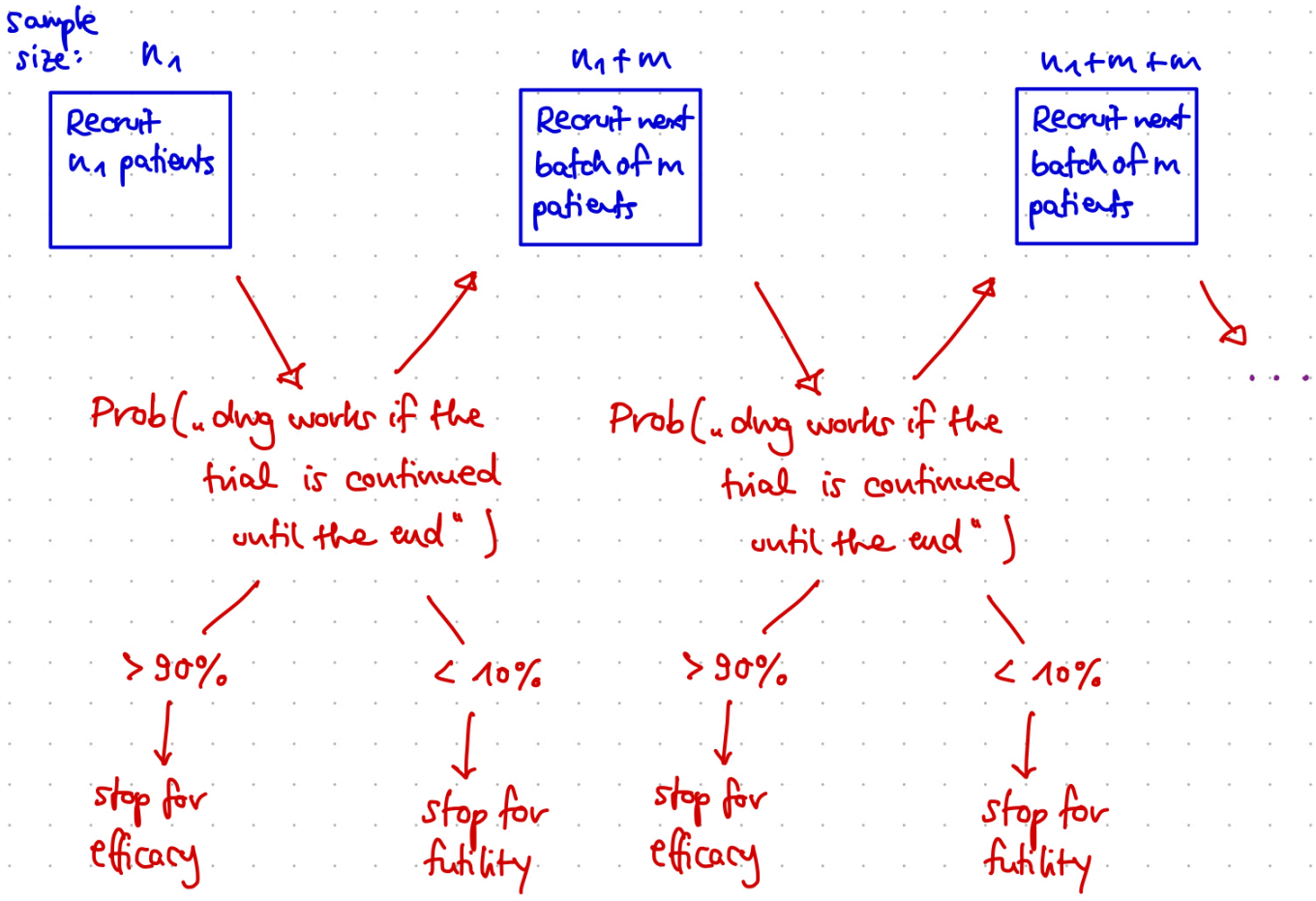
- **Challenge:** Filter promising candidates

- From 2015-2016, a total of 9985 clinical and regulatory phase transitions were recorded and analyzed from 7455 development programs, across 1103 companies in the Biomedtracker database (Biotechnology Innovation Organization, 2015):

  > ***"Phase II clinical programs continue to experience the lowest success rate of the four development phases, with only 30.7% of developmental candidates advancing to Phase III."***

- **Challenge:** Filter promising candidates
- **One solution:** stop trials for efficacy or futility early

- (Bayesian) group-sequential designs for phase IIA *'proof-of-concept'* studies are often based on the **predictive probability approach**:



Publications:

- Seamless phase II/III designs (J. Yuan et al., 2016), seamless phase I/II designs (Chen, Zhang, Jianga, & Yan, 2022; Guo & Yuan, 2017; Liu, Guo, & Yuan, 2018; Y. Yuan & Yin, 2011)

Publications:

- Seamless phase II/III designs (J. Yuan et al., 2016), seamless phase I/II designs (Chen et al., 2022; Guo & Yuan, 2017; Liu et al., 2018; Y. Yuan & Yin, 2011)
- phase II designs for binary endpoints (Stallard, 2012; Guo & Zang, 2020; Leung & Wang, 2001; Tan & Machin, 2002)

Publications:

- Seamless phase II/III designs (J. Yuan et al., 2016), seamless phase I/II designs (Chen et al., 2022; Guo & Yuan, 2017; Liu et al., 2018; Y. Yuan & Yin, 2011)
- phase II designs for binary endpoints (Stallard, 2012; Guo & Zang, 2020; Leung & Wang, 2001; Tan & Machin, 2002)
- phase II designs for time-to-survival endpoint (Wu et al., 2021; Zhao et al., 2012)

What about software?

- CRAN task view for clinical trials: `BayesDesign` (survival endpoints), `MinEDfind` (minimum effective dose finding),

Publications:

- Seamless phase II/III designs (J. Yuan et al., 2016), seamless phase I/II designs (Chen et al., 2022; Guo & Yuan, 2017; Liu et al., 2018; Y. Yuan & Yin, 2011)
- phase II designs for binary endpoints (Stallard, 2012; Guo & Zang, 2020; Leung & Wang, 2001; Tan & Machin, 2002)
- phase II designs for time-to-survival endpoint (Wu et al., 2021; Zhao et al., 2012)

What about software?

- CRAN task view for clinical trials: `BayesDesign` (survival endpoints), `MinEDfind` (minimum effective dose finding),
- R packages `ph2bayes` and `ph2bye`, based on (Thall & Wathen, 2007) and (Lee & Liu, 2008)

Bayesian Efficacy Monitoring Via Predictive Probability

J. Jack Lee, Ying-Wei Kuo, Diane Liu and Nan Chen

Department of Biostatistics, MD Anderson Cancer Center, Houston, TX 77030

PID:901; v1.1.4.0 ; Last Updated: 10/19/2020

Calculate Stopping Boundaries

Trial Name (Optional)

mytrial

Design Parameters:

Probability of Response (θ)

Declare efficacious if

$$PP = \sum \{ Prob(future\ data) * [Prob\ (\theta > \theta_0 \mid current\ and\ future\ data) \geq P_T\ ] \}$$

Reference response rate (θ_0)

0,3

Threshold for declaring efficacy at the end of the trial (P_T)

0,9

Early Stopping for Futility: $PP < P_L$

P_L

Stopping Boundaries | Operating Characteristics | Trial Monitoring | Support Document

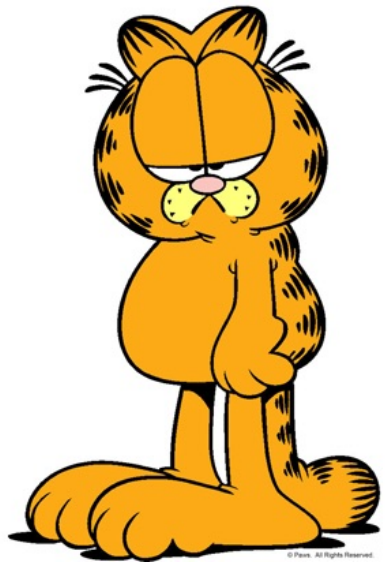
CSV | Excel | PDF | Print

Table SB1: Futility Early Stopping Boundaries

# Patients (inclusive)	# Responses (inclusive) are considered futile	Actions
5	0	Early stopping
10	0 - 2	Early stopping
15	0 - 3	Early stopping
20	0 - 5	Early stopping
25	0 - 7	Early stopping
30	0 - 9	Early stopping
35	0 - 12	Early stopping

Previous 1 Next

CSV | Excel | PDF | Print


© Paws. All Rights Reserved.

Plan for today

1) Propose a novel group-sequential design based on Bayesian evidence values

Plan for today

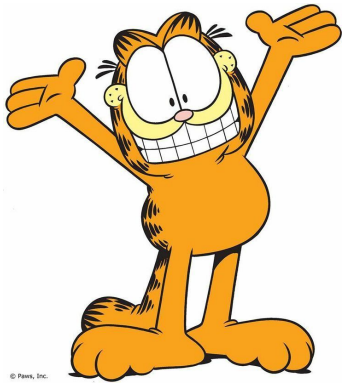
1) Propose a novel group-sequential design based on Bayesian evidence values
2) Show that the predictive probability approach is a special case of our new design

Plan for today

1) Propose a novel group-sequential design based on Bayesian evidence values
2) Show that the predictive probability approach is a special case of our new design
3) Provide theoretical results and show how the new solution outperforms existing ones

Plan for today

1) Propose a novel group-sequential design based on Bayesian evidence values
2) Show that the predictive probability approach is a special case of our new design
3) Provide theoretical results and show how the new solution outperforms existing ones
4) Provide a ready-to-use software implementation, the R package `brada`



Development of the new design

Step 1:

Theory of Bayesian evidence values (Kelter, 2021b, 2022; Pereira & Stern, 2020; Diniz, Pereira, Polpo, Stern, & Wechsler, 2012; Borges & Stern, 2007; Madruga, Pereira, & Stern, 2003; Madruga, Esteves, & Wechsler, 2001; Pereira, Stern, & Wechsler, 2008; Pereira & Stern, 1999)

Step 2:

Fuse the theory of Bayesian evidence values with the predictive probability approach

Step 1: Theory of Bayesian evidence values

Step 1: Theory of Bayesian evidence values

Definition (Bayesian Evidence Interval)

The Bayesian evidence interval $\mathrm{EI}_r(\nu)$ is defined as

$$\mathrm{EI}_r(\nu) := \left\{ \theta \in \Theta \middle| \frac{p(\theta|y)}{r(\theta)} \geq \nu \right\} \tag{1}$$

for a given reference function $r : \Theta \to [0, \infty)$, $\theta \mapsto r(\theta)$ and observed data $y \in \mathcal{Y}$, where Θ denotes the parameter space.

- Highest-posterior-density intervals, credible intervals, and even confidence intervals (asymptotically) are special cases of Bayesian evidence interval (Kelter, 2022)

Step 1: Theory of Bayesian evidence values

Definition (Bayesian Evidence Interval)

The Bayesian evidence interval $\mathrm{EI}_r(\nu)$ is defined as

$$\mathrm{EI}_r(\nu) := \left\{ \theta \in \Theta \middle| \frac{p(\theta|y)}{r(\theta)} \geq \nu \right\} \tag{1}$$

for a given reference function $r : \Theta \to [0, \infty)$, $\theta \mapsto r(\theta)$ and observed data $y \in \mathcal{Y}$, where Θ denotes the parameter space.

- Highest-posterior-density intervals, credible intervals, and even confidence intervals (asymptotically) are special cases of Bayesian evidence interval (Kelter, 2022)
- Example: $r(\theta) :\equiv 1 \Rightarrow \mathrm{EI}_r(\nu) := \{\theta \in \Theta | p(\theta|y) \geq \nu\}$
  $\to$ pick ν to obtain HPD-intervals

Step 1: Theory of Bayesian evidence values

Definition (Bayesian Evidence Interval)

The Bayesian evidence interval $\mathrm{EI}_r(\nu)$ is defined as

$$\mathrm{EI}_r(\nu) := \left\{ \theta \in \Theta \middle| \frac{p(\theta|y)}{r(\theta)} \geq \nu \right\} \tag{1}$$

for a given reference function $r : \Theta \to [0, \infty)$, $\theta \mapsto r(\theta)$ and observed data $y \in \mathcal{Y}$, where Θ denotes the parameter space.

- Highest-posterior-density intervals, credible intervals, and even confidence intervals (asymptotically) are special cases of Bayesian evidence interval (Kelter, 2022)
- Example: $r(\theta) :\equiv 1 \Rightarrow \mathrm{EI}_r(\nu) := \{\theta \in \Theta | p(\theta|y) \geq \nu\}$
  $\to$ pick ν to obtain HPD-intervals

Definition (Bayesian Evidence Value)

Let $H_0 := \Theta_0$ and $H_1 := \Theta \setminus \Theta_0$ be a null and alternative hypothesis with $\Theta_0 \in \Theta$. For a given Bayesian evidence interval $\mathrm{EI}_r(\nu)$ with reference function $r(\theta)$ to level ν, the Bayesian evidence value $\mathrm{Ev}_{\mathrm{EI}_r(\nu)}(H_0)$ for the (interval) null hypothesis H_i, $i = 0, 1$ is

$$\mathrm{Ev}_{\mathrm{EI}_r(\nu)}(H_i) := \int_{\mathrm{EI}_r(\nu) \cap \Theta_i} p(\theta|y) d\theta \tag{2}$$

Importantly, for $r(\theta) :\equiv 1$ and $\nu := 0$ we have

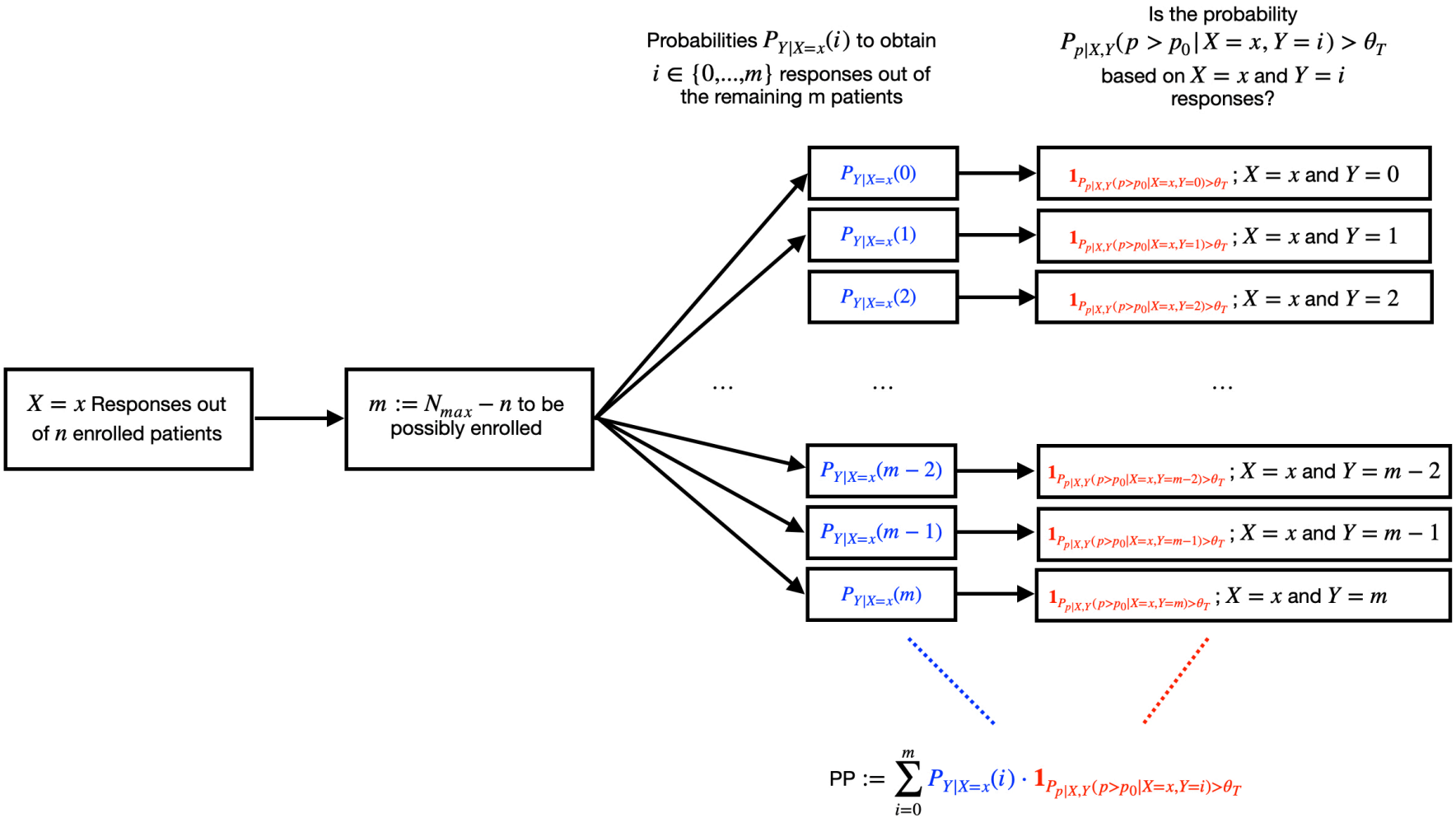
$$\mathrm{EI}_r(\nu) := \left\{ \theta \in \Theta \middle| \theta \in \frac{p(\theta|y)}{r(\theta)} \geq \nu \right\} = \left\{ \theta \in \Theta \middle| \theta \in \frac{p(\theta|y)}{1} \geq 0 \right\} =: \Theta$$

so

$$\mathrm{Ev}_{\mathrm{EI}_r(\nu)}(H_i) := \int_{\mathrm{EI}_r(\nu) \cap \Theta_i} p(\theta|y) d\theta = \int_{\Theta \cap \Theta_i} p(\theta|y) d\theta = P_{\vartheta|Y}(\Theta_i) = P_{\vartheta|Y}(H_i)$$

Step 2: Fuse the theory with the predictive probability approach

The predictive probability approach in detail



The predictive probability approach in detail

- Setting: single-arm phase IIA trial with the goal to evaluate the response rate $p \geq 0$ for a new drug or treatment

The predictive probability approach in detail

- Setting: single-arm phase IIA trial with the goal to evaluate the response rate $p \geq 0$ for a new drug or treatment
- The null hypothesis $H_0 : p \leq p_0$ is tested against the alternative $H_1 : p > p_1$, where $p_0, p_1 \in [0, 1]$, $p_0 \leq p_1$ and p_0 is a predefined threshold for determining the minimum clinically important effect (Kelter, 2021a; Chuang-Stein et al., 2011; Cook et al., 2018).

The predictive probability approach in detail

- Setting: single-arm phase IIA trial with the goal to evaluate the response rate $p \geq 0$ for a new drug or treatment
- The null hypothesis $H_0 : p \leq p_0$ is tested against the alternative $H_1 : p > p_1$, where $p_0, p_1 \in [0, 1]$, $p_0 \leq p_1$ and p_0 is a predefined threshold for determining the minimum clinically important effect (Kelter, 2021a; Chuang-Stein et al., 2011; Cook et al., 2018).
- For simplicity, assume a Beta prior $p \sim \mathcal{B}(a_0, b_0)$ is selected for the response rate p, which offers a broad range of flexibility in terms of modelling the prior beliefs about p.

The predictive probability approach in detail

- $N_{\max}$ = maximum number of patients

The predictive probability approach in detail

- $N_{\max}$ = maximum number of patients
- X = number of responses in the current n enrolled patients, $n \leq N_{\max}$.

The predictive probability approach in detail

- $N_{\max}$ = maximum number of patients
- X = number of responses in the current n enrolled patients, $n \leq N_{\max}$.
- Reasonable assumption: $X \sim \text{Bin}(n, p)$
- If n patients are enrolled in the trial out of which $X = x$ show a response, there remain $m = N_{\max} - n$ patients which can be enrolled in the trial.

The predictive probability approach in detail

- Y = responses in the $m = N_{max} - n$ remaining patients which can be enrolled in the future

The predictive probability approach in detail

- Y = responses in the $m = N_{max} - n$ remaining patients which can be enrolled in the future
- If out of these remaining m exactly i respond to the treatment, and

$$P_{p|X,Y}(p > p_0|X = x, Y = i) > \theta_T \tag{3}$$

for, say, $\theta_T = 0.95$, this will be interpreted as the drug being effective

The predictive probability approach in detail

- Y = responses in the $m = N_{max} - n$ remaining patients which can be enrolled in the future
- If out of these remaining m exactly i respond to the treatment, and

$$P_{p|X,Y}(p > p_0|X = x, Y = i) > \theta_T \tag{3}$$

  for, say, $\theta_T = 0.95$, this will be interpreted as the drug being effective
- However, we do not know the number Y of responses in the future $m = N_{\max} - n$ patients

The predictive probability approach in detail

- Y = responses in the $m = N_{max} - n$ remaining patients which can be enrolled in the future
- If out of these remaining m exactly i respond to the treatment, and

$$P_{p|X,Y}(p > p_0 | X = x, Y = i) > \theta_T \tag{3}$$

  for, say, $\theta_T = 0.95$, this will be interpreted as the drug being effective
- However, we do not know the number Y of responses in the future $m = N_{\max} - n$ patients
- Marginalizing out p of the binomial likelihood yields the prior predictive distribution which is Beta-Binomial:

$$Y \sim \text{Beta-Binom}(m, a_0 + x, b_0 + n - x) \tag{4}$$

The predictive probability approach in detail

- The expected predictive probability of trial success – henceforth abbreviated PP – can now be calculated as follows:

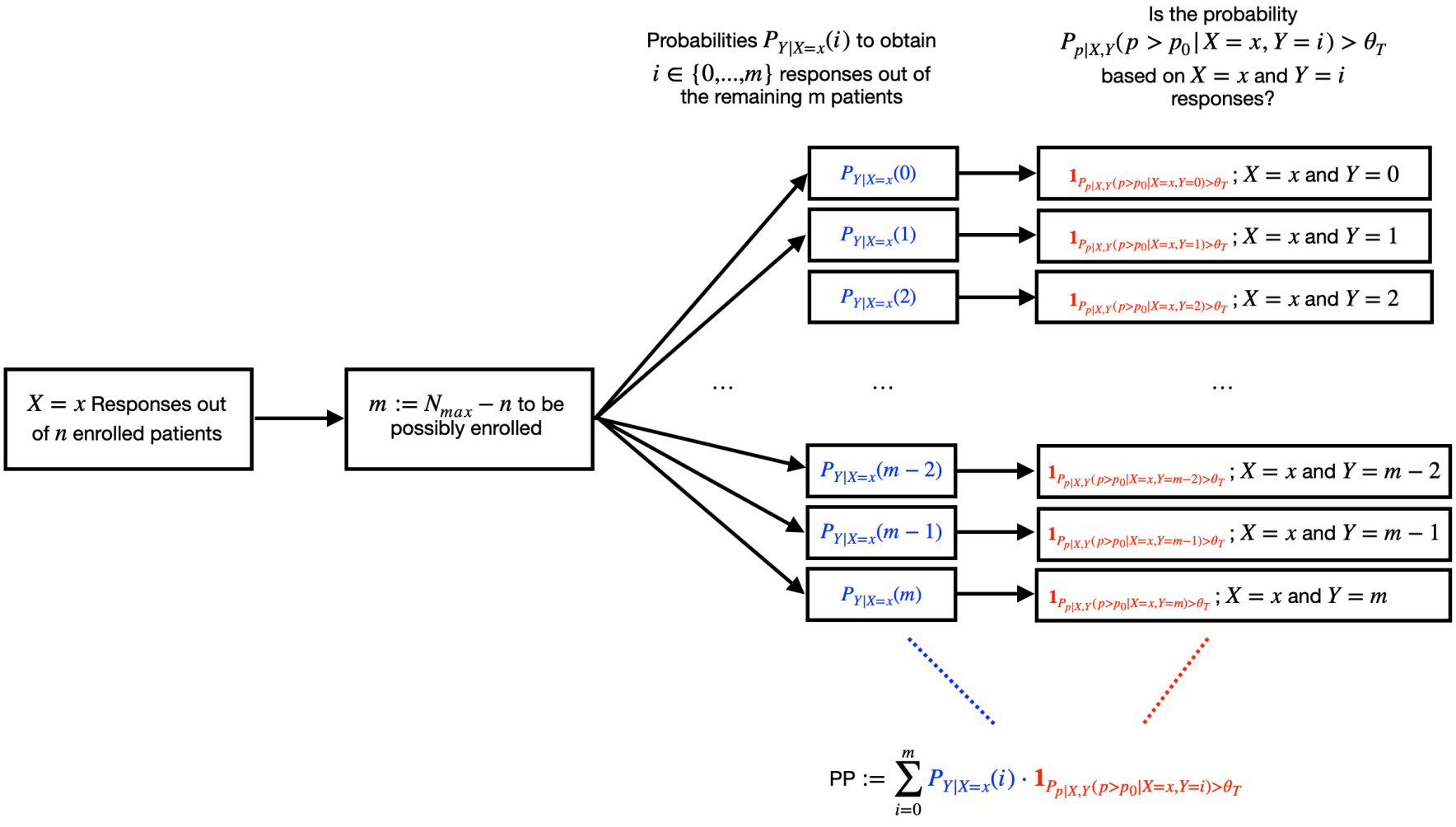
$$\begin{aligned} \text{PP} &= \mathbb{E}\left[\mathbb{1}_{P_{p|X,Y}(p>p_0|X,Y)>\theta_T}|x\right] = \int_{\mathcal{Y}} \mathbb{1}_{P_{p|X,Y}(p>p_0|X,Y)>\theta_T} dP_{Y|X=x} \\ &= \sum_{i=0}^{m} P_{Y|X=x}(i) \cdot \mathbb{1}_{P_{p|X,Y}(p>p_0|X=x,Y=i)>\theta_T} \end{aligned} \tag{5}$$

where

$$\mathbb{1}_{P_{p|X,Y}(p>p_0|X=x,Y=i)>\theta_T} := \begin{cases} 1, & \text{if } P_{p|X,Y}(p > p_0|X = x, Y = i) > \theta_T \\ 0, & \text{if } P_{p|X,Y}(p > p_0|X = x, Y = i) \leq \theta_T \end{cases}$$

is an indicator which measures whether the evidence against $H_0 : p \leq p_0$ is large enough – that is, $P_{p|X,Y}(p > p_0|X = x, Y = i) > \theta_T$ – conditional on $X = x$ and $Y = i$ or not.

The predictive probability approach in detail



The predictive probability approach in detail

- Recruit $n < N_{\max}$ patients and observe $X = x$ responses
- If PP $< \theta_L$, stop the trial and reject the alternative hypothesis $H_1 : p > p_0$
- If PP $> \theta_U$, stop the trial and reject the null hypothesis $H_0 : p \leq p_0$
- Otherwise recruit the next patient until reaching $N_{\max}$ patients

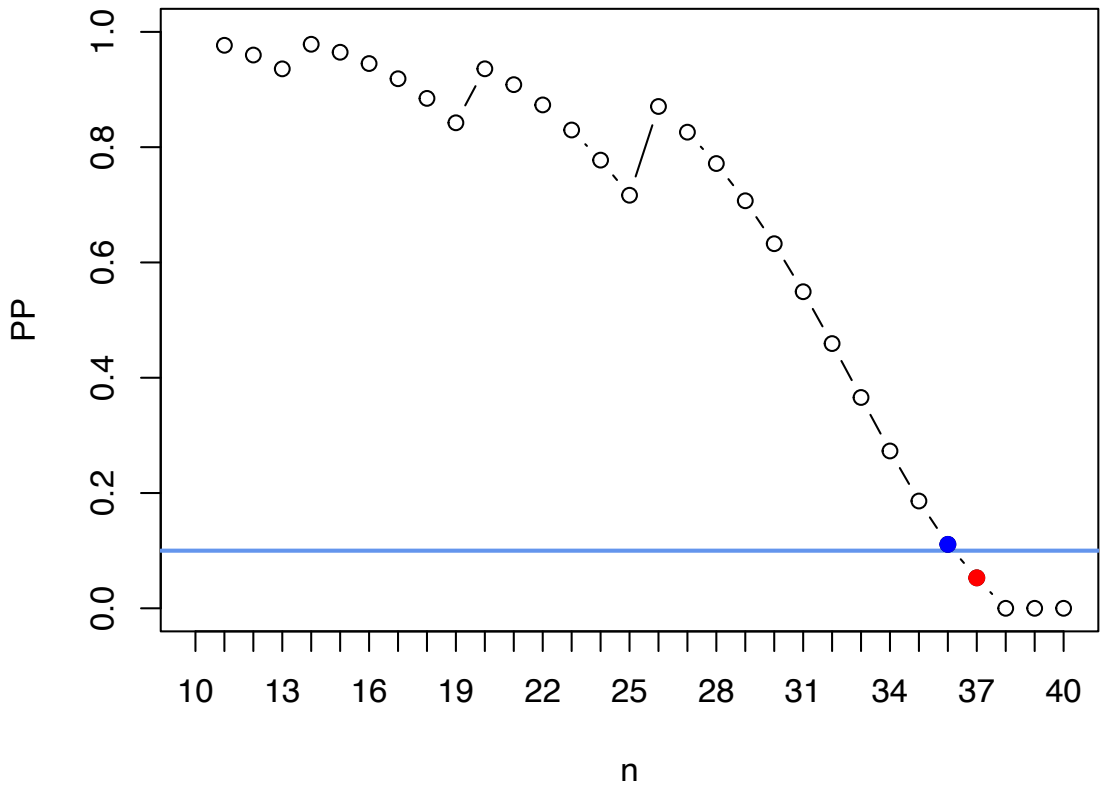


Figure 1: Bayesian group-sequential stopping for futility based on predictive probabilities using PP

The predictive evidence value (PEV) design

The predictive evidence value (PEV) design

The novel Bayesian group-sequential design based on Bayesian evidence values modifies PP as follows into PP_e:

$$\begin{aligned}\text{PP}_e &= \mathbb{E}\left[\mathbb{1}_{\text{Ev}_{\text{EI}_r(\nu)}(H_1)>\theta_T}|x\right] = \int_{\mathcal{Y}} \mathbb{1}_{\text{Ev}_{\text{EI}_r(\nu)}(H_1)>\theta_T} dP_{Y|X=x} \\ &= \sum_{i=0}^{m} P_{Y|X=x}(i) \cdot \mathbb{1}_{\text{Ev}_{\text{EI}_r(\nu)}(H_1)>\theta_T} \end{aligned} \tag{6}$$

where

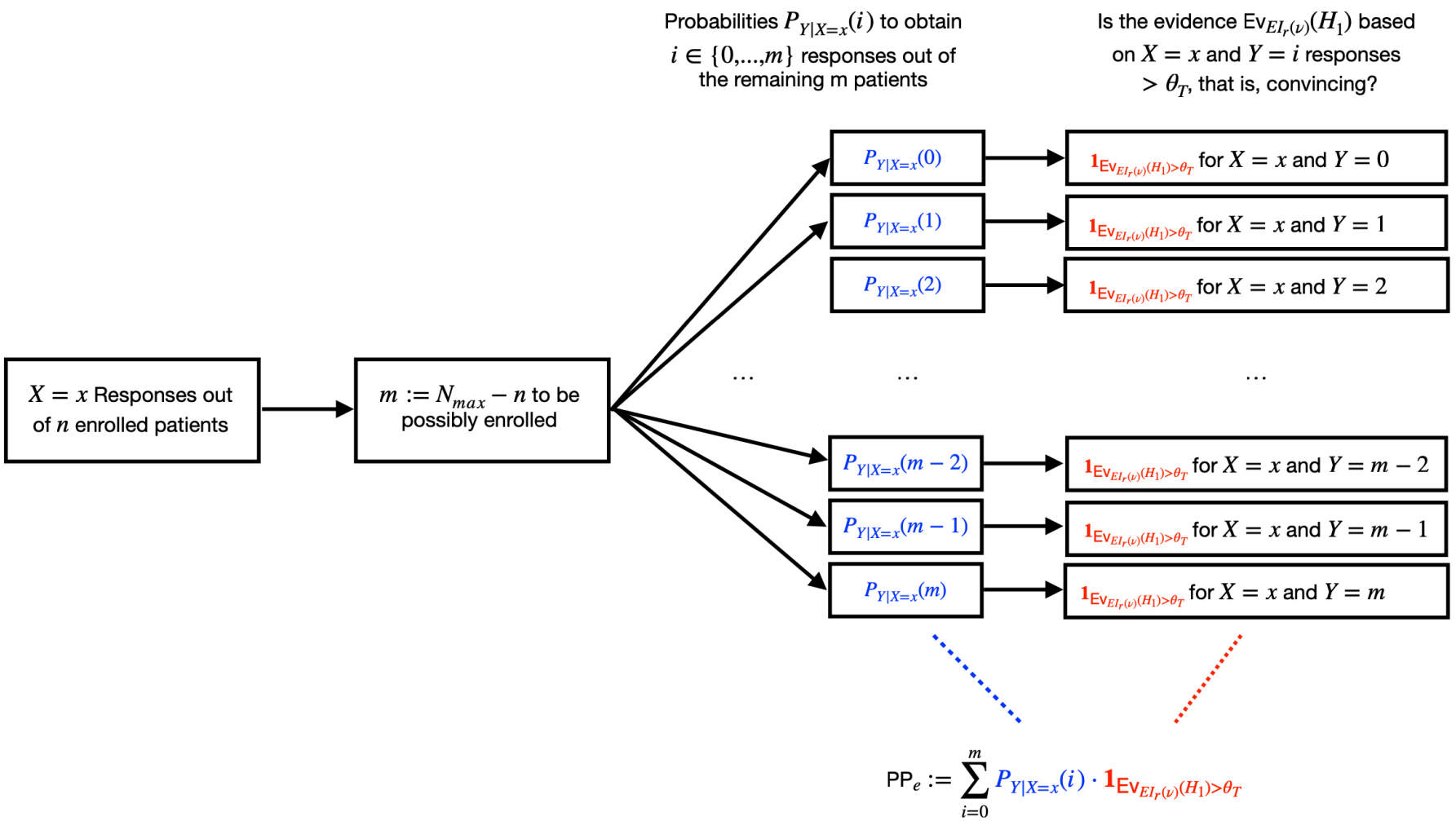
$$\mathbb{1}_{\text{Ev}_{\text{EI}_r(\nu)}(H_1)>\theta_T} := \begin{cases} 1, & \text{if } \text{Ev}_{\text{EI}_r(\nu)}(H_1) > \theta_T \\ 0, & \text{if } \text{Ev}_{\text{EI}_r(\nu)}(H_1) \leq \theta_T \end{cases}$$

The predictive evidence value (PEV) design

Now, the PEV design based on PP_e differs from the basic predictive probability approach as follows:

- The reference function r and the evidence threshold $\nu \geq 0$ for $\mathrm{Ev}_{\mathrm{EI}_r(\nu)}(H_1)$ influence the result
- The posterior probability $P_{p|\mathrm{X},Y}(p > p_0|\mathrm{X} = x, Y = i) > \theta_T$ condition for effectivity is replaced by the Bayesian evidence value condition $\mathrm{Ev}_{\mathrm{EI}_r(\nu)}(H_1) > \theta_T$ for effectivity.

The predictive evidence value design



The predictive evidence value design

- Recruit $n < N_{\max}$ patients and observe $X = x$ responses
- If $\text{PP}_e < \theta_L$, stop the trial and reject the alternative hypothesis $H_1 : p > p_0$
- If $\text{PP}_e > \theta_U$, stop the trial and reject the null hypothesis $H_0 : p \leq p_0$
- Otherwise recruit the next patient until reaching $N_{\max}$ patients

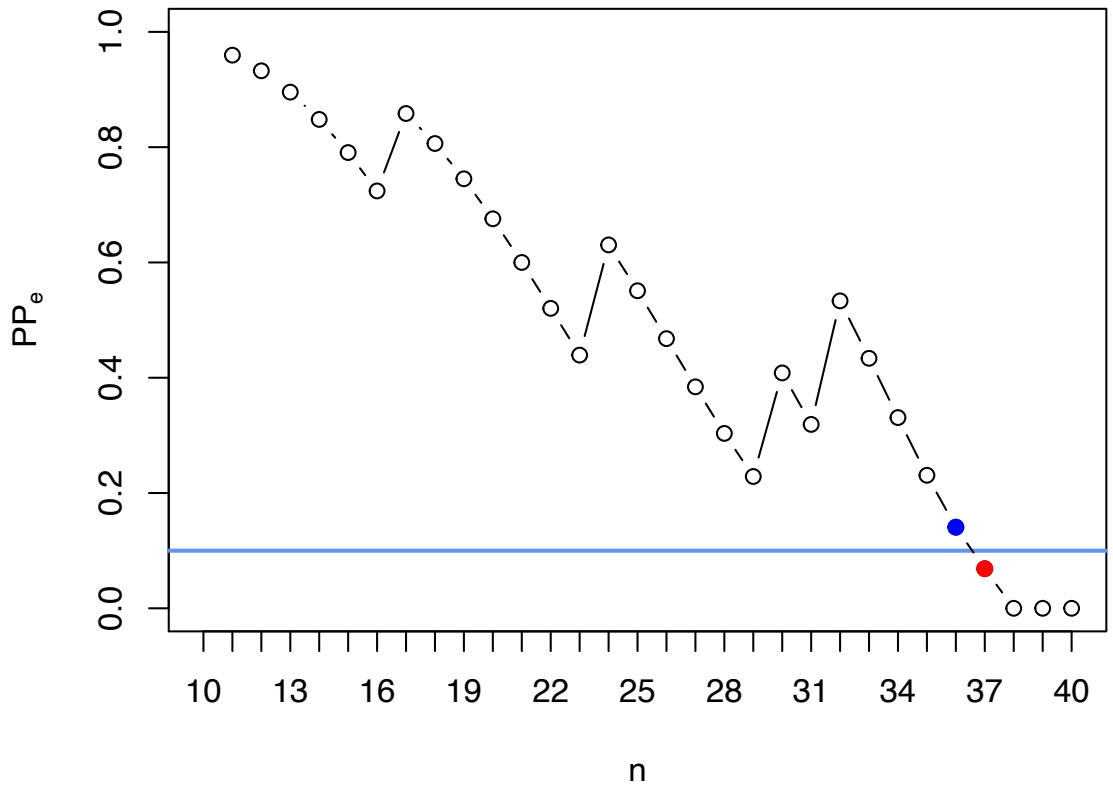


Figure 2: Bayesian group-sequential stopping for futility based on the predictive evidence value (PEV) design using PP_e instead of PP

Theoretical results

Theoretical results

Theorem

If $\nu := 0$ and $r(p) := 1$, then the predictive evidence value design and predictive probability design are equivalent.

- The PEV design a generalization of the predictive probability approach.
- The consequence:

Corollary

Let $\nu := 0$ and $r(p) := 1$ and denote α_{PP} and α_{PP_e} and β_{PP} and β_{PP_e} as the false-positive and false-negative error rates under $H_0 : p \leq p_0$ for the prior predictive and predictive evidence value designs. Then,

$$\alpha_{PP} = \alpha_{PP_e} \qquad \textit{and} \qquad \beta_{PP} = \beta_{PP_e}$$

Theoretical results

► The consequence in full generality:

Corollary

Under the conditions of Theorem 3, the operating characteristics of the PP and PEV designs are identical. The latter include the probability of early stopping (PET) and the expected sample size until early stopping as well as their associated variances.

► The predictive probability design is a special case of the PEV design **with respect to all trial operating characteristics!**

Theoretical results

- What can be said about reducing the false-positive rate?

Theorem

If $r(\theta) := 1$ *and if* $\{\theta \in \Theta | \frac{p(\theta|y)}{r(\theta)} \geq \nu\} \subseteq \{\theta \in \Theta | p(\theta|y) > 0\}$ *then*

$$\alpha_{PP_e} \leq \alpha_{PP}$$

where a sharp inequality can hold if and only if

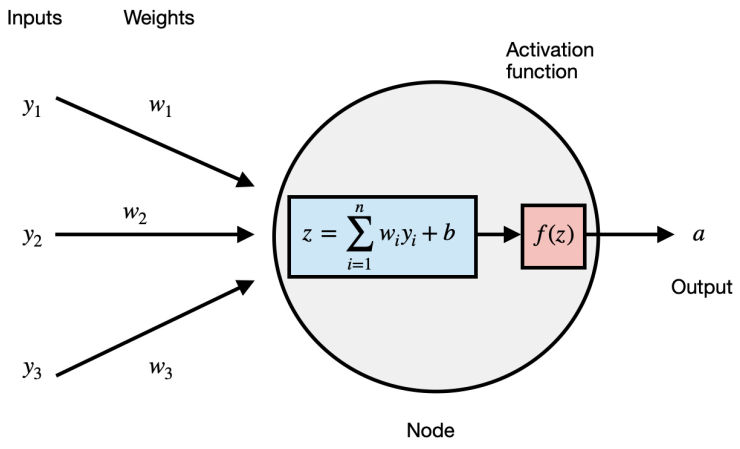
$$\{\theta \in \Theta | \frac{p(\theta|y)}{r(\theta)} \geq \nu\} \subset \{\theta \in \Theta | p(\theta|y) > 0\}$$

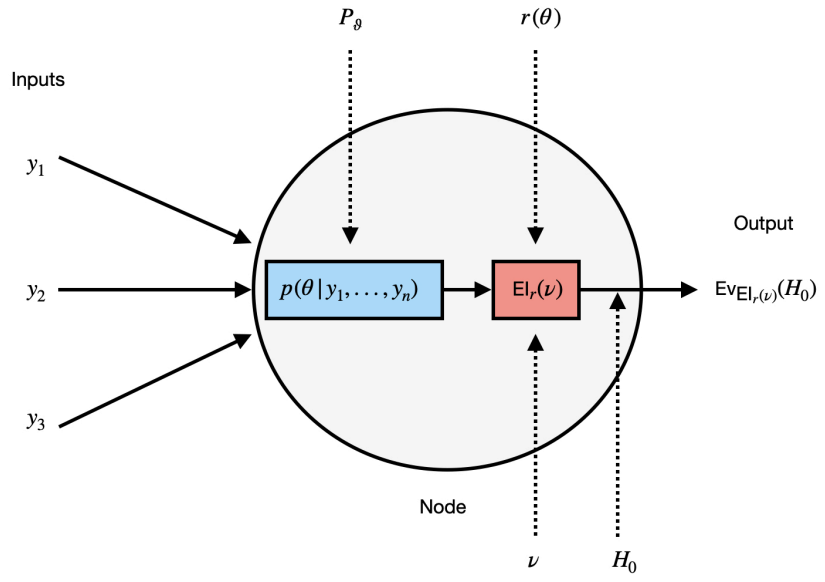
and $EI_r(\nu) \cap H_1 \neq \emptyset \neq EI_r(\nu) \neq H_1$*, where* Θ *denotes the parameter space.*

- $\mathrm{EI}_r(\nu) \cap H_1 \neq \emptyset \neq \mathrm{EI}_r(\nu) \neq H_1$ depends crucially on the choice of $r(p)$ and $\nu \geq 0$!
- We will quickly see how to turn this into our advantage...

Calibrating the PEV design

Calibrating the PEV design



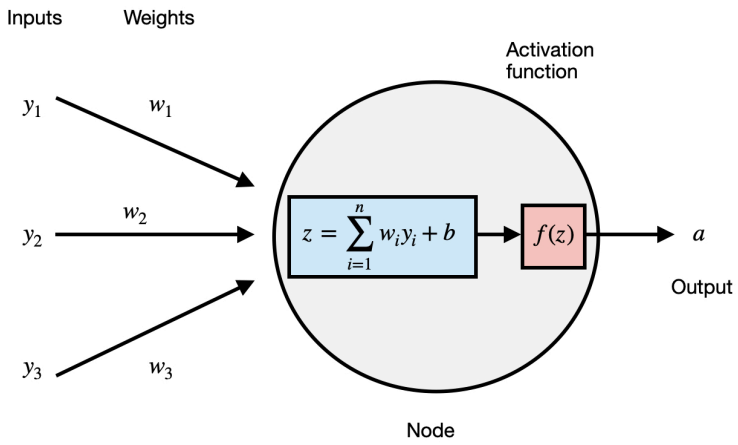


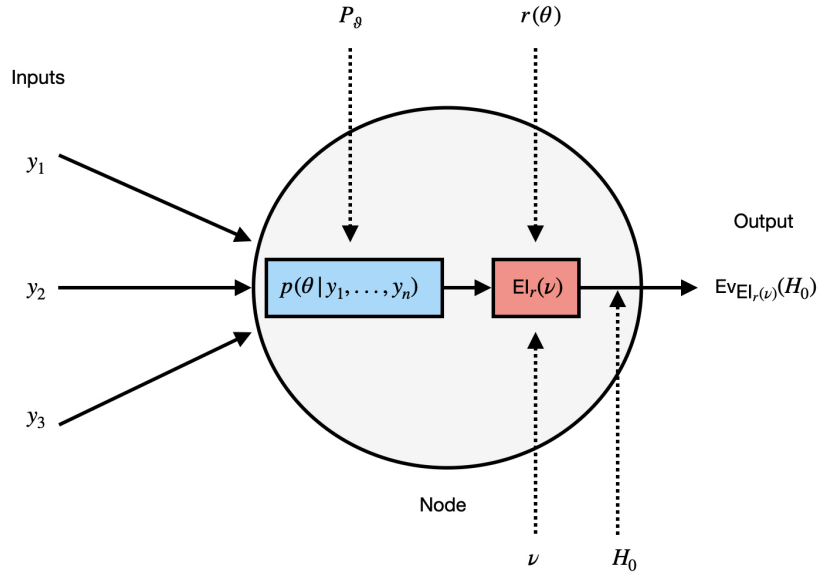
$$\text{Evidence interval} \quad \mathrm{EI}_r(\nu) := \left\{ \theta \in \Theta \middle| \theta \in \frac{p(\theta|y)}{r(\theta)} \geq \nu \right\}$$

$$\overset{r(p):=1}{=} \left\{ \theta \in \Theta \middle| \theta \in p(\theta|y) \geq \nu \right\}$$

$$\text{Evidence value} \qquad \mathrm{Ev}_{\mathrm{EI}_r(\nu)}(H_i) := \int_{\mathrm{EI}_r(\nu) \cap \Theta_i} p(\theta|y) d\theta$$

Calibrating the PEV design





Idea: Fix a flat reference function $r(p) := 1$ and vary $\nu \geq 0$ to calibrate the trial's operating characteristics!

Ready-to-use software implementation

The R package brada

- brada = **B**ayesian **R**esponse **A**daptive **D**esign **A**nalysis

- Available on CRAN in before christmas!
- 4 vignettes:
    1) Getting started
    2) power analyses and sample size calculations
    3) PEV design in the brada package
    4) monitoring a trial with the brada package

The R package brada

Example: Phase IIa lung cancer trial; treatment: vascular endothelial growth factor antibody + epidermal growth factor receptor tyrosine kinase inhibitor (Lee & Liu, 2008)

- Frequentist solution: Simon's two-stage design
- The current standard treatment yields a response rate of $\approx 20\%$, so we have $p_0 = 0.2$. The target response rate of the new regimen is 40%, so $p_1 = 0.4$.
- Requirement 1: $\alpha \leq 10\%$
- Requirement 2: $\beta \leq 10\%$ for $p \geq 0.4$

Table 2 Operating characteristics of Simon's two-stage designs and the PP design with type I and type II error rates ≤ 0.10, prior for $p = \text{beta}(0.2,0.8)$, $p_0 = 0.2$, $p_1 = 0.4$

Simon's minimax/optimal 2-satge						
	r_1/n_1	r/N_{max}	PET (p_0)	$E(N\|p_0)$	α	β
Minimax	3/19	10/36	0.46	28.26	0.086	0.098
Optimal	3/17	10/37	0.55	26.02	0.095	0.097
Predictive probability						
θ_L	θ_T	r/N_{max}	PET (p_0)	$E(N\|p_0)$	α	β
0.001	**[0.852,0.922]***	**10/36**	**0.86**	**27.67**	**0.088**	**0.094**
0.011	[0.830,0.908]	10/37	0.85	25.13	0.099	0.084
0.001	[0.876,0.935]	11/39	0.88	29.24	0.073	0.092
0.001	[0.857,0.923]	11/40	0.86	30.23	0.086	0.075
0.003	[0.837,0.910]	11/41	0.85	30.27	0.100	0.062
0.043	[0.816,0.895]	11/42	0.86	23.56	0.099	0.083
0.001	[0.880,0.935]	12/43	0.88	32.13	0.072	0.074
0.001	[0.862,0.924]	12/44	0.87	33.71	0.085	0.059
0.001	[0.844,0.912]	12/45	0.85	34.69	0.098	0.048
0.032	[0.824,0.898]	12/46	0.86	26.22	0.098	0.068
0.001	[0.884,0.936]	13/47	0.89	35.25	0.071	0.058
0.001	[0.868,0.925]	13/48	0.87	36.43	0.083	0.047
0.001	[0.850,0.914]	13/49	0.86	37.86	0.095	0.038
0.020	[0.832,0.901]	13/50	0.86	30.60	0.100	0.046

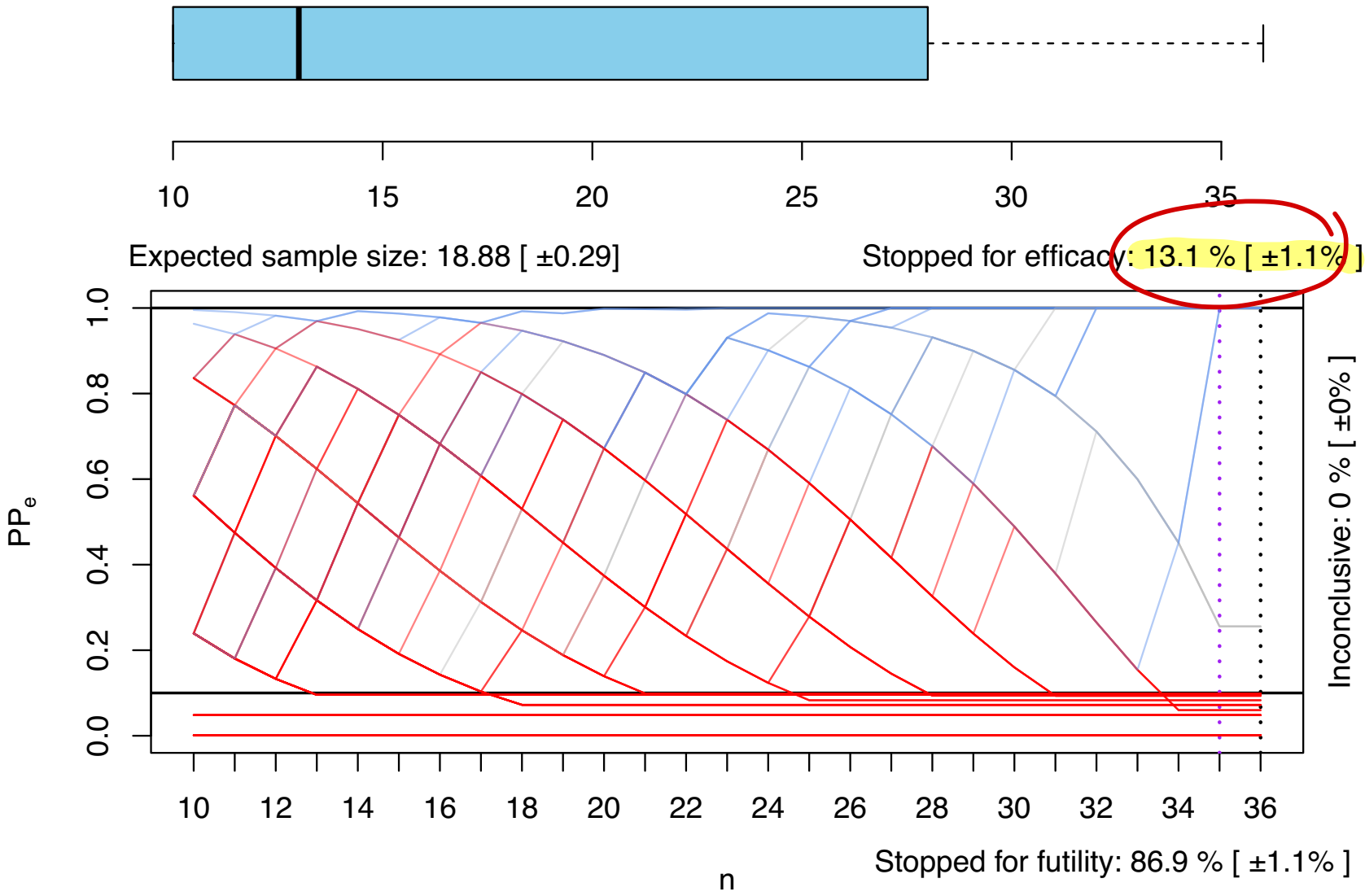
Figure 3: Source: (Lee & Liu, 2008)

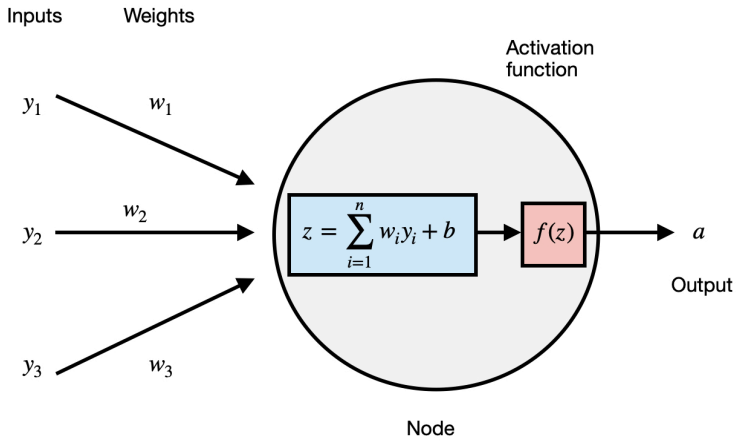
The R package brada

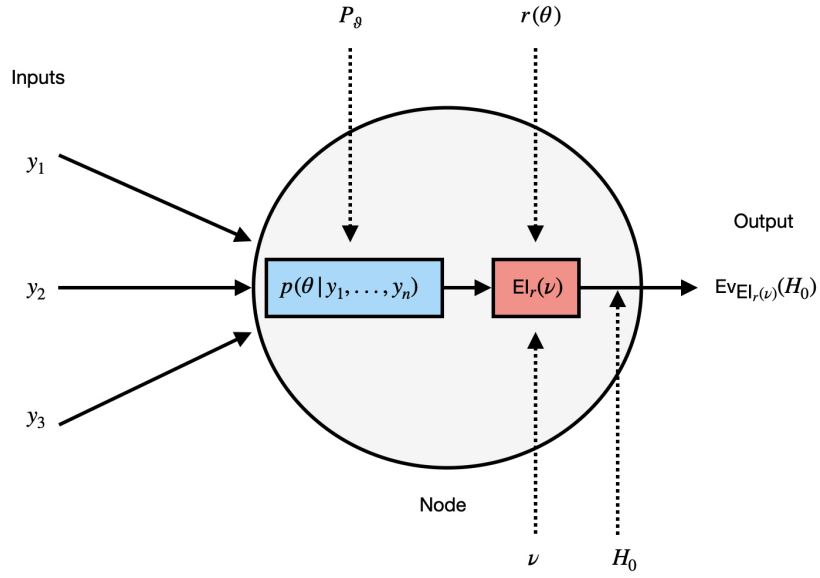
```r
library(brada)

pev = brada(a0 = 0.2, b0 = 0.8, # vague prior
            Nmax = 36, # maximum trial size
            batchsize = 1, # continuous monitoring
            nInit = 10, # first interim at 10 patients
            p_true = 0.2, # true response probability
            p0 = 0.2, # probability of null hypothesis
            p1 = 0.2,
            theta_T = 0.8, # threshold to declare efficacy
            theta_L = 0.1, # threshold to stop for futility
            theta_U = 1.0, # threshold to stop for efficacy
            nsim = 1000, # number of Monte Carlo iterations
            method = "PPe", # PEV design...
            refFunc = "flat", # ...flat reference function
            nu = 0, # ...and evidence threshold zero
            cores = 12) # parallelization

plot(ex_pev)
```

flat reference function $r(p) := 1$, evidence threshold $\nu := 0$

Operating characteristics under $H_0 : p \leq 0.2$:







$$\text{Evidence interval} \quad \mathrm{EI}_r(\nu) := \left\{ \theta \in \Theta \middle| \theta \in \frac{p(\theta|y)}{r(\theta)} \geq \nu \right\}$$

$$\overset{r(p):=1}{=} \left\{ \theta \in \Theta \middle| \theta \in p(\theta|y) \geq \nu \right\}$$

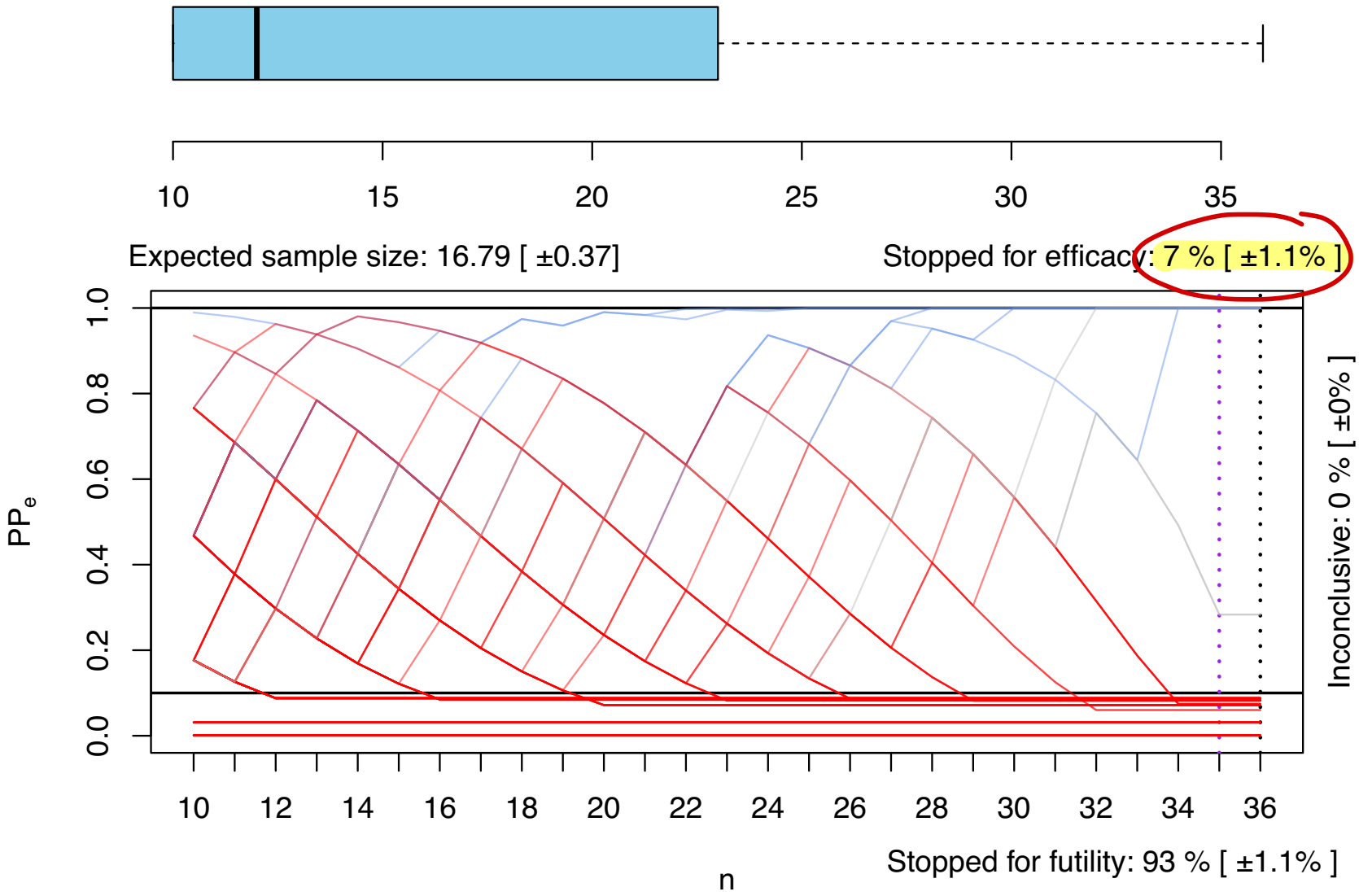
$$\text{Evidence value} \qquad \mathrm{Ev}_{\mathrm{EI}_r(\nu)}(H_i) := \int_{\mathrm{EI}_r(\nu) \cap \Theta_i} p(\theta|y) d\theta$$

```
calibrate(ex_pev, cores = 12, seq = seq(0,2,by=0.1), alpha =
    0.1, calibration = "nu", nsim = 500)

False-positive rate is below specified threshold for nu equal to
     1.3
```

$\Rightarrow$ proceed with flat reference function $r(p) := 1$ and evidence threshold $\nu := 1.3$

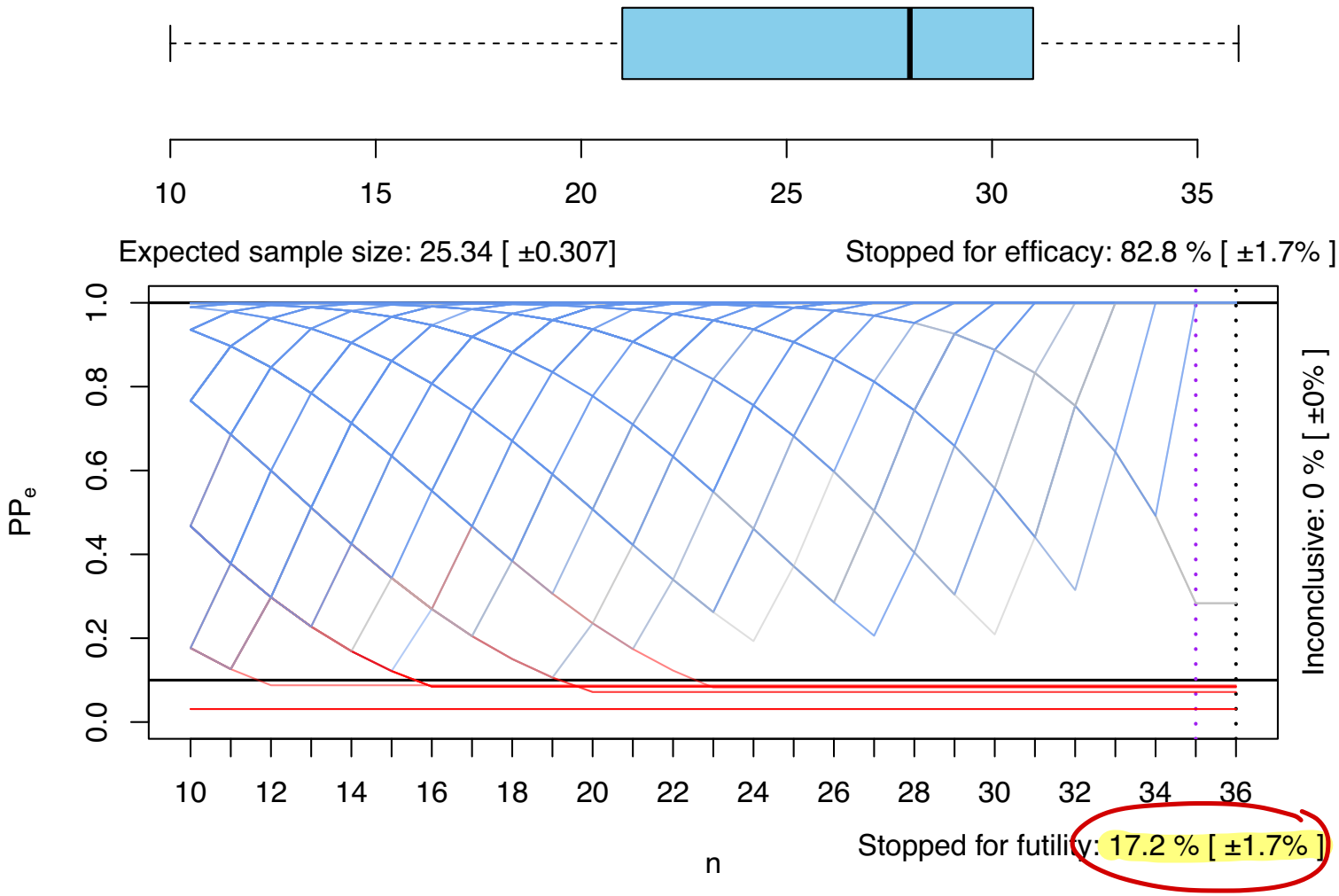
```r
pev_updated = brada(a0 = 0.2, b0 = 0.8,
                    Nmax = 36,
                    batchsize = 1,
                    nInit = 10,
                    p_true = 0.2,
                    p0 = 0.2,
                    p1 = 0.2,
                    theta_T = 0.8,
                    theta_L = 0.1,
                    theta_U = 1.0,
                    nsim = 500,
                    method = "PPe",
                    refFunc = "flat",
                    nu = 1.3,
                    cores = 6)

plot(pev_updated)
```

Operating characteristics under $H_0 : p \leq 0.2$:



What about the power under $H_1 : p \geq 0.4$?

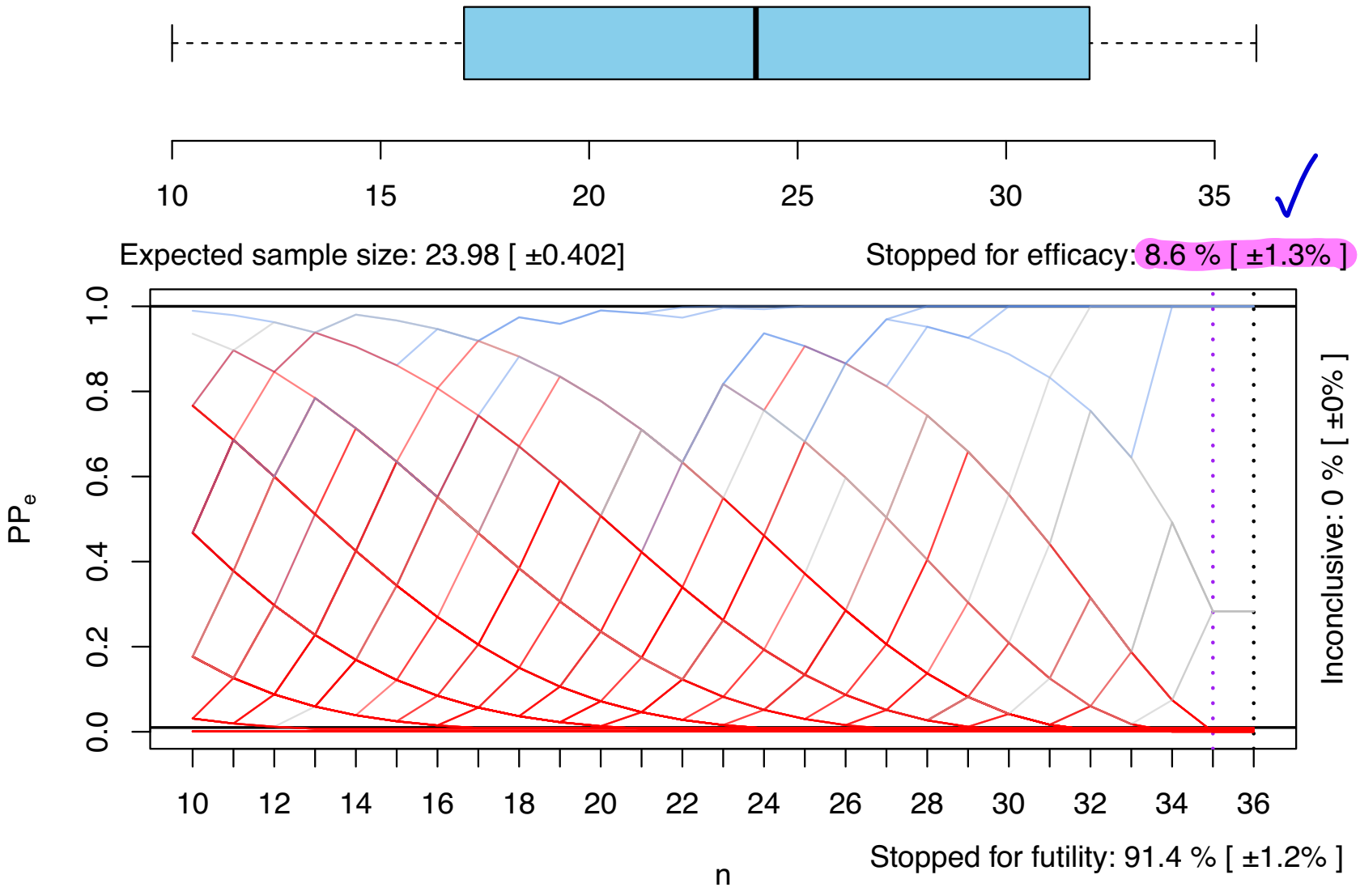
```
power_pev_updated = power(pev_updated, p_true = 0.4, nsim = 500,
     cores = 12)
plot(power_pev_updated)
```



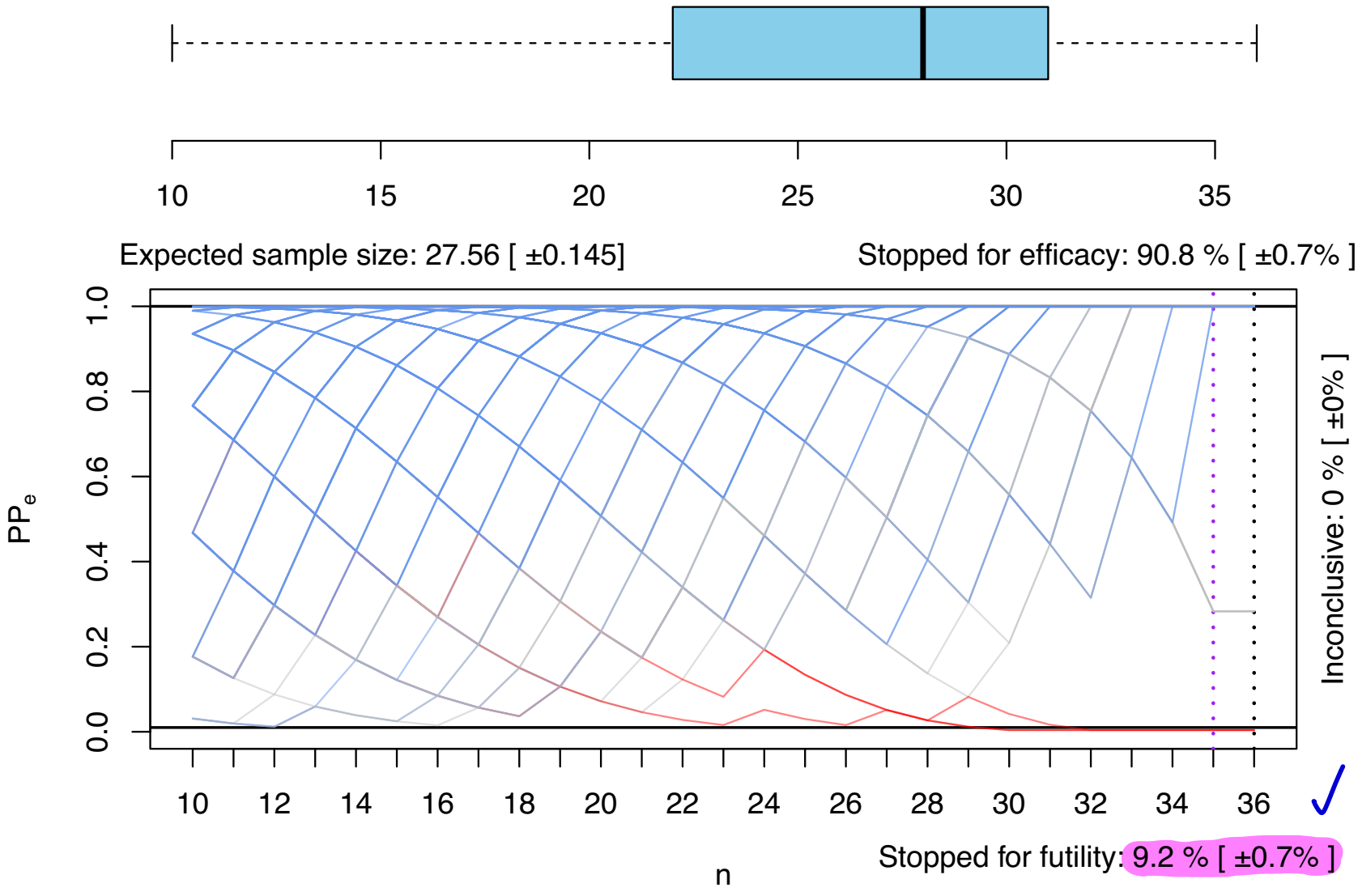
```r
ex_pev_updated2 = brada(a0 = 0.2, b0 = 0.8,
             Nmax = 36,
             batchsize = 1,
             nInit = 10,
             p_true = 0.2,
             p0 = 0.2,
             p1 = 0.2,
             theta_T = 0.8,
             theta_L = 0.01,
             theta_U = 1.0,
             nsim = 1500,
             method = "PPe",
             refFunc = "flat",
             nu = 1.3,
             cores = 6)
plot(ex_pev_updated2)
power_ex_pev_updated2 = power(ex_pev_updated2, p_true = 0.4,
    nsim = 1500, cores = 12)
plot(power_ex_pev_updated2)
```

We changed only $\theta_L = 0.1$ to $\theta_L = 0.01$

Operating characteristics under $H_0 : p \leq 0.2$:



Operating characteristics under $H_1 : p \geq 0.4$:



Comparison

Simon's two-stage design

α	β	$\mathbb{E}[N\|H_0]$	PET$[H_0]$	Power for $p \geq 0.4$
0.086	0.098	29.26	0.46	≥ 0.80

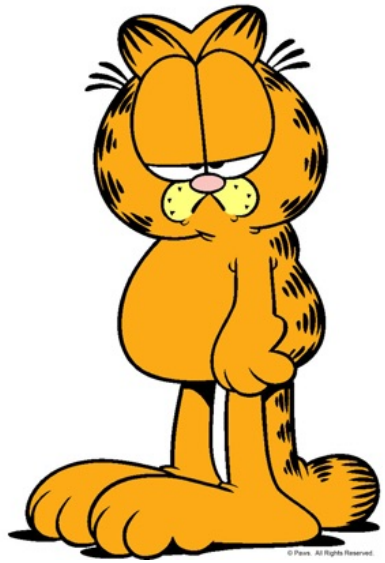
Calibrated PP design

α	β	$\mathbb{E}[N\|H_0]$	PET$[H_0]$	Power for $p \geq 0.4$
0.088	0.094	27.67	0.86	≥ 0.906

Calibrated PEV design

α	β	$\mathbb{E}[N\|H_0]$	PET$[H_0]$	Power for $p \geq 0.4$
0.099	0.099	23.98	0.91	≥ 0.908

What about runtimes...?



What about runtimes...?

Not as bad as you think! $\approx 10 - 15$ minutes to calibrate a design

Is such a design in accordance with FDA requirements?

"Finally, stopping or adaptation rules can be specified on a variety of different scales (...). ***The choice of scale is relatively unimportant as long as the operating characteristics of the designs are adequately evaluated.***"
(U.S. Food and Drug Administration Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, 2019)

Summary

Summary

The PEV design...

- is based on the theory of Bayesian evidence values and the predictive probability approach
- generalizes the predictive probability approach
- is more flexible due to the choice of the reference function $r(p)$ and evidence threshold $\nu \geq 0$
- Calibration: flat reference function, increase ν to decrease false-positive rate, decrease θ_L to decrease false-negative rate
- R package `brada` will be released shortly

Thanks for your attention!

Questions?

Further reading:

▶ Kelter, R. (2022). The Evidence Interval and the Bayesian Evidence Value - On a unified theory for Bayesian hypothesis testing and interval estimation. *British Journal of Mathematical and Statistical Psychology*, online first, `https://doi.org/10.1111/bmsp.12267`

▶ Kelter, R. (2021). fbst: An R package for the Full Bayesian Significance Test for testing a sharp null hypothesis against its alternative via the e value. *Behavior Research Methods*, 54, 11141130

▶ CRAN package **`fbst`**: https://cran.r-project.org/web/packages/fbst/index.html

Simulation study

Test $H_0 : p < 0.2$ against $H_1 : p > 0.4$, simulate data according to H_0 and under H_1 according to $p = 0.4$.

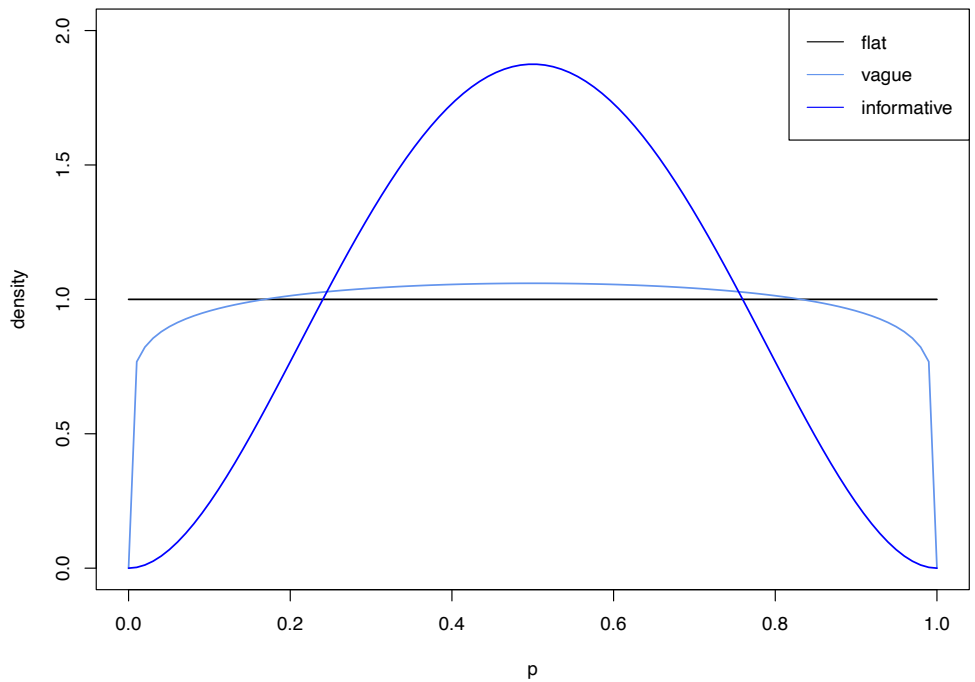
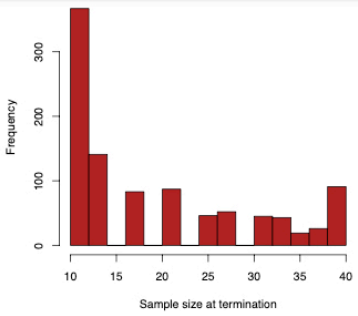


Figure 4: Prior distributions considered for the simulation study; flat $\mathcal{B}(1, 1)$ prior, vague $\mathcal{B}(1.1, 1.1)$ prior, and an informative $\mathcal{B}(3, 3)$ prior

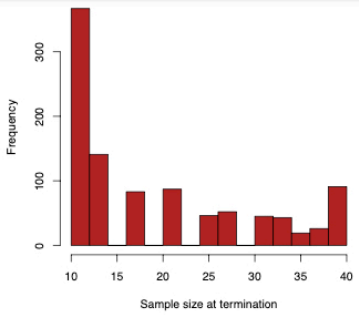
Prior	Ref. function	N_{max}	n	θ_T	θ_L	θ_U	PET	Var[PET]	$\mathbb{E}[N\|n,x]$	Var$[N\|n,x]$	α	β
Predictive evidence value (PEV) approach based on PP_e												
flat	flat	40	10	0.90	0.10	1.0	0.909	0.08	19.697	109.09	0.063	0.107
flat	vague	40	10	0.90	0.10	1.0	0.909	0.08	19.697	109.09	0.063	0.107
flat	informative	40	10	0.90	0.10	1.0	1	0	10.332	0.74	0	0.972
vague	flat	40	10	0.90	0.10	1.0	0.84	0.13	22.842	123.66	0.114	0.069
vague	vague	40	10	0.90	0.10	1.0	0.909	0.08	19.697	109.09	0.063	0.107
vague	informative	40	10	0.90	0.10	1.0	0.985	0.01	14.139	52.32	0.009	0.935
informative	flat	40	10	0.90	0.10	1.0	0.709	0.21	29.704	90.94	0.215	0.019
informative	vague	40	10	0.90	0.10	1.0	0.828	0.14	26.014	99.99	0.12	0.046
informative	informative	40	10	0.90	0.10	1.0	0.922	0.072	21.99	89.30	0.05	0.773
Predictive probability approach (PP) based on PP												
flat	-	40	10	0.90	0.10	1.0	0.909	0.083	19.697	109.09	0.063	0.107
vague	-	40	10	0.90	0.10	1.0	0.84	0.135	22.85	123.82	0.114	0.068
informative	-	40	10	0.90	0.10	1.0	0.706	0.208	29.76	90.36	0.216	0.018
Simon's Two-Stage design (Minimax and Optimal)												
Minimax		36	19				0.4551		28.3		0.0861	0.0976
Optimal		37	17				0.5489		26		0.0948	0.0967

Table 1: Simulation results for the phase IIA group-sequential Bayesian designs. N_{max} = maximum trial size, n = number of patients included in the first stage, θ_T = threshold for efficacy, θ_L and θ_U = thresholds for early stopping for futility or efficacy based on PP respectively PP_e, PET = probability of early termination, α and β = false-positive and false-negative rates; a flat prior or reference function is defined as $\mathcal{B}(1,1)$, a vague prior is $\mathcal{B}(1.1,1.1)$ and an informative prior is $\mathcal{B}(3,3)$.

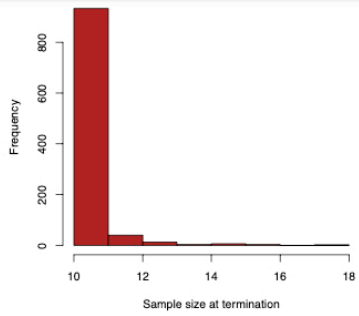
Results



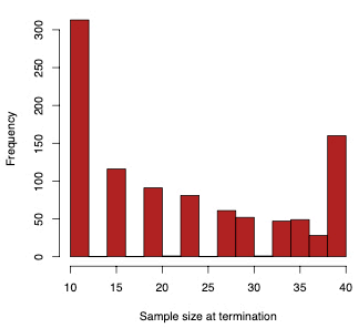
(a) PEV approach, flat prior, flat reference function



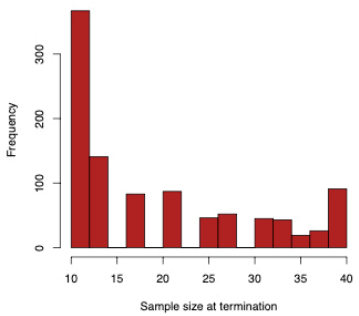
(b) PEV approach, flat prior, vague $B(1.1, 1.1)$ reference function



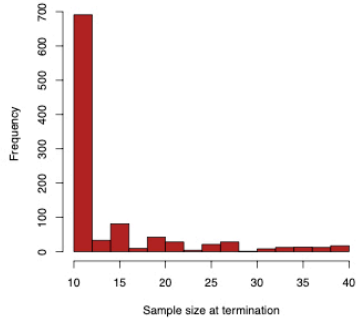
(c) PEV approach, flat prior, informative $B(3, 3)$ reference function



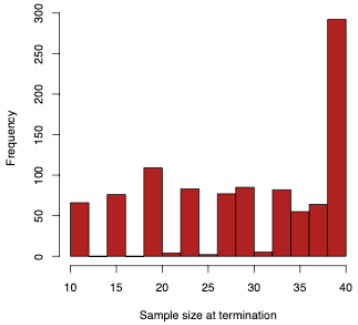
(d) PEV approach, vague $B(1.1, 1.1)$ prior, flat reference function

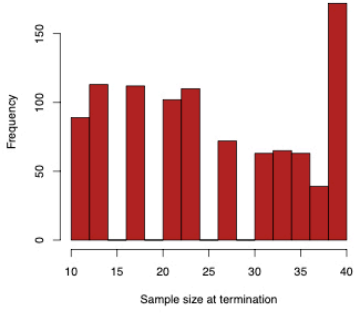


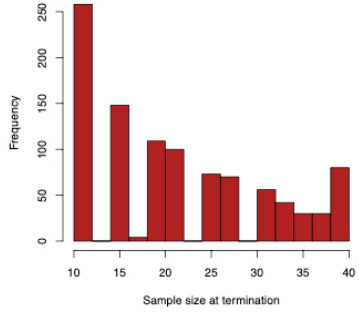
(e) PEV approach, vague $B(1.1, 1.1)$ prior, vague $B(1.1, 1.1)$ reference function



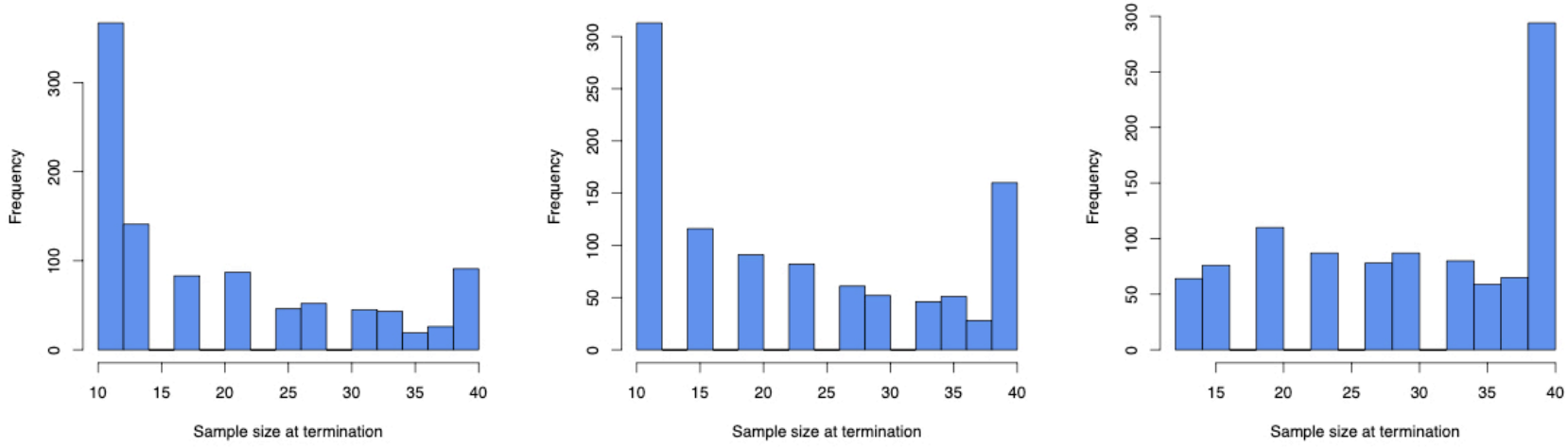
(f) PEV approach, vague $B(1.1, 1.1)$ prior, informative $B(3, 3)$ reference function







Results



(a) PP approach, flat prior

(b) PP approach, vague $\mathcal{B}(1.1, 1.1)$ prior

(c) PP approach, informative $\mathcal{B}(3, 3)$ prior

Figure 6: Sample sizes at termination for the PP group-sequential design under flat, vague and informative priors; a flat prior or reference function is $\mathcal{B}(1, 1)$, a vague prior or reference function is $\mathcal{B}(1.1, 1.1)$ and an informative prior or reference function is $\mathcal{B}(3, 3)$.

References

Arrowsmith, J. (2011, feb). Phase III and submission failures: 20072010. *Nature Reviews Drug Discovery 2011 10:2*, *10*(2), 87. Retrieved from https://www.nature.com/articles/nrd3375 doi: 10.1038/nrd3375

Biotechnology Innovation Organization. (2015). *Clinical Development Success Rates 2006-2015* (Tech. Rep.). Biotechnology Innovation Organization. Retrieved from https://www.bio.org/sites/default/files/legacy/bioorg/docs/ClinicalDevelopmentSuccessRates2006-2015-BIO,Biomedtracker,Amplion2016.pdf

Borges, W., & Stern, J. M. (2007). The Rules of Logic Composition for the Bayesian Epistemic e-Values. *Logic Journal of the IGPL*, *15*(5-6), 401–420. doi: 10.1093/jigpal/jzm032

Chen, X., Zhang, J., Jianga, L., & Yan, F. (2022, oct). Shotgun-2: A Bayesian phase I/II basket trial design to identify indication-specific optimal biological doses. *Statistical Methods in Medical Research*. Retrieved from `https://journals.sagepub.com/doi/10.1177/09622802221129049` doi: 10.1177/09622802221129049

Chuang-Stein, C., Kirby, S., Hirsch, I., & Atkinson, G. (2011, may). The role of the minimum clinically important difference and its impact on designing a trial. *Pharmaceutical statistics*, *10*(3), 250–256. Retrieved from `https://pubmed.ncbi.nlm.nih.gov/20936625/` doi: 10.1002/PST.459

Cook, J., Julious, S., Sones, W., Hampson, L., Hewitt, C., Berlin, J., ... Vale, L. (2018). DELTA 2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *Trials*, *19*(1), 1–6. doi: 10.1136/bmj.k3750

Diniz, M., Pereira, C. A. B., Polpo, A., Stern, J. M., & Wechsler, S. (2012). Relationship between Bayesian and frequentist significance indices. *International Journal for Uncertainty Quantification*, *2*(2), 161–172.

Guo, B., & Yuan, Y. (2017, apr). Bayesian Phase I/II Biomarker-Based Dose Finding for Precision Medicine With Molecularly Targeted Agents. *Journal of the American Statistical Association*, *112*(518), 508–520. doi: 10.1080/01621459.2016.1228534

Guo, B., & Zang, Y. (2020, dec). BILITE: A Bayesian randomized phase II design for immunotherapy by jointly modeling the longitudinal immune response and time-to-event efficacy. *Statistics in Medicine*, *39*(29), 4439–4451. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1002/sim.8733https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8733https://onlinelibrary.wiley.com/doi/10.1002/sim.8733 doi: 10.1002/SIM.8733

Kelter, R. (2021a). Bayesian Hodges-Lehmann tests for statistical equivalence in the two-sample setting: Power analysis, type I error rates and equivalence boundary selection in biomedical research. *BMC Medical Research Methodology*, *21*(171). doi: 10.1186/s12874-021-01341-7

Kelter, R. (2021b, sep). fbst: An R package for the Full Bayesian Significance Test for testing a sharp null hypothesis against its alternative via the e value. *Behavior Research Methods*(online first), 1–23. Retrieved from https://link.springer.com/10.3758/s13428-021-01613-6 doi: 10.3758/s13428-021-01613-6

Kelter, R. (2022). The Evidence Interval and the Bayesian Evidence Value - On a unified theory for Bayesian hypothesis testing and interval estimation. *British Journal of Mathematical and Statistical Psychology*(online first). Retrieved from https://doi.org/10.1111/bmsp.12267 doi: https://doi.org/10.1111/bmsp.12267

Lee, J. J., & Liu, D. D. (2008, mar). A predictive probability design for phase II cancer clinical trials. *Clinical Trials*, *5*(2), 93–106. Retrieved from https://journals.sagepub.com/doi/10.1177/1740774508089279 doi: 10.1177/1740774508089279

Leung, D. H. Y., & Wang, Y. G. (2001). A Bayesian decision approach for sample size determination in phase II trials. *Biometrics*, *57*(1), 309–312. doi: 10.1111/j.0006-341X.2001.00309.x

Liu, S., Guo, B., & Yuan, Y. (2018, jul). A Bayesian Phase I/II Trial Design for Immunotherapy. *https://doi.org/10.1080/01621459.2017.1383260*, *113*(523), 1016–1027. Retrieved from `https://www.tandfonline.com/doi/abs/10.1080/01621459.2017.1383260` doi: 10.1080/01621459.2017.1383260

Madruga, M. R., Esteves, L. G., & Wechsler, S. (2001). On the Bayesianity of Pereira-Stern tests. *Test*, *10*(2), 291–299. doi: 10.1007/BF02595698

Madruga, M. R., Pereira, C. A. d. B., & Stern, J. M. (2003, dec). Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference*, *117*(2), 185–198. doi: 10.1016/S0378-3758(02)00368-3

Matthews, J. N. (2006). *Introduction to Randomized Controlled Clinical Trials* (2nd Editio ed.). Boca Raton, FL: CRC Press.

Pereira, C. A. d. B., & Stern, J. M. (1999). Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy*, *1*(4), 99–110. doi: 10.3390/e1040099

Pereira, C. A. d. B., & Stern, J. M. (2020). The e-value: a fully Bayesian significance measure for precise statistical hypotheses and its research program. *São Paulo Journal of Mathematical Sciences*. doi: 10.1007/s40863-020-00171-7

Pereira, C. A. d. B., Stern, J. M., & Wechsler, S. (2008). Can a Significance Test be genuinely Bayesian? *Bayesian Analysis*, *3*(1), 79–100. doi: 10.1214/08-BA303

Stallard, N. (2012, may). Optimal sample sizes for phase II clinical trials and pilot studies. *Statistics in medicine*, *31*(11-12), 1031–1042. Retrieved from `https://pubmed.ncbi.nlm.nih.gov/22052407/` doi: 10.1002/SIM.4357

Tan, S. B., & Machin, D. (2002, jul). Bayesian two-stage designs for phase II clinical trials. *Statistics in Medicine*, *21*(14), 1991–2012. doi: 10.1002/SIM.1176

Thall, P. F., & Wathen, J. K. (2007, mar). Practical Bayesian Adaptive Randomization in Clinical Trials. *European journal of cancer (Oxford, England : 1990)*, *43*(5), 859. doi: 10.1016/J.EJCA.2007.01.006

U.S. Food and Drug Administration Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. (2019). Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry. *Web archive, accessed 01/02/2022*. Retrieved from `https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry`

Wu, J., Pan, H., & Hsu, C. W. (2021, nov). Bayesian single-arm phase II trial designs with time-to-event endpoints. *Pharmaceutical statistics*, *20*(6), 1235–1248. Retrieved from https://pubmed.ncbi.nlm.nih.gov/34085764/ doi: 10.1002/PST.2143

Yuan, J., Pang, H., Tong, T., Xi, D., Guo, W., & Mesenbrink, P. (2016, sep). Seamless Phase IIa/IIb and Enhanced Dose Finding Adaptive Design. *Journal of biopharmaceutical statistics*, *26*(5), 912. Retrieved from /pmc/articles/PMC5025390//pmc/articles/PMC5025390/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5025390/ doi: 10.1080/10543406.2015.1094807

Yuan, Y., & Yin, G. (2011, jan). Bayesian pase I/II adaptively randomized oncology trials with combined drugs. *The Annals of Applied Statistics*, *5*(2A), 924. Retrieved from /pmc/articles/PMC3286607//pmc/articles/PMC3286607/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3286607/ doi: 10.1214/10-AOAS433

Zhao, L., Taylor, J. M., & Schuetze, S. M. (2012, jul). Bayesian decision theoretic two-stage design in phase II clinical trials with survival endpoint. *Statistics in Medicine*, *31*(17), 1804–1820. doi: 10.1002/SIM.4511