Bayesian Designs for Phase II Clinical Trials

Yalin Zhu
Biostatistics and Data Management
Regeneron Pharmaceuticals Inc.
yalin.zhu@regeneron.com

August 9, 2016

Introduction for Bayesian designs

Overview of the current clinical trials

Average Cost per Patient: Oncology vs. All Rx categories (2011)

- Phase I: \$73000 (vs. \$36000)
- Phase II: \$57000 (vs. \$47500)
- Phase III: \$66000 (vs. \$47000)

Overall Success Rates (2003-2011)

- 6.7% of Phase I oncology entries were approved
- 10% of Phase I entries in all Rx categories were approved

Phase III Success Rates (2003-2010)

- 34% of trials achieved statistical significance in primary endpoints

[1]

[1] Dirk et al. (2016)

Why Using Bayesian Methods in Phase II trials?

Phase II clinical trials goals:

1. obtain a precise estimate of the response rate of the new drug
2. make GO/NO GO decisions for further testing in a phase III trial.

Advantages of **B**ayesian over **F**requentist:

1. Bayesian methods are natrual and easy to interpret.
2. Bayesian designs allow us to sequentially monitor the trials.
3. There is a bigger opportunity to stop trials earlier. Save money!

Note: In the early phase II development of the oncology drugs, most trials are **open label**, **single-arm** studies.

Traditional Approaches in Phase II Oncology Study

$$H_0 : p \leq p_0 \text{ versus } H_1 : p \geq p_1 = p_0 + \delta.$$

$$X_1 \sim Bin(n_1, p),\ X_2 \sim Bin(n_2, p) \text{ and } n = n_1 + n_2.$$

- $PET(p) = Pr(X_1 \leq r_1) = \sum_{x_1=0}^{r_1} \binom{n_1}{x_1} p^x (1-p)^{n_1 - x_1}$
- $E(N|p) = PET(p)n_1 + (1 - PET(p))n$
- Power function: $\beta(p) = Pr(X_1 > r_1 \cap X_1 + X_2 > r)$

1. **Gehan's** Two-stage Design (1961) *Cited by 664*
   - "14+11". Simplicity.
2. **Simon's** Two-stage Design (1989) *Cited by* **2612**
   - Early stop for futility only, two criteria (minimax and optimal) for selecting sample sizes and stop boundaries. Simon's design is the most popular design method in phase II oncology study.
3. **Jung's** Admissible Design (2004)
   - Aim to minimize *linear combination of the expected and maximum*

Search Criteria

Under the constraints on **type I error** ($\beta(p_0) \leq \alpha$) and **power** ($\beta(p_1) \geq power$), the design parameters (n_1, r_1, n, r) can be searched to meet one of the following criteria:

- Optimal: minimize the expected sample size $E(N|p_0)$

Search Criteria

Under the constraints on **type I error** ($\beta(p_0) \leq \alpha$) and **power** ($\beta(p_1) \geq power$), the design parameters (n_1, r_1, n, r) can be searched to meet one of the following criteria:

- Optimal: minimize the expected sample size $E(N|p_0)$
- Minimax: minimize the maxumum sample size $n_1 + n_2$.

Search Criteria

Under the constraints on **type I error** ($\beta(p_0) \leq \alpha$) and **power** ($\beta(p_1) \geq power$), the design parameters (n_1, r_1, n, r) can be searched to meet one of the following criteria:

- Optimal: minimize the expected sample size $E(N|p_0)$
- Minimax: minimize the maxumum sample size $n_1 + n_2$.
- Admissible: minimize the Bayes risk via convex hull (contains Optimal and Minimax designs).

Toolkits for Simon's Design (I):

- R function `ph2simon` in package `clinfun`.

```
library(clinfun)
ph2simon(pu=0.15, pa=0.3, ep1 = 0.05, ep2=0.1, 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
```

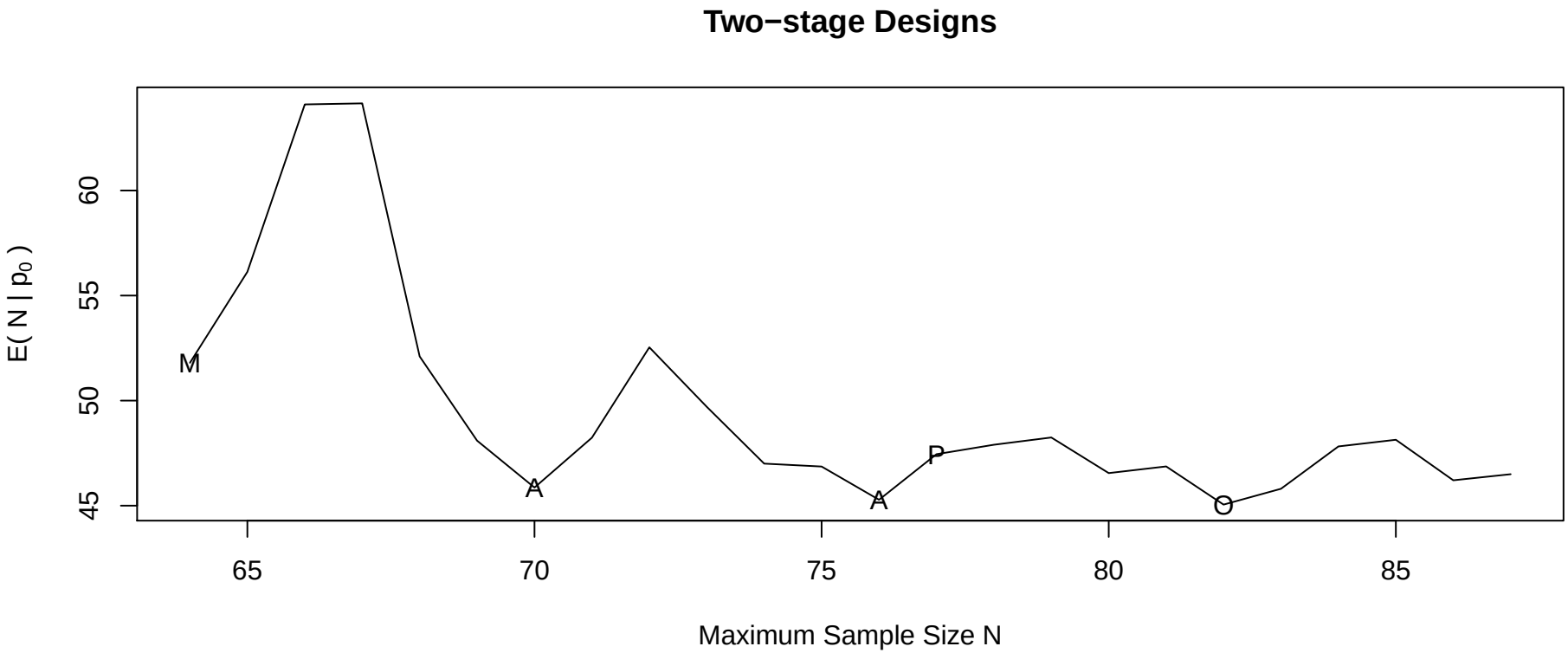
Toolkits for Simon's Design (II):

However, the function does not provide the information of type I error, power. We also want to find the admissible designs.

- We build a function `binom.design` including the class of admissible designs, which provides the **operating characteristic** information and **visualization options** for the design selections.

```
source("simon_admissible.R")
binom.design(output="admissible", p0=0.15, p1=0.30,
      signif.level = 0.05, power.level = 0.9, plot.out = T)
```

Example output




```
##                 r1 n1  r  n EN.p0. PET.p0.   error  power
## Optimal          5 30 17 82  45.05  0.7106 0.04609 0.9007
## Admissible       5 31 16 76  45.28  0.6827 0.04695 0.9037
## Admissible.1     6 36 15 70  45.86  0.7099 0.04655 0.9001
## Minimax          6 42 14 64  51.80  0.5545 0.04846 0.9003
```

Bayesian: Sequential Monitor

- Prior:

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

- Observed Data (likelihood):

$$x \sim Binomial(n, p).$$

- Posterior:

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

- Predictive:

$$Y|x \sim Beta - Binomial(N_{max} - n, a + x, b + n - x)$$

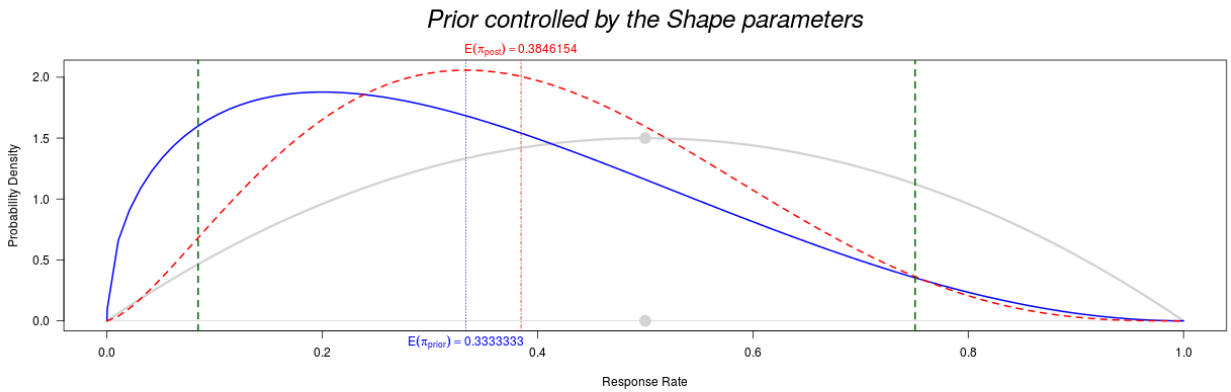
Prior and posterior distributions (Web tool)

We build a website using Shiny, it help look into the relationship among prior, observed data and posteror distribution.

`https://allen.shinyapps.io/Beta_Bayes_Prior/`

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 .



	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.

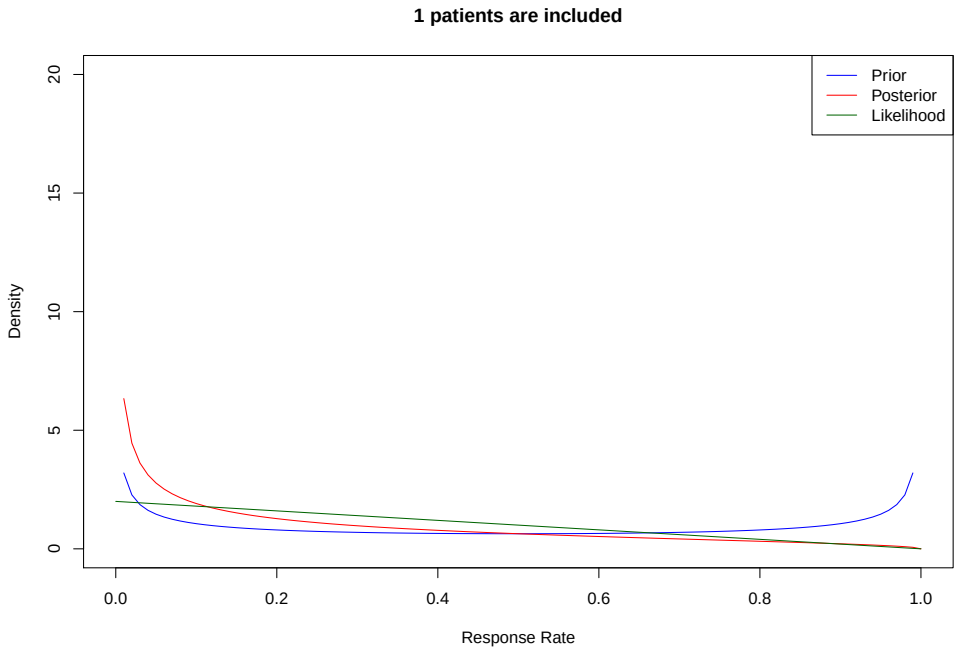
The likelihood curve is scaled so that the area under the curve = 1. This is done by multiplying the output from dbinom by the number of trials + 1.

Simulation Example

Consider *objective response rate* (ORR) as the primary endpoint with the following hypotheses:

$$H_0 : ORR \leq 15\% \quad versus \quad H_1 : ORR > 15\%.$$

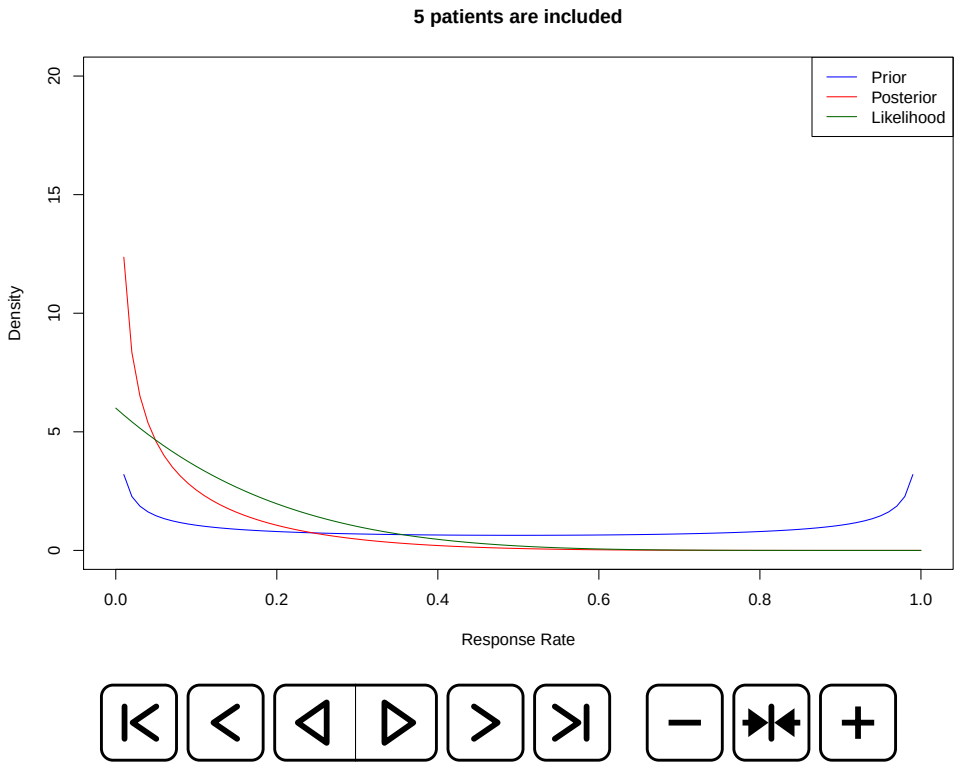
```
## Prior: Beta(0.5,0.5)
```



Simulation Example (II)

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

```
## Prior: Beta(0.5,0.5)
```



Prior selection: Non-informative priors

1. ***Jeffery prior***:
   - *a=0.5*
   - *b=0.5*

- *The prior sample size is 1, mean = 0.5, variance = 0.125 (Data dominate)*
- ***Remark***: *Compared Unif(0,1) prior, whose variance is 1/12=0.0833, Jeffery prior is more flat and less informative.*

2. ***Optimist's prior***:
   - $a = ORR_{prior} + 1$
   - $b = (1 - ORR_{prior}) + 1$

- *The prior sample size = 3, centered around* ORR_{prior}.
- *Such a prior distribution is sufficiently vague to allow for the possibility that ORR may take any value in the range* $0 < ORR < 1$.

Prior selection: Informative priors:

3. ***Based on prior ORR and Sample size ("ORR+N")***:
   - $a = ORR_{prior} + 1 + N_{prior} ORR_{prior}$
   - $b = (1 - ORR_{prior}) + 1 + N_{prior}(1 - ORR_{prior})$

- *The prior sample size* $= 3 + N_{prior}$.
- *It is similar to use mean*$=ORR_{prior}$ *and variance*$=\frac{ORR_{prior}(1-ORR_{prior})}{N_{prior}}$, *where the sample size* $= N_{prior} - 1$

4. ***Based on prior ORR and Width of confidence interval ("ORR+W")***:
   - $ORR_{prior} = \frac{a}{a+b}$
   - $W_{95} = F^{-1}(0.975; a, b) - F^{-1}(0.025; a, b)$

- *Solve the non-linear equations for a and b*

GO / NO GO Decision: Posterior Probability

$PostP = Pr(p > p_0|x)$

Algorithm 1

- **Step 1:** Specify the upper and lower probability cutoffs θ_U and θ_L. Typically, $\theta_U \in [0.9, 1]$ for efficacy and $\theta_L \in [0, 0.05]$ for futility, true null response rate p_0.
- **Step 2:** Let

$$S_U = \min\{x \in \mathbb{N} : PostP > \theta_U\}$$

and

$$S_L = \max\{x \in \mathbb{N} : PostP < \theta_L\}$$

- **Step 3:** Make decisions after observing another x responses out of n patients:
  - If $x \geq S_U$, then stop the trial for efficacy;
  - if $x \leq S_L$, then stop the trial for futility;
  - otherwise, continue the trial until N_{max} reached.

GO / NO GO Decision: Predictive Probability

$$PredP = Pr_{Y|x}\{Pr(p > p_0|x, Y) \geq \theta_T\}$$

Algorithm 2

- **Step 1:** Specified the upper and lower probability cutoffs θ_U and θ_L, typically, $\theta_U \in [0.9, 1]$ for efficacy and $\theta_L \in [0, 0.05]$ for futility. 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.

GO / NO GO Decision: Predictive Probability

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

Algorithm 2 (Cont.)

- **Step 3:** Make decisions after observing another x responses out of n patients:
  - If $x \geq S_U$, then stop the trial for efficacy ;
  - if $x \leq S_L$, then stop the trial for futility;
  - otherwise, continue the trial until N_{max} reached.

Remark: If there were no indicator function in the formula, the PredP simply reduces to the PostP after averaging out the unobserved Y.

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

Practical Examples

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs
  - **Sequential monitor**:

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs
  - **Sequential monitor**:
    1. Posterior probability

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs
  - **Sequential monitor**:
    1. Posterior probability
    2. Predictive probability.

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs
  - **Sequential monitor**:
    1. Posterior probability
    2. Predictive probability.
- **Outcome measure**: **ORR** is determined by the proportion of patients with best overall response of CR or PR among patients in SAF.

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs
  - **Sequential monitor**:
    1. Posterior probability
    2. Predictive probability.
- **Outcome measure**: **ORR** is determined by the proportion of patients with best overall response of CR or PR among patients in SAF.
- **Prior information**:

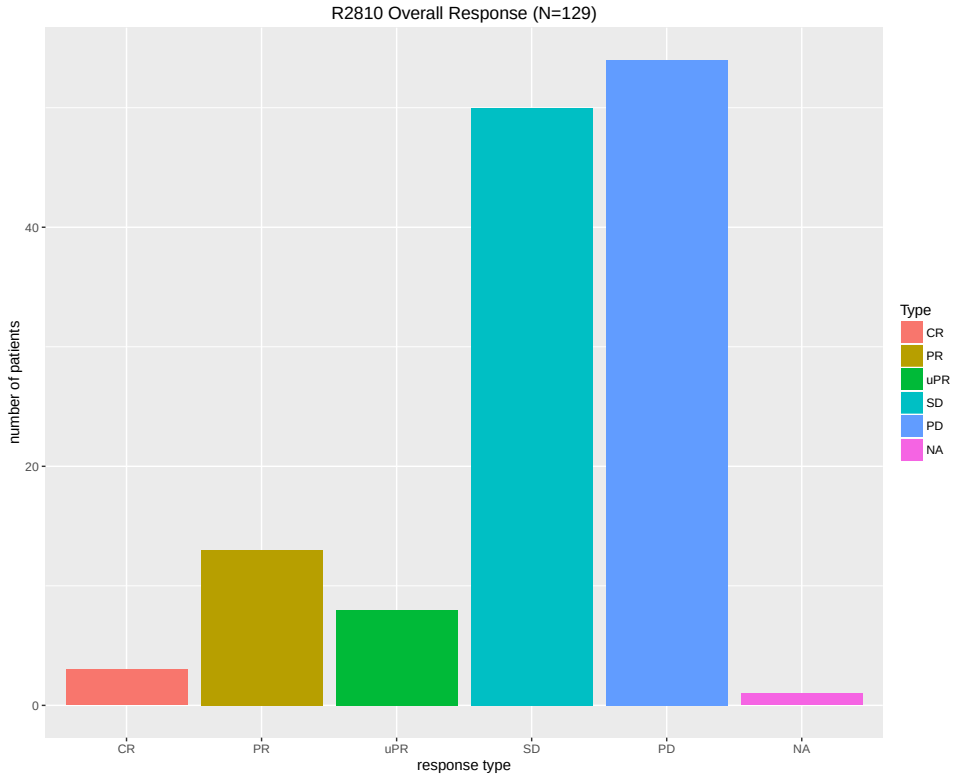
Design for R2810 Phase II Clinical Trials

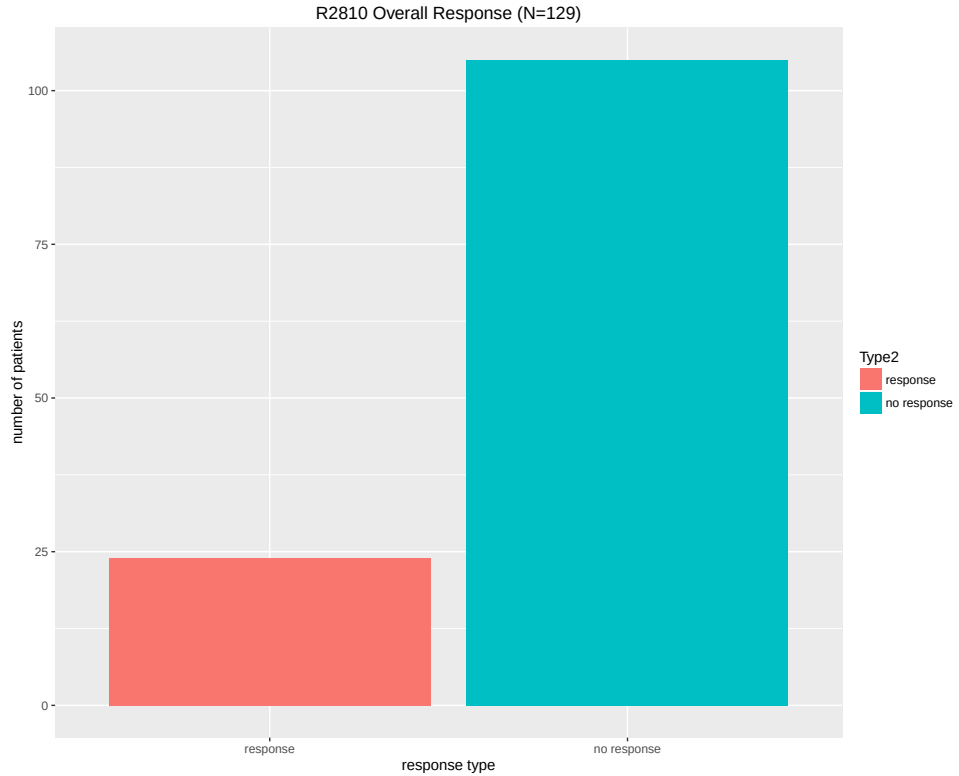
- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs
  - **Sequential monitor**:
    1. Posterior probability
    2. Predictive probability.
- **Outcome measure**: **ORR** is determined by the proportion of patients with best overall response of CR or PR among patients in SAF.
- **Prior information**:
  1. *KEYTRUDA* ($ORR = 0.236, N = 173, W_{95} = 0.186$)

Design for R2810 Phase II Clinical Trials

- **Data reference**: R2810 - RECIST (verson 1.1) Overall Response Investigator (PD, SD, CR or PR)
- **Aim of study**: According to current data and prior competitors' information, make interim analysis, determine suitable stopping rule.
- **Study design**:
  - **Two-stage Design**: Admissible designs
  - **Sequential monitor**:
    1. Posterior probability
    2. Predictive probability.
- **Outcome measure**: **ORR** is determined by the proportion of patients with best overall response of CR or PR among patients in SAF.
- **Prior information**:
  1. *KEYTRUDA* ($ORR = 0.236, N = 173, W_{95} = 0.186$)
  2. *OPDIVO* ($ORR = 0.317, N = 120, W_{95} = 0.173$)

Data Summary






```
##
##  Exact binomial test
##
## data:  sum(r01) and length(r01)
## number of successes = 20, number of trials = 100, p-value = 0.7
## alternative hypothesis: true probability of success is not equal to 0.2
## 95 percent confidence interval:
##  0.123 0.264
## sample estimates:
## probability of success
##                  0.186
```

Exact Binomial Test

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

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

Because of a relationship between the cumulative binomial distribution and the beta distribution, the Clopper-Pearson interval is sometimes presented in an alternate format that uses quantiles from the beta distribution.

$$B\left(\frac{\alpha}{2}; x, n-x+1\right) < \theta < B\left(1-\frac{\alpha}{2}; x+1, n-x\right)$$

Processing Data By Time-to-event

```
order.data <- response[order(response$TRTSDTM),c("TRTSDTM","AVALC")]
print(cbind(head(order.data,5), tail(order.data,5)),row.names=FALSE)
```

```
##              TRTSDTM AVALC             TRTSDTM AVALC
##  2015-02-04 09:50:00     SD 2016-04-13 11:07:00     SD
##  2015-02-06 09:08:00     PD 2016-04-18 11:05:00     PD
##  2015-02-09 11:15:00     SD 2016-04-18 12:20:00     SD
##  2015-03-16 09:30:00     PD 2016-04-19 08:40:00     PD
##  2015-03-18 09:55:00     PD 2016-04-19 10:42:00     PD
```

```
rorder <- r01[order(response$TRTSDTM)]
rorder
```

```
##   [1] 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 1 0
##  [36] 0 0 1 1 0 0 1 0 0 1 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [71] 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
## [106] 0 1 0 0 1 1 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
```

```
rtotal <- cumsum(r01[order(response$TRTSDTM)])
rtotal
```

```
##   [1]  0  0  0  0  0  0  0  1  2  2  2  2  2  2  2  2  2  2  3  3  3  3  3
##  [24]  3  3  3  3  4  5  5  5  6  6  7  7  7  7  8  9  9  9 10 10 10 11 12
##  [47] 12 12 12 13 13 13 13 13 13 13 14 14 14 14 14 14 14 14 14 14 14 14 14
##  [70] 14 14 14 15 15 15 15 15 15 15 15 15 15 15 15 15 15 16 16 16 16 16 17 17
##  [93] 18 18 18 18 18 18 18 18 18 18 18 18 18 18 19 19 19 20 21 21 21 22 22
## [116] 22 22 23 23 23 24 24 24 24 24 24 24 24 24
```

Two-stage Design Analysis

We can use the following program to run the two-stage designs (Optimal, Minimax and Admissible)

```
binom.design(output="admissible", p0=0.1, p1=0.3, signif.level = 0.05, power.level = 0.9)
```

```
##            r1 n1 r  n EN.p0. PET.p0.  error power
## Optimal     2 18 6 35   22.5   0.734 0.0474 0.902
## Admissible  2 19 6 34   23.4   0.705 0.0438 0.901
## Minimax     2 22 6 33   26.2   0.620 0.0409 0.902
```

		r1(data)	n1	r(data)	n	EN.p0.	PET.p0.	error	power
Scenario 1	Optimal	2(2)	18	6(7)	35	22.5255	0.7338	0.0474	0.9016
$p_0 = 0.1$	Admissible	2(3)	19	6(7)	34	23.4183	0.7054	0.0438	0.9014
$p_1 = 0.3$	Minimax	2(3)	22	6(6)	33	26.1795	0.6200	0.0409	0.9018
Scenario 2	Optimal	4(3)	19	15(13)	54	30.4349	0.6733	0.0482	0.9045
$p_0 = 0.2$	Admissible	4(3)	20	14(12)	49	30.7402	0.6296	0.0457	0.9030
$p_1 = 0.4$	Minimax	5(3)	24	13(11)	45	31.2263	0.6559	0.0483	0.9001
Scenario 3	Optimal	13(3)	24	36(14)	61	34.0132	0.7294	0.0487	0.9014
$p_0 = 0.5$	Admissible	12(3)	23	34(14)	57	34.5199	0.6612	0.0482	0.9046
$p_1 = 0.7$	Minimax	14(3)	27	32(13)	53	36.1144	0.6494	0.0461	0.9004

Table 1: Illustration of Simon's two-stage designs with three scenarios of design parameters, under the constraints on $\alpha = 0.05$ and $1 - \beta = 0.9$.

Results for two-stage designs

Except the **admissible** and **minimax** design under **Scenario 1 ($p_0 = 0.1$, $p_1 = 0.3$)**, other designs will be early terminated for futility for our data.
We have not used any prior information, although we have!

Seqential Monitor: PostP Design

The Bayesian designs provide a seqence of stopping boundary for response with corresponding sample size. Theoretically we can sequentially monitor the trial based on these rules.

Programming Examples: **Jeffery prior** $a = b = 0.5$

```
library(dplyr)
library(ph2bye)
PostP.design(type = "futility", nmax = 129, a=0.5, b=0.5, p0=0.4, theta = 0.05) %>% filter(bound %in% c(4,5,13,
```

```
##    n bound
## 1 19     4
## 2 22     5
## 3 46    13
## 4 49    14
## 5 52    15
```

```
simon.power(r1 = 4, n1=19, r=15, n=52, p = 0.2)
```

```
## [1] 0.0366
```

```
PostP.design(type = "efficacy", nmax = 129, a=0.5, b=0.5, p0=0.2, theta = 0.9)%>% filter(bound %in% c(4,5,13,14
```

```
##    n bound
## 1  8     4
## 2 12     5
## 3 43    13
## 4 48    14
## 5 52    15
```

Optimist's prior: KEYTRUDA : $a = 1.24, b = 1.76$

```
KT.prior <- optprior(KT.mean); KT.prior # KEYTRUDA optimist's prior
```

```
##    a    b
## 1.24 1.76
```

```
PostP.design(type = "futility", nmax = 129, a=1.24, b=1.76, p0=0.4, theta = 0.05) %>% filter(bound %in% c(4,5,1
```

```
##    n bound
## 1 19     4
## 2 23     5
## 3 47    13
## 4 50    14
## 5 52    15
```

```
PostP.design(type = "efficacy", nmax = 129, a=1.24, b=1.76, p0=0.2, theta = 0.9)%>% filter(bound %in% c(4,5,13,
```

```
##    n bound
## 1  9     4
## 2 12     5
## 3 45    13
## 4 49    14
## 5 53    15
```

```
bayes.design(a = 1.24, b = 1.76, r = rorder, stop.rule = "futility", p0 = 0.4, time.interval = 0.1)
bayes.design(a = 1.24, b = 1.76, r = rorder, stop.rule = "efficacy", p0 = 0.2, time.interval = 0.1)
```

Optimist's prior: OPDIVO : $a = 1.32, b = 1.68$

```
OD.prior <- optprior(OD.mean); OD.prior # OPDIVO optimist's prior
```

```
##    a    b
## 1.32 1.68
```

```
PostP.design(type = "futility", nmax = 129, a=1.32, b=1.68, p0=0.4, theta = 0.05) %>% filter(bound %in% c(4,5,1
```

```
##    n bound
## 1 20     4
## 2 23     5
## 3 47    13
## 4 50    14
## 5 53    15
```

```
PostP.design(type = "efficacy", nmax = 129, a=1.32, b=1.68, p0=0.2, theta = 0.9)%>% filter(bound %in% c(4,5,13,
```

```
##    n bound
## 1  9     4
## 2 13     5
## 3 45    13
## 4 49    14
## 5 53    15
```

```
bayes.design(a=1.32, b=1.68, r = rorder, stop.rule = "futility", p0 = 0.4, time.interval = 0.1)
bayes.design(a=1.32, b=1.68, r = rorder, stop.rule = "efficacy", p0 = 0.2, time.interval = 0.1)
```

ORR+N prior: KEYTRUDA : $a = 42.1, b = 133.9$

```
KT.prior <- ORRNprior(KT.mean, KT.n); KT.prior # KEYTRUDA optimist's prior
```

```
##     a     b
##  42.1 133.9
```

```
PostP.design(type = "futility", nmax = 129, a=42.1, b=133.9, p0=0.4, theta = 0.05) %>% filter(bound %in% c(4,5,
```

```
##    n bound
## 1  4     4
## 2  5     5
## 3 13    13
## 4 14    14
## 5 15    15
```

```
PostP.design(type = "efficacy", nmax = 129, a=42.1, b=133.9, p0=0.2, theta = 0.9)%>% filter(bound %in% c(4,5,13
```

```
##    n bound
## 1 13     4
## 2 18     5
## 3 54    13
## 4 59    14
## 5 64    15
```

```
bayes.design(a=42.1, b=133.9, r = rorder, stop.rule = "futility", p0 = 0.4, time.interval = 0.1)
bayes.design(a=42.1, b=133.9, r = rorder, stop.rule = "efficacy", p0 = 0.2, time.interval = 0.1)
```

ORR+N prior: KEYTRUDA : $a = 39.4, b = 83.6$

```
OD.prior <- ORRNprior(OD.mean, OD.n); OD.prior # OPDIVO optimist's prior
```

```
##    a    b
## 39.4 83.6
```

```
PostP.design(type = "futility", nmax = 129, a=39.4, b=83.6, p0=0.4, theta = 0.05) %>% filter(bound %in% c(4,5,1
```

```
##    n bound
## 1  9     4
## 2 11     5
## 3 33    13
## 4 36    14
## 5 39    15
```

```
PostP.design(type = "efficacy", nmax = 129, a=39.4, b=83.6, p0=0.2, theta = 0.9)%>% filter(bound %in% c(4,5,13,
```

```
##     n bound
## 1  54     4
## 2  58     5
## 3  95    13
## 4 100    14
## 5 104    15
```

```
bayes.design(a=39.4, b=83.6, r = rorder, stop.rule = "futility", p0 = 0.4, time.interval = 0.1)
bayes.design(a=39.4, b=83.6, r = rorder, stop.rule = "efficacy", p0 = 0.2, time.interval = 0.1)
```

ORR+W prior: KEYTRUDA: $a = 18.5, b = 59.9$

```
KT.prior <- ORRWprior(KT.mean, KT.W);  # KEYTRUDA optimist's prior
```

```
PostP.design(type = "futility", nmax = 129, a=18.5, b=59.9, p0=0.4, theta = 0.05) %>% filter(bound %in% c(4,5,1
```

```
##    n bound
## 1  4     4
## 2  5     5
## 3 20    13
## 4 23    14
## 5 26    15
```

```
PostP.design(type = "efficacy", nmax = 129, a=18.5, b=59.9, p0=0.2, theta = 0.9)%>% filter(bound %in% c(4,5,13,
```

```
##    n bound
## 1  5     4
## 2  9     5
## 3 45    13
## 4 49    14
## 5 54    15
```

```
bayes.design(a=18.5, b=59.9, r = rorder, stop.rule = "futility", p0 = 0.4, time.interval = 0.1)

bayes.design(a=18.5, b=59.9, r = rorder, stop.rule = "futility", p0 = 0.25, time.interval = 0.1)
# [1] "Stop the trial for futility after the inclusion of 83 patients."
bayes.design(a=18.5, b=59.9, r = rorder, stop.rule = "efficacy", p0 = 0.2, time.interval = 0.1)
```

ORR+W prior: OPDIVO: $a = 34.7, b = 74.9$

```
OD.prior <- ORRWprior(OD.mean, OD.W);  # OPDIVO optimist's prior
```

```
PostP.design(type = "futility", nmax = 129, a=34.7, b=74.9, p0=0.4, theta = 0.05) %>% filter(bound %in% c(4,5,1
```

```
##    n bound
## 1  9     4
## 2 12     5
## 3 34    13
## 4 36    14
## 5 39    15
```

```
PostP.design(type = "efficacy", nmax = 129, a=34.7, b=74.9, p0=0.2, theta = 0.9)%>% filter(bound %in% c(4,5,13,
```

```
##    n bound
## 1 46     4
## 2 50     5
## 3 87    13
## 4 91    14
## 5 96    15
```

```
bayes.design(a=34.7, b=74.9, r = rorder, stop.rule = "futility", p0 = 0.4, time.interval = 0.1)
bayes.design(a=34.7, b=74.9, r = rorder, stop.rule = "efficacy", p0 = 0.2, time.interval = 0.1)
```

Predictive Monitor: PredP Design

In R2810 trials, we set $\theta_U = 0.9$ for efficacy stop, $\theta_L = 0.05$ for futility stop, $\theta_T = 0.9$ for future patients cutoff. For example,

```
PredP.design(type = "futility", nmax = 129, a = 0.5, b = 0.5, p0 = 0.2 , theta_t = 0.9, delta = 0.2, theta = 0.
```

```
##    n bound
## 1 16     4
## 2 19     5
## 3 38    13
## 4 41    14
## 5 43    15
```

```
PredP.design(type = "futility", nmax = 129,  a=1.24, b=1.76, p0 = 0.2 , theta_t = 0.9, delta = 0.2, theta = 0.0
```

```
##    n bound
## 1 16     4
## 2 19     5
## 3 38    13
## 4 41    14
## 5 43    15
```

Operating characteristics (OC) of PostP design

- Compared with Simon's admissible designs, we take Scenario 2 ($p_0 = 0.2, p_1 = 0.4$) to illustrate the OC performance.
- Since $ORR + N$ and $ORR + W$ priors are too informative, the observed data hardly affects the posteriors, thus we only compare Jeffery's and optimist's prior. priors.
- Both PostP and PredP stop for futility only.
- Choose the design based on type I error and power constraints.

		r1(data)	n1	r(data)	n	error	power
Scenario 2	Optimal	4(3)	19	15(13)	54	0.0482	0.9045
$p_0 = 0.2$	Admissible	4(3)	20	14(12)	49	0.0457	0.9030
$p_1 = 0.4$	Minimax	5(3)	24	13(11)	45	0.0483	0.9001
PostP	Jeffery	4(3)	19	15(13)	52	0.0366	0.889
Design	Opt.KT	5(3)	23	15(13)	52	0.0379	0.903
	Opt.OD	5(3)	23	15(13)	53	0.0437	0.913
PredP	Jeffery	5(3)	19	14(12)	41	0.0088	0.677
Design	Opt.KT	5(3)	19	14(12)	41	0.0088	0.677
	Opt.OD	5(3)	19	13(11)	39	0.0130	0.700

Table 2: Stopping boundary and operating characteristics comparisons for two-stage designs, PostP and PredP designs with different priors. $\alpha = 0.05$ and $1 - \beta = 0.9$.

Conclusions and Further Discussions

Conclusions

- Under the same boundaries as frequentist two-stage designs, both PostP and PredP designs use smaller sample sizes for the first and second stage. Particularly, PredP design sample sizes are much smaller than two-stage designs. **Cost: lose power, but still control type I error**
- We can also do the similar comparisons for other scenarios. Stopping for efficacy can also be considered.

Further Discussions

1. Similar search design constrained by type I error and power can be constructed to find the optimal boundaries for PostP/PredP design.
2. Based on the subgroup property of R2810 expansion cohorts, hierarchical Bayesian models (MCMC) can be considered.

Computing Tools

- We have wrapped the functions for PostP and PredP designs into an R pacakge `ph2bye`, which is now available on CRAN.
  - `https://cran.r-project.org/web/packages/ph2bye`
- Useful R package in clinical trial research
- `haven`: read SAS dataset rapidly;
- `clinfun`, `gsDesign`: a bunch of design functions;
- `MCMCpack`: simulate complicated posterior models;
- `dplyr`: Manipulate big data.

References

1. Ashley, Dirk Reitsma MD Austin Combest, and Simmons Jürgen Hummel. "Improving Oncology Trials Through Adaptive Designs." Applied Clinical Trials 35.3 (2015).
2. Simon, Richard. "Optimal two-stage designs for phase II clinical trials." Controlled clinical trials 10.1 (1989): 1-10.
3. Jung, Sin-Ho, et al. "Admissible two-stage designs for phase II cancer clinical trials." Statistics in medicine 23.4 (2004): 561-569.
4. Tan, Say-Beng, and David Machin. "Bayesian two-stage designs for phase II clinical trials." Statistics in medicine 21.14 (2002): 1991-2012.
5. Mayo, Matthew S., and Byron J. Gajewski. "Bayesian sample size calculations in phase II clinical trials using informative conjugate priors." Controlled clinical trials 25.2 (2004): 157-167.
6. Thall, Peter F., and Richard Simon. "Practical Bayesian guidelines for phase IIB clinical trials." Biometrics (1994): 337-349.
7. Lee, J. Jack, and Diane D. Liu. "A predictive probability design for phase II cancer clinical trials." Clinical Trials 5.2 (2008): 93-106.

Acknowledge

- Questions?

- Questions?
- Thank you!