An example to illustrate boundary re-calculations during the trial

Marcel Wolbers, Gernot Wassmer, and Friedrich Pahlke

Last change: 26 Oktober, 2021

Contents

Summary 1

1 Introduction 1

2 Original trial design 2

3 Boundary and power update at the first interim analysis 3

4 Boundary and power update at the second interim analysis 5

5 Boundary and power update at the final analysis 7

6 Overview of all updates 8

6.1 Original design .......................... 8

6.2 Updated boundaries and power at the first interim analysis .................. 9

6.3 Updated boundaries and power at the second interim analysis .............. 9

6.4 Updated boundaries and power at the final analysis ........................... 10

Summary

This R Markdown document provides an example for updating the group sequential boundaries when using an alpha spending function approach based on observed information rates in rpact.

1 Introduction

Group-sequential designs based on $\alpha$-spending functions protect the Type I error exactly even if the pre-planned interim schedule is not exactly adhered to. However, this requires re-calculation of the group-sequential boundaries at each interim analysis based on actually observed information fractions. Unless deviations from the planned information fractions are substantial, the re-calculated boundaries are quite similar to the pre-planned boundaries and the re-calculation will affect the actual test decision only on rare occasions.

Importantly, it is not allowed that the timing of future interim analyses is “motivated” by results from earlier interim analyses as this could inflate the Type I error rate. Deviations from the planned information fractions should thus only occur due to operational reasons (as it is difficult to hit the exact number of events exactly in a real trial) or due to external evidence.

The general principles for these boundary re-calculation are as follows (see also, Wassmer & Brannath, 2016, p78f):

- Updates at interim analyses prior to the final analysis:
Information fractions are updated according to the actually observed information fraction at the interim analysis relative to the planned maximum information.

- The planned \( \alpha \)-spending function is then applied to these updated information fractions.

- Updates at the final analysis in case the observed information at the final analysis is larger (“over-running”) or smaller (“under-running”) than the planned maximum information:
  - Information fractions are updated according to the actually observed information fraction at all interim analyses relative to the observed maximum information. \( \Rightarrow \) Information fraction at final analysis is re-set to 1 but information fractions for earlier interim analyses are also changed.
  - The originally planned \( \alpha \)-spending function cannot be applied to these updated information fractions because this would modify the critical boundaries of earlier interim analyses which is clearly not allowed. Instead, one uses the \( \alpha \) that has actually been spent at earlier interim analyses and spends all remaining \( \alpha \) at the final analysis.

This general principle be implemented via a user-defined \( \alpha \)-spending function and is illustrated for an example trial with a survival endpoint below.

**First, load the rpact package**

```r
library(rpact)
packageVersion("rpact") # version should be version 2.0.5 or later
```

```shell
[1] '3.1.1'
```

## 2 Original trial design

The original trial design for this example is based on a standard O'Brien & Fleming type \( \alpha \)-spending function with planned efficacy interim analyses after 50% and 75% of information as specified below.

### Initial design

```r
design <- getDesignGroupSequential(sided = 1, alpha = 0.025, beta = 0.2,
                                   informationRates = c(0.5, 0.75, 1), typeOfDesign = "asOF")
```

### Initial sample size calculation

```r
sampleSizeResult <- getSampleSizeSurvival(
  design = design,
  lambda2 = log(2)/60, hazardRatio = 0.75,
  dropoutRate1 = 0.025, dropoutRate2 = 0.025, dropoutTime = 12,
  accrualTime = 0, accrualIntensity = 30,
  maxNumberOfSubjects = 1000)
```

### Summarize design

```r
summary(sampleSizeResult)
```

### Sample size calculation for a survival endpoint

### Sequential analysis with a maximum of 3 looks (group sequential design), overall significance level 2.5% (one-sided).

### The sample size was calculated for a two-sample logrank test,

### \( H_0: \) hazard ratio = 1, \( H_1: \) hazard ratio = 0.75, control lambda(2) = 0.012,

### maximum number of subjects = 1000, accrual time = 33.333, dropout rate(1) = 0.025,

### dropout rate(2) = 0.025, dropout time = 12, power 80%.

### Stage 1 2 3

### Information rate 50% 75% 100%

### Efficacy boundary (z-value scale) 2.963 2.359 2.014

### Overall power 0.1680 0.5400 0.8000
3 BOUNDARY AND POWER UPDATE AT THE FIRST INTERIM ANALYSIS

Assume that the first interim was conducted after 205 rather than the planned 194 events.

The updated design is calculated as per the code below. Note that for the calculation of boundary values on the treatment effect scale, we use the function `getPowerSurvival()` with the updated design rather than the function `getSampleSizeSurvival()` as we are only updating the boundary, not the sample size or the maximum number of events.

```r
# Update design using observed information fraction at first interim.
# Information fraction of later interim analyses is not changed.
designUpdate1 <- getDesignGroupSequential(sided = 1, alpha = 0.025, beta = 0.2,
informationRates = c(205/387, 0.75, 1), typeOfDesign = "asOF")

# Recalculate the power to get boundary values on the effect scale
# (Use original maxNumberOfEvents and sample size)
powerUpdate1 <- getPowerSurvival(
  design = designUpdate1,
  lambda2 = log(2)/60, hazardRatio = 0.75,
dropoutRate1 = 0.025, dropoutRate2 = 0.025, dropoutTime = 12,
accrualTime = 0, accrualIntensity = 30,
  maxNumberOfSubjects = 1000, maxNumberOfEvents = 387, directionUpper = FALSE)

powerUpdate1
```

### Design plan parameters and output for survival data:

#### Design parameters:
- **Information rates**: 0.530, 0.750, 1.000
- **Critical values**: 2.867, 2.366, 2.015
- **Futility bounds (non-binding)**: -Inf, -Inf
- **Cumulative alpha spending**: 0.002073, 0.009649, 0.025000
- **Local one-sided significance levels**: 0.002073, 0.008998, 0.021968
- **Significance level**: 0.0250
- **Test**: one-sided

#### User defined parameters:
- **Direction upper**: FALSE
- **lambda(2)**: 0.0116
- **Hazard ratio**: 0.750
- **Maximum number of subjects**: 1000
- **Maximum number of events**: 387
## Accrual time
33.33

## Accrual intensity
30

## Drop-out rate (1)
0.025

## Drop-out rate (2)
0.025

## Default parameters:

### Theta H0
1

### Type of computation
Schoenfeld

### Planned allocation ratio
1

### kappa
1

### Piecewise survival times
0.00

## Drop-out time
12.00

### Sample size and output:

#### median(1)
80.0

#### median(2)
60.0

#### lambda(1)
0.00866

#### Follow up time
35.81

#### Expected number of events
317.0

#### Overall reject
0.8001

#### Reject per stage [1]
0.2097

#### Reject per stage [2]
0.3294

#### Reject per stage [3]
0.2610

#### Early stop
0.5391

#### Analysis times [1]
40.60

#### Analysis times [2]
52.73

#### Analysis times [3]
69.14

#### Expected study duration
57.75

#### Maximal study duration
69.14

#### Cumulative number of events [1]
205.0

#### Cumulative number of events [2]
290.2

#### Cumulative number of events [3]
387.0

#### Number of subjects [1]
1000.0

#### Number of subjects [2]
1000.0

#### Number of subjects [3]
1000.0

#### Expected number of subjects
1000.0

#### Critical values (treatment effect scale) [1]
0.670

#### Critical values (treatment effect scale) [2]
0.758

#### Critical values (treatment effect scale) [3]
0.815

### Legend:

#### (i): values of treatment arm i

#### [k]: values at stage k

The updated information rates and corresponding boundaries as per the output above are summarized as follows:

### Power calculation for a survival endpoint

Sequential analysis with a maximum of 3 looks (group sequential design), overall significance level 2.5% (one-sided).

The results were calculated for a two-sample logrank test, H0: hazard ratio = 1, power directed towards smaller values, H1: hazard ratio = 0.75, control lambda(2) = 0.012, maximum number of subjects = 1000, maximum number of events = 387,
## accrual time = 33.333, accrual intensity = 30, dropout rate(1) = 0.025, 
dropout rate(2) = 0.025, dropout time = 12.

---

### Stage

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>53%</td>
<td>75%</td>
<td>100%</td>
</tr>
<tr>
<td>Efficacy boundary (z-value scale)</td>
<td>2.867</td>
<td>2.366</td>
<td>2.015</td>
</tr>
<tr>
<td>Overall power</td>
<td>0.2097</td>
<td>0.5391</td>
<td>0.8001</td>
</tr>
<tr>
<td>Expected number of subjects</td>
<td>1000.0</td>
<td>1000.0</td>
<td>1000.0</td>
</tr>
<tr>
<td>Number of subjects</td>
<td>317.0</td>
<td>317.0</td>
<td>317.0</td>
</tr>
<tr>
<td>Expected number of events</td>
<td>205</td>
<td>291</td>
<td>387</td>
</tr>
<tr>
<td>Analysis time</td>
<td>40.6</td>
<td>52.7</td>
<td>69.1</td>
</tr>
<tr>
<td>Expected study duration</td>
<td>57.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative alpha spent</td>
<td>0.0021</td>
<td>0.0096</td>
<td>0.0250</td>
</tr>
<tr>
<td>One-sided local significance level</td>
<td>0.0021</td>
<td>0.0090</td>
<td>0.0220</td>
</tr>
<tr>
<td>Efficacy boundary (t)</td>
<td>0.670</td>
<td>0.758</td>
<td>0.815</td>
</tr>
<tr>
<td>Exit probability for efficacy (under H0)</td>
<td>0.0021</td>
<td>0.0076</td>
<td></td>
</tr>
<tr>
<td>Exit probability for efficacy (under H1)</td>
<td>0.2097</td>
<td>0.3294</td>
<td></td>
</tr>
</tbody>
</table>

### Legend:

- (t): treatment effect scale

---

## 4 Boundary and power update at the second interim analysis

Assume that the efficacy boundary was not crossed at the first interim analysis and the trial continued to the second interim analysis which was conducted after 285 rather than the planned 291 events. The updated design is calculated in the same way as for the first interim analysis as per the code below. The idea is to use the cumulative alpha spent from the first stage and an updated cumulative alpha that is spent for the second stage. For the second stage, this can be obtained with the original O'Brien & Fleming alpha spending function:

```r
# Update design using observed information fraction at first and second interim.
designUpdate2 <- getDesignGroupSequential(sided = 1, alpha = 0.025, beta = 0.2,
                                          informationRates = c(205/387,285/387,1), typeOfDesign = "asOF")

# Recalculate power to get boundary values on effect scale
# (Use original maxNumberOfEvents and sample size)
powerUpdate2 <- getPowerSurvival(
  design = designUpdate2,
  lambda2 = log(2)/60,hazardRatio = 0.75,
  dropoutRate1 = 0.025, dropoutRate2 = 0.025, dropoutTime = 12,
  accrualTime = 0,accrualIntensity = 30,
  maxNumberOfSubjects = 1000, maxNumberOfEvents = 387, directionUpper = FALSE)
powerUpdate2
```

### Design plan parameters and output for survival data:

```r
## Design parameters:

### Information rates : 0.530, 0.736, 1.000
### Critical values : 2.867, 2.393, 2.011
### Futility bounds (non-binding) : -Inf, -Inf
### Cumulative alpha spending : 0.002073, 0.009005, 0.025000
### Local one-sided significance levels : 0.002073, 0.008356, 0.022154
### Significance level : 0.0250
```
Test : one-sided

User defined parameters:
- Direction upper : FALSE
- lambda(2) : 0.0116
- Hazard ratio : 0.750
- Maximum number of subjects : 1000
- Maximum number of events : 387
- Accrual time : 33.33
- Accrual intensity : 30
- Drop-out rate (1) : 0.025
- Drop-out rate (2) : 0.025

Default parameters:
- Theta H0 : 1
- Type of computation : Schoenfeld
- Planned allocation ratio : 1
- kappa : 1
- Piecewise survival times : 0.00
- Drop-out time : 12.00

Sample size and output:
- median(1) : 80.0
- median(2) : 60.0
- lambda(1) : 0.00866
- Follow up time : 35.81
- Expected number of events : 317.2
- Overall reject : 0.8004
- Reject per stage [1] : 0.2097
- Reject per stage [2] : 0.3101
- Reject per stage [3] : 0.2806
- Early stop : 0.5198
- Analysis times [1] : 40.60
- Expected study duration : 57.82
- Maximal study duration : 69.14
- Cumulative number of events [1] : 205.0
- Cumulative number of events [2] : 285.0
- Cumulative number of events [3] : 387.0
- Number of subjects [1] : 1000.0
- Number of subjects [2] : 1000.0
- Number of subjects [3] : 1000.0
- Expected number of subjects : 1000.0
- Critical values (treatment effect scale) [1] : 0.670
- Critical values (treatment effect scale) [2] : 0.753
- Critical values (treatment effect scale) [3] : 0.815

Legend:
- (i): values of treatment arm i
- [k]: values at stage k
5 Boundary and power update at the final analysis

Assume that the efficacy boundary was also not crossed at the second interim analysis and the trial continued to the final analysis which was conducted after 393 rather than the planned 387 events. The updated design is calculated as per the code below. The idea here to use the cumulative alpha spent from the first and the second stage stage and the final alpha that is spent for the last stage. An updated correlation has to be used and the original O’Brien & Fleming alpha spending function cannot be used anymore. Instead, the alpha spending function needs to be user defined as follows:

```
# Update boundary with information fractions as per actually observed event numbers
# !! use user-defined alpha-spending and spend alpha according to actual alpha spent
# according to the second interim analysis
designUpdate3 <- getDesignGroupSequential(sided = 1, alpha = 0.025, beta = 0.2,
                                           informationRates = c(205,285,393)/393,
                                           typeOfDesign = "asUser",
                                           userAlphaSpending = designUpdate2$alphaSpent)

# Recalculate power to get boundary values on effect scale
# (Use planned sample size and **observed** maxNumberOfEvents)
powerUpdate3 <- getPowerSurvival(
  design = designUpdate3,
  lambda2 = log(2)/60, hazardRatio = 0.75,
  dropoutRate1 = 0.025, dropoutRate2 = 0.025, dropoutTime = 12,
  accrualTime = 0, accrualIntensity = 30,
  maxNumberOfSubjects = 1000, maxNumberOfEvents = 393, directionUpper = FALSE)
```

```
# Design plan parameters and output for survival data:
## Design parameters:
## Information rates : 0.522, 0.725, 1.000
## Critical values : 2.867, 2.393, 2.014
## Futility bounds (non-binding) : -Inf, -Inf
## Cumulative alpha spending : 0.002073, 0.009005, 0.025000
## Local one-sided significance levels : 0.002073, 0.008356, 0.022021
## Significance level : 0.0250
## Test : one-sided
##
## User defined parameters:
## Direction upper : FALSE
## lambda(2) : 0.0116
## Hazard ratio : 0.750
## Maximum number of subjects : 1000
## Maximum number of events : 393
## Accrual time : 33.33
## Accrual intensity : 30
## Drop-out rate (1) : 0.025
## Drop-out rate (2) : 0.025
##
## Default parameters:
## Theta H0 : 1
## Type of computation : Schoenfeld
## Planned allocation ratio : 1
## kappa : 1
## Piecewise survival times : 0.00
```
6. OVERVIEW OF ALL UPDATES

For easier comparison, all discussed boundary updates and power calculations are summarized more conveniently below. Note that each update only affects boundaries for the current or later analyses, i.e., earlier boundaries are never retrospectively modified.

6.1 Original design

## Sample size calculation for a survival endpoint

## Sequential analysis with a maximum of 3 looks (group sequential design), overall significance level 2.5% (one-sided).

## The sample size was calculated for a two-sample logrank test,
## H0: hazard ratio = 1, H1: hazard ratio = 0.75, control lambda(2) = 0.012,
## maximum number of subjects = 1000, accrual time = 33.333, dropout rate(1) = 0.025,
## dropout rate(2) = 0.025, dropout time = 12, power 80%.

## Stage
## Information rate 50% 75% 100%
## Efficacy boundary (z-value scale) 2.963 2.359 2.014
## Overall power 0.1680 0.5400 0.8000
6.2 Updated boundaries and power at the first interim analysis

## Expected number of subjects 1000.0
## Number of subjects 1000.0 1000.0 1000.0
## Cumulative number of events 194 291 387
## Analysis time 39.1 52.7 69.1
## Expected study duration 58.0
## Cumulative alpha spent 0.0015 0.0096 0.0250
## One-sided local significance level 0.0015 0.0092 0.0220
## Efficacy boundary (t) 0.653 0.758 0.815
## Exit probability for efficacy (under H0) 0.0015 0.0081
## Exit probability for efficacy (under H1) 0.1680 0.3720
##
## Legend:
## (t): treatment effect scale

## Updated boundaries and power at the first interim analysis

## Power calculation for a survival endpoint
##
## Sequential analysis with a maximum of 3 looks (group sequential design), overall
## significance level 2.5% (one-sided).
## The results were calculated for a two-sample logrank test,
## H0: hazard ratio = 1, power directed towards smaller values,
## H1: hazard ratio = 0.75, control lambda(2) = 0.012,
## maximum number of subjects = 1000, maximum number of events = 387,
## accrual time = 33.333, accrual intensity = 30, dropout rate(1) = 0.025,
## dropout rate(2) = 0.025, dropout time = 12.
##
## Stage 1 2 3
## Information rate 53% 75% 100%
## Efficacy boundary (z-value scale) 2.867 2.366 2.015
## Overall power 0.2097 0.5391 0.8001
## Expected number of subjects 1000.0
## Number of subjects 1000.0 1000.0 1000.0
## Expected number of events 317.0
## Cumulative number of events 205 291 387
## Analysis time 40.6 52.7 69.1
## Expected study duration 57.8
## Cumulative alpha spent 0.0021 0.0096 0.0250
## One-sided local significance level 0.0021 0.0090 0.0220
## Efficacy boundary (t) 0.670 0.758 0.815
## Exit probability for efficacy (under H0) 0.0021 0.0076
## Exit probability for efficacy (under H1) 0.2097 0.3294
##
## Legend:
## (t): treatment effect scale

6.3 Updated boundaries and power at the second interim analysis

## Power calculation for a survival endpoint
##
## Sequential analysis with a maximum of 3 looks (group sequential design), overall
## significance level 2.5% (one-sided).
## The results were calculated for a two-sample logrank test,
## H0: hazard ratio = 1, power directed towards smaller values,
# 6.4 Updated boundaries and power at the final analysis

Power calculation for a survival endpoint

Sequential analysis with a maximum of 3 looks (group sequential design), overall significance level 2.5% (one-sided).

The results were calculated for a two-sample logrank test,

## H0: hazard ratio = 1, power directed towards smaller values,
## H1: hazard ratio = 0.75, control lambda(2) = 0.012,
## maximum number of subjects = 1000, maximum number of events = 393,
## accrual time = 33.333, accrual intensity = 30, dropout rate(1) = 0.025,
## dropout rate(2) = 0.025, dropout time = 12.

## Stage 1 2 3
## Information rate 52.2% 72.5% 100%
## Efficacy boundary (z-value scale) 2.867 2.393 2.014
## Overall power 0.2097 0.5198 0.8060
## Expected number of subjects 1000.0 1000.0 1000.0
## Number of subjects 1000.0 1000.0 1000.0
## Expected number of events 320.1
## Cumulative number of events 206 285 393
## Analysis time 40.6 51.9 70.3
## Expected study duration 58.4
## Cumulative alpha spent 0.0021 0.0090 0.0250
## One-sided local significance level 0.0021 0.0084 0.0220
## Efficacy boundary (t) 0.670 0.753 0.816
## Exit probability for efficacy (under H0) 0.0021 0.0069
## Exit probability for efficacy (under H1) 0.2097 0.3101

## Legend:
## (t): treatment effect scale
## (t): treatment effect scale

System: rpact 3.1.1, R version 4.1.0 (2021-05-18), platform: x86_64-w64-mingw32

To cite R in publications use:


To cite package ‘rpact’ in publications use:


License

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/.