Comparing sample size and power calculation results for a group-sequential trial with a survival endpoint: rpact vs. gsDesign

Gernot Wassmer, Friedrich Pahlke, and Marcel Wolbers

28 Juni, 2019

The design

- 1:1 randomized
- Two-sided log-rank test; 80% power at the 5% significance level (or one-sided at 2.5%)
- Target HR for primary endpoint (PFS) is 0.75
- PFS in the control arm follows a piece-wise exponential distribution, with the hazard rate h(t) estimated using historical controls as follows:
  - h(t) = 0.025 for t between 0 and 6 months;
  - h(t) = 0.04 for t between 6 and 9 months;
  - h(t) = 0.015 for t between 9 and 15 months;
  - h(t) = 0.01 for t between 15 and 21 months;
  - h(t) = 0.007 for t beyond 21 months.
- An annual dropout probability of 20%
- Interim analyses at 33% and 70% of total information
- Alpha-spending version of O'Brien-Fleming boundary for efficacy
- No futility interim
- 1405 subjects recruited in total
- Staggered recruitment:
  - 15 pt/month during first 12 months;
  - subsequently, increase of # of sites and ramp up of recruitment by +6 pt/month each month until a maximum of 45 pt/month

Calculation with gsDesign

```
library(gsDesign)
options(warn = -1) # avoid warnings generated by gsDesign
x <- gsSurv(k = 3,test.type = 1, alpha = 0.025, beta = 0.2,
  timing = c(0.33,0.7), sfu = sfLDOF,           # boundary
  hr = 0.75,
  lambdaC = c(0.025, 0.04, 0.015, 0.01, 0.007), # piecewise lambdas
  S = c(6, 3, 6, 6),                            # piecewise survival times
  eta= -log(1 - 0.2) / 12,                      # dropout
  gamma = c(15, 21, 27, 33, 39, 45),            # recruitment, pt no
  R = c(12, 1, 1, 1, 1, (1405 - 300) / 45),     # recruitment duration
  minfup = NULL)
print(x,digits=5)
```

```
## Time to event group sequential design with HR= 0.75
## Equal randomization:          ratio=1
## One-sided group sequential design with
## 80 % power and 2.5 % Type I Error.
##
```

```
##   Analysis  N   Z   Nominal p  Spend
##          1 128 3.73    0.0001 0.0001
##          2 271 2.44    0.0074 0.0073
##          3 386 2.00    0.0227 0.0176
##      Total                   0.0250
##
## ++ alpha spending:
##  Lan-DeMets O'brien-Fleming approximation spending function.
##
## Boundary crossing probabilities and expected sample size
## assume any cross stops the trial
##
## Upper boundary (power or Type I Error)
##           Analysis
##    Theta      1      2      3 Total  E{N}
##   0.0000 0.0001 0.0073 0.0176 0.025 385.0
##   0.1437 0.0175 0.4517 0.3309 0.800 329.1
##              T         n   Events HR efficacy
## IA 1  26.78703  785.4162 127.3407       0.516
## IA 2  38.62360 1318.0620 270.1171       0.743
## Final 50.80093 1405.0000 385.8810       0.816
## Accrual rates:
##          Stratum 1
## 0-12            15
## 12-13           21
## 13-14           27
## 14-15           33
## 15-16           39
## 16-40.56        45
## Control event rates (H1):
##        Stratum 1
## 0-6        0.025
## 6-9        0.040
## 9-15       0.015
## 15-21      0.010
## 21-Inf     0.007
## Censoring rates:
##        Stratum 1
## 0-6       0.0186
## 6-9       0.0186
## 9-15      0.0186
## 15-21     0.0186
## 21-Inf    0.0186
```

Calculation with rpact

Design

```
# Load the package `rpact`
library(rpact)
packageVersion("rpact")
```

```
## [1] '2.0.1.9003'
```

```
design <- getDesignGroupSequential(
  sided = 1, alpha = 0.025, beta = 0.2,
  informationRates = c(0.33, 0.7, 1),
  typeOfDesign="asOF")
design
```

```
## Design parameters and output of group sequential design:
##
## User defined parameters:
##   Type of design                              : asOF
##   Information rates                           : 0.330, 0.700, 1.000
##
## Derived from user defined parameters:
##   Maximum number of stages                    : 3
##
## Default parameters:
##   Stages                                      : 1, 2, 3
##   Significance level                          : 0.0250
##   Type II error rate                          : 0.2
##   Two-sided power                             : FALSE
##   Test                                        : one-sided
##   Tolerance                                   : 1e-08
##   Type of beta spending                       : none
##
## Output:
##   Cumulative alpha spending                   : 9.549e-05, 7.384e-03, 2.500e-02
##   Critical values                             : 3.731, 2.440, 2.000
##   Stage levels                                : 9.549e-05, 7.351e-03, 2.274e-02
```

Sample size/timing of interim

```
piecewiseSurvivalTime <- list(
  "0 - <6"   = 0.025,
  "6 - <9"   = 0.04,
  "9 - <15"  = 0.015,
  "15 - <21" = 0.01,
  ">=21"     = 0.007)

accrualTime <- list(
  "0  - <12" = 15,
  "12 - <13" = 21,
  "13 - <14" = 27,
  "14 - <15" = 33,
  "15 - <16" = 39,
  ">=16"     = 45)

y <- getPowerSurvival(
  design = design, typeOfComputation = "Schoenfeld",
  thetaH0 = 1, directionUpper = FALSE,
  dropoutRate1 = 0.2, dropoutRate2 = 0.2, dropoutTime = 12,
  allocationRatioPlanned = 1,
  accrualTime = accrualTime,
  piecewiseSurvivalTime = piecewiseSurvivalTime,
```

```
  hazardRatio = 0.75,
  maxNumberOfEvents = x$n.I[3],
  maxNumberOfSubjects = 1405)
y
```

```
## Design plan parameters and output for survival data:
##
## Design parameters:
##   Significance level                          : 0.0250
##   Test                                        : one-sided
##
## User defined parameters:
##   Direction upper                             : FALSE
##   Lambda (2)                                  : 0.025, 0.040, 0.015, 0.010, 0.007
##   Hazard ratio                                : 0.750
##   Maximum number of subjects                  : 1405.0
##   Maximum number of events                    : 385.9
##   Accrual intensity                           : 15.0, 21.0, 27.0, 33.0, 39.0, 45.0
##   Piecewise survival times                    : 0.00, 6.00, 9.00, 15.00, 21.00
##   Drop-out rate (1)                           : 0.200
##   Drop-out rate (2)                           : 0.200
##
## Default parameters:
##   Type of computation                         : Schoenfeld
##   Theta H0                                    : 1
##   Planned allocation ratio                    : 1
##   Event time                                  : 12
##   Drop-out time                               : 12.00
##
## Sample size and output:
##   Lambda (1)                                  : 0.01875, 0.03000, 0.01125, 0.00750, 0.00525
##   Accrual time                                : 12.00, 13.00, 14.00, 15.00, 16.00, 40.56
##   Total accrual time                          : 40.56
##   Follow up time                              : 10.25
##   Analysis times [1]                          : 26.79
##   Analysis times [2]                          : 38.62
##   Analysis times [3]                          : 50.80
##   Expected study duration                     : 44.87
##   Maximal study duration                      : 50.80
##   Number of events by stage [1]               : 127.3
##   Number of events by stage [2]               : 270.1
##   Number of events by stage [3]               : 385.9
##   Expected number of events                   : 328.9
##   Number of subjects [1]                      : 785.4
##   Number of subjects [2]                      : 1318.1
##   Number of subjects [3]                      : 1405.0
##   Expected number of subjects                 : 1354.8
##   Reject per stage [1]                        : 0.0175
##   Reject per stage [2]                        : 0.4526
##   Reject per stage [3]                        : 0.3307
##   Overall reject                              : 0.801
##   Early stop                                  : 0.47
##   Critical values (effect scale) [1]          : 0.516
##   Critical values (effect scale) [2]          : 0.743
```

```
##   Critical values (effect scale) [3]           : 0.816
##   Local one-sided significance levels [1]      : 9.549e-05
##   Local one-sided significance levels [2]      : 7.351e-03
##   Local one-sided significance levels [3]      : 2.274e-02
##
## Legend:
##   (i): values of treatment arm i
##   [k]: values at stage k
```

Comparison: analysis time of rpact vs. gsDesign

Absolute differences:

```
timeDiff <- round(x$T - y$analysisTime,4)
rownames(timeDiff) <- c("Stage 1", "Stage 2", "Stage 3")
colnames(timeDiff) <- "Difference analysis time"
timeDiff
```

```
##         Difference analysis time
## Stage 1                     0e+00
## Stage 2                     0e+00
## Stage 3                    -1e-04
```

Remark

Obviously, there is a difference in the calculation of the necessary number of events which are, in rpact, calculated as

```
(qnorm(0.975) + qnorm(0.8))^2 / log(0.75)^2 * 4 *
        getDesignCharacteristics(getDesignGroupSequential(sided = 1, alpha = 0.025,
            kMax = 3, typeOfDesign = "asOF", informationRates = c(0.33,0.7,1)))$inflationFactor
```

```
## [1] 385.0479
```

which is slightly different to the maximum number of events in gsDesign which is

```
x$n.I[3]
```

```
## [1] 385.881
```

Therefore, running

```
getSampleSizeSurvival(
        design = design, typeOfComputation = "Schoenfeld",
        thetaH0 = 1,
        dropoutRate1 = 0.2, dropoutRate2 = 0.2, dropoutTime = 12,
        allocationRatioPlanned = 1,
        accrualTime = accrualTime,
        piecewiseSurvivalTime = piecewiseSurvivalTime,
        hazardRatio = 0.75,
        maxNumberOfSubjects = 1405)$analysisTime
```

```
##          [,1]
## [1,] 26.76183
## [2,] 38.57834
## [3,] 50.63114
```

is not exactly equal to getPowerSurvival from above. This, however, has absolutely no consequences in practice but explains the differences in rpact and gsDesign.

System: rpact 2.0.1.9003, R version 3.5.3 (2019-03-11), platform: x86_64-w64-mingw32

To cite package 'rpact' in publications use:

Gernot Wassmer and Friedrich Pahlke (2019). rpact: Confirmatory Adaptive Clinical Trial Design and Analysis. R package version 2.0.1.9003. https://www.rpact.org

License