Bayesian design using Beta-binomial model for single-arm clinical trials

Yalin Zhu

July 25, 2016

1 Description

The purpose of a phase I trial is to study the drug's toxicity in humans and to identify the 'best' dose to be used (this is usually the highest dose which does not result in excessive toxicity). Following this, a phase II trial using this 'best' dose is then conducted. The goal in such a trial is to assess the efficacy of the drug (often demonstrated by using tumour response as the indicator) to determine if it should be further tested in a large-scale randomized phase III trial.

Phase II clinical trials thus play an important role in the development and testing of a new drug. The main goal of a phase II trial is not to obtain a precise estimate of the response rate of the new drug, but rather to accept or reject the drug for further testing in a phase III trial. A commonly used primary endpoint in phase II cancer clinical trials is the clinical response to a treatment, which is a binary endpoint defined as the patient achieving complete or partial response within a predefined treatment course. In the early phase II development of new drugs, most trials are open label, single-arm studies, while late phase II trials tend to be multiarm, randomized studies.

Nevertheless, the analysis of the trial results typically include the obtaining of an estimate of the true response proportion, along with the associated 95% (frequentist) confidence interval. Such an analysis does not always answer the question of interest to the investigator. For example, the investigator might wish to know the probability that the true response proportion exceeds the prespecified target value, or he may wish to identify the (credible) interval that has a 95% probability of containing the true response proportion (this is not the same as the 95% frequentist confidence interval). Such questions can be answered using a Bayesian approach. With a Bayesian approach, we can obtain the posterior probability distribution of the true response proportion. This allows us to compute the probability that the response proportion falls within any prespecified region of interest, including the region above the target proportion. The bayesian credible interval is the interval that has a 95% probability of containing the true response proportion. A Bayesian design also allows for the formal incorporation of relevant information from other sources of evidence in the monitoring and analysis of the trial.

In this project, we use Beta-binomial conjugate to develop some bayesian deisgn methods. First of all, we explore what the Bayesian prior and posterior distribution looks like. Then we explore single-arm design methods using posterior probability and predictive probability. Some R functions and web applications are developed as well.

2 Prior Elicitation and Posterior Construction

A strength of Bayesian design and analysis is the ability to formally incorporate available information. But choosing a prior distribution requires careful consideration and work. This aspect of the Bayesian approach is more art than science. Several meetings between clinical investigators and statisticians may be necessary for assembling, evaluating, and quantifying the evidence based on literature or prior experience. In the process of selecting a prior distribution, the statistician should evaluate its sensitivity on the design's operating characteristics. Using a non-informative prior may be appropriate. Such a prior imitates a frequentist approach at the analysis stage but does not take existing information into consideration. And because a non-informative prior is artificial, it can lead to a poor design by overreacting to early results. When incorporating historical information into the prior, we almost always down-weight it in comparison to data collected in the actual trial, as described above.

For a phase II single-arm trial, suppose our goal is to evaluate the *response rate* p for a new drug by testing the hypothesis $H_0 : p \leq p_0$ versus $H_1 : p \geq p_1 = p_0 + \delta$, which implies $p > p_0$. We assume that the prior distribution of the response rate follows a Beta distribution,

$$p \sim Beta(a, b).$$

In the Bayesian methods, the p is regarded as a random variable (which is fixed parameter for frequentist). The quantity $\frac{a}{a+b}$ and $\frac{ab}{(a+b)^2(a+b+1)}$ gives the prior mean and variance, while the magnitude of $a+b$ indicates how informative the prior is. Since the quantities a and b can be considered as the numbers of effective prior **responses** and **non-responses**, respectively, $a+b$ can be thought of as a measure of prior precision: the larger this sum, the more informative the prior and the stronger the belief it contains.

2.1 Choose the prior distribution and parameters

Let us look into the beta-binomial bayesian frame. First of all, we need to select parameter of $Beta(a, b)$. There are several methods to choose the prior parameter in the exsiting literatures.

2.1.1 Simply choose a non-inormative prior.

For example, $Beta(1, 1)$ (equivalent to $U niform(0,1)), since this kind of prior provides very little information, it is also called vague prior.

We can look at how the posterior proabability changes with more patients enrolled (prior updated) under the $Beta(1, 1)$ prior.

2.1.2 Plot the prior, likelihood and posterior density functions

We develop an R function which not only plots the posterior tendency with updating previous posterior as a new prior, but also provide a full list of inference information (indlucing outcome cohort data, posterior mean, credible interval)

```r
source("animation_update.R")
```

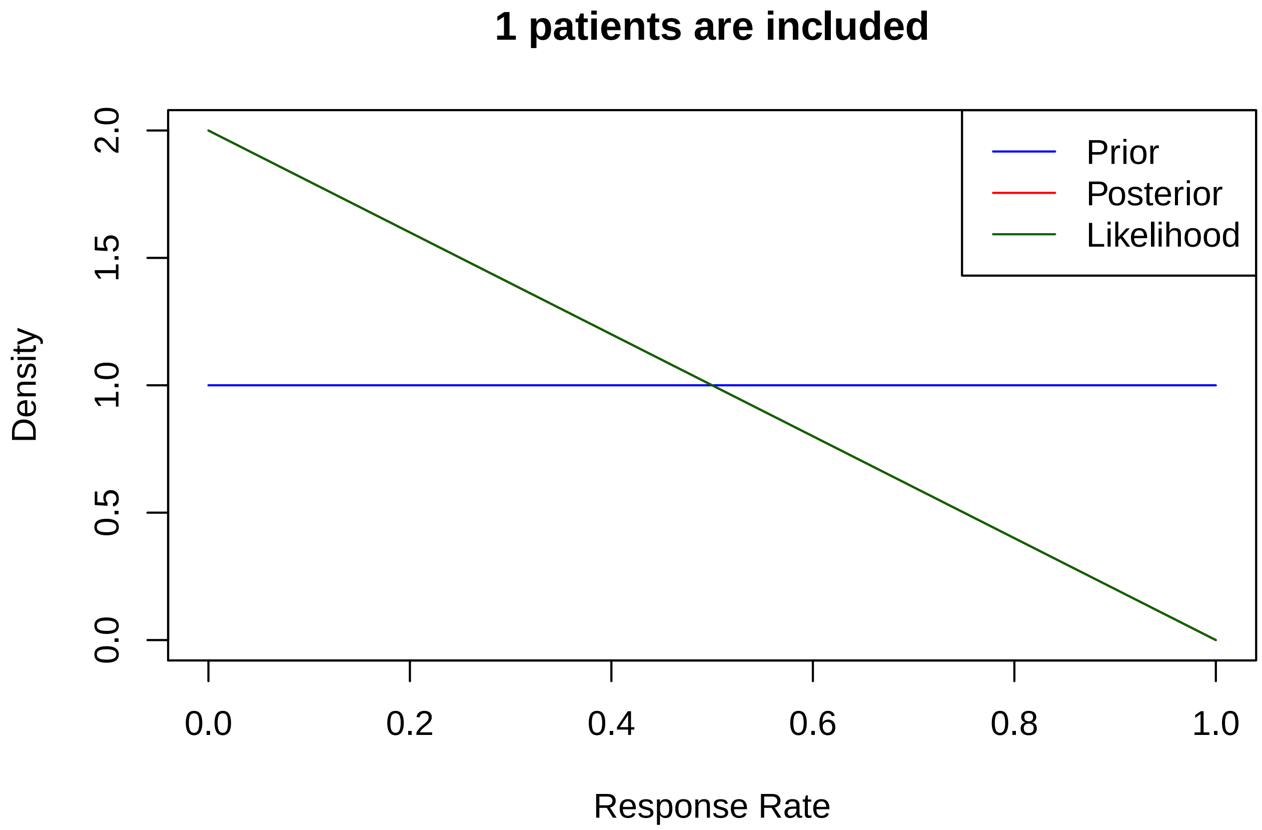
After CSCC patients receiving the cancer treatment neo-adjuvant therapy and surgery, consider the single primary endpoint: pathological complete response (pCR) with the following hypotheses:

$$H_0 : pCR \leq 15\% \quad versus \quad H_1 : pCR > 15\%$$

If the study sequtially monitors the prior and posteror, we can plot the distributions and likelihood as animations with the simulated data.

```r
BB.sim(M=20,N=1,p=0.15)
```

```
## Prior: Beta(1,1)
##
## ======== Cohort Number: 1 ========
## Observations -- Sample Size: 1(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 1(1)
##   Observed Response Rate: 0
## Posterior: Beta(1,2)
##   Posterior Mean: 0.333,  Difference between posterior and true response rate: 0.333
##   95% Credible Interval: (0.0126,0.842)

## ======== Cohort Number: 2 ========
## Observations -- Sample Size: 2(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 2(1)
##   Observed Response Rate: 0
## Posterior: Beta(1,3)
##   Posterior Mean: 0.25,  Difference between posterior and true response rate: 0.25
##   95% Credible Interval: (0.0084,0.708)

## ======== Cohort Number: 3 ========
## Observations -- Sample Size: 3(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 3(1)
##   Observed Response Rate: 0
## Posterior: Beta(1,4)
##   Posterior Mean: 0.2,  Difference between posterior and true response rate: 0.2
##   95% Credible Interval: (0.00631,0.602)

## ======== Cohort Number: 4 ========
## Observations -- Sample Size: 4(1)  ||  Number of Response: 1(1)  ||  Number of Failure: 3(0)
##   Observed Response Rate: 0.25
## Posterior: Beta(2,4)
##   Posterior Mean: 0.333,  Difference between posterior and true response rate: 0.0833
##   95% Credible Interval: (0.0527,0.716)

## ======== Cohort Number: 5 ========
## Observations -- Sample Size: 5(1)  ||  Number of Response: 2(1)  ||  Number of Failure: 3(0)
##   Observed Response Rate: 0.4
## Posterior: Beta(3,4)
##   Posterior Mean: 0.429,  Difference between posterior and true response rate: 0.0286
##   95% Credible Interval: (0.118,0.777)

## ======== Cohort Number: 6 ========
## Observations -- Sample Size: 6(1)  ||  Number of Response: 2(0)  ||  Number of Failure: 4(1)
##   Observed Response Rate: 0.333
## Posterior: Beta(3,5)
##   Posterior Mean: 0.375,  Difference between posterior and true response rate: 0.0417
##   95% Credible Interval: (0.099,0.71)

## ======== Cohort Number: 7 ========
## Observations -- Sample Size: 7(1)  ||  Number of Response: 2(0)  ||  Number of Failure: 5(1)
##   Observed Response Rate: 0.286
## Posterior: Beta(3,6)
##   Posterior Mean: 0.333,  Difference between posterior and true response rate: 0.0476
##   95% Credible Interval: (0.0852,0.651)

## ======== Cohort Number: 8 ========
## Observations -- Sample Size: 8(1)  ||  Number of Response: 3(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.375
## Posterior: Beta(4,6)
##   Posterior Mean: 0.4,  Difference between posterior and true response rate: 0.025
##   95% Credible Interval: (0.137,0.701)
## ======== Cohort Number: 9 ========
## Observations -- Sample Size: 9(1)  ||  Number of Response: 3(0)  ||  Number of Failure: 6(1)
##   Observed Response Rate: 0.333
## Posterior: Beta(4,7)
##   Posterior Mean: 0.364,  Difference between posterior and true response rate: 0.0303
##   95% Credible Interval: (0.122,0.652)

## ======== Cohort Number: 10 ========
## Observations -- Sample Size: 10(1)  ||  Number of Response: 3(0)  ||  Number of Failure: 7(1)
##   Observed Response Rate: 0.3
## Posterior: Beta(4,8)
##   Posterior Mean: 0.333,  Difference between posterior and true response rate: 0.0333
##   95% Credible Interval: (0.109,0.61)

## ======== Cohort Number: 11 ========
## Observations -- Sample Size: 11(1)  ||  Number of Response: 4(1)  ||  Number of Failure: 7(0)
##   Observed Response Rate: 0.364
## Posterior: Beta(5,8)
##   Posterior Mean: 0.385,  Difference between posterior and true response rate: 0.021
##   95% Credible Interval: (0.152,0.651)

## ======== Cohort Number: 12 ========
## Observations -- Sample Size: 12(1)  ||  Number of Response: 4(0)  ||  Number of Failure: 8(1)
##   Observed Response Rate: 0.333
## Posterior: Beta(5,9)
##   Posterior Mean: 0.357,  Difference between posterior and true response rate: 0.0238
##   95% Credible Interval: (0.139,0.614)

## ======== Cohort Number: 13 ========
## Observations -- Sample Size: 13(1)  ||  Number of Response: 4(0)  ||  Number of Failure: 9(1)
##   Observed Response Rate: 0.308
## Posterior: Beta(5,10)
##   Posterior Mean: 0.333,  Difference between posterior and true response rate: 0.0256
##   95% Credible Interval: (0.128,0.581)

## ======== Cohort Number: 14 ========
## Observations -- Sample Size: 14(1)  ||  Number of Response: 4(0)  ||  Number of Failure: 10(1)
##   Observed Response Rate: 0.286
## Posterior: Beta(5,11)
##   Posterior Mean: 0.312,  Difference between posterior and true response rate: 0.0268
##   95% Credible Interval: (0.118,0.551)

## ======== Cohort Number: 15 ========
## Observations -- Sample Size: 15(1)  ||  Number of Response: 4(0)  ||  Number of Failure: 11(1)
##   Observed Response Rate: 0.267
## Posterior: Beta(5,12)
##   Posterior Mean: 0.294,  Difference between posterior and true response rate: 0.0275
##   95% Credible Interval: (0.11,0.524)

## ======== Cohort Number: 16 ========
## Observations -- Sample Size: 16(1)  ||  Number of Response: 5(1)  ||  Number of Failure: 11(0)
##   Observed Response Rate: 0.312
## Posterior: Beta(6,12)
##   Posterior Mean: 0.333,  Difference between posterior and true response rate: 0.0208
##   95% Credible Interval: (0.142,0.56)

## ======== Cohort Number: 17 ========
## Observations -- Sample Size: 17(1)  ||  Number of Response: 5(0)  ||  Number of Failure: 12(1)
##   Observed Response Rate: 0.294
## Posterior: Beta(6,13)
##   Posterior Mean: 0.316,  Difference between posterior and true response rate: 0.0217
##   95% Credible Interval: (0.133,0.535)

## ======== Cohort Number: 18 ========
## Observations -- Sample Size: 18(1)  ||  Number of Response: 5(0)  ||  Number of Failure: 13(1)
##   Observed Response Rate: 0.278
## Posterior: Beta(6,14)
##   Posterior Mean: 0.3,  Difference between posterior and true response rate: 0.0222
##   95% Credible Interval: (0.126,0.512)

## ======== Cohort Number: 19 ========
## Observations -- Sample Size: 19(1)  ||  Number of Response: 5(0)  ||  Number of Failure: 14(1)
##   Observed Response Rate: 0.263
## Posterior: Beta(6,15)
##   Posterior Mean: 0.286,  Difference between posterior and true response rate: 0.0226
##   95% Credible Interval: (0.119,0.491)

## ======== Cohort Number: 20 ========
## Observations -- Sample Size: 20(1)  ||  Number of Response: 6(1)  ||  Number of Failure: 14(0)
##   Observed Response Rate: 0.3
## Posterior: Beta(7,15)
##   Posterior Mean: 0.318,  Difference between posterior and true response rate: 0.0182
##   95% Credible Interval: (0.146,0.522)
```

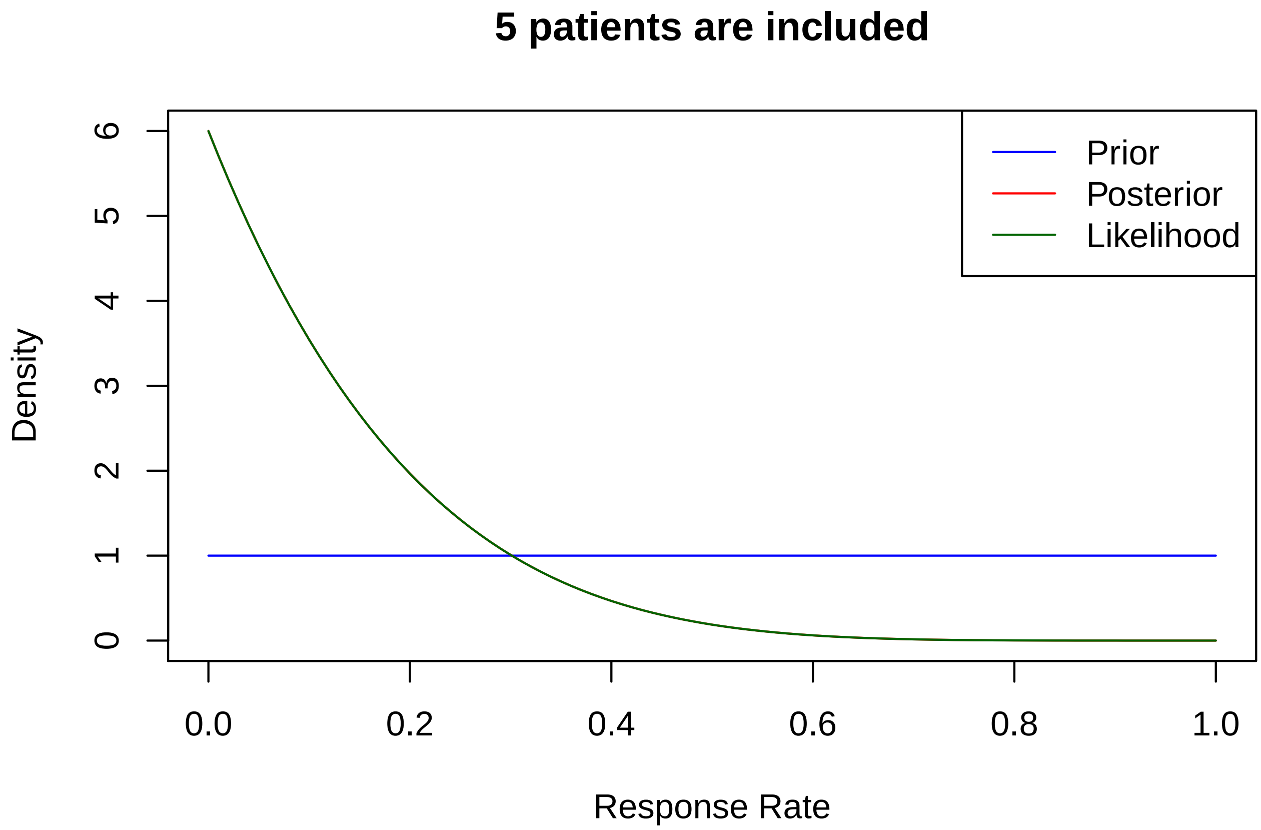


We can observe after 10 patients enrolled, the difference between posterior and true response rate reduced to a stable level below 3%.

We can also monitor the patients by cohort. Simulate each cohort contains 5 patients, the results are shown as follows:

```
BB.sim(M=10,N=5,p=0.15)
```

```
## Prior: Beta(1,1)
##
## ======== Cohort Number: 1 ========
## Observations -- Sample Size: 5(5)  ||  Number of Response: 0(0)  ||  Number of Failure: 5(5)
##   Observed Response Rate: 0
## Posterior: Beta(1,6)
##   Posterior Mean: 0.143,  Difference between posterior and true response rate: 0.143
##   95% Credible Interval: (0.00421,0.459)

## ======== Cohort Number: 2 ========
## Observations -- Sample Size: 10(5)  ||  Number of Response: 1(1)  ||  Number of Failure: 9(4)
##   Observed Response Rate: 0.1
## Posterior: Beta(2,10)
##   Posterior Mean: 0.167,  Difference between posterior and true response rate: 0.0667
##   95% Credible Interval: (0.0228,0.413)
## ======== Cohort Number: 3 ========
## Observations -- Sample Size: 15(5)  ||  Number of Response: 1(0)  ||  Number of Failure: 14(5)
##   Observed Response Rate: 0.0667
## Posterior: Beta(2,15)
##   Posterior Mean: 0.118,  Difference between posterior and true response rate: 0.051
##   95% Credible Interval: (0.0155,0.302)

## ======== Cohort Number: 4 ========
## Observations -- Sample Size: 20(5)  ||  Number of Response: 3(2)  ||  Number of Failure: 17(3)
##   Observed Response Rate: 0.15
## Posterior: Beta(4,18)
##   Posterior Mean: 0.182,  Difference between posterior and true response rate: 0.0318
##   95% Credible Interval: (0.0545,0.363)

## ======== Cohort Number: 5 ========
## Observations -- Sample Size: 25(5)  ||  Number of Response: 5(2)  ||  Number of Failure: 20(3)
##   Observed Response Rate: 0.2
## Posterior: Beta(6,21)
##   Posterior Mean: 0.222,  Difference between posterior and true response rate: 0.0222
##   95% Credible Interval: (0.0897,0.394)

## ======== Cohort Number: 6 ========
## Observations -- Sample Size: 30(5)  ||  Number of Response: 5(0)  ||  Number of Failure: 25(5)
##   Observed Response Rate: 0.167
## Posterior: Beta(6,26)
##   Posterior Mean: 0.188,  Difference between posterior and true response rate: 0.0208
##   95% Credible Interval: (0.0745,0.337)

## ======== Cohort Number: 7 ========
## Observations -- Sample Size: 35(5)  ||  Number of Response: 6(1)  ||  Number of Failure: 29(4)
##   Observed Response Rate: 0.171
## Posterior: Beta(7,30)
##   Posterior Mean: 0.189,  Difference between posterior and true response rate: 0.0178
##   95% Credible Interval: (0.0819,0.328)

## ======== Cohort Number: 8 ========
## Observations -- Sample Size: 40(5)  ||  Number of Response: 8(2)  ||  Number of Failure: 32(3)
##   Observed Response Rate: 0.2
## Posterior: Beta(9,33)
##   Posterior Mean: 0.214,  Difference between posterior and true response rate: 0.0143
##   95% Credible Interval: (0.106,0.349)

## ======== Cohort Number: 9 ========
## Observations -- Sample Size: 45(5)  ||  Number of Response: 9(1)  ||  Number of Failure: 36(4)
##   Observed Response Rate: 0.2
## Posterior: Beta(10,37)
##   Posterior Mean: 0.213,  Difference between posterior and true response rate: 0.0128
##   95% Credible Interval: (0.109,0.339)

## ======== Cohort Number: 10 ========
## Observations -- Sample Size: 50(5)  ||  Number of Response: 10(1)  ||  Number of Failure: 40(4)
##   Observed Response Rate: 0.2
## Posterior: Beta(11,41)
##   Posterior Mean: 0.212,  Difference between posterior and true response rate: 0.0115
##   95% Credible Interval: (0.113,0.331)
```



After only 4 cohort (20 patients) enrolled, the difference between posterior and true response rate reduced to a stable level below 3%.

2.1.3 Mean and Variance prior

Based on the mean μ and the variance σ^2, we can derive the prior parameter of $Beta(a, b)$ distribution with

$$a = \mu \left\{ \frac{\mu(1-\mu)}{\sigma^2} - 1 \right\}$$

and

$$b = (1-\mu) \left\{ \frac{\mu(1-\mu)}{\sigma^2} - 1 \right\}.$$

Notations

μ_0: mean of prior response probability;

σ_0: standard deviation of prior response probability;

n: total sample size;

x: number of response subjects;

For the mean and variance prior, We reproduce the arsenic trioxide trials example (Zohar, Teramukai, and Zhou 2008), with MM and APL data. The results are shown as follows.

```
source("simple_design.R")
```

```
## create MM data and set the prior mean and variance
MM.r = rep(0, 12)
MM.mean = 0.1
MM.var = 0.0225
post.mean(mu0 = MM.mean, sigma0 = sqrt(MM.var), r = MM.r)
```

```
## $para.a
## [1] 0.3
##
## $para.b
## [1] 2.7
##
## $`poterior mean`
##  [1] 0.0750 0.0600 0.0500 0.0429 0.0375 0.0333 0.0300 0.0273 0.0250 0.0231
## [11] 0.0214 0.0200
```

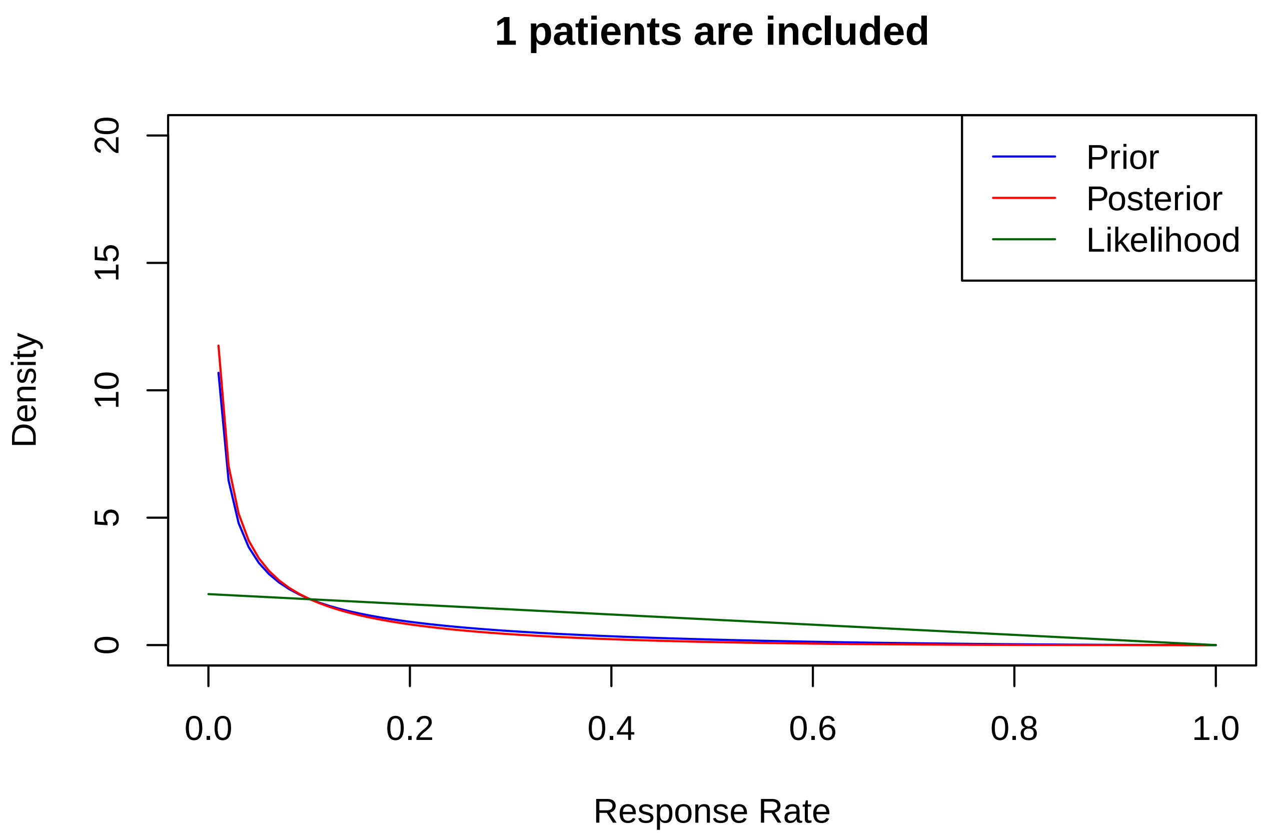
```
## create APL data and set the prior mean and variance
APL.r <- c(0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1)
APL.mean = 0.3
APL.var = 0.0191
post.mean(mu0 = APL.mean, sigma0 = sqrt(APL.var), r = APL.r)
```

```
## $para.a
## [1] 3
##
## $para.b
## [1] 7
##
## $`poterior mean`
##  [1] 0.273 0.333 0.308 0.286 0.333 0.375 0.412 0.444 0.421 0.450 0.476
## [12] 0.500 0.478 0.500 0.520 0.539 0.556 0.571 0.586 0.600
```

We also create R function to plot animations for the MM and APL data

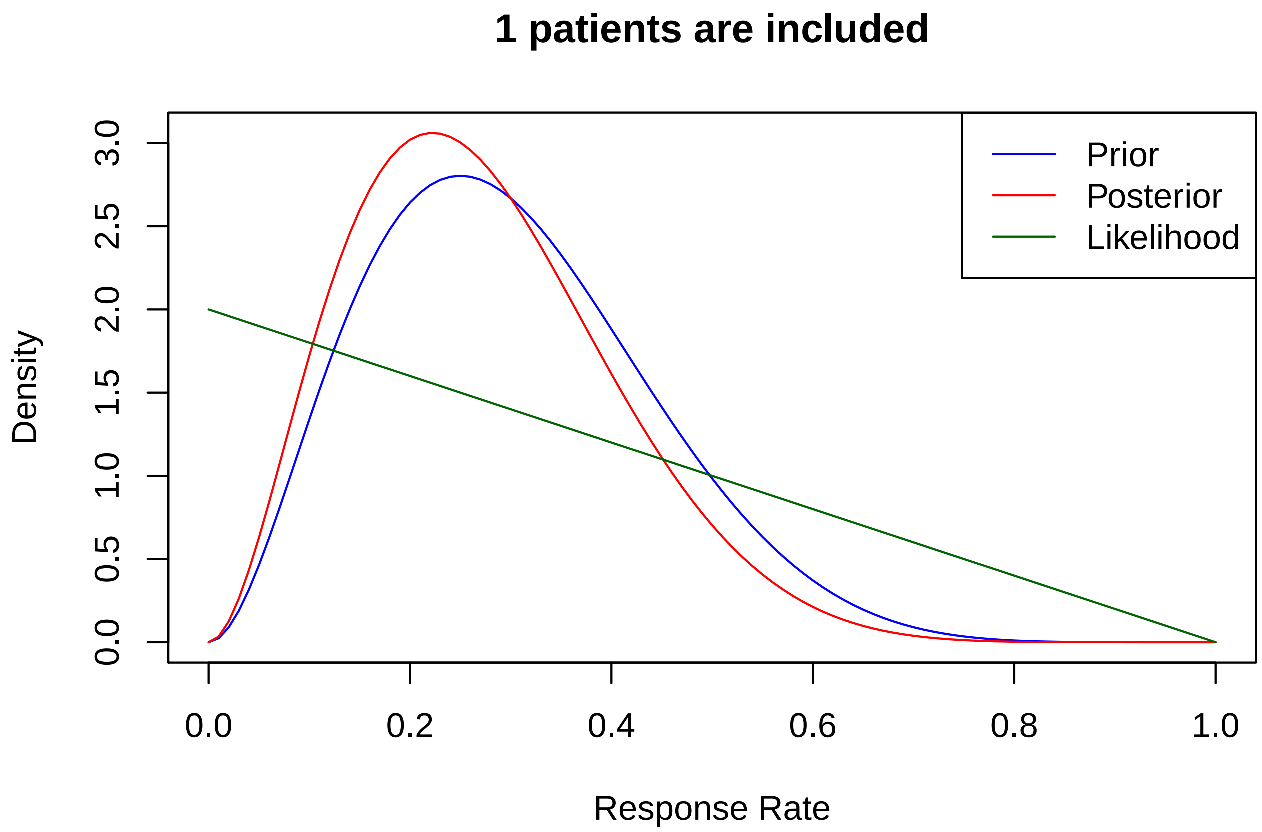
```
MM.r <- rep(0,12)
BB.plot(0.3,2.7,r=MM.r)
```

```
## Prior: Beta(0.3,2.7)
##
## ======== Cohort Number: 1 ========
## Observations -- Sample Size: 1(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 1(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,3.7)
##   Posterior Mean: 0.075
##   95% Credible Interval: (9.48e-07,0.43)
## ======== Cohort Number: 2 ========
## Observations -- Sample Size: 2(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 2(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,4.7)
##   Posterior Mean: 0.06
##   95% Credible Interval: (7.31e-07,0.353)

## ======== Cohort Number: 3 ========
## Observations -- Sample Size: 3(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 3(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,5.7)
##   Posterior Mean: 0.05
##   95% Credible Interval: (5.95e-07,0.298)

## ======== Cohort Number: 4 ========
## Observations -- Sample Size: 4(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 4(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,6.7)
##   Posterior Mean: 0.0429
##   95% Credible Interval: (5.01e-07,0.258)

## ======== Cohort Number: 5 ========
## Observations -- Sample Size: 5(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 5(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,7.7)
##   Posterior Mean: 0.0375
##   95% Credible Interval: (4.33e-07,0.227)

## ======== Cohort Number: 6 ========
## Observations -- Sample Size: 6(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 6(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,8.7)
##   Posterior Mean: 0.0333
##   95% Credible Interval: (3.81e-07,0.203)

## ======== Cohort Number: 7 ========
## Observations -- Sample Size: 7(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 7(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,9.7)
##   Posterior Mean: 0.03
##   95% Credible Interval: (3.41e-07,0.184)

## ======== Cohort Number: 8 ========
## Observations -- Sample Size: 8(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 8(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,10.7)
##   Posterior Mean: 0.0273
##   95% Credible Interval: (3.08e-07,0.168)

## ======== Cohort Number: 9 ========
## Observations -- Sample Size: 9(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 9(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,11.7)
##   Posterior Mean: 0.025
##   95% Credible Interval: (2.81e-07,0.154)

## ======== Cohort Number: 10 ========
## Observations -- Sample Size: 10(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 10(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,12.7)
##   Posterior Mean: 0.0231
##   95% Credible Interval: (2.58e-07,0.143)

## ======== Cohort Number: 11 ========
## Observations -- Sample Size: 11(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 11(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,13.7)
##   Posterior Mean: 0.0214
##   95% Credible Interval: (2.39e-07,0.133)

## ======== Cohort Number: 12 ========
## Observations -- Sample Size: 12(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 12(1)
##   Observed Response Rate: 0
## Posterior: Beta(0.3,14.7)
##   Posterior Mean: 0.02
##   95% Credible Interval: (2.22e-07,0.124)
```



```
APL.r <- c(0,1,0,0,1,1,1,1,0,1,1,1,0,1,1,1,1,1,1,1)
BB.plot(3,7,r=APL.r)
```

```
## Prior: Beta(3,7)
##
## ======== Cohort Number: 1 ========
## Observations -- Sample Size: 1(1)  ||  Number of Response: 0(0)  ||  Number of Failure: 1(1)
##   Observed Response Rate: 0
## Posterior: Beta(3,8)
##   Posterior Mean: 0.273
##   95% Credible Interval: (0.0667,0.556)

## ======== Cohort Number: 2 ========
## Observations -- Sample Size: 2(1)  ||  Number of Response: 1(1)  ||  Number of Failure: 1(0)
##   Observed Response Rate: 0.5
## Posterior: Beta(4,8)
##   Posterior Mean: 0.333
##   95% Credible Interval: (0.109,0.61)

## ======== Cohort Number: 3 ========
## Observations -- Sample Size: 3(1)  ||  Number of Response: 1(0)  ||  Number of Failure: 2(1)
##   Observed Response Rate: 0.333
## Posterior: Beta(4,9)
##   Posterior Mean: 0.308
##   95% Credible Interval: (0.0992,0.572)

## ======== Cohort Number: 4 ========
## Observations -- Sample Size: 4(1)  ||  Number of Response: 1(0)  ||  Number of Failure: 3(1)
##   Observed Response Rate: 0.25
## Posterior: Beta(4,10)
##   Posterior Mean: 0.286
##   95% Credible Interval: (0.0909,0.538)

## ======== Cohort Number: 5 ========
## Observations -- Sample Size: 5(1)  ||  Number of Response: 2(1)  ||  Number of Failure: 3(0)
##   Observed Response Rate: 0.4
## Posterior: Beta(5,10)
##   Posterior Mean: 0.333
##   95% Credible Interval: (0.128,0.581)

## ======== Cohort Number: 6 ========
## Observations -- Sample Size: 6(1)  ||  Number of Response: 3(1)  ||  Number of Failure: 3(0)
##   Observed Response Rate: 0.5
## Posterior: Beta(6,10)
##   Posterior Mean: 0.375
##   95% Credible Interval: (0.163,0.616)

## ======== Cohort Number: 7 ========
## Observations -- Sample Size: 7(1)  ||  Number of Response: 4(1)  ||  Number of Failure: 3(0)
##   Observed Response Rate: 0.571
## Posterior: Beta(7,10)
##   Posterior Mean: 0.412
##   95% Credible Interval: (0.198,0.646)

## ======== Cohort Number: 8 ========
## Observations -- Sample Size: 8(1)  ||  Number of Response: 5(1)  ||  Number of Failure: 3(0)
##   Observed Response Rate: 0.625
## Posterior: Beta(8,10)
##   Posterior Mean: 0.444
##   95% Credible Interval: (0.23,0.671)

## ======== Cohort Number: 9 ========
## Observations -- Sample Size: 9(1)  ||  Number of Response: 5(0)  ||  Number of Failure: 4(1)
##   Observed Response Rate: 0.556
## Posterior: Beta(8,11)
##   Posterior Mean: 0.421
##   95% Credible Interval: (0.215,0.643)

## ======== Cohort Number: 10 ========
## Observations -- Sample Size: 10(1)  ||  Number of Response: 6(1)  ||  Number of Failure: 4(0)
##   Observed Response Rate: 0.6
## Posterior: Beta(9,11)
##   Posterior Mean: 0.45
##   95% Credible Interval: (0.244,0.665)
## ======== Cohort Number: 11 ========
## Observations -- Sample Size: 11(1)  ||  Number of Response: 7(1)  ||  Number of Failure: 4(0)
##   Observed Response Rate: 0.636
## Posterior: Beta(10,11)
##   Posterior Mean: 0.476
##   95% Credible Interval: (0.272,0.685)

## ======== Cohort Number: 12 ========
## Observations -- Sample Size: 12(1)  ||  Number of Response: 8(1)  ||  Number of Failure: 4(0)
##   Observed Response Rate: 0.667
## Posterior: Beta(11,11)
##   Posterior Mean: 0.5
##   95% Credible Interval: (0.298,0.702)

## ======== Cohort Number: 13 ========
## Observations -- Sample Size: 13(1)  ||  Number of Response: 8(0)  ||  Number of Failure: 5(1)
##   Observed Response Rate: 0.615
## Posterior: Beta(11,12)
##   Posterior Mean: 0.478
##   95% Credible Interval: (0.282,0.678)

## ======== Cohort Number: 14 ========
## Observations -- Sample Size: 14(1)  ||  Number of Response: 9(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.643
## Posterior: Beta(12,12)
##   Posterior Mean: 0.5
##   95% Credible Interval: (0.306,0.694)

## ======== Cohort Number: 15 ========
## Observations -- Sample Size: 15(1)  ||  Number of Response: 10(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.667
## Posterior: Beta(13,12)
##   Posterior Mean: 0.52
##   95% Credible Interval: (0.328,0.709)

## ======== Cohort Number: 16 ========
## Observations -- Sample Size: 16(1)  ||  Number of Response: 11(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.688
## Posterior: Beta(14,12)
##   Posterior Mean: 0.538
##   95% Credible Interval: (0.349,0.722)

## ======== Cohort Number: 17 ========
## Observations -- Sample Size: 17(1)  ||  Number of Response: 12(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.706
## Posterior: Beta(15,12)
##   Posterior Mean: 0.556
##   95% Credible Interval: (0.369,0.734)

## ======== Cohort Number: 18 ========
## Observations -- Sample Size: 18(1)  ||  Number of Response: 13(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.722
## Posterior: Beta(16,12)
##   Posterior Mean: 0.571
##   95% Credible Interval: (0.388,0.745)

## ======== Cohort Number: 19 ========
## Observations -- Sample Size: 19(1)  ||  Number of Response: 14(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.737
## Posterior: Beta(17,12)
##   Posterior Mean: 0.586
##   95% Credible Interval: (0.406,0.755)

## ======== Cohort Number: 20 ========
## Observations -- Sample Size: 20(1)  ||  Number of Response: 15(1)  ||  Number of Failure: 5(0)
##   Observed Response Rate: 0.75
## Posterior: Beta(18,12)
##   Posterior Mean: 0.6
##   95% Credible Interval: (0.423,0.765)
```

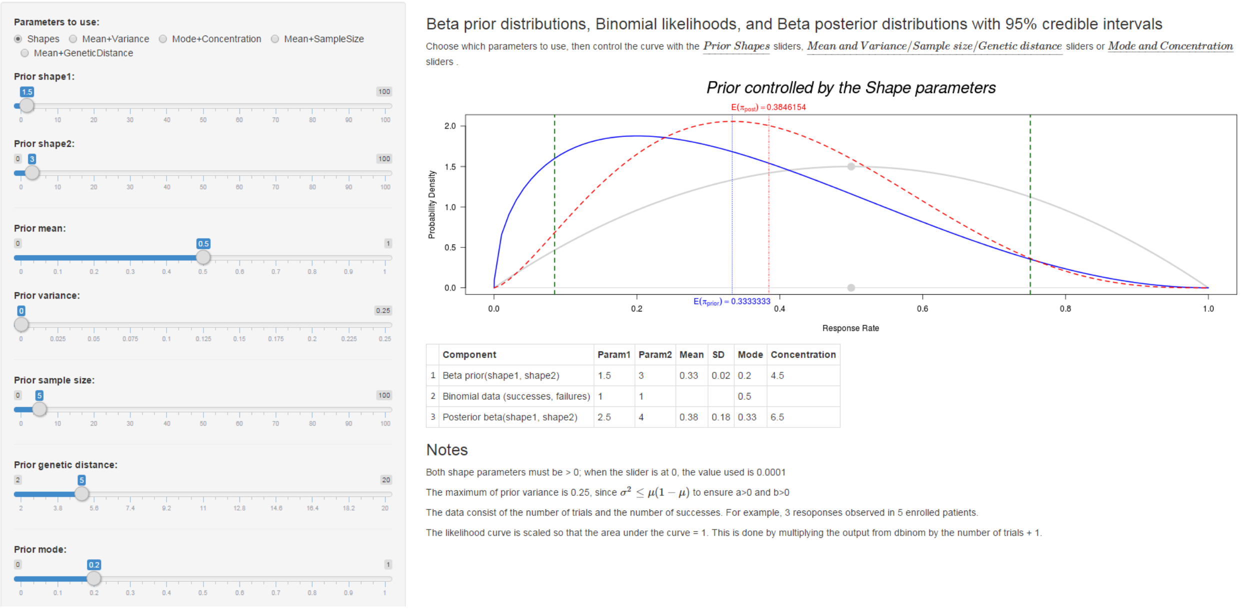


The results is identical with original paper’s results.

2.2 A Web application for choosing prior and simulating posterior distributions

For users' convenience, we also develop a web application to provide the prior information and posterior distributions inferences. Please refer to the following link: https://allen.shinyapps.io/Beta_Bayes_Prior/

The interface is display in the Figure 1:



Parameters to use:

● Shapes ○ Mean+Variance ○ Mode+Concentration ○ Mean+SampleSize ○ Mean+GeneticDistance

Prior shape1: 1.5

Prior shape2: 3

Prior mean: 0.5

Prior variance: 0

Prior sample size: 5

Prior genetic distance: 5

Prior mode: 0.2

Beta prior distributions, Binomial likelihoods, and Beta posterior distributions with 95% credible intervals

Choose which parameters to use, then control the curve with the *Prior Shapes* sliders, *Mean and Variance/Sample size/Genetic distance* sliders or *Mode and Concentration* sliders .

Prior controlled by the Shape parameters
$E(\pi_{post}) = 0.3846154$
$E(\pi_{prior}) = 0.3333333$
Probability Density
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Response Rate

	Component	Param1	Param2	Mean	SD	Mode	Concentration
1	Beta prior(shape1, shape2)	1.5	3	0.33	0.02	0.2	4.5
2	Binomial data (successes, failures)	1	1			0.5	
3	Posterior beta(shape1, shape2)	2.5	4	0.38	0.18	0.33	6.5

Notes

Both shape parameters must be > 0; when the slider is at 0, the value used is 0.0001

The maximum of prior variance is 0.25, since $\sigma^2 \leq \mu(1-\mu)$ to ensure a>0 and b>0

The data consist of the number of trials and the number of successes. For example, 3 resoponses observed in 5 enrolled patients.

The likelihood curve is scaled so that the area under the curve = 1. This is done by multiplying the output from dbinom by the number of trials + 1.

Figure 1: Display of the web application.

Remark Credible intervals are not unique on a posterior distribution. There are two main methods to define a suitble credible interval:

- **Equal-tailed (ET)** interval is the interval where the probability of being below the interval is as likely as being above it. This interval will include the median.
- **Highest Posterior Density (HPD)** interval is the narrowest interval, which for a unimodal distribution will involve choosing those values of highest probability density including the mode.

How to choose the credible interval is also our interest.

3 Single-arm Design Studies

We mainly want to determin the following quantities:

- Sample Size Determination (N_{max})
- Stopping Boundary (θ_L, θ_U)

We consider the following two bayesian methods: Posterior Probability (abbr. "PostP") and Predictive Probability (abbr. "PredP"). For Phase IIA (earlier) trials, we prefer to allow early stopping due to futility, but not due to efficacy, and final stop due to efficacy or pre-spicified N_{max}.

3.1 Sequential Stopping based on Posterior Probability design (`PostP`)

A simple and practical method is to use the posterior probability to monitor the trials. Dring each trial, the data are monitored continuously, and decisions are made adaptively until the pre-specified maximum sample size N_{max} is reached.(Thall and Simon 1994)

Suppose we have decieded a maximum number of accrued patients N_{max}, and assume that the number of responses X among the current n patients ($n \leq N_{max}$) follows a $Binomial(n, p)$ distribution. By the conjugacy of the beta prior and binomial likelihood, the posterior distribution of the response rate $p|X = x$ is still a beta distribution,

$$p|x \sim Beta(a + x, b + n - x).$$

Then the posterior probability
$PostP = Pr(p > p_0|x)$ can be used to decide the sample size and stopping boundary.

3.1.1 Algorithm 1

- **Step 1:** Specified the upper and lower probability cutoffs θ_U and θ_L. Typically, $\theta_U \in [0.9, 1]$ and $\theta_L \in [0, 0.05]$, set true null response rate p_0 a pre-specified value.
- **Step 2:** Let
$$S_U = \min\{x \in \mathbb{N} : PostP > \theta_U\}$$
and
$$S_L = \max\{x \in \mathbb{N} : PostP < \theta_L\}$$
be the upper and lower decision boudries based on the number of observed responses.
- **Step 3:** Make decisions after observing another x responses out of n patients:
  - If $x \geq S_U$, then stop the trial for efficacy; (could be ignored for futility only)
  - if $x \leq S_L$, then stop the trial for futility;
  - otherwise, continue the trial until N_{max} reached.

Although the stopping rule requires that the trial be terminated to declare the experimental drug promising if $x \geq S_U$, investigators rarely stop the trial in such a case so that more patients are allowed to benefit from the "good" drug. Therefore, the stopping rule for superiority of the drug is often not implemented in a single-arm phase II trial.

3.1.2 the cancer treatment neo-adjuvant therapy and surgery with single pCR endpoint using `PostP` method

First of all, we apply Simon's two-stage design:

```r
library(clinfun)
ph2simon(pu=0.15, pa=0.30, ep1=0.05, ep2=0.10, nmax=100)
```

```
##
##  Simon 2-stage Phase II design
##
## Unacceptable response rate:  0.15
## Desirable response rate:  0.3
## Error rates: alpha =  0.05 ; beta =  0.1
##
```

```
##         r1 n1  r  n EN(p0) PET(p0)
## Optimal  5 30 17 82  45.05  0.7106
## Minimax  6 42 14 64  51.80  0.5545
```

Then we apply the PostP bayesian design by using the vague prior $Beta(1,1)$. (In the future, we can use informative priors)

```
source("postp.R")
PostP.design(type = "futility", nmax=100, a=1, b=1, p0=0.15, delta=0.15, theta=0.05)
```

```
##     n bound
## 1   1    NA
## 8   8     0
## 13 13     1
## 18 18     2
## 23 23     3
## 27 27     4
## 32 32     5
## 36 36     6
## 40 40     7
## 44 44     8
## 48 48     9
## 52 52    10
## 56 56    11
## 60 60    12
## 64 64    13
## 68 68    14
## 72 72    15
## 76 76    16
## 80 80    17
## 84 84    18
## 88 88    19
## 92 92    20
## 95 95    21
## 99 99    22
```

```
PostP.design(type = "efficacy", nmax=100, a=1, b=1, p0=0.15, delta=0.15, theta=0.9)
```

```
##     n bound
## 1   1     1
## 3   3     2
## 7   7     3
## 12 12     4
## 17 17     5
## 22 22     6
## 27 27     7
## 32 32     8
## 37 37     9
## 42 42    10
## 48 48    11
## 53 53    12
## 59 59    13
## 64 64    14
## 70 70    15
## 76 76    16
## 81 81    17
## 87 87    18
## 93 93    19
## 99 99    20
```

Theoretically, We can choose any one row as the early stopping. Here we can compare the Simon's Optimal design result, and select r1/n1=5/32 as the stopping rule for futility, r/n=17/81 as final stop for efficacy. (or r1/n1/r/n = 6/36/14/64 sompared with Simon's Minimax).

We can also use Jeffrey prior: $Beta(0.5, 0.5)$.

```
PostP.design(type = "futility", nmax=100, a=0.5, b=0.5, p0=0.15, delta=0.15, theta=0.05)
```

```
##     n bound
## 1   1    NA
## 6   6     0
## 12 12     1
## 17 17     2
## 22 22     3
## 26 26     4
## 30 30     5
## 35 35     6
## 39 39     7
## 43 43     8
## 47 47     9
## 51 51    10
## 55 55    11
## 59 59    12
## 63 63    13
## 67 67    14
## 71 71    15
## 75 75    16
## 79 79    17
## 83 83    18
## 87 87    19
## 91 91    20
## 94 94    21
## 98 98    22
```

```
PostP.design(type = "efficacy", nmax=100, a=0.5, b=0.5, p0=0.15, delta=0.15, theta=0.9)
```

```
##     n bound
## 1   1     1
## 3   3     2
## 6   6     3
## 11 11     4
## 15 15     5
## 20 20     6
## 25 25     7
## 30 30     8
## 35 35      9
## 41 41     10
## 46 46     11
## 52 52     12
## 57 57     13
## 63 63     14
## 68 68     15
## 74 74     16
## 80 80     17
## 85 85     18
## 91 91     19
## 97 97     20
```

3.2 Predictive Probability Design (`PredP`)

The predictive probability approach looks into the future objectives based on the current observed data to project whether a positive conclusion at the end of study is likely or not, and then makes a sensible decision at the present time accordingly.(Lee and Liu 2008)

Let Y be the number of responses in the rest of $m = N_{max} - n$ future patients. Suppose our design is to declare efficacy if the posterior probability of p exceeding some pre-specified level p_0 is greater than some threshold θ_T . Marginalizing p out of the binomial likelihood, it is well known that Y|X=x follows a beta-binomial distribution, i.e. $Y|x \sim Beta - Binomial(m, a + x, b + n - x)$.

By the end of the trial, suppose we observe additional $Y = y$ response, then the posterior distribution including future future y patient $p|(X = x, Y = y)$ is also

$$Beta(a + x + y; b + N_{max} - x - y)$$

. The predictive probability (PredP) of trial success can then be calculated as follows. Denote the posterior probability with the future data by $B_y = Pr(p > p0|x, Y = y)$ and $I_y = I(B_y > \theta_T)$, then we have

$$\begin{aligned}
PredP &= Pr_{Y|x}\{Pr(p > p_0|x, Y) \geq \theta_T\} \\
&= E\{I[Pr(p > p_0|x, Y) \geq \theta_T]\Big|x\} \\
&= \sum_{y=0}^{m} I[Pr(p > p_0|x, Y = y) \geq \theta_T] \times Pr(Y = y|x) \\
&= \sum_{y=0}^{N_{max}-n} I_y \times Pr(Y = y|x).
\end{aligned}$$

Note that if there were no indicator function in 3.2, the PredP simply reduces to the PostP after averaging out the unobserved Y.

$$\sum_{y=0}^{N_{max}-n} Pr(p > p_0|x, Y = y) \times Pr(Y = y|x) = Pr(p > p_0|x)$$

Then we can decide the sample size and stop boundary by using the following algorithm:

3.2.1 Algorithm 2

- **Step 1:** Specified the upper and lower probability cutoffs θ_U and θ_L, typically, $\theta_U \in [0.9, 1]$ and $\theta_L \in [0, 0.05]$. Specified cutoff θ_T for the future y patients, typically, $\theta_T \in [0.8, 1]$. Set true null response rate p_0 a pre-specified value.

- **Step 2:** Given x obwervations, let

  $$S_U = \min\{x + y \in \mathbb{N} : PredP > \theta_U\}$$

  and

  $$S_L = \max\{x + y \in \mathbb{N} : PredP < \theta_L\}$$

  be the upper and lower decision boudries based on the number of observed responses.
- **Step 3:** Make decisions after observing another x responses out of n patients:
  - If $x \geq S_U$, then stop the trial for efficacy (could be ignored for futility only);
  - if $x \leq S_L$, then stop the trial for futility;
  - otherwise, continue the trial until N_{max} reached.

3.2.2 the cancer treatment neo-adjuvant therapy and surgery with single pCR endpoint using `PostP` method

Now we can apply the PostP bayesian design to the the cancer treatment pCR case, still use the vague prior $Beta(1, 1)$.

```
source("predp.R")
PredP.design(type = "futility", nmax=100, a=1, b=1, p0=0.15, delta=0.15, theta=0.05)
```

```
##       n bound
## 1      1    NA
## 6      6     0
## 10    10     1
## 14    14     2
## 18    18     3
## 21    21     4
## 24    24     5
## 28    28     6
## 31    31     7
## 34    34     8
## 37    37     9
## 40    40    10
## 43    43    11
## 46    46    12
## 48    48    13
## 51    51    14
## 54    54    15
## 57    57    16
## 60    60    17
## 62    62    18
## 65    65    19
## 67    67    20
## 70    70    21
## 73    73    22
## 75    75    23
## 78    78    24
## 80    80    25
## 82    82    26
## 85    85    27
## 87    87    28
## 89   89    29
## 92   92    30
## 94   94    31
## 96   96    32
## 97   97    33
## 99   99    34
## 100 100    35
```

```
PredP.design(type = "efficacy", nmax=100, a=1, b=1, p0=0.15, delta=0.15, theta=0.9)
```

```
##       n bound
## 1    1      1
## 3    3      2
## 6    6      3
## 9    9      4
## 13  13      5
## 17  17      6
## 21  21      7
## 26  26      8
## 30  30      9
## 35  35     10
## 40  40     11
## 45  45     12
## 50  50     13
## 55  55     14
## 60  60     15
## 66  66     16
## 71  71     17
## 77  77     18
## 83  83     19
## 91  91     20
```

Based on the Simon's Optimal design result, we can select r1/n1=5/24 as the stopping rule for futility, r/n=17/71 as final stop for efficacy. (or r1/n1/r/n = 6/28/14/55 sompared with Simon's Minimax). Compared with Simon's design and PostP design, under the same stopping boudaries, PredP design allows smaller sample sizes.

Using Jeffrey prior $Beta(0.5, 0.5)$ the results shown as follws:

```
PredP.design(type = "futility", nmax=100, a=0.5, b=0.5, p0=0.15, delta=0.15, theta=0.05)
```

```
##         n bound
## 1       1    NA
## 4       4     0
## 9       9     1
## 13     13     2
## 17     17     3
## 20     20     4
## 24     24     5
## 27     27     6
## 30     30     7
## 33     33     8
## 36     36     9
## 39   39     10
## 42   42     11
## 45   45     12
## 48   48     13
## 51   51     14
## 54   54     15
## 57   57     16
## 59   59     17
## 62   62     18
## 65   65     19
## 67   67     20
## 70   70     21
## 72   72     22
## 75   75     23
## 78   78     24
## 80   80     25
## 82   82     26
## 85   85     27
## 87   87     28
## 89   89     29
## 91   91     30
## 94   94     31
## 96   96     32
## 97   97     33
## 99   99     34
## 100 100     35
```

```
PredP.design(type = "efficacy", nmax=100, a=0.5, b=0.5, p0=0.15, delta=0.15, theta=0.9)
```

```
##      n bound
## 1    1     1
## 2    2     2
## 5    5     3
## 9    9     4
## 12  12     5
## 16  16     6
## 21  21     7
## 25  25     8
## 30  30     9
## 34  34    10
## 39  39    11
## 44  44    12
## 49  49    13
## 54  54    14
## 60  60    15
## 65  65    16
## 71  71    17
## 77  77    18
## 83  83    19
## 90  90    20
```

3.2.3 MM & APL data examples for PostP and PredP design

We also used the MM & APL data to illustrate two bayesian methods by animations.

Suppose we want to monitor the patients one-by-one, that is, the patient outcome follows $Bernoulli(p)$, and the stopping boundary can be deciede based on updating posteriors. We can create the animation plots to search the stopping boundary by taking the MM and APL data examples.

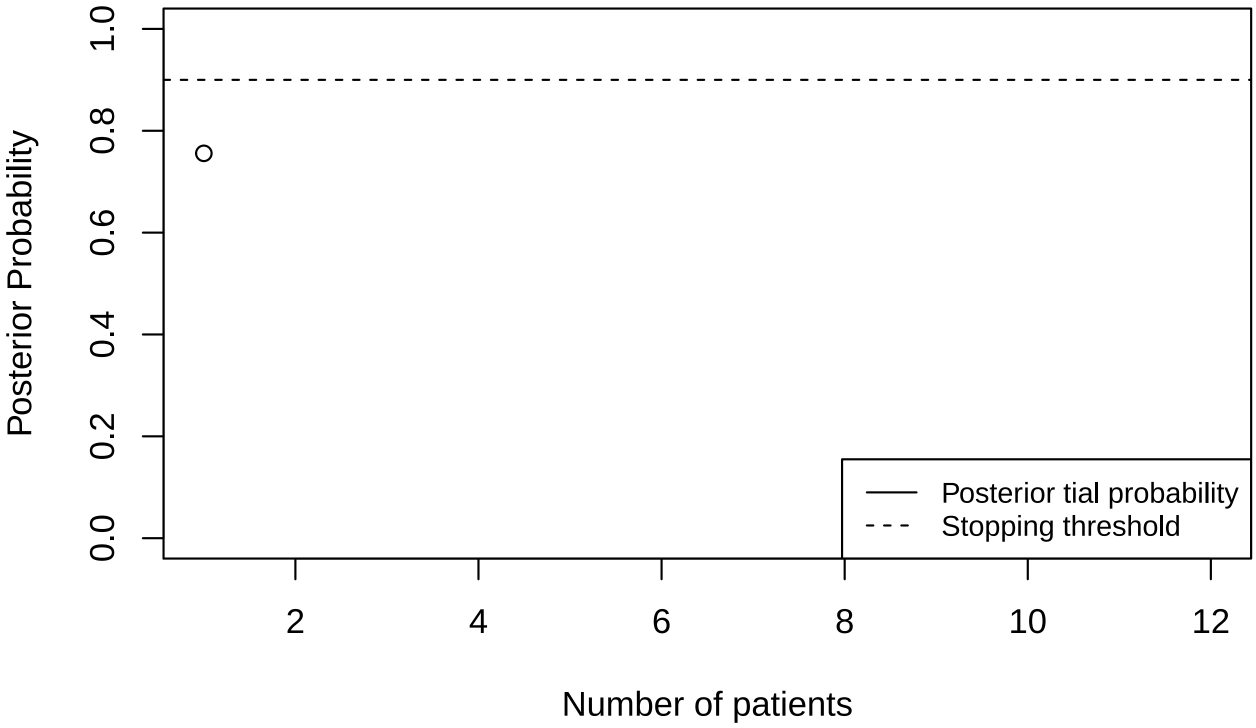
```
## Test MM data from the above examples
bayes.desgin(mu0 = MM.mean, sigma = sqrt(MM.var), r = MM.r, stop.rule = "futility",
    p0 = 0.1, ymax = 18)
```

```
## $para.a
## [1] 0.3
##
## $para.b
## [1] 2.7
##
## $`poterior mean`
##  [1] 0.0750 0.0600 0.0500 0.0429 0.0375 0.0333 0.0300 0.0273 0.0250 0.0231
## [11] 0.0214 0.0200
```

```
## Loading required package: stats4
```

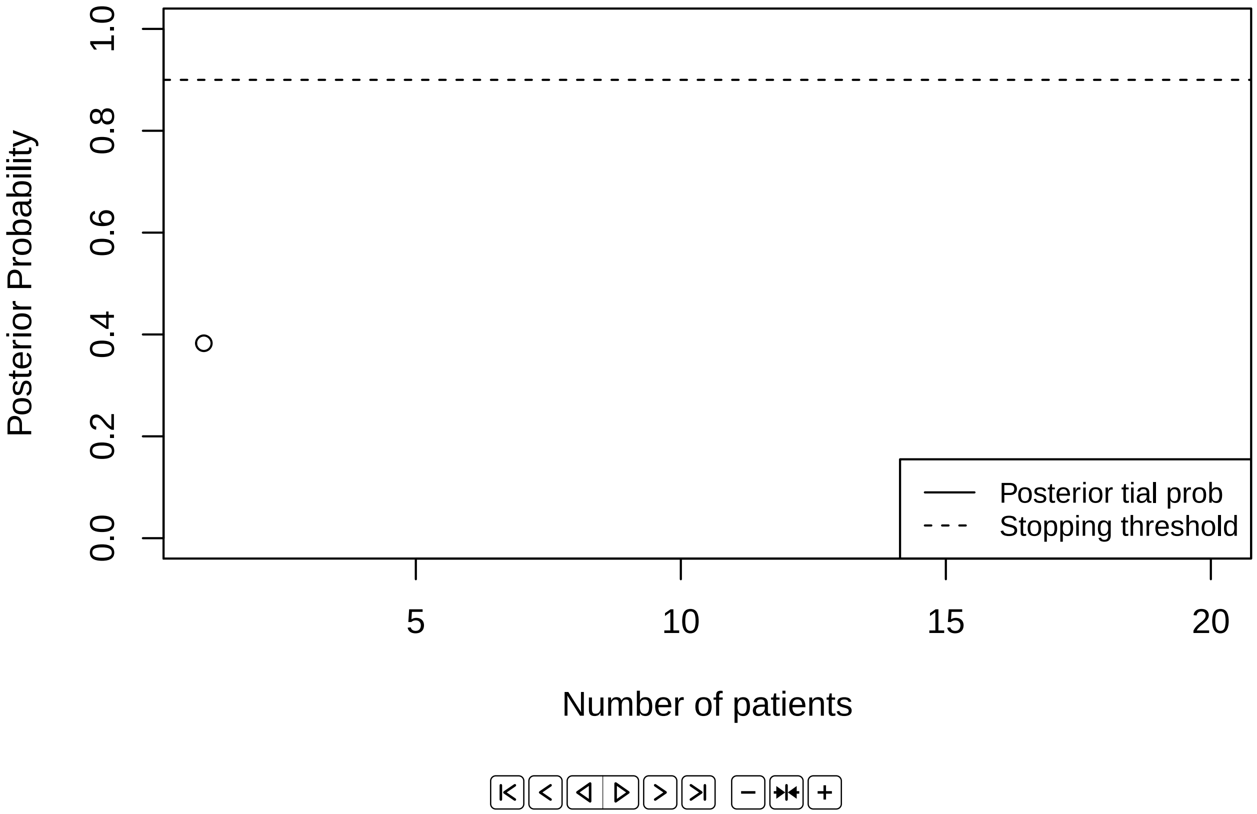
```
## Loading required package: splines
```

```
## [1] "Stop the trial for futility after the inclusion of 7 patients."
```



```
## Test APL data from the above examples
bayes.desgin(mu0 = APL.mean, sigma = sqrt(APL.var), r = APL.r, stop.rule = "efficacy",
    p0 = 0.1, ymax = 4.5)
```

```
## $para.a
## [1] 3
##
## $para.b
## [1] 7
##
## $`poterior mean`
##  [1] 0.273 0.333 0.308 0.286 0.333 0.375 0.412 0.444 0.421 0.450 0.476
## [12] 0.500 0.478 0.500 0.520 0.539 0.556 0.571 0.586 0.600
##
## [1] "Stop the trial for efficacy after the inclusion of 10 patients."
```



4 Summary and Future Works

The frameworks of both *PostP* and *PredP* methods allow the researcher to monitor the trial continously or by any cohort size. Compared to Simon's minimax and optimal two-stage design, the PostP and PredP designs monitor the data more frequently, both two designs have a larger probability of early termination and a smaller expected sample size in the null case than Simon's. All designs have the same maximum sample size with controlled Type I and Type II error rates.

We can consider *pCR* not only binary level, but also multinomial level, then we can use **Direchlet-Multinomial** prior and similar idea to develop the single-arm design (Thall, Simon, and Estey 1995). Another potential work is to extend the single primary endpoint to co-primary endpoint by using similar Bayesian idea (or hierarchical Bayesian design if the endpoint have some hierarchical structures)

References

Lee, J Jack, and Diane D Liu. 2008. "A Predictive Probability Design for Phase II Cancer Clinical Trials." *Clinical Trials* 5 (2). SAGE Publications: 93–106.

Thall, Peter F, and Richard Simon. 1994. "Practical Bayesian Guidelines for Phase IIB Clinical Trials." *Biometrics*. JSTOR, 337–49.

Thall, Peter F, Richard M Simon, and Elihu H Estey. 1995. "Bayesian Sequential Monitoring Designs for

Single-Arm Clinical Trials with Multiple Outcomes." *Statistics in Medicine* 14 (4). Wiley Online Library: 357–79.

Zohar, Sarah, Satoshi Teramukai, and Yinghui Zhou. 2008. "Bayesian Design and Conduct of Phase II Single-Arm Clinical Trials with Binary Outcomes: A Tutorial." *Contemporary Clinical Trials* 29 (4). Elsevier: 608–16.