Bayesian Methods for Clinical Trials

by Libby Daniells & Pavel Mozgunov & Thomas Jaki

24th – 25th of February 2025

Practical 3: Bayesian sequential designs

You are asked to design a Phase II singe-agent PoC clinical trial of a new anti-cancer agent. The team would like to evaluate a particular dose of the new agent in terms of the objective response rate (ORR) at Week 12 after the start of the treatment. The uninteresting response rate that would not suggest a promising activity is $p_0 = 0.30$ and an interesting treatment effect that would trigger the further development of the compound is $p_1 = 0.50$. The currently planned sample size is $N = 40$ patients.
You are asked to design this study using a Bayesian design and consider inclusion of the interim analysis/analyses during the course of the trial.

(a) Firstly, consider a design with no interim analysis. Apply posterior probability criterion to claim the efficacy at the end of the trial. Find the critical probability threshold for the posterior probability that will ensure the control of the type I error at 5%. What power is achieved for the proposed sample size and the found value of the probability threshold? Would you recommend to change the sample size?

The critical value that controls the type I error below 5% is $c = 0.959$. The type I error is 3.1% due to the discreteness of the data. The power under alternative hypothesis is 78.6%. One could increase the sample size to achieve 80% or alternatively see how different design options affect the power.

```
N<-40
p0<-0.30
p1<-0.50
nsims<-100000

> X<-rbinom(n=nsims,size=N,prob=p0)
> post.prob<-pbeta(p0,shape1=X+1,shape2=N-X+1,lower.tail=F)
>
> c<-0.959
> mean(post.prob>c)
[1] 0.0309
>
> c<-0.958
```

```
> mean(post.prob>c)
[1] 0.06178
>
> X<-rbinom(n=nsims,size=N,prob=p1)
> post.prob<-pbeta(p0,shape1=X+1,shape2=N-X+1,lower.tail=F)
>
> c<-0.959
> mean(post.prob>c)
[1] 0.78649
```

(b) Consider now an inclusion of the interim analysis after the half of the total number of patients has been recruited in the trial. At the time of the interim analysis, one can stop the trial earlier for efficacy. Use the same probability threshold found in (a) for the interim and final analysis. What effect will it have on the type I error? Do you need to adjust the critical probability threshold. If yes, find the new value that control the type I error at the desirable level. What is the power of the study now?

If the critical value of 0.959 will not be adjusted then the type I error will be inflated to 6.5% as we will have more possibilities to reject the null hypothesis

```
> X1<-rbinom(n=nsims,size=N/2,prob=p0)
> X2<-rbinom(n=nsims,size=N/2,prob=p0)
> post.prob.1<-pbeta(p0,shape1=X1+1,shape2=N/2-X1+1,lower.tail=F)
> post.prob.2<-pbeta(p0,shape1=X1+X2+1,shape2=N-X1-X2+1,lower.tail=F)
> c<-0.959
> mean(post.prob.1>c | post.prob.2>c)
[1] 0.06509
```

We need to adjust the critical value to ensure the control of the type I error.

```
> c<-0.974
> mean(post.prob.1>c | post.prob.2>c)
[1] 0.04127
```

The power under the new design is 79.6%. This is due to a higher type I error (but still controlled). An early look gives us more possibilities and hence we get closer to the desirable type I error level.

```
> X1<-rbinom(n=nsims,size=N/2,prob=p1)
> X2<-rbinom(n=nsims,size=N/2,prob=p1)
> post.prob.1<-pbeta(p0,shape1=X1+1,shape2=N/2-X1+1,lower.tail=F)
```

```
> post.prob.2<-pbeta(p0,shape1=X1+X2+1,shape2=N-X1-X2+1,lower.tail=F)
> c<-0.974
> mean(post.prob.1>c | post.prob.2>c)
[1] 0.79602
```

(c) Consider now an inclusion of the interim analysis after the half of the total number of patients has been recruited in the trial. At the time of the interim analysis, one can stop the trial earlier for futility only. The decision rule for the futility is based on the posterior probability. Evaluate several design options with different futility bounds and propose a design option that achieves a balance in the operating characteristics. Why have you chosen this design?

We evaluate 3 design options with the futility bound for the posterior probability being 75%, 50%, and 25%. The results are given in the table below and the example code for 75% is following.

Design Option	Futility Bound	Type I	P(Stop) (under null)	Power	P(Stop) (under alt)
Fixed	–	3.2%	–	78.6%	–
Adaptive	75%	3.2%	23.6%	78.4%	0.6%
Adaptive	50%	3.2%	41.6%	78.6%	2.1%
Adaptive	25%	2.8%	77.4%	74.4%	13.0%

```
### 75% critical value
> X1<-rbinom(n=nsims,size=N/2,prob=p0)
> X2<-rbinom(n=nsims,size=N/2,prob=p0)
> post.prob.futility<-pbeta(p0,shape1=X1+1,shape2=N/2-X1+1,lower.tail=T)
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+1,shape2=N-X1-X2+1,lower.tail=F)
> c<-0.959
> t<-0.75
> mean(post.prob.futility<t & post.prob.efficacy>c)
[1] 0.03142
> mean(post.prob.futility>t)
[1] 0.23904
>
>
> X1<-rbinom(n=nsims,size=N/2,prob=p1)
> X2<-rbinom(n=nsims,size=N/2,prob=p1)
> post.prob.futility<-pbeta(p0,shape1=X1+1,shape2=N/2-X1+1,lower.tail=T)
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+1,shape2=N-X1-X2+1,lower.tail=F)
> mean(post.prob.futility<t & post.prob.efficacy>c)
[1] 0.78666
> mean(post.prob.futility>t)
[1] 0.00602
```

One can argue that the rule with 75% is not aggresive enough as it leads only to 23% stopping trial earlier under the null. On the other hand, 25% result in a very high proportion of early stopping but also in approximately 4% loss in power and 13% stopping the trial earlier for futility when we should not have. The critical bound of 50% achieves a good places with nearly 42% of trials stopped earlier under the null and no/little power loss. The options with the futility bound of 50% can be recommended for further evaluation.

(d) Consider now the futility rule based on the conditional power that uses the planned response rate (i.e. assumed under the alternative). Evaluate several design options using different futility bounds. Find a value of the futility bound that would match as closely as possible the operating characteristics of the design in point (c). Have you managed to find one or the operating characteristics could not be matched? Why?

For the futility rule based on the conditional power, the critical value is 20% ensure control of type I error at 5%. The operating characteristics are matched exactly of the rule of 50% futility bound for the posterior probability as the rules on different scales can be matched exactly.

```
> ###############   Point (d)  ####################
> nsims<-10^5
> X1<-rbinom(n=nsims,size=N/2,prob=p0)
> X2<-rbinom(n=nsims,size=N/2,prob=p0)
> CP<-c()
> for(i in 1:nsims){
+   CP[i]<-sum(dbinom(x=(18-X1[i]):20,size=20,prob=0.5))
+ }
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+1,shape2=N-X1-X2+1,lower.tail=F)
> c<-0.959
> l<-0.20
> mean(CP>l & post.prob.efficacy>c)
[1] 0.03275
> mean(CP<l)
[1] 0.41612
> X1<-rbinom(n=nsims,size=N/2,prob=p1)
> X2<-rbinom(n=nsims,size=N/2,prob=p1)
> CP<-c()
> for(i in 1:nsims){
+   CP[i]<-sum(dbinom(x=(18-X1[i]):20,size=20,prob=0.5))
+ }
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+1,shape2=N-X1-X2+1,lower.tail=F)
> mean(CP>l & post.prob.efficacy>c)
[1] 0.78288
> mean(CP<l)
[1] 0.02097
```

Design Option	Futility Bound	Type I	P(Stop) (under null)	Power	P(Stop) (under alt)
Fixed	–	3.2%	–	78.6%	–
Posterior	50%	3.2%	41.6%	78.6%	2.1%
CP(H1)	20%	3.2%	41.7%	78.3%	2.1%

(e) Assume now that there is some prior knowledge about the response rate. Specifically, the most likely value is 40% and we are 80% confident that the response rate is between 20% and 65%. Find the Beta prior matching these characteristics. For the new prior distribution, evaluate the decision rules proposed in (c). How does the new prior change the operating characteristics? Why? Does one need to make any changes to the design under the new prior? Why?

The Beta prior corresponding to this knowledge is $Beta(3.2, 7.5 - 3.2)$. Using this prior in the posterior futility criterion of 50%, one can obtain that the type I error is now inflated to 6.3%. This is due the prior knowledge shifting our prior (and hence posterior) to the right from the null hypothesis of 0.3. However, as the result, the power is also increased to 86%

```
> X1<-rbinom(n=nsims,size=N/2,prob=p0)
> X2<-rbinom(n=nsims,size=N/2,prob=p0)
> post.prob.futility<-pbeta(p0,shape1=X1+3.2,shape2=N/2-X1+(7.5-3.2),lower.tail=T)
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+3.2,shape2=N-X1-X2+(7.5-3.2),lower.tail=F)
> c<-0.959
> t<-0.50
> mean(post.prob.futility<t & post.prob.efficacy>c)
[1] 0.06294 # Type I error
> mean(post.prob.futility>t)
[1] 0.41557
> X1<-rbinom(n=nsims,size=N/2,prob=p1)
> X2<-rbinom(n=nsims,size=N/2,prob=p1)
> post.prob.futility<-pbeta(p0,shape1=X1+3.2,shape2=N/2-X1+(7.5-3.2),lower.tail=T)
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+3.2,shape2=N-X1-X2+(7.5-3.2),lower.tail=F)
> mean(post.prob.futility<t & post.prob.efficacy>c)
[1] 0.86139 # Power
> mean(post.prob.futility>t)
[1] 0.02054
```

To control type I error at 5%, one needs to find the new critical value. This is $c = 0.965$. However, for the new critical value the power goes down back to 78.3%

```
> X1<-rbinom(n=nsims,size=N/2,prob=p0)
> X2<-rbinom(n=nsims,size=N/2,prob=p0)
> post.prob.futility<-pbeta(p0,shape1=X1+3.2,shape2=N/2-X1+(7.5-3.2),lower.tail=T)
```

```
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+3.2,shape2=N-X1-X2+(7.5-3.2),lower.tail=F)
> c<-0.965
> t<-0.50
> mean(post.prob.futility<t & post.prob.efficacy>c)
[1] 0.03132
> mean(post.prob.futility>t)
[1] 0.41768
>
> X1<-rbinom(n=nsims,size=N/2,prob=p1)
> X2<-rbinom(n=nsims,size=N/2,prob=p1)
> post.prob.futility<-pbeta(p0,shape1=X1+3.2,shape2=N/2-X1+(7.5-3.2),lower.tail=T)
> post.prob.efficacy<-pbeta(p0,shape1=X1+X2+3.2,shape2=N-X1-X2+(7.5-3.2),lower.tail=F)
> mean(post.prob.futility<t & post.prob.efficacy>c)
[1] 0.78333
> mean(post.prob.futility>t)
[1] 0.02053
```