Introduction to the Design and Evaluation of Group Sequential Clinical Trials

Session 3 - Evaluation of Group Sequential Designs

Presented July 27, 2016

Daniel L. Gillen
Department of Statistics
University of California, Irvine

John M. Kittelson
Department of Biostatistics & Informatics
University of Colorado Denver

Statistical basis for stopping criteria

Recall: reasons to monitor trial endpoints

- To maintain the validity of the informed consent for:
  - Subjects currently enrolled in the study.
  - New subjects entering the study.
- To ensure the ethics of randomization.
  - Randomization is only ethical under equipoise.
  - If there is not equipoise, then the trial should stop.
- To identify the best treatment as quickly as possible:
  - For the benefit of all patients (i.e., so that the best treatment becomes standard practice).
  - For the benefit of study participants (i.e., so that participants are not given inferior therapies for any longer than necessary).

SISCR
UW - 2016

Statistical basis for stopping

When do we have enough information to make a decision?

- Sepsis trial example:
  - Statistical standards for evidence in the fixed-sample trial
  - How might we implement those same standards at an interim analysis?

Statistical basis for stopping criteria

SISCR
UW - 2016

Recall sepsis trial fixed-sample design

- Primary outcome (28-day mortality):
  - $Y_{ki} \sim \mathcal{B}(1, \theta_k)$ for *ith* patient in treatment group $k = 0, 1$
- Within-group summary measure: θ_k
- Between-group contrast: $\theta = \theta_1 - \theta_0$
- Design hypotheses (1-sided superiority test):

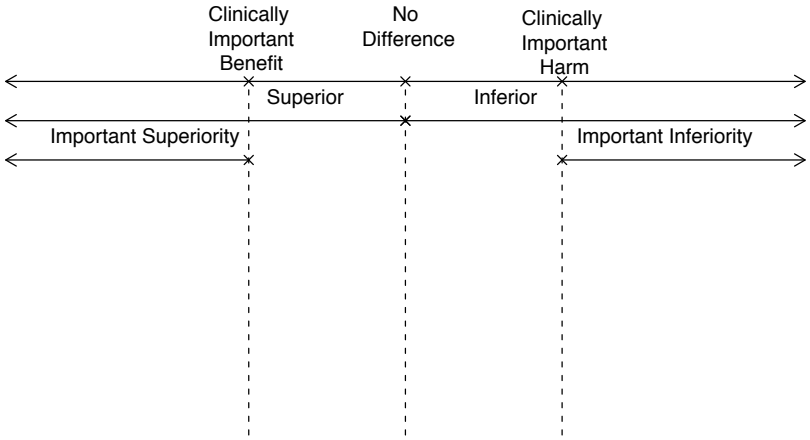
$$\begin{aligned} \text{Null:} \quad & \theta \geq 0 \\ \text{Alternative:} \quad & \theta \leq -0.07 \end{aligned}$$

- Sample size: 1700 patients (850 per group) gives:
  - $\beta = 0.907$ for $\theta = -0.07$ if $\theta_0 = 0.3$.

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

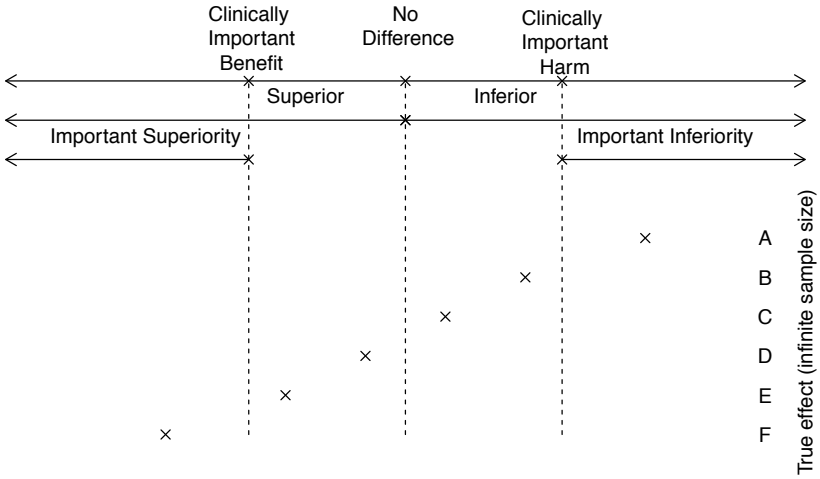
Statistical basis for stopping criteria
Example: sepsis trial

- Scientific/clinical structuring of parameter space



Statistical basis for stopping criteria
Example: sepsis trial

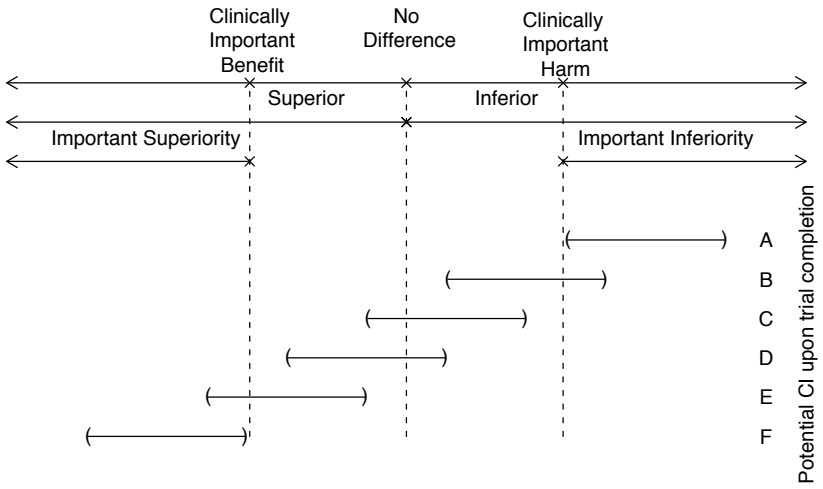
- Inference with an infinite sample size



- E, F $\Rightarrow$ Use new antibody
- D $\Rightarrow$ Is it worthwhile if benefits are unimportant?
- A, B, C $\Rightarrow$ Do not use new antibody

Statistical basis for stopping criteria
Example: sepsis trial

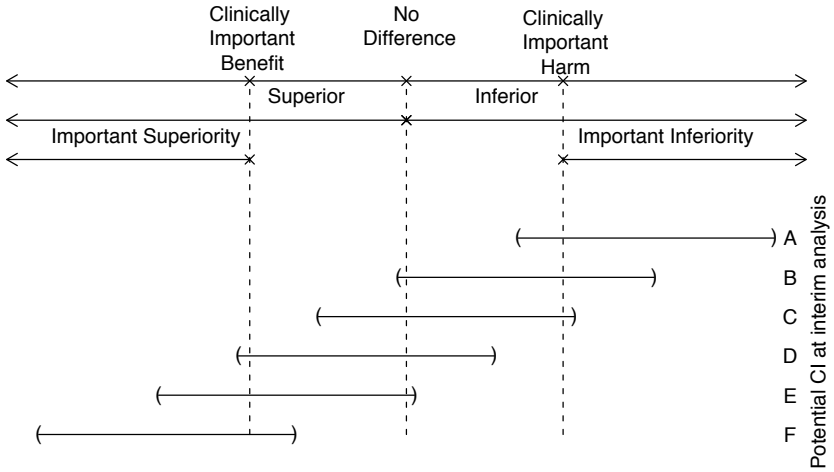
- Possible conclusions upon trial completion



- E, F $\Rightarrow$ Use new antibody
- A, B, C, D $\Rightarrow$ Do not use new antibody

Statistical basis for stopping criteria
Example: sepsis trial

- Possible conclusions at interim analysis



- F $\Rightarrow$ Stop?: use new antibody
- D, E $\Rightarrow$ Continue trial
- A, B, C $\Rightarrow$ Stop?: do not use new antibody

Fixed-sample design in RCTdesign

Sepsis design from session 2 (but using $\theta_+ = -0.07$ instead of -0.05):

```
> SepsisFixed <- seqDesign( prob.model = "proportions", arms = 2,
+        null.hypothesis = .3, alt.hypothesis = 0.23, alpha = 0.025,
+         ratio = c(1., 1.), nbr.analyses = 1, test.type = "less",
+                 sample.size=1700, power = "calculate",)

> SepsisFixed

Call:
seqDesign(prob.model = "proportions", arms = 2, null.hypothesis = 0.3,
    alt.hypothesis = 0.23, ratio = c(1, 1), nbr.analyses = 1,
    sample.size = 1700, test.type = "less", power = "calculate",
    alpha = 0.025)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in probabilities (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=  0.00    (size  = 0.0250)
    Alternative hypothesis : Theta <= -0.07    (power = 0.9066)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N= 1700)  -0.0418  -0.0418
```


Adding interim analyses in RCTdesign

SISCR
UW - 2016

Sepsis trial: adding interim analyses

- RCTdesgn will automatically add interim analyses
- Defaults:
  - Equally-spaced analyses
  - Emerson-Fleming symmetric designs
  - O'Brien-Fleming boundary shape

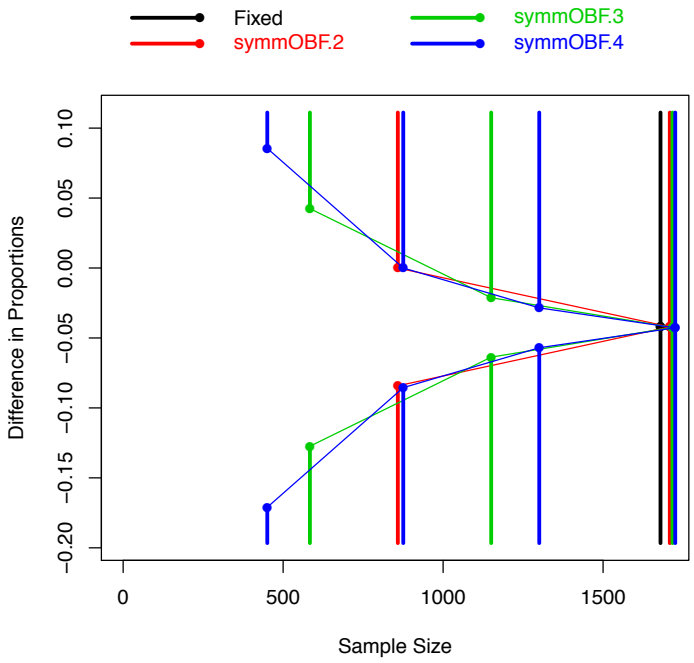
```
> symmOBF.2 <- update(binomFixed,nbr.analyses=2)
> symmOBF.3 <- update(binomFixed,nbr.analyses=3)
> symmOBF.4 <- update(binomFixed,nbr.analyses=4)
```

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Sepsis trial: adding interim analyses

Stopping bounds for `symmOBF.2, symmOBF.3, symmOBF.4`:

```
> seqPlotBoundary(symmOBF.2, symmOBF.3, symmOBF.4)
```



Sepsis trial: adding interim analyses

Stopping bounds for `symmOBF.2, symmOBF.3, symmOBF.4`:

Interim Analysis	Stop for Efficacy	Stop for Futility
`symmOBF.2:`		
N= 850	-0.0842	0.0000
N=1700	-0.0421	-0.0421
`symmOBF.3:`		
N= 567	-0.1274	0.0425
N= 850	-0.0637	-0.0212
N=1700	-0.0425	-0.0425
`symmOBF.4:`		
N= 425	-0.1710	0.0855
N= 567	-0.0855	0.0000
N= 850	-0.0570	-0.0285
N=1700	-0.0427	-0.0427

SISCR
UW - 2016

Sepsis trial: adding interim analyses

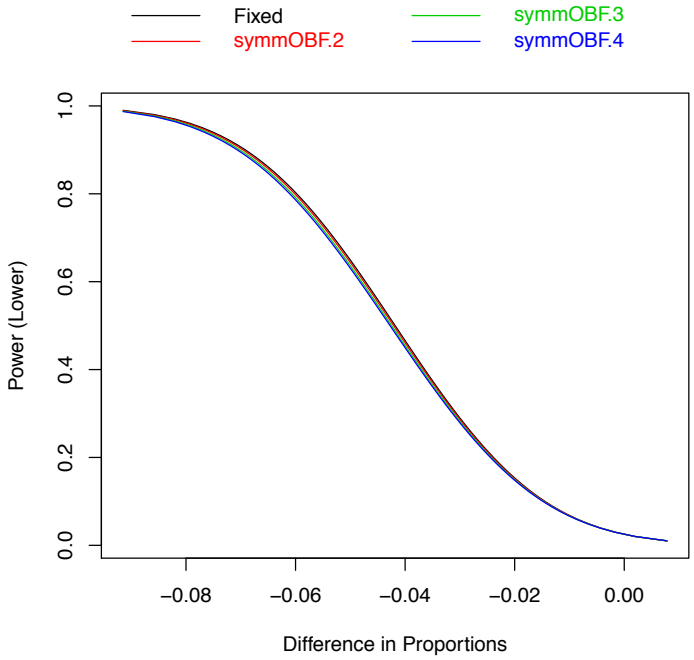
Effect of adding interim analyses

- Power decreases (unless sample size is increased)
- Expected sample size gets smaller

Effect of interim analyses on trial power

Does the number of interim analyses affect trial power?

```
> seqPlotPower(symmOBF.2,symmOBF.3,symmOBF.4)
```



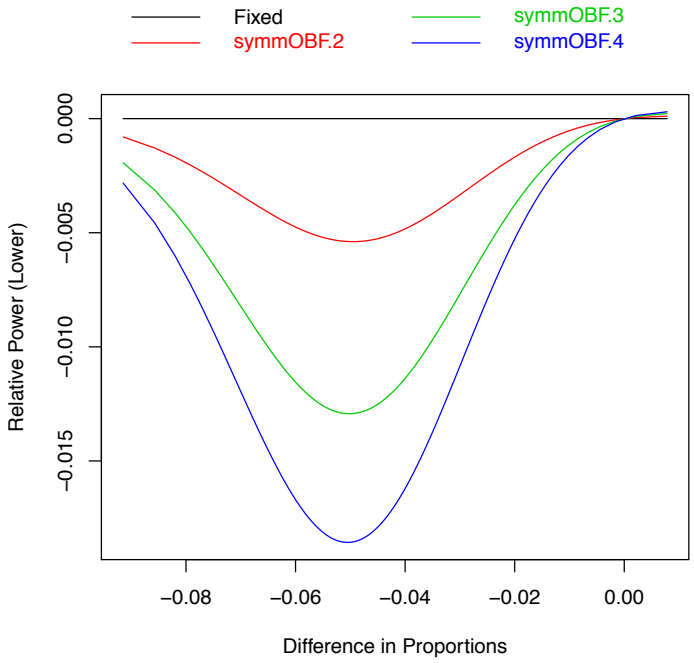
SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
***Sepsis trial:** add interim analyses
***Sepsis trial:** number of boundaries
***Sepsis trial:** early conservatism
***Sepsis trial:** power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Effect of interim analyses on trial power

Power difference from fixed-sample design

```
> seqPlotPower(symmOBF.2,symmOBF.3,symmOBF.4,reference=T)
```

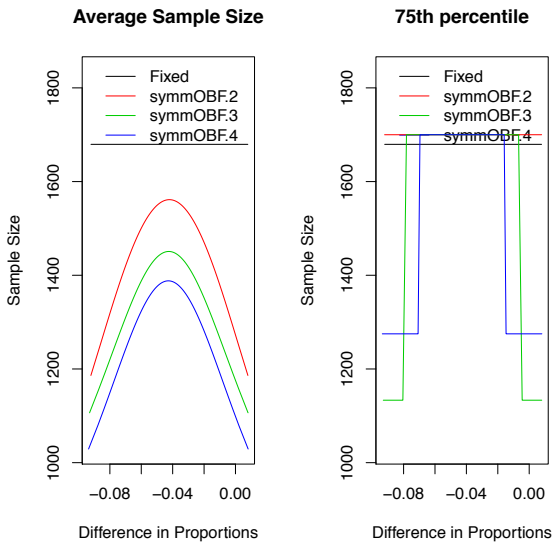


Effect of interim analyses on sample size

Does the number of interim analyses affect the sample size?

- Number of patients is a random variable summaries:
  - Average sample number (ASN)
  - 75th percentile of sample size distribution

```
> seqPlotASN(symmOBF.2,symmOBF.3,symmOBF.4)
```



Sepsis trial: reasons for stopping

Selecting reasons for early termination

- Stop for either efficacy or futility (e.g., `symmOBF.4`).
- Stop only for futility:

```
> futOnlyOBF.4 <- update(binomFixed,nbr.analyses=4,
        early.stopping="null")
```

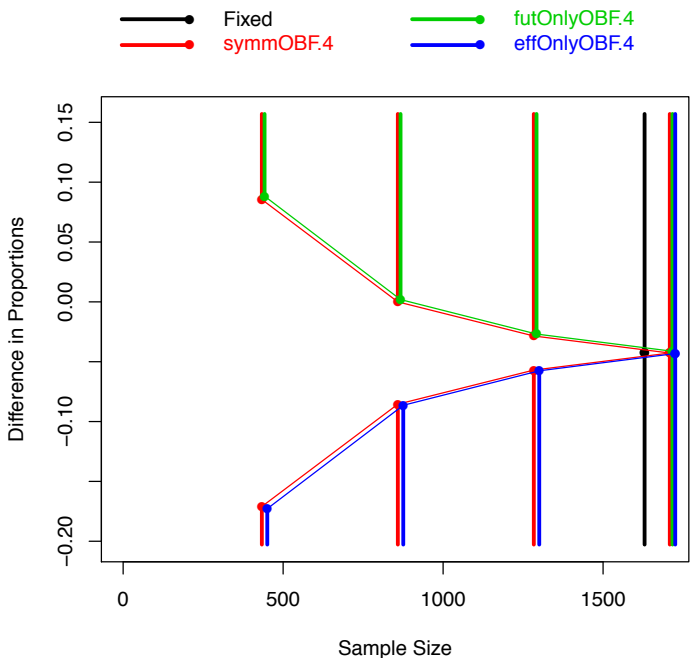
- Stop only for efficacy:

```
> effOnlyOBF.4 <- update(binomFixed,nbr.analyses=4,
        early.stopping="alt")
```


Sepsis trial: reasons for stopping

Stopping bounds for `symmOBF.4, futOnlyOBF.3, effOnlyOBF.4`:

```
> seqPlotBoundary(symmOBF.4,futOnlyOBF.4,effOnlyOBF.4)
```



Sepsis trial: reasons for stopping

Stopping bounds for `symmOBF.4, futOnlyOBF.3, effOnlyOBF.4:`

Interim Analysis	Stop for Efficacy	Stop for Futility
`symmOBF.4:`		
N= 425	-0.1710	0.0855
N= 567	-0.0855	0.0000
N= 850	-0.0570	-0.0285
N=1700	-0.0427	-0.0427
`futOnlyOBF.4:`		
N= 425	-Inf	0.0883
N= 567	-Inf	0.0019
N= 850	-Inf	-0.0269
N=1700	-0.0413	-0.0413
`effOnlyOBF.4:`		
N= 425	-0.1728	Inf
N= 567	-0.0864	Inf
N= 850	-0.0576	Inf
N=1700	-0.0432	-0.0432

Sepsis trial: reasons for stopping

Effect of stopping for one or more hypothesis

- Stopping for both null and alternative hypothesis:
  - Symmetric power for futility and efficacy decisions
  - Symmetric ASN for futility and efficacy decisions
- Stopping for futility (null hypothesis):
  - Power for efficacy *may* decrease
  - ASN reduced for futility, but not for efficacy
- Stopping for efficacy (alternative hypothesis):
  - Power for efficacy *may* decrease
  - ASN reduced for efficacy, but not for futility

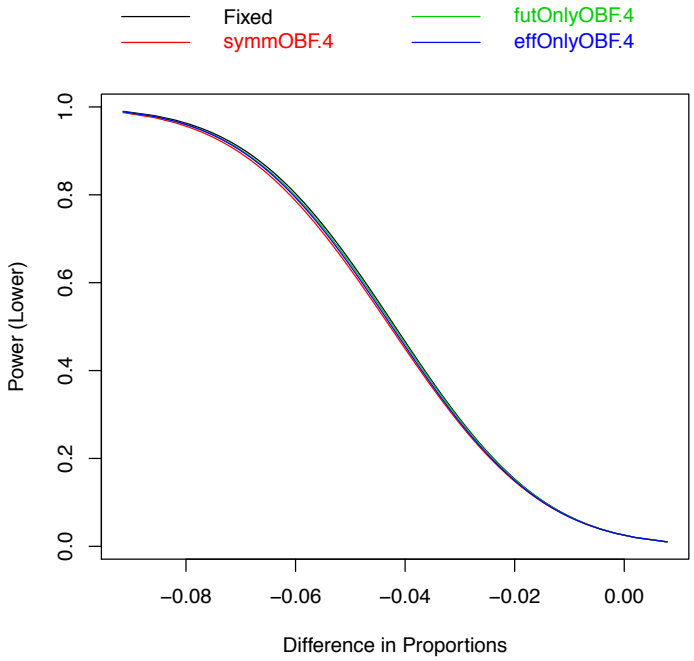
SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
***Sepsis trial:** add interim analyses
***Sepsis trial:** number of boundaries
***Sepsis trial:** early conservatism
***Sepsis trial:** power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Effect of number of boundaries on trial power

Does the number of boundaries affect trial power?

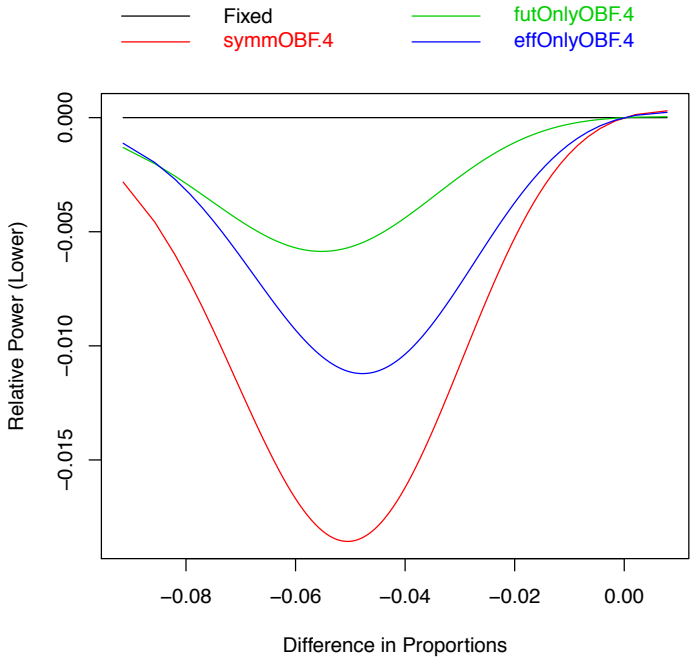
```
> seqPlotPower(symmOBF.4,futOnlyOBF.4, effOnlyOBF.4)
```



Effect of number of boundaries on trial power

Power difference from fixed-sample design

```
> seqPlotPower(symmOBF.4,futOnlyOBF.4, effOnlyOBF.4,reference=T)
```

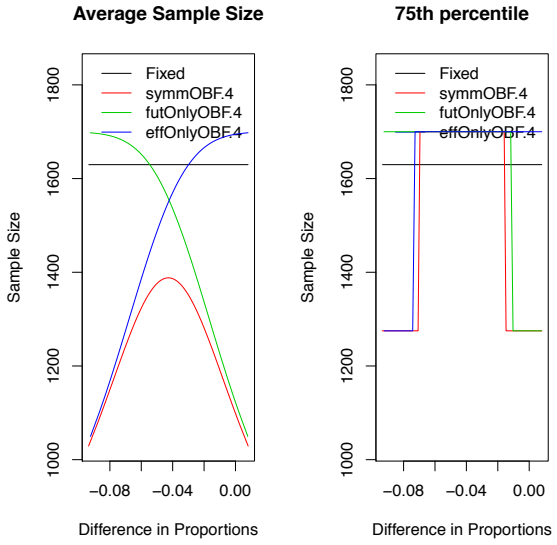


Effect of number of boundaries on sample size

Does the number of boundaries affect the sample size?

- Number of patients is a random variable summaries:
  - Average sample number (ASN)
  - 75th percentile of sample size distribution

```
> seqPlotASN(symmOBF.4,futOnlyOBF.4, effOnlyOBF.4)
```



SISCR
UW - 2016

Sepsis trial: early conservatism

Selecting degree of early conservatism

- An important design consideration is whether it should be relatively easy or hard to stop at an early interim analysis:
  - O'Brien-Fleming design shows early conservatism: (i.e., relatively difficult to stop at early interim analyses).

    The following give identical designs (due to default settings):

```
> symmOBF.4 <- update(binomFixed,nbr.analyses=4)
> symmOBF.4 <- update(binomFixed,nbr.analyses=4,
        P=c(1,1))
```

  - Pocock design is not conservative in early decisions. (i.e., relatively easy to stop at early interim analyses).

```
> symmPOC.4 <- update(binomFixed,nbr.analyses=4,
        P=c(0.5,0.5))
```

  - Degree of conservatism does not have to be symmetric.

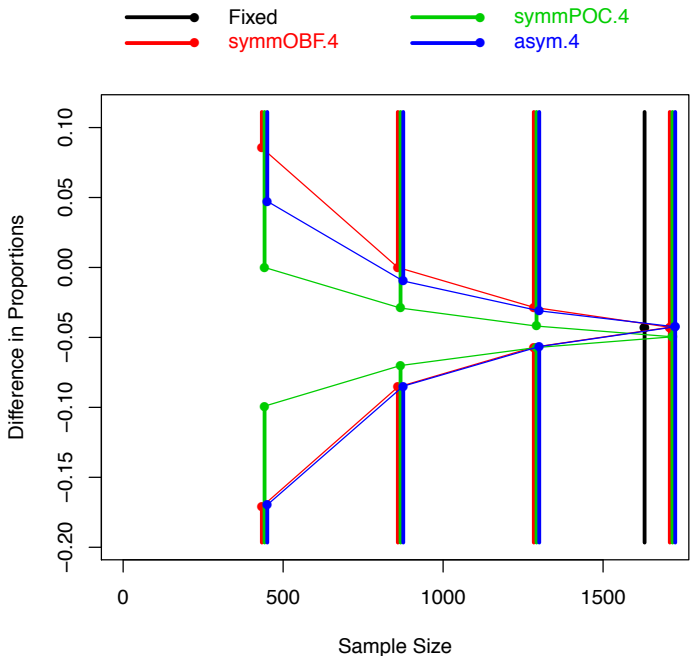
```
> asym.4 <- update(binomFixed,nbr.analyses=4,
        P=c(1,0.8))
```

SISCR
UW - 2016

Sepsis trial: early conservatism

Stopping bounds for `symmOBF.4, symmPOC.4, asym.4:`

```
> seqPlotBoundary(symmOBF.4,symmPOC.4,asym.4)
```



Sepsis trial: early conservatism

Stopping bounds for `symmOBF.4, symmPOC.4,asym.4:`

Interim Analysis	Stop for Efficacy	Stop for Futility
`symmOBF.4:`		
N= 425	-0.1710	0.0855
N= 567	-0.0855	0.0000
N= 850	-0.0570	-0.0285
N=1700	-0.0427	-0.0427
`symmPOC.4:`		
N= 425	-0.0991	0.0000
N= 567	-0.0701	-0.0290
N= 850	-0.0572	-0.0419
N=1700	-0.0496	-0.0496
`asym.4:`		
N= 425	-0.1697	0.0473
N= 567	-0.0848	-0.0097
N= 850	-0.0566	-0.0310
N=1700	-0.0424	-0.0424

Sepsis trial: early conservatism

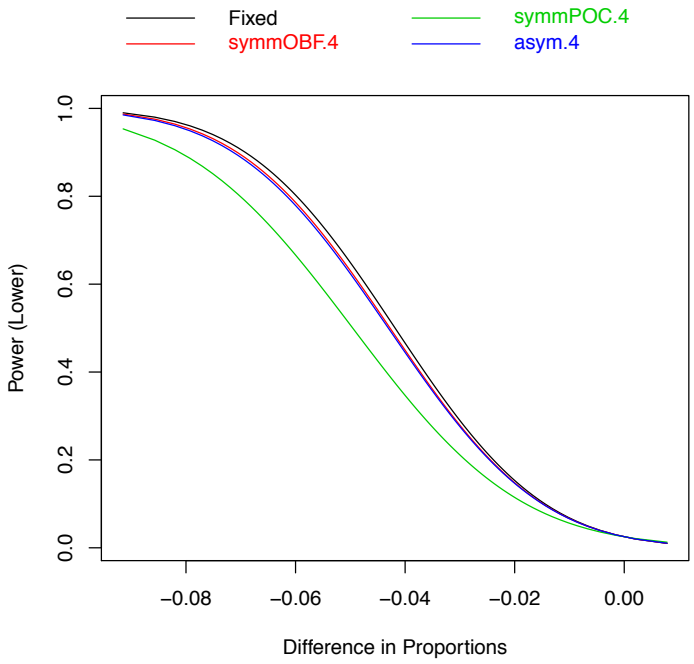
Effect of early conservatism

- More conservatism (harder to stop at early analyses:
  - Tends to give higher power
  - Tends to give larger ASN
- Less conservatism (easier to stop):
  - Tends to decrease power
  - Tends to reduce ASN
- Asymmetric conservatism:
  - Often need early sensitivity for harm, but conservatism for efficacy

Effect of early conservatism on trial power

Does the degree of early conservatism affect trial power?

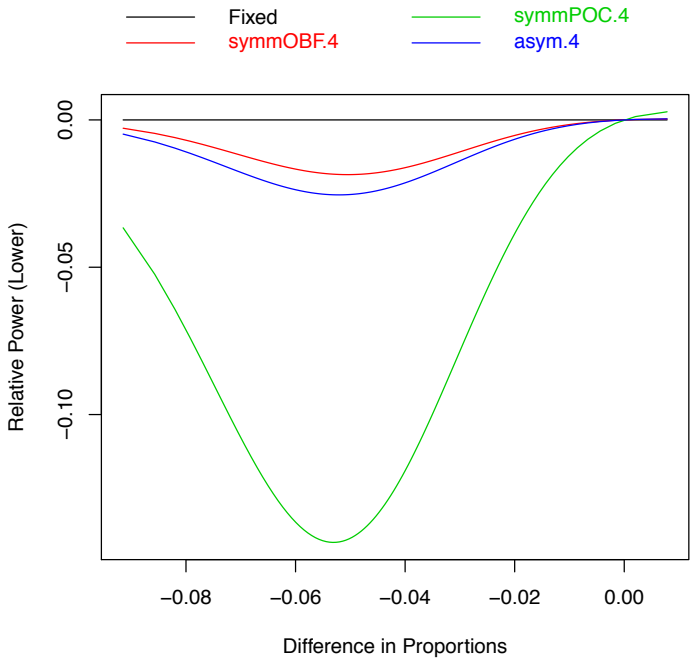
```
> seqPlotPowerseqPlotPower(symmOBF.4,symmPOC.4,asym.4)
```



Effect of early conservatism on trial power

Power difference from fixed-sample design

```
> seqPlotPower(symmOBF.4,symmPOC.4,asym.4,reference=T)
```



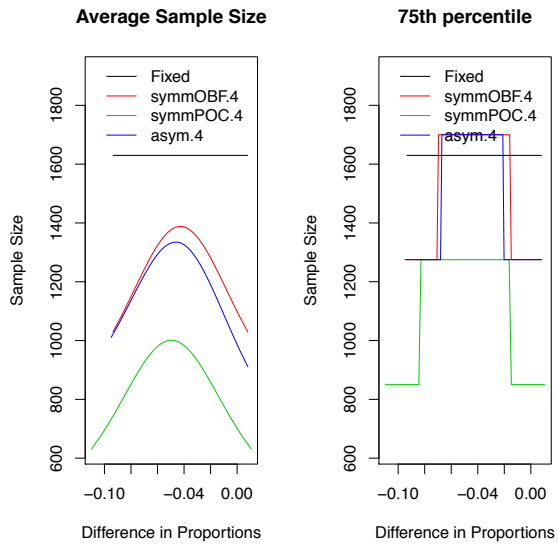
SISCR
UW - 2016

Effect of early conservatism on sample size

Does early conservatism affect the sample size?

- Number of patients is a random variable summaries:
  - Average sample number (ASN)
  - 75th percentile of sample size distribution

```
> seqPlotASN(symmOBF.4,symmPOC.4,asym.4)
```



SISCR
UW - 2016

Sepsis trial: power vs maximal sample size

SISCR
UW - 2016

Boundary shape

- Above designs use $N = 1700$:
  - Different group sequential designs have different power
- N can be chosen to give equal power
- For example, compare
  `symmOBF.4, symmPOC.4,symmPOCpower.4`:

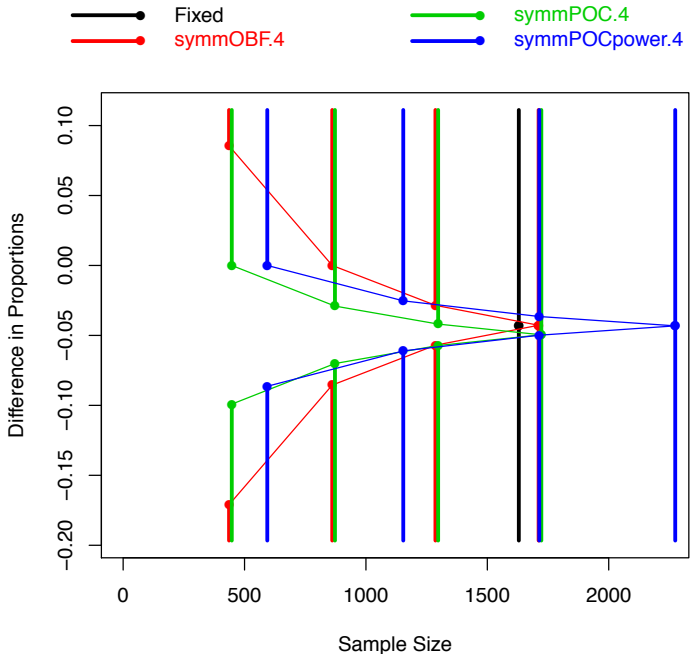
```
> symmPOCpower.4 <- update(symmPOC.4,power=0.8945)
```


Sepsis trial: power vs maximal sample size

SISCR
UW - 2016

Stopping bounds for `symmOBF.4, symmPOC.4, symmPOCpower.4`:

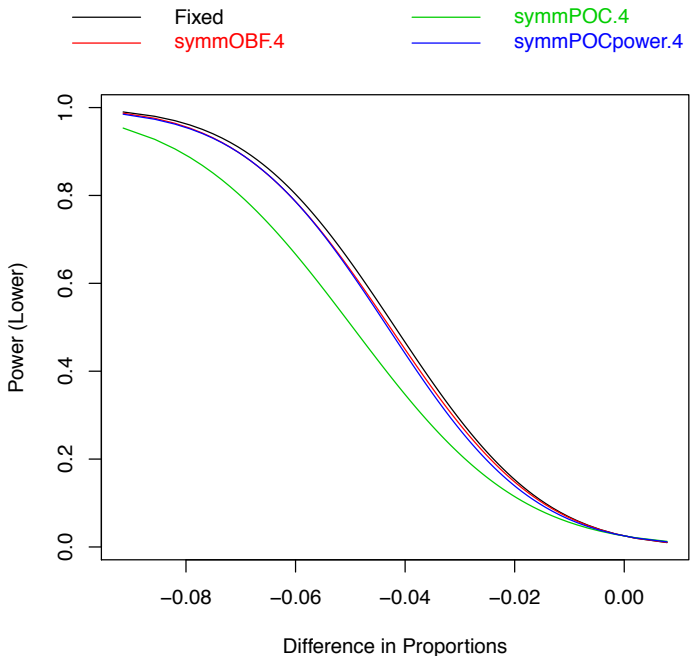
```
> seqPlotBoundary(symmOBF.4,symmPOC.4,symmPOCpower.4)
```



Sepsis trial: power vs maximal sample size

Power for `symmOBF.4, symmPOC.4, symmPOCpower.4`:

```
> seqPlotPower(symmOBF.4,symmPOC.4,symmPOCpower.4)
```

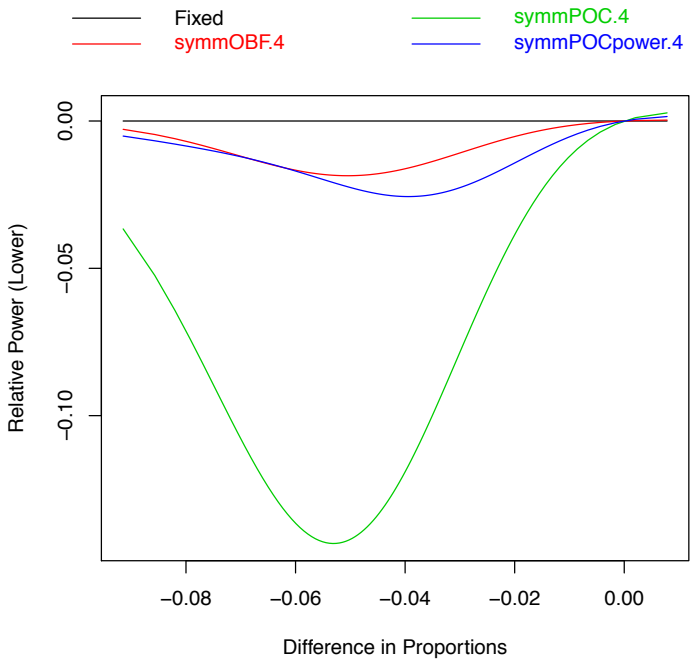


Sepsis trial: power vs maximal sample size

Power difference from fixed-sample design

:

```
> seqPlotPower(symmOBF.4,symmPOC.4,symmPOCpower.4,reference=T)
```



General characteristics of group sequential designs

Specifying interim decision criteria

- Key considerations (illustrated in sepsis example):
  - Boundary structure
    - Boundary scale
    - Number and timing of interim analyses
    - Boundary shape
    - Number of boundaries: reasons for early termination
  - Statistical operating characteristics
  - Design properties (ASN, stopping probabilities)

Boundary structure

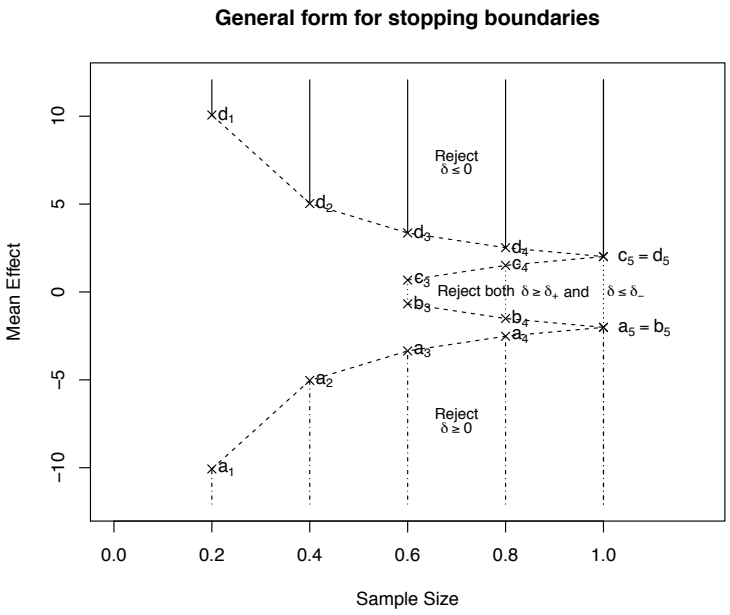
General structure for stopping rules

- Number and timing of analyses
  - N counts the sampling units accrued to the study (with outcome measurements)
  - Up to N analyses of the data to be performed
  - Analyses performed after accruing sample sizes of $N_1 < N_2 < \cdots N_J$
  - (More generally, N measures statistical information)
- Boundaries (decision criteria) at the analyses
  - $a_j \leq b_j \leq c_j \leq d_j$ where the a, b, c and d are boundaries at the i-the analysis (when N_j observations)
  - At the final (J-th) analysis $a_J = b_J$ and $c_J = d_J$ to guarantee stopping

Boundary structure

General structure for stopping rules

Illustration of general structure:



General structure: boundary scales

Boundary scales

- Stopping boundaries can be defined on a variety of scales
  - Sum of observations
  - Point estimate of treatment effect
  - Normalized (Z) statistic
  - Fixed-sample P value
  - Error spending function
  - Conditional probability
  - Predictive probability
  - Bayesian posterior probability

General structure: boundary scales

Utility of scales when evaluating designs

- Several of the boundary scales have interpretations that are useful in evaluating the operating characteristics of a design
  - Sample mean scale
  - Conditional probability futility scales
  - Predictive probability futility scale
  - Bayesian posterior probability scale
  - (Error spending scale)

General structure: boundary shape and location

Boundary shape functions

- Π_j measures the proportion of total information accrued at the jth analysis
  - Often $\Pi_j = \frac{N_j}{N_J}$
- Boundary shape function $f(\Pi_j)$ is a monotonic function used to relate the dependence of boundaries at successive analyses on the information accrued to the study at that analysis

General structure: boundary shape and location

General structure of decision boundaries

- Stopping boundaries for the sample mean statistic:
  - $a_j = \theta_a - f_a(\Pi_j)$
  - $b_j = \theta_b + f_b(\Pi_j)$
  - $c_j = \theta_c - f_c(\Pi_j)$
  - $d_j = \theta_d + f_d(\Pi_j)$

  where θ_* represents the hypothesis rejected by the corresponding boundary:

$$\begin{aligned} \hat{\theta}_j \leq a_j \quad & \text{rejects} \quad \theta \geq \theta_a \\ \hat{\theta}_j \geq b_j \quad & \text{rejects} \quad \theta \leq \theta_b \\ \hat{\theta}_j \leq c_j \quad & \text{rejects} \quad \theta \geq \theta_c \\ \hat{\theta}_j \geq d_j \quad & \text{rejects} \quad \theta \leq \theta_d \end{aligned}$$

General structure: boundary shape and location

Boundary shape function (unified family)

- Parameterization of boundary shape (unified family):

$$f_*(\Pi_j) = \left[A_* + \Pi_j^{-P_*}(1-\Pi_j)^{-R_*}\right] \times G_*$$

- Distinct parameters possible for each boundary
- Parameters A_*, P_*, and R_* are typically specified by trialist
- Critical value G_* usually calculated by computer search using sequential sampling density

General structure: boundary shape and location

Unified design family

- Choice of P parameter ($P \geq 0$):
  - Larger values of P make early stopping more difficult (impossible when P infinite)
  - When $A = R = 0$:

$$f_*(\Pi_j) = G_* \Pi_j^{-P_*}$$

    - $P = 0.5$ gives Pocock (1977) type boundary shapes (constant on Z scale)
    - $P = 1.0$ gives O'Brien-Fleming (1979) type boundary shapes (constant on partial sum scale)
    - $0.5 < P < 1$ corresponds to power family (Δ) in Wang and Tsiatis (1987): $P = 1 - \Delta$
    - Reasonable range of values: $0 < P < 2.5$
    - $P = 0$ with $A = R = 0$ possible for some (not all) boundaries, but not particularly useful
  - Illustrations to follow...

SISCR
UW - 2016

General structure: finite termination constraint

Constraints to assure termination at the Jth interim analysis and appropriate operating characteristics:

- Finite termination constraint:

$$\begin{aligned} a_J = b_J \quad &\Rightarrow \quad \theta_a - \theta_b = f_a(1) + f_b(1) \\ c_J = d_J \quad &\Rightarrow \quad \theta_c - \theta_d = f_c(1) + f_d(1) \\ a_J \leq d_J \quad &\Rightarrow \quad \theta_a - \theta_d \leq f_a(1) + f_d(1) \end{aligned}$$

SISCR
UW - 2016

General structure: finite termination constraint

Constraints to assure termination at the Jth interim analysis and appropriate operating characteristics:

- We then select G_a, G_b, G_c, G_d in a 4-parameter search to satisfy the following operating characteristics:

$$\begin{aligned} P[\hat{\theta}_M \leq a_M | \theta = \theta_a] &= \beta_\ell \\ P[\hat{\theta}_M \geq b_M | \theta = \theta_b] &= 1 - \alpha_\ell \\ \\ P[\hat{\theta}_M \leq c_M | \theta = \theta_c] &= 1 - \alpha_u \\ P[\hat{\theta}_M \geq d_M | \theta = \theta_d] &= \beta_u \end{aligned}$$

where:

- M denotes the random time at which the trial stopped
- α_ℓ, β_ℓ denote the size and power for the lower test
- α_u, β_u denote the size and power for the upper test

Stopping rules: Unified family

Example: symmetric tests (Emerson & Fleming (1989)

- Symmetric tests are an important class of designs with
  - Symmetric operating characteristics:

$$\alpha_\ell = \alpha_u = (1 - \beta_\ell) = (1 - \beta_u)$$

  - Symmetric boundary shapes
    (less important, but useful for illustration)

$$f_a(\Pi_j) = f_b(\Pi_j) = f_c(\Pi_j) = f_d(\Pi_j) = f(\Pi_j)$$

  - It then follows that
    $G_a = G_b = G_c = G_d = G$
  - So that symmetric designs have the form:

$$\begin{aligned} a_j &= -f(\Pi_j) \\ b_j &= -\theta_* + f(\Pi_j) \\ \\ c_j &= \theta_* - f(\Pi_j) \\ d_j &= f(\Pi_j) \end{aligned}$$

where $\theta_* = 2G$

SISCR
UW - 2016

Common designs: JK's canonical classes

- There are an infinite number of group sequential designs for any particular trial
- Unified family provides general framework
- There are some natural classes that help to organize the possibilities
  - Why stop early (revisited):
    - Superiority study
    - Approximate equivalence study
    - Non-inferiority study
    - Equivalence (2-sided hypothesis) study
  - Standardized design scale
  - Common boundary forms:
    - Superiority study
    - Approximate equivalence study
    - Non-inferiority study
    - Equivalence (2-sided hypothesis) study

Reasons for early termination

- Setting (parameterization of the problem)
  - Treatment effect measure: θ
  - Suppose:
    - Larger θ means that active treatment is superior.
    - $\theta = 0$ denotes no difference between active and control treatment.
    - $\theta \geq \theta_+$ denotes clinically important superiority of active treatment.
    - $\theta \leq \theta_-$ denotes clinically important inferiority of active treatment.
      [Where $\theta_- < 0 < \theta_+$]

Common design classes

Reasons for early termination

- Why would you want to stop a study early?
  - Superiority study:
    - For superiority (reject $H_0 : \theta \leq 0$)
    - For lack of superiority (reject $H_A : \theta \geq \theta_+$)
  - Approximate equivalence study:
    - For lack of inferiority (reject $H_0 : \theta \leq \theta_-$)
    - For lack of superiority (reject $H_A : \theta \geq \theta_+$)
  - Non-inferiority study:
    - For lack of inferiority (reject $H_0 : \theta \leq \theta_-$)
    - For inferiority (reject $H_A : \theta \geq 0$)
  - Equivalence (2-sided) study:
    - For superiority (reject $\theta \leq 0$)
    - For inferiority (reject $\theta \geq 0$)
    - For both non-inferiority and non-superiority (reject both $\theta \leq \theta_-$ and $\theta \geq \theta_+$)

SISCR
UW - 2016

Common design classes

Standardized scale

In what follows I present a standardized design. It can be mapped to any specific design.

- Standardization:
  - Without interim stopping, but with sample sizes $N_1 < N_2, ..., < N_J$):

$$\hat{\theta}_j \dot{\sim} \mathcal{N}\left(\theta, \frac{V}{N_j}\right)$$

    where V is the variance (follows from probability model)
  - Let:

$$\hat{\delta}_j = \frac{\hat{\theta}_j - \theta_\emptyset}{\sqrt{V/N_J}}$$

  - Thus:

$$\hat{\delta}_j \dot{\sim} \mathcal{N}\left(\delta, \frac{1}{\Pi_j}\right)$$

    where $\Pi_j = \frac{N_j}{N_J}$.

SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
***Sepsis trial:** add interim analyses
***Sepsis trial:** number of boundaries
***Sepsis trial:** early conservatism
***Sepsis trial:** power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Common design classes

SISCR
UW - 2016

Boundary form in standardized scale

- In general there are 4 potential boundaries in a group sequential design which I denote by $a_j \leq b_j \leq c_j \leq d_j$ $(j = 1, ..., J)$:

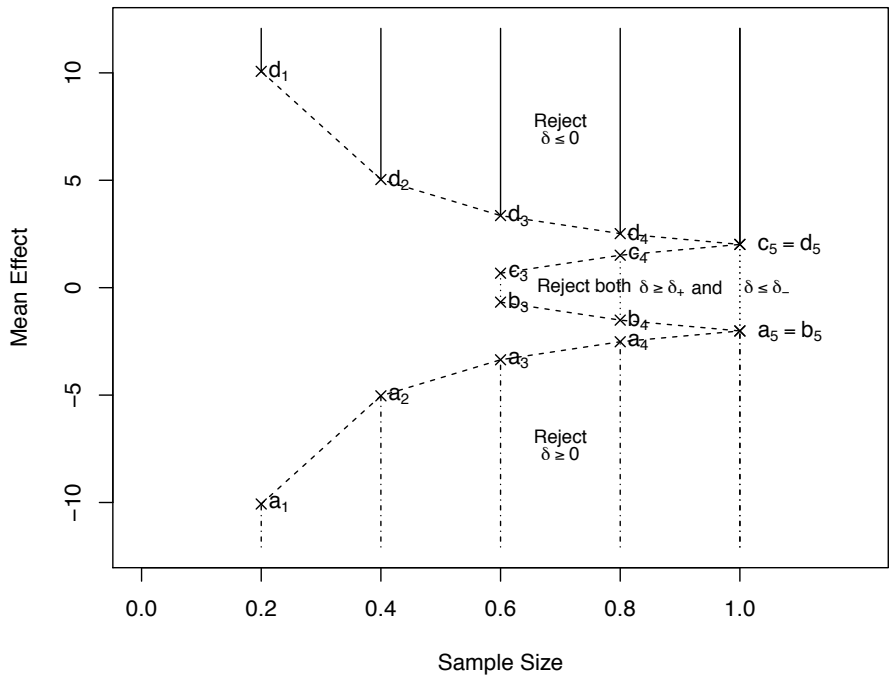
$$\begin{array}{lcll} \hat{\delta}_j \geq d_j & \rightarrow & \text{Reject } \delta \leq \delta_d & (\text{usually } \delta_d = 0) \\ \hat{\delta}_j \leq c_j & \rightarrow & \text{Reject } \delta \geq \delta_c & (\text{usually } \delta_c = \delta_+) \\ \hat{\delta}_j \geq b_j & \rightarrow & \text{Reject } \delta \leq \delta_b & (\text{usually } \delta_b = \delta_-) \\ \hat{\delta}_j \leq a_j & \rightarrow & \text{Reject } \delta \geq \delta_a & (\text{usually } \delta_a = 0) \end{array}$$

  with $\delta_- < 0 < \delta_+$ (often $\delta_- = -\delta_+$).
- Set $d_J = c_J$ and $a_J = b_J$ so that the trial has to terminate by analysis J.

Common design classes
Boundary form (number and location)

SISCR
UW - 2016

General form for stopping boundaries



Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Boundary form (number and location)

Superiority study

- Stop for superiority:

$$\hat{\delta}_j \geq d_j \rightarrow \text{ Reject } \delta \leq 0$$

- Stop for non-superiority:

$$\hat{\delta}_j \leq a_j \rightarrow \text{ Reject } \delta \geq \delta_+$$

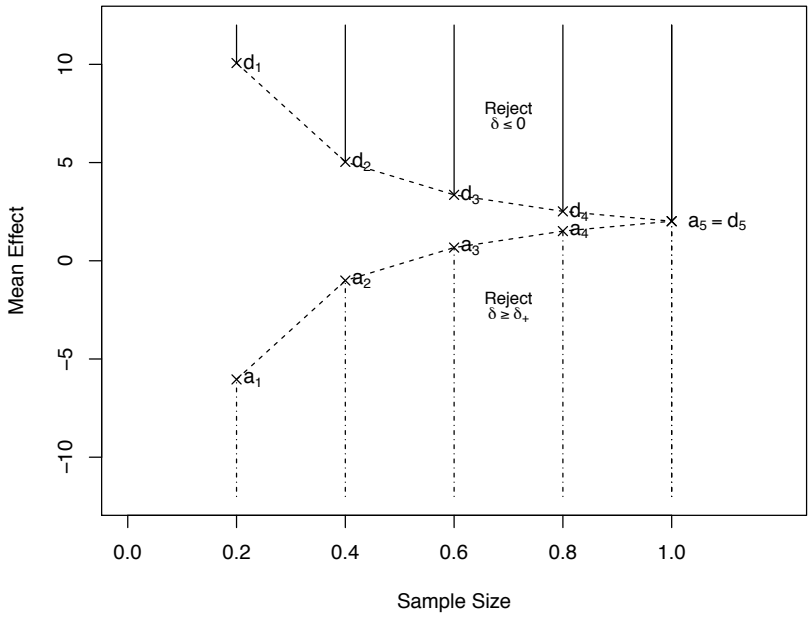
- Stop for either superiority or non-superiority:

$$\begin{array}{lll} \hat{\delta}_j \geq d_j & \rightarrow & \text{Reject } \delta \leq 0 \\ \hat{\delta}_j \leq a_j & \rightarrow & \text{Reject } \delta \geq \delta_+ \end{array}$$

Boundary form (number and location)

A superiority design is obtained by an upward shift of the *a*- and *b*-boundaries.

General form for superiority boundaries



Boundary form (number and location)

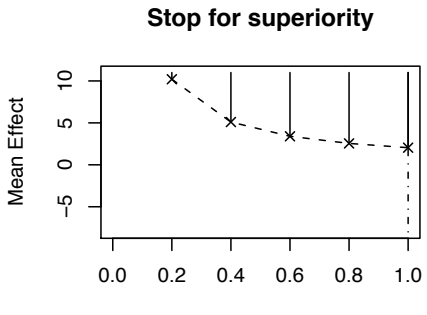
Superiority study

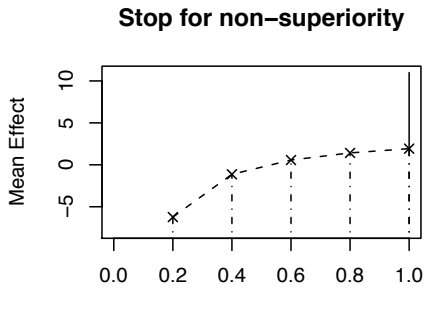
- RCTdesign:

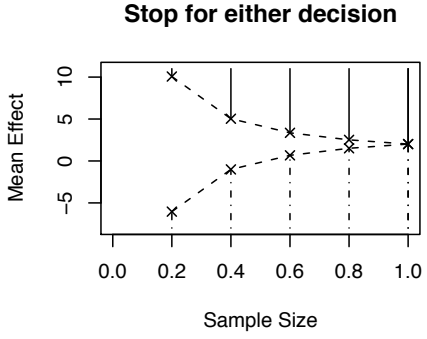
```
> sup.D <- seqDesign(prob.model = "normal", arms = 1,
+         null.hypothesis = 0., alt.hypothesis = 3.92,
+         variance = 1., sample.size = 1, test.type = "greater",
+         nbr.analyses = 5, power = "calculate",alpha = 0.025,
+         epsilon = c(0., 1.),early.stopping = "alternative",
+         display.scale = seqScale(scaleType = "X"))
> sup.A <- update(sup.D,early.stopping="null")
> sup.DA <- update(sup.D,early.stopping="both")
```


Boundary form (number and location)

Superiority study designs







Boundary form (number and location)

SISCR
UW - 2016

Non-inferiority study

- Stop for non-inferiority:

$$\hat{\delta}_j \geq d_j \rightarrow \text{ Reject } \delta \leq \delta_-$$

- Stop for inferiority:

$$\hat{\delta}_j \leq a_j \rightarrow \text{ Reject } \delta \geq 0$$

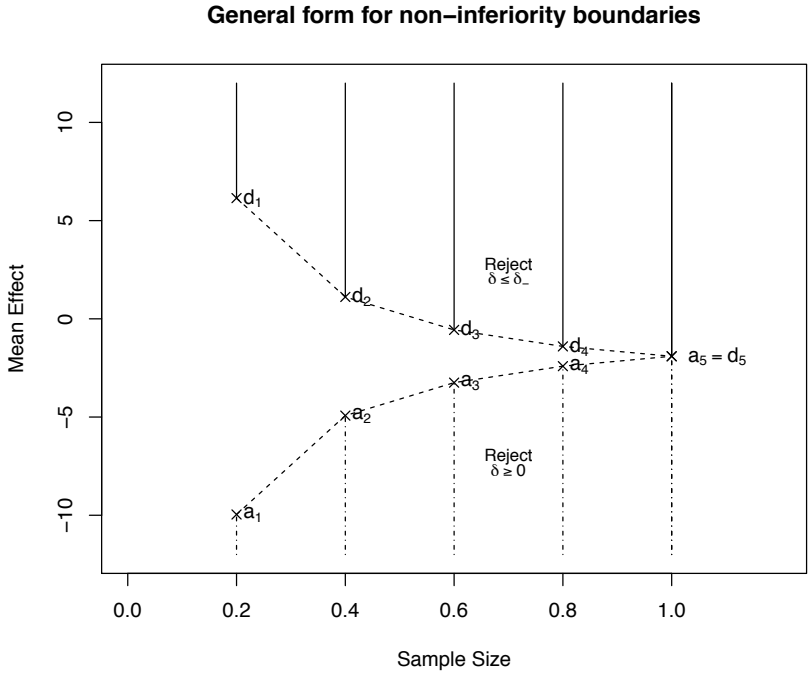
- Stop for either inferiority or non-inferiority:

$$\begin{aligned} \hat{\delta}_j \geq d_j &\rightarrow \text{ Reject } \delta \leq \delta_- \\ \hat{\delta}_j \leq a_j &\rightarrow \text{ Reject } \delta \geq 0 \end{aligned}$$

Boundary form (number and location)

SISCR
UW - 2016

A non-inferiority design is obtained by a downward shift of the *c*- and *d*-boundaries.



Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Boundary form (number and location)

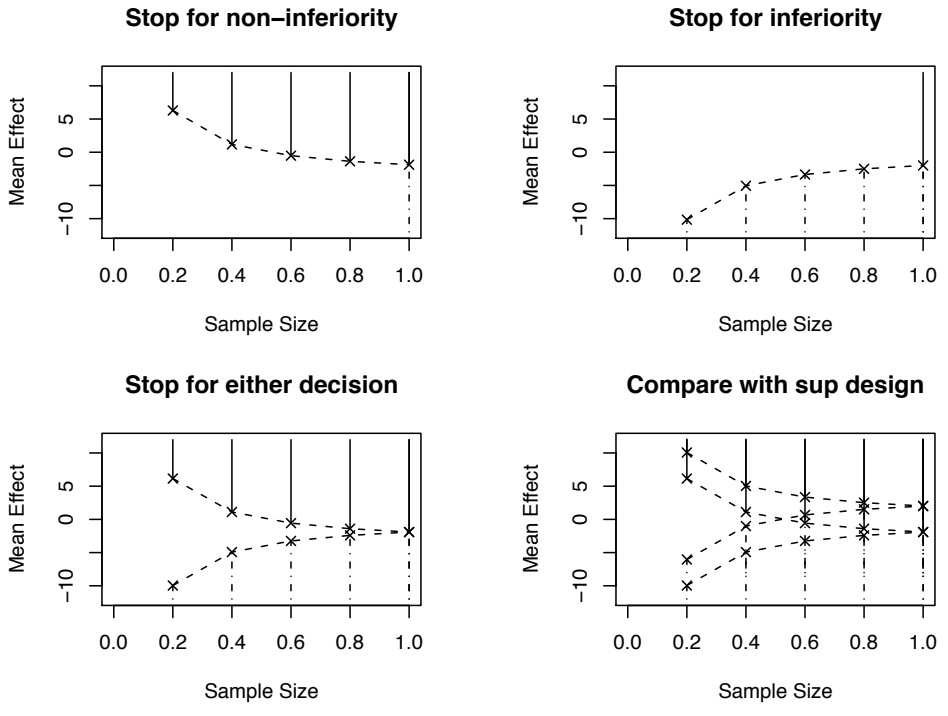
Non-inferiority study

- RCTdesign:

```
> nonInf.D <- update(sup.D,null.hypothesis=-3.92,
+         alt.hypothesis=0)
>  nonInf.A <- update(nonInf.D,early.stopping="null")
>  nonInf.DA <- update(nonInf.D,early.stopping="both")
```


Boundary form (number and location)

Non-inferiority study designs



Boundary form (number and location)

Equivalence study

- Stop for superiority (of A over B or B over A):

$$\hat{\delta}_j \geq d_j \rightarrow \text{Reject } \delta \leq 0$$
$$\hat{\delta}_j \leq a_j \rightarrow \text{Reject } \delta \geq 0$$

- Stop for equivalence:

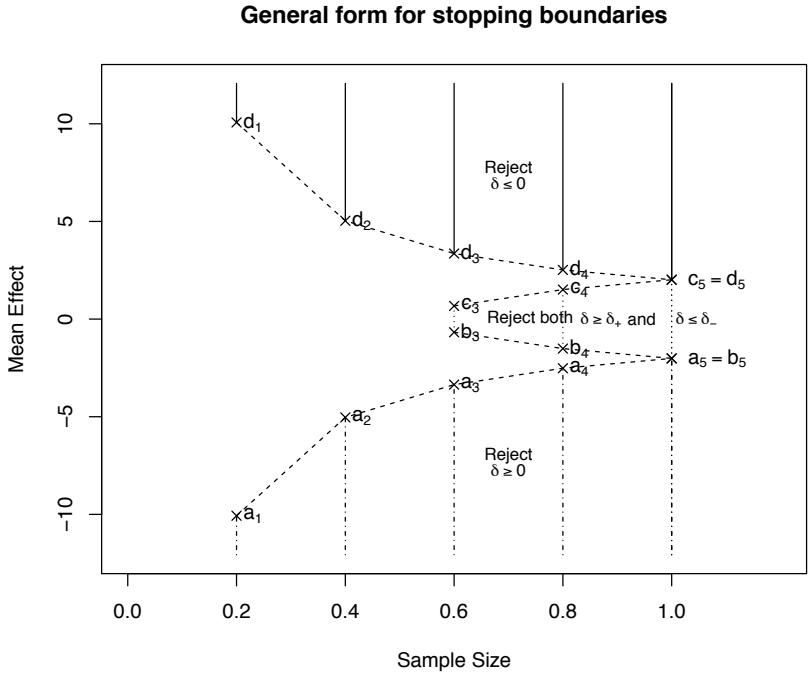
$$b_j \leq \hat{\delta}_j \leq c_j \rightarrow \text{Reject } \delta \leq \delta_- \text{ and } \delta \geq \delta_+$$

- Stop for either superiority or equivalence:

$$\begin{aligned} \hat{\delta}_j \geq d_j &\rightarrow \text{Reject } \delta \leq 0 \\ b_j \leq \hat{\delta}_j \leq c_j &\rightarrow \text{Reject } \delta \leq \delta_- \text{ and } \delta \geq \delta_+ \\ \hat{\delta}_j \leq a_j &\rightarrow \text{Reject } \delta \geq 0 \end{aligned}$$

Boundary form (number and location)

(Illustrated earlier).



Boundary form (number and location)

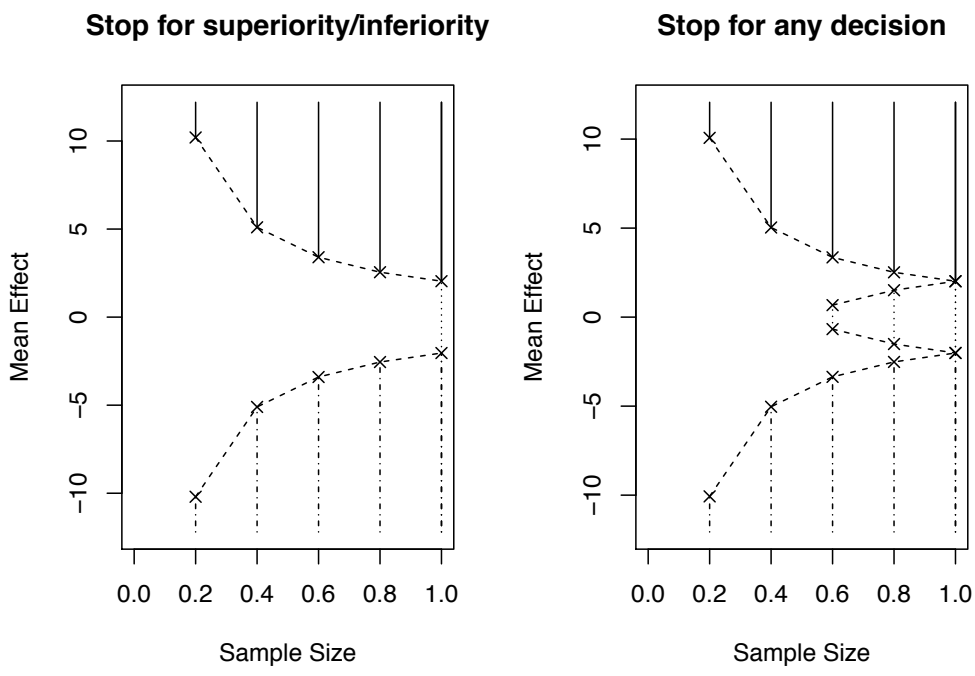
Equivalence study

- RCTdesign:

```
eq.Alt <- update(sup.D,test.type="two.sided",
   epsilon=c(1,1))
eq.Both <- update(eq.Alt,early.stopping="both")
```


Boundary form (number and location)

Equivalence study designs



Design evaluation

Design evaluation

- Interim analyses are used to address ethical and efficiency considerations
  - Scientific objectives are developed in the fixed-sample design
  - The monitoring plan (sequential design) should not alter the science
    - Maintain design hypotheses
    - Maintain design operating characteristics (PPV)
- Sequential sampling density

  Required to evaluate/maintain statistical properties
- Design characteristics and evaluation
- Examples

Sampling density for sequentially-sampled statistic

Historic context

- Wald (1947?): Sequential probability ratio test. Continuous monitoring; non-finite sample size.
- Armitage, McPherson, and Rao (1969): Recursive form for a sequentially sampled statistic
- Pocock (1977): Application in clinical trials; small sample consistency (t-statistic); decision criteria that are constant on Z-scale.
- O'Brien-Fleming (1979): Consistency for χ^2 statistic; decision criteria that are constant on partial sum scale; (early conservatism).
- Wang and Tsiatis (1987): Group sequential designs for 1-sided versus 2-sided hypothesis tests; parameterization of early conservatism.
- Emerson and Fleming (1989): Symmetric group sequential test designs.
- Kittelson and Emerson (1999): Unified family of group sequential test designs.

SISCR
UW - 2016

Sampling density for sequentially-sampled statistic

Uses/need for sampling density

- Same applications as sampling density for non-sequential statistic
  - Inference: point, interval estimation, p-value
  - Search for boundaries that satisfy operating characteristics
  - Sample size/power of sequential test
  - Bias-adjustment for sequentially-sampled statistic
- We seek the bivariate sampling density (M, S) where
  - M denotes the stopping time ($1 \leq M \leq J$), and
  - $S = S_M$ denotes the value of the partial sum statistic at the stopping time
  - This density is determined by:
    - Nature of the outcome: probability model, functional, and contrast
    - Nature of the stopping rules (boundary shape)
    - Number of stopping boundaries
    - Timing of the interim analyses (in information time)
    - Notes: the density does *not* depend on the boundary scale. Boundaries from most other scales can be mapped to stopping criteria for $\hat{\theta}$

Sampling density for sequentially-sampled statistic

Group sequential sampling density

- Let $\mathcal{S}_j$ and $\mathcal{C}_j = \mathcal{S}_j^c$ denote, respectively, the stopping and continuation sets at the jth interim analysis.
- The sampling density for the observation $(M = m, S = s)$ is:

$$p(m, s; \theta) = \begin{cases} f(m, s; \theta) & s \notin \mathcal{C}_m \\ 0 & else \end{cases}$$

where the (sub)density function $f(j, s; \theta)$ is recursively defined as

$$\begin{aligned} f(1, s; \theta) &= \frac{1}{\sqrt{n_1 V}} \phi\left(\frac{s - n_1\theta}{\sqrt{n_1 V}}\right) \\ f(j, s; \theta) &= \int_{\mathcal{C}_{(j-1)}} \frac{1}{\sqrt{n_j V}} \phi\left(\frac{s - u - n_j\theta}{\sqrt{n_j V}}\right) f(j-1, u; \theta)\, du, \\ & \quad j = 2, \ldots, m \end{aligned}$$

with $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ denoting the density for the standard normal distribution.

Design Evaluation: properties

Design properties

- There is no uniformly most powerful group sequential test; thus,
  - The unified family (RCTdesign) contains the full complement of possibilities
  - General classes (JK canonical classes) help structure the possibilities
  - There are continuua between classes that enables design iterations to begin in one class and move to a more suitable design
  - But, what properties should we be considering as we iterate?

Design Evaluation: properties

Design properties

- Elements that are established in the fixed-sample design:
  - Endpoint, prob model, functional, contrast
  - Maximal information (sample size, N_J; design alternative hypothesis)
  - Statistical standard for evidence (α level)
- Evaluation of group sequential design:
  - Sample size is a random variable; characteristics of interest:
    - Mean (Average Sample Number - ASN)
    - Quantiles (median, 25th, 75th percentiles)
    - power curve
    - Power for fixed N_J
    - N_J for fixed power
    - Stopping probability at each interim analysis
    - Inference at the boundary: What is the statistical inference (point estimate, interval estimate, and p-value) that would be reported if the trial is stopped?
- Iterate: modify the stopping rules until an acceptable mix of properties is found.

SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Design Evaluation: properties

SISCR
UW - 2016

Design properties

- RCTdesign (Suppose you have two designs: `dsgnA`, `dsgnB`):
  - Plot designs:
    ```
    plot(dsgnA,dsgnB,superpose=T)
    ```
  - Plot ASN:
    ```
    seqPlotASN(dsgnA,dsgnB)
    ```
  - Plot power:
    ```
    seqPlotPower(dsgnA,dsgnB)
    seqPlotPower(dsgnA,dsgnB,
    reference=dsgnA)
    ```
  - Plot inference:
    ```
    seqPlotInference(dsgnA,dsgnB)
    ```
  - Plot Stopping Probabilities
    ```
    seqPlotStopProb(dsgnA)
    ```

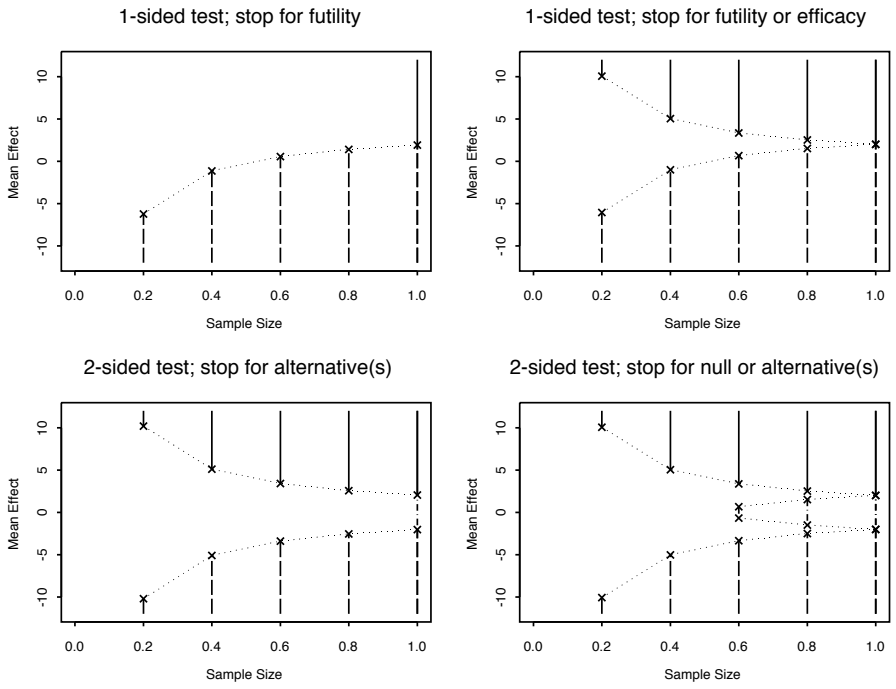

Illustration of general design properties

SISCR
UW - 2016

Four classes of designs

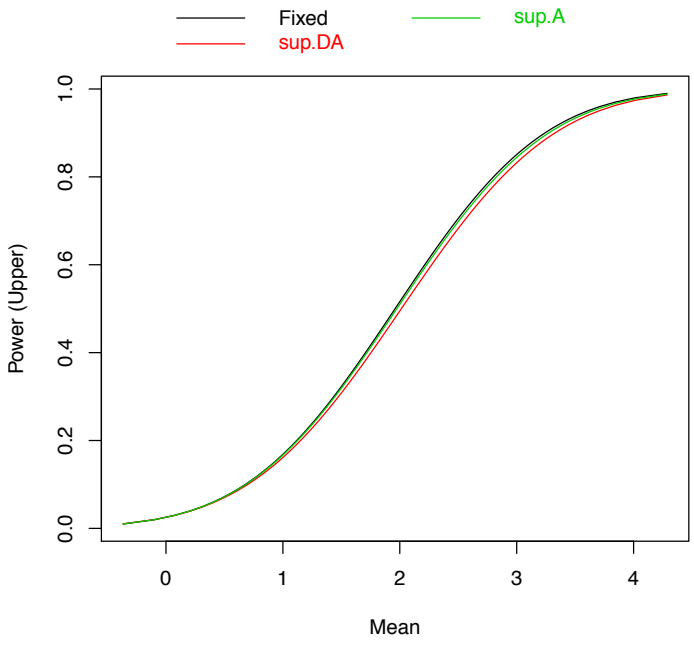
- One-sided test; One-sided stopping
  (allow stopping for efficacy *or* futility, but not both)
- One-sided test; Two-sided stopping
  (allow stopping for either efficacy or futility)
- Two-sided test; One-sided stopping
  (allow stopping only for the alternative(s))
- Two-sided test; Two-sided stopping
  (allow stopping for either the null or the alternative)

Illustration of general design properties
Four design classes



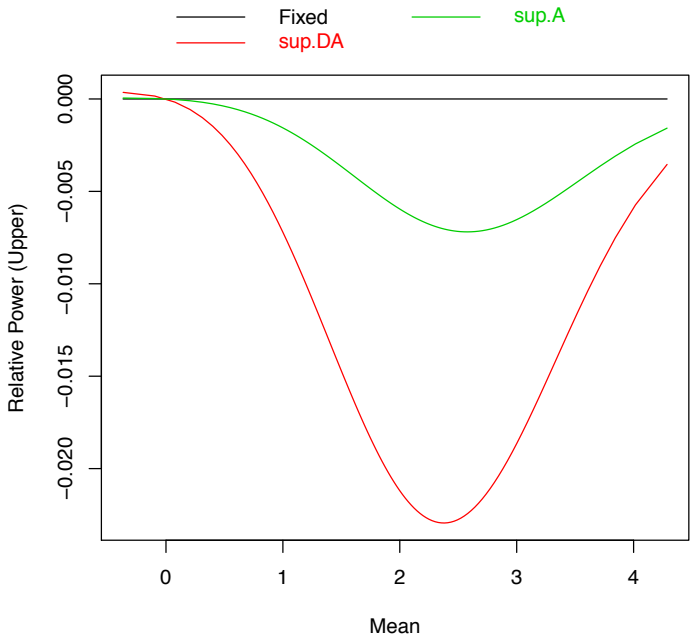
Power of one-sided tests

```
> seqPlotPower(sup.DA,sup.A)
```



Power of one-sided tests relative to fixed-sample test

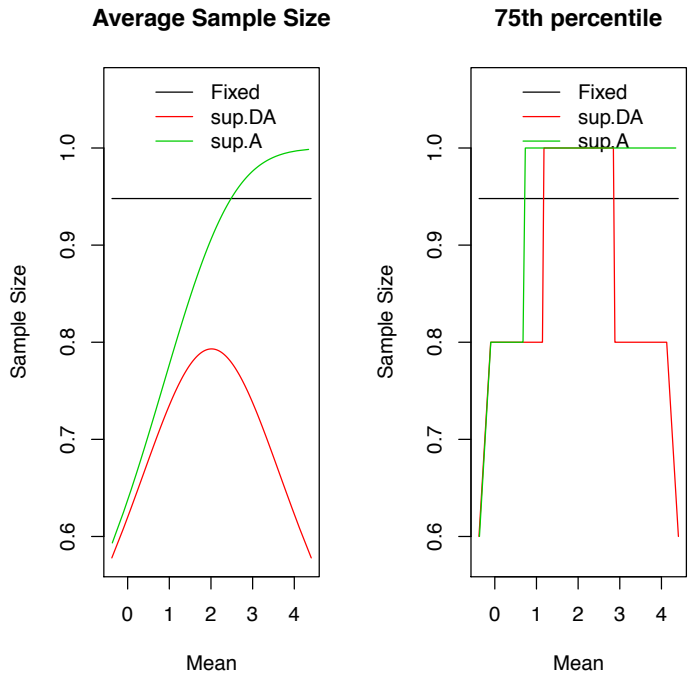
```
> seqPlotPower(sup.DA,sup.A)
```



SISCR
UW - 2016

ASN for one-sided tests

```
> seqPlotASN(sup.DA,sup.A)
```

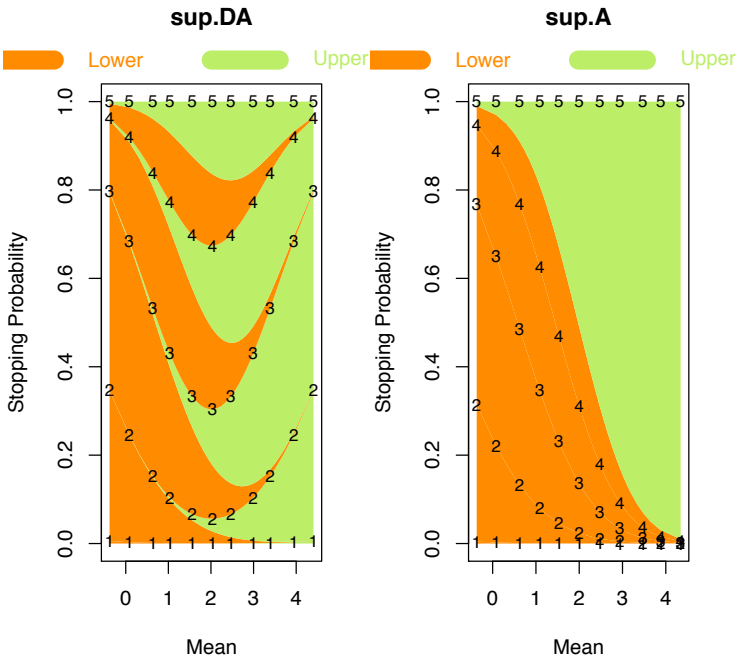


SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
***Sepsis trial:** add interim analyses
***Sepsis trial:** number of boundaries
***Sepsis trial:** early conservatism
***Sepsis trial:** power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

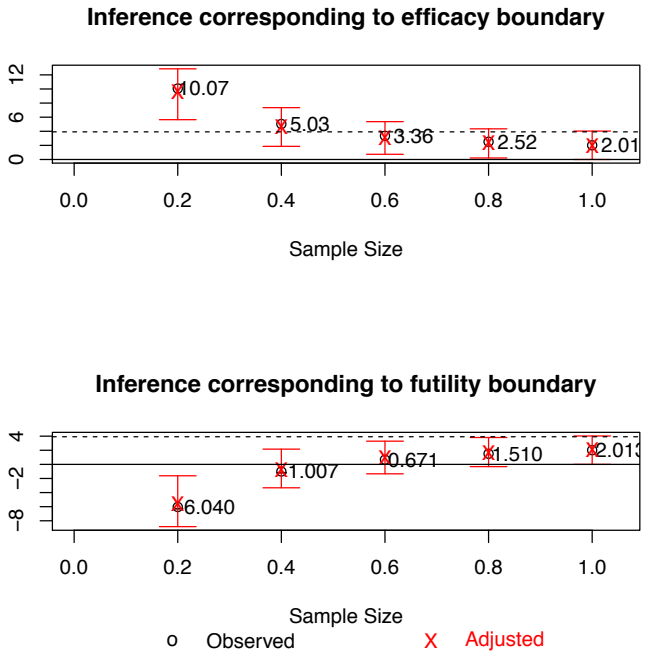
Stopping probabilities for one-sided tests

```
> seqPlotStopProb(sup.DA,sup.A)
```



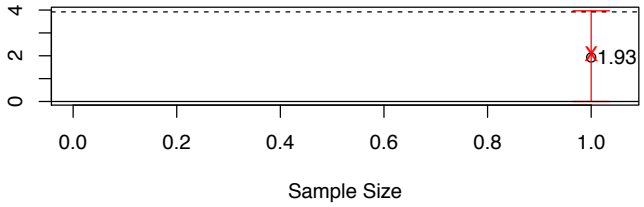
Inference at the boundary for `sup.DA`

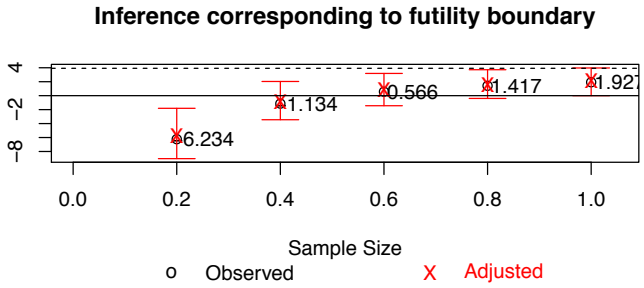
```
> seqPlotInference(sup.DA)
```



Inference at the boundary for `sup.A`

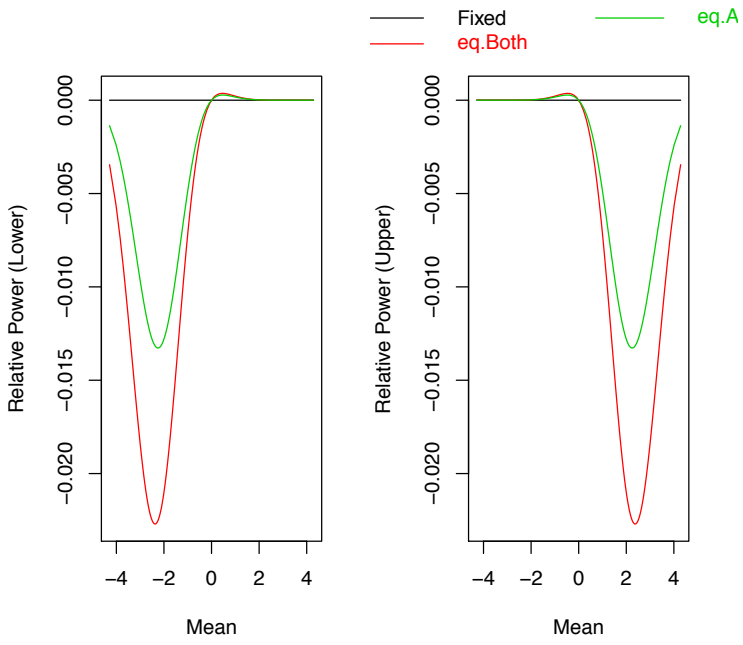
```
> seqPlotInference(sup.A)
```





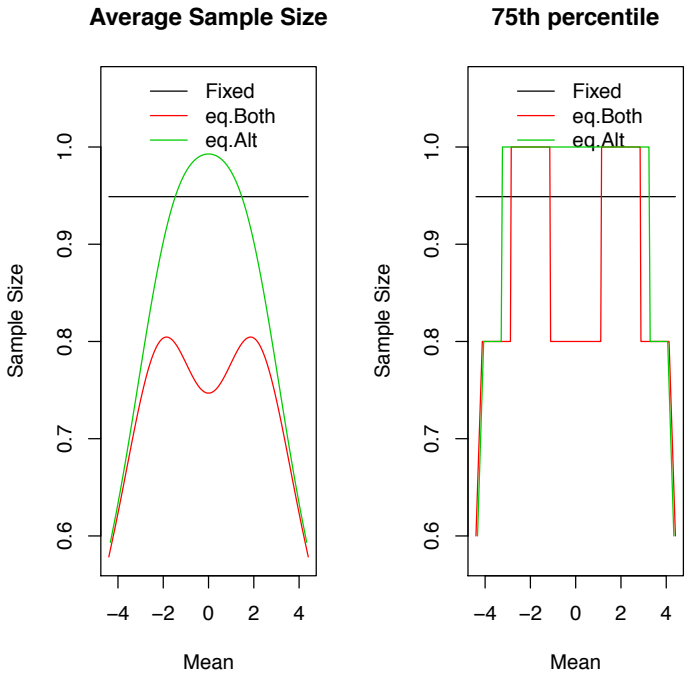
Power of two-sided tests relative to fixed-sample test

```
> seqPlotPower(eq.Both,eq.Alt,reference=T)
```



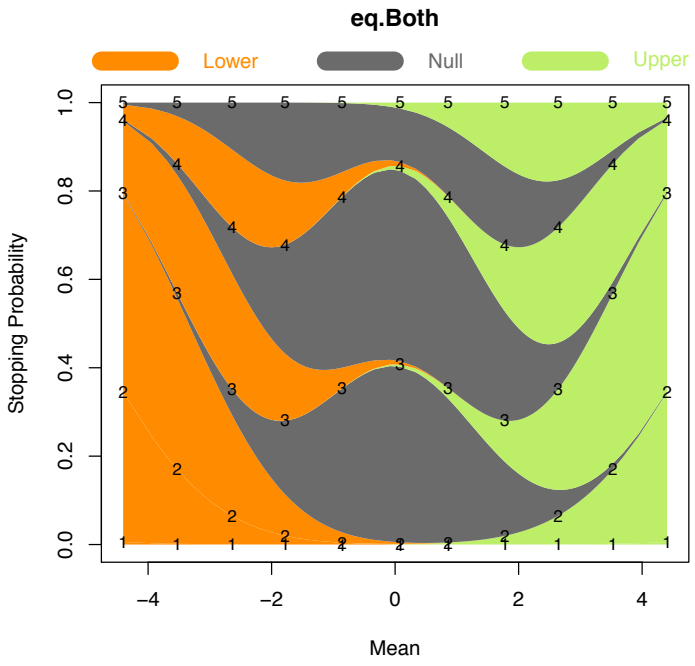
ASN for two-sided tests

```
> seqPlotASN(eq.Both,eq.Alt)
```



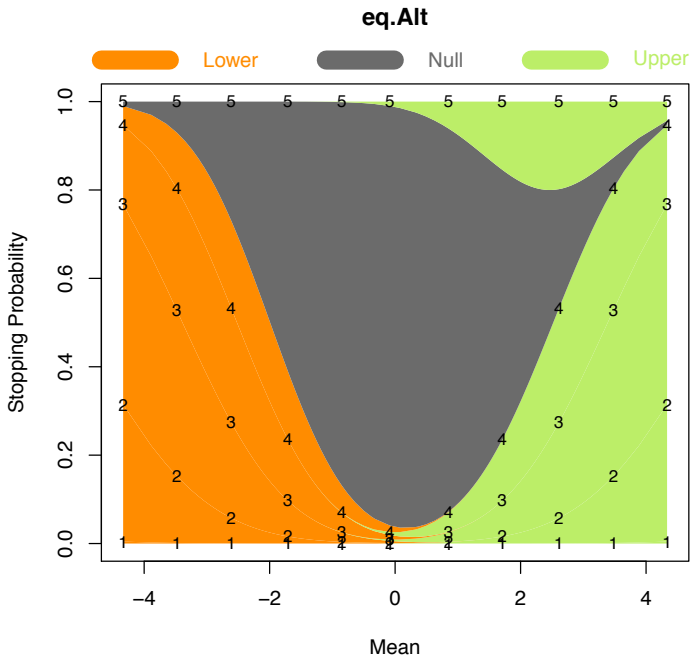
Stopping probabilities for `eq.Both`

```
> seqPlotStopProb(eq.Both)
```



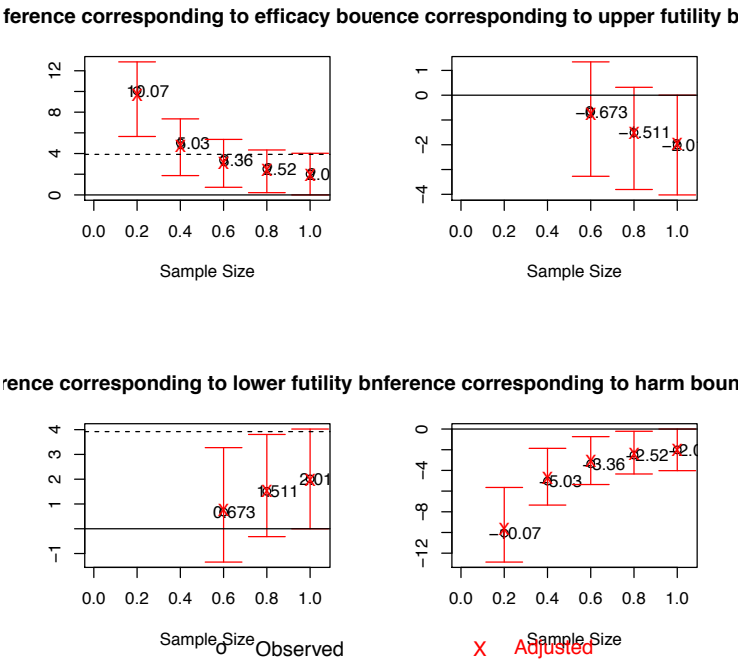
Stopping probabilities for `eq.Alt`

```
> seqPlotStopProb(eq.Alt)
```



Inference at the boundary for `eq.Both`

```
> seqPlotInference(eq.Both)
```



Inference at the boundary for `eq.Alt`

```
> seqPlotInference(eq.Alt)
```

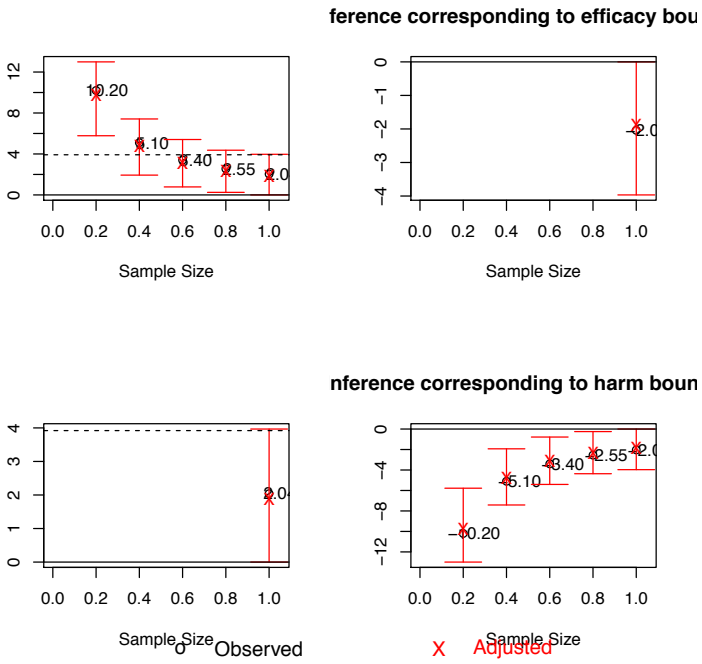


Illustration of general design properties

So what is the general behavior?

- For any given sample size, adding interim analyses reduces power.
- For any given power, adding interim analyses increases the sample size.
- Having fewer interim analyses:
  - Leads to properties (maximal sample size, power, etc) that are closer to those of a fixed sample study.
  - However, ASN may be larger and stopping probabilities lower.
- Having more early conservatism:
  - Leads to properties (maximal sample size, power, etc) that are closer to those of a fixed sample study.
  - However, ASN may be larger and stopping probabilities lower.

SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
***Sepsis trial:** add interim analyses
***Sepsis trial:** number of boundaries
***Sepsis trial:** early conservatism
***Sepsis trial:** power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Background

- Hodgkin's lymphoma represents a class of neoplasms that start in lymphatic tissue
- Approximately 7,350 new cases of Hodgkin's are diagnosed in the US each year (nearly equally split between males and females)
- 5-year survival rate among stage IV (most severe) cases is approximately 60-70%

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Background (cont.)

- Common treatments include the use of chemotherapy, radiation therapy, immunotherapy, and possible bone marrow transplantation
- Treatment typically characterized by high rate of initial response followed by relapse
- Hypothesize that experimental monoclonal antibody in addition to standard of care will increase time to relapse among patients remission

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Definition of Treatment

- Administered via IV once a week for 4 weeks
- Patients randomized to receive standard of care plus active treatment or placebo (administered similarly)
- Treatment discontinued in the event of grade 3 or 4 AEs
- Primary analysis based upon intention-to-treat

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Refinement of the primary endpoint

Primary endpoint: Comparison of hazards for event (censored continuous data)

- Duration of followup
  - Wish to compare relapse-free survival over 4 years
  - Patients accrued over 3 years in order to guarantee at least one year of followup for all patients
- Measures of treatment effect (comparison across groups)
  - Hazard ratio (Cox estimate; implicitly weighted over time)
  - No adjustment for covariates
  - Statistical information dictated by number of events (under proportional hazards, statistical information is approximately D/4)

Case Study : Hodgkin's Trial

Definition of statistical hypotheses

Null hypothesis

- Hazard ratio of 1 (no difference in hazards)
- Estimated baseline survival
  - Median progression-free survival approximately 9 months
  - (needed in this case to estimate variability)

Alternative hypothesis

- One-sided test for decreased hazard
  - Unethical to prove increased mortality relative to comparison group in placebo controlled study (always??)
- 33% decrease in hazard considered clinically meaningful
  - Corresponds to a difference in median survival of 4.4 months assuming exponential survival

Case Study : Hodgkin's Trial

Criteria for statistical evidence

- Type I error: Probability of falsely rejecting the null hypothesis Standards:
  - Two-sided hypothesis tests: 0.050
  - One-sided hypothesis test: 0.025
- Power: Probability of correctly rejecting the null hypothesis (1-type II error) Popular choice:
  - 80% power

SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Determination of sample size

- Sample size chosen to provide desired operating characteristics
  - Type I error : 0.025 when no difference in mortality
  - Power : 0.80 when 33% reduction in hazard
- Expected number of events determined by assuming
  - Exponential survival in placebo group with median survival of 9 months
  - Uniform accrual of patients over 3 years
  - Negligible dropout

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Specification of fixed sample design using RCTdesign

- Definition of original design

```
> survFixed <- seqDesign( prob.model = "hazard", arms = 2,
                          null.hypothesis = 1, alt.hypothesis = 0.67,
                          ratio = c(1, 1), nbr.analyses = 1,
                          test.type = "less",
                          power = 0.80, alpha = 0.025 )


> survFixed
Call:
seqDesign(prob.model = "hazard", arms = 2, null.hypothesis = 1,
    alt.hypothesis = 0.67, ratio = c(1, 1), nbr.analyses = 1,
    test.type = "less", power = 0.8, alpha = 0.025)

PROBABILITY MODEL and HYPOTHESES:
   Theta is hazard ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.00    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.67    (power = 0.800)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                            a      d
    Time 1 (N= 195.75) 0.7557 0.7557
```

Case Study : Hodgkin's Trial

Determination of sample size (cont.)

- Interpretation:
  - In order to desire the required number of patients we found in Session 2 that we would need to accrue:
    - N=76 patients per year for 3 years if the null hypothesis were true (Total of 228 patients)
    - N=81 patients per year for 3 years if the alternative hypothesis were true (Total of 243 patients)

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Re-designing the study

- Sponsor felt that attaining 75-80 patients per year would be unrealistic
- Wished to consider design operating characteristics assuming approximately uniform accrual of 50 patients per year while maintaining the same accrual time and follow up
- Problem: Need to determine the expected number of events if 50 subjects were accrued per year
- Solution: Solve backwards using the `nEvents` argument in `seqPHSubjects()`, substituting various numbers of events (see Session 2)

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
***Sepsis trial:** add interim analyses
***Sepsis trial:** number of boundaries
***Sepsis trial:** early conservatism
***Sepsis trial:** power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

Re-designing the study

- After a (manual) iterative search, we found that if roughly 50 patients are accrued yearly (under the alternative), 121 events would be expected

```
> seqPHSubjects( survFixed, controlMedian = 0.75, accrualTime = 3,
                 followupTime = 1, nEvents = 121 )

  accrualTime followupTime   rate hazardRatio controlMedian nSubjects
1           3            1 46.584        1.00          0.75    139.75
2           3            1 49.757        0.67          0.75    149.27
```


Case Study : Hodgkin's Trial

SISCR
UW - 2016

Re-designing the study

- Use the `update()` function in RCTdesign to update to the new sample size and compare operating characteristics

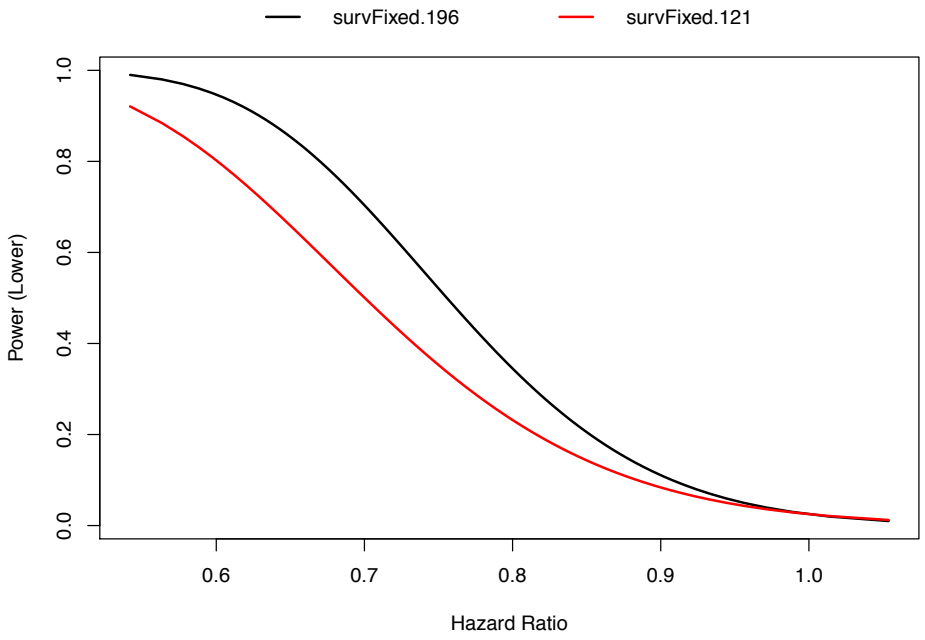
```
> survFixed.121 <- update( survFixed, sample.size=121,
                           power="calculate" )
> survFixed.121
Call:
seqDesign(prob.model = "hazard", arms = 2, null.hypothesis = 1,
    alt.hypothesis = 0.67, ratio = c(1, 1), nbr.analyses = 1,
    sample.size = 121, test.type = "less", power = "calculate",
    alpha = 0.025)

PROBABILITY MODEL and HYPOTHESES:
   Theta is hazard ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.00    (size  = 0.0250)
    Alternative hypothesis : Theta <= 0.67    (power = 0.5959)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                        a      d
    Time 1 (N= 121) 0.7002 0.7002
```

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

Statistical power using RCTdesign

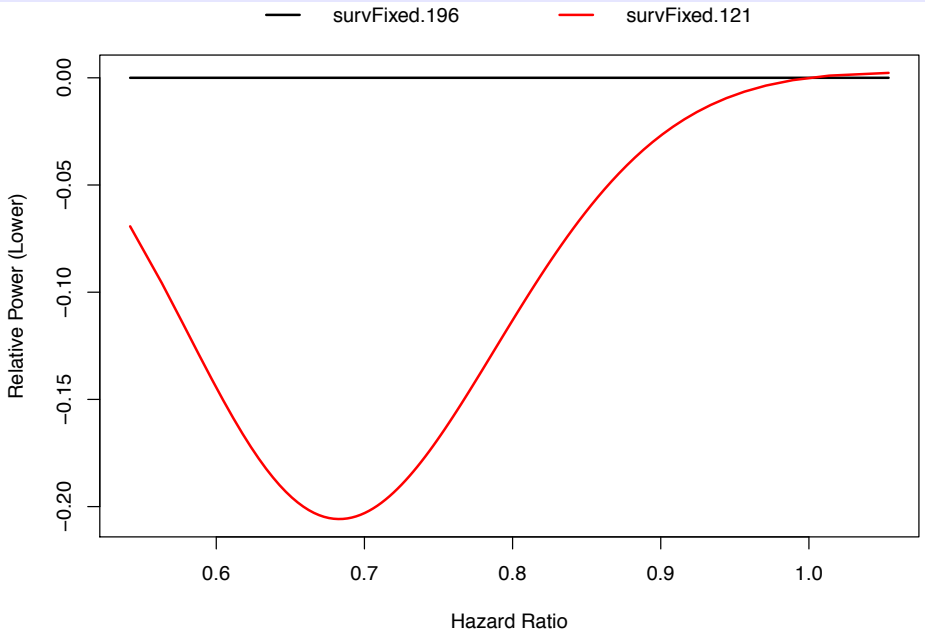
- Compare power curves using `seqPlotPower()`



Case Study : Hodgkin's Trial

Statistical power using RCTdesign

- Often more useful to compare differences between power curves
- Use the `reference` argument in `seqPlotPower()`



Case Study : Hodgkin's Trial

Candidate group sequential designs

- Principles in guiding initial choice of stopping rule
  - Early conservatism
    - Long-term benefit of high importance
    - Early stopping precludes the observation of long-term safety data
  - Ability to stop early for futility
    - Safety concerns
    - Logistical considerations (monetary)
  - Number and timing of interim analyses
    - Trade-off between power and sample size
    - Determined by information accrual (events) but ultimately scheduled on calendar time

Case Study : Hodgkin's Trial

Candidate group sequential designs

- `SymmOBF.2, SymmOBF.3, SymmOBF.4`
  - One-sided symmetric stopping rules with O'Brien-Fleming boundary relationships having 2, 3, and 4 equally spaced analyses,respectively, and a max sample size of 196 events
- `SymmOBF.Power`
  - One-sided symmetric stopping rule with O'Brien-Fleming boundary having 4 equally spaced analyses, and 80% under the alternative hypothesis (HR=0.67)
- `Futility.5, Futility.8, Futility.9`
  - One-sided stopping rules from the unified family [5] with a total of 4 equally spaced analyses, with a maximal sample size of 196 events, and having O'Brien-Fleming lower (efficacy) boundary relationships and upper (futility) boundary relationships corresponding to boundary shape parameters P = 0.5, 0.8, and 0.9, respectively. P = 0.5 corresponds to Pocock boundary shape functions, and P = 1.0 corresponds to O'Brien-Fleming boundary relationships

SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
***Sepsis trial:** add interim analyses
***Sepsis trial:** number of boundaries
***Sepsis trial:** early conservatism
***Sepsis trial:** power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

Candidate group sequential designs

- `Eff11.Fut8, Eff11.Fut9`
  - One-sided stopping rules from the unified family with a total of 4 equally spaced analyses, with a maximal sample size of 196 events, and having lower (efficacy) boundary relationships corresponding to boundary shape parameter P = 1.1 and upper (futility) boundary relationships corresponding to boundary shape parameters P = 0.8, and 0.9, respectively. P = 0.5 corresponds to Pocock boundary shape functions, and P = 1.0 corresponds to O'Brien-Fleming boundary relationships
- `Fixed.Power`
  - A fixed sample study which provides the same power to detect the alternative (HR=0.67) as the `Futility.8` trial design

Case Study : Hodgkin's Trial

SISCR
UW - 2016

Candidate group sequential designs

- Specification of candidate designs using `update()`

```
> Fixed <- survFixed
>
> SymmOBF.2 <- update( Fixed, nbr.analyses=2, P=c(1,1),
                       sample.size=196, power="calculate" )
> SymmOBF.3 <- update( SymmOBF.2, nbr.analyses = 3, P=c(1,1) )
> SymmOBF.4 <- update( SymmOBF.2, nbr.analyses = 4, P=c(1,1) )
> SymmOBF.Power <- update( SymmOBF.4, power = 0.80 )
>
> Futility.5 <- update( SymmOBF.4, P=c(1,.5) )
> Futility.8 <- update( SymmOBF.4, P=c(1,.8) )
> Futility.9 <- update( SymmOBF.4, P=c(1,.9) )
>
> Eff11.Fut8 <- update( SymmOBF.4, P=c(1.1,.8) )
> Eff11.Fut9 <- update( SymmOBF.4, P=c(1.1,.9) )
>
> Fixed.Power <- update( SymmOBF.2, nbr.analyses=1, power=0.7767 )
```

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

Candidate group sequential designs

- Stopping boundaries for `SymmOBF.4`

```
> SymmOBF.4
Call:
seqDesign(prob.model = "hazard", arms = 2, null.hypothesis = 1,
    alt.hypothesis = 0.67, ratio = c(1, 1), nbr.analyses = 4,
    sample.size = 196, test.type = "less", power = "calculate",
    alpha = 0.025, P = c(1, 1))

PROBABILITY MODEL and HYPOTHESES:
   Theta is hazard ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >= 1.00    (size  = 0.0250)
    Alternative hypothesis : Theta <= 0.67    (power = 0.7837)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                         a      d
    Time 1 (N=  49) 0.3183 1.7724
    Time 2 (N=  98) 0.5642 1.0000
    Time 3 (N= 147) 0.6828 0.8263
    Time 4 (N= 196) 0.7511 0.7511
```

SISCR
UW - 2016

Case Study : Hodgkin's Trial

Boundaries on various design scales

- Normalized Z statistic: $Z_j = z_j = (\hat{\theta}_j - \theta_0)/se(\hat{\theta}_j)$

```
> seqBoundary( SymmOBF.4, scale="Z" )
STOPPING BOUNDARIES: Normalized Z-value scale
                         a       d
    Time 1 (N=  49) -4.0065  2.0032
    Time 2 (N=  98) -2.8330  0.0000
    Time 3 (N= 147) -2.3131 -1.1566
    Time 4 (N= 196) -2.0032 -2.0032
```

SISCR
UW - 2016

Case Study : Hodgkin's Trial

Boundaries on various design scales

- Fixed sample P value statistic: $P_j = \Phi(z_j)$

```
> 1-seqBoundary( SymmOBF.4, scale="P" )
STOPPING BOUNDARIES: Fixed Sample P-value scale
                          a      d
    Time 1 (N=  49) 0.0000 0.9774
    Time 2 (N=  98) 0.0023 0.5000
    Time 3 (N= 147) 0.0104 0.1237
    Time 4 (N= 196) 0.0226 0.0226
```


Case Study : Hodgkin's Trial

SISCR
UW - 2016

Boundaries on various design scales

- Error spending statistic:

$$E_{aj} = \frac{1}{\alpha_L}\left(\Pr\left[S_j \le s_j, \bigcap_{k=1}^{j-1} S_k \in C_k \mid \theta = \theta_0\right] + \sum_{\ell=1}^{j-1} \Pr\left[S_\ell \le a_\ell, \bigcap_{k=1}^{\ell-1} S_k \in C_k \mid \theta = \theta_0\right]\right),$$

where α_L is the lower type I error of the stopping rule defined by

$$\alpha_L = \sum_{\ell=1}^{J} \Pr\left[S_\ell \le a_\ell, \bigcap_{k=1}^{\ell-1} S_k \in C_k | \theta = \theta_0\right].$$

```
> seqBoundary( SymmOBF.4, scale="E" )
STOPPING BOUNDARIES: Error Spending Function scale
                          a      d
    Time 1 (N=  49) 0.0012 0.0012
    Time 2 (N=  98) 0.0927 0.0927
    Time 3 (N= 147) 0.4470 0.4470
    Time 4 (N= 196) 1.0000 1.0000
```

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

Boundaries on various design scales

- Error spending statistic:

$$E_{aj} = \frac{1}{\alpha_L}\left(\Pr\left[S_j \le s_j, \bigcap_{k=1}^{j-1} S_k \in C_k \mid \theta = \theta_0\right] + \sum_{\ell=1}^{j-1} \Pr\left[S_\ell \le a_\ell, \bigcap_{k=1}^{\ell-1} S_k \in C_k \mid \theta = \theta_0\right]\right),$$

  where α_L is the lower type I error of the stopping rule defined by

$$\alpha_L = \sum_{\ell=1}^{J} \Pr\left[S_\ell \le a_\ell, \bigcap_{k=1}^{\ell-1} S_k \in C_k | \theta = \theta_0\right].$$

```
> seqBoundary( SymmOBF.4, scale="E" )*.025
STOPPING BOUNDARIES: Error Spending Function scale
                          a      d
    Time 1 (N=  49) 0.0000 0.0000
    Time 2 (N=  98) 0.0023 0.0023
    Time 3 (N= 147) 0.0112 0.0112
    Time 4 (N= 196) 0.0250 0.0250
```


Case Study : Hodgkin's Trial

Boundaries on various design scales

- `RCTdesign` also has the ability to incorporate prior distributions for treatment effects in order to evaluate:
  - Bayesian posterior probabilities
  - Bayesian predictive probabilities
- More to come later...

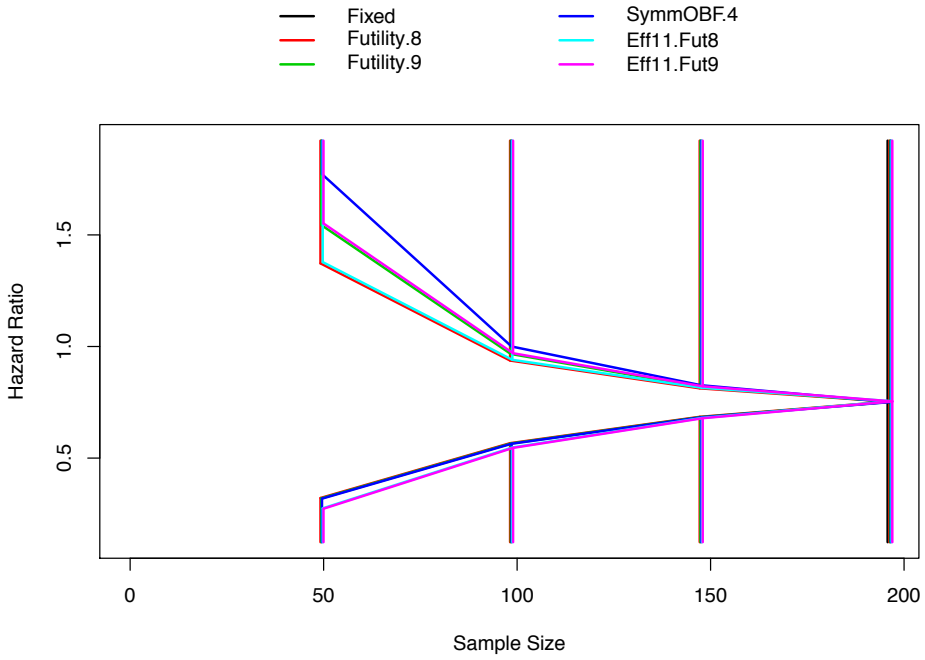
SISCR
UW - 2016

Design of Group Sequential Trials
Group sequential design for sepsis trial
*Statistical basis for stopping criteria
*Sepsis trial: add interim analyses
*Sepsis trial: number of boundaries
*Sepsis trial: early conservatism
*Sepsis trial: power vs maximal sample size
General characteristics group sequential designs
*Boundary structure
*Boundary scales
*Boundary shape
*Four canonical classes
Design evaluation
Group sequential sampling density
Design evaluation criteria
Properties of canonical classes
Case Study: Design of Hodgkin's Trial
Background
Fixed sample design
Group sequential design evaluations

Case Study : Hodgkin's Trial

Visual comparison of stopping boundaries

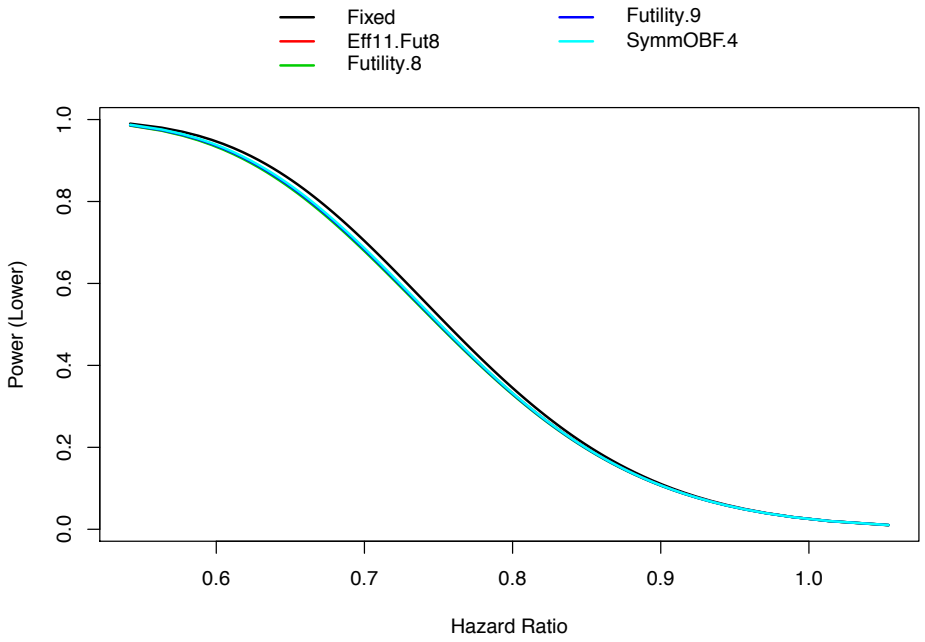
- Stopping boundaries can be plotted using `seqPlotBoundary()`



Case Study : Hodgkin's Trial

Visual comparison of statistical power for selected designs

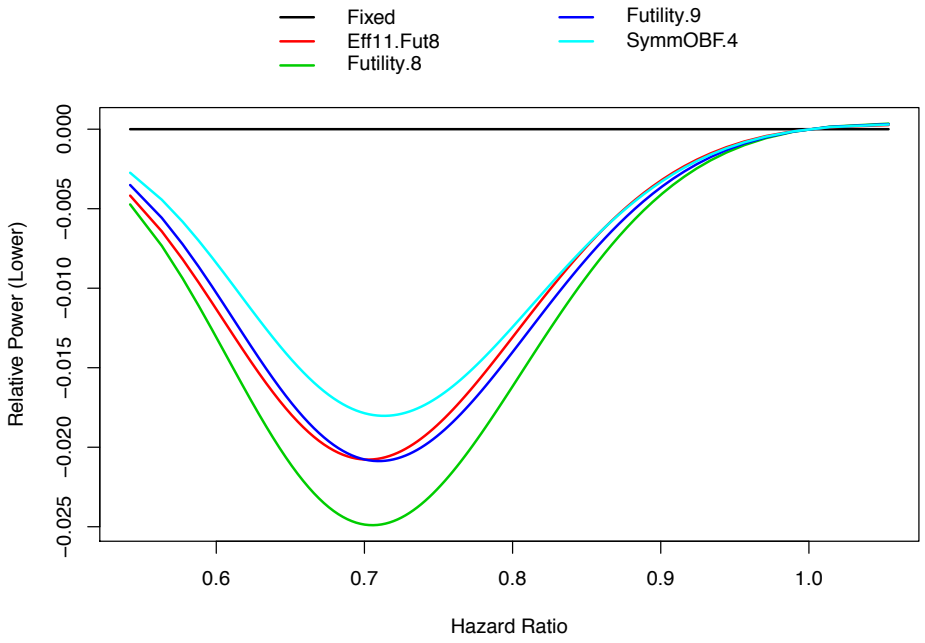
- Power curves (or differences) can be plotted with `seqPlotPower()`



Case Study : Hodgkin's Trial

Visual comparison of statistical power for selected designs

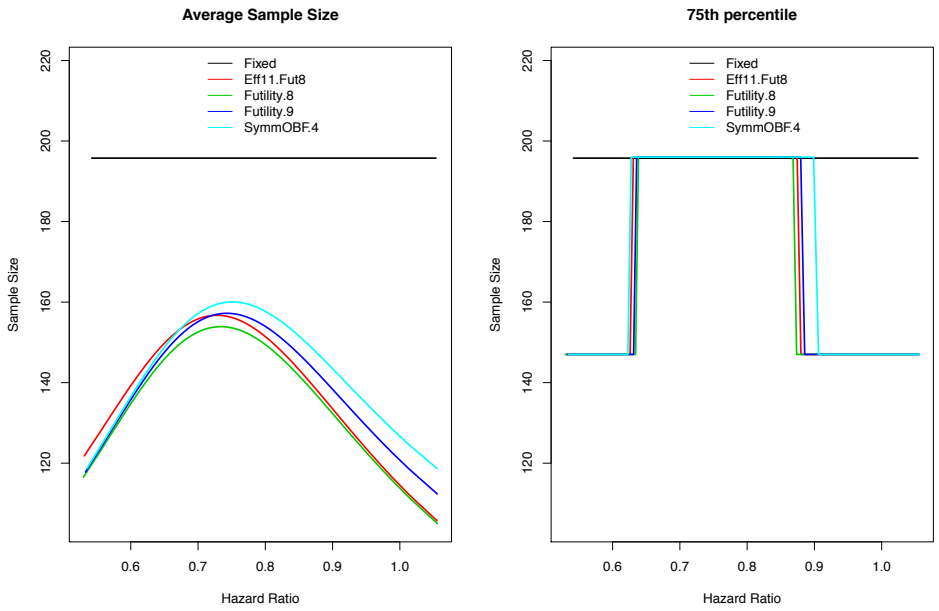
- As before, power curves (or differences) can be plotted with `seqPlotPower()`



Case Study : Hodgkin's Trial

Comparison of sample size distributions

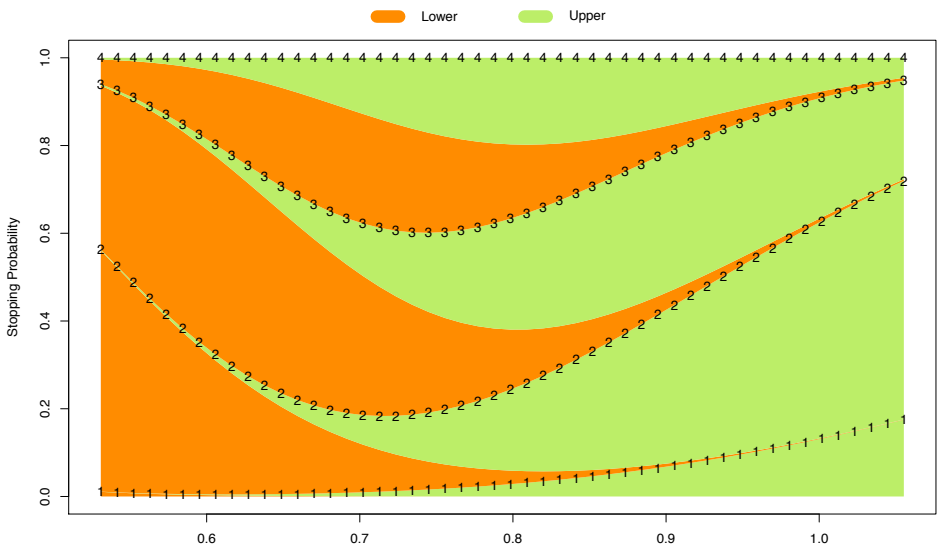
- Mean and quantiles of the sample size distribution can be plotted with `seqPlotASN()`



Case Study : Hodgkin's Trial

Stopping probabilities at each analysis for design `Eff11.Fut8`

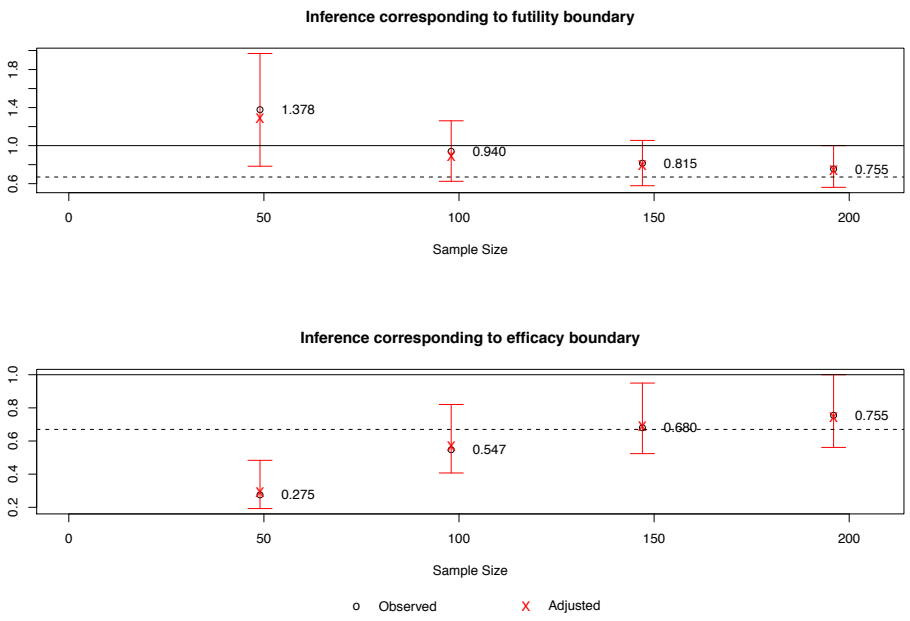
- Plot stopping probabilities using the `seqPlotStopProb()` function



Case Study : Hodgkin's Trial

Inference at each analysis for design `Eff11.Fut8`

- Plot inference on the boundaries using the `seqPlotStopProb()` function



Case Study : Hodgkin's Trial

Tabulation of operating characteristics for design `Eff11.Fut8`

- Computed operating characteristics can be obtained with the `seqOC()` function

```
> seqOC( Eff11.Fut8,  theta=seq(.6,1,by=.2) )
Operating characteristics
 Theta     ASN Power.lower
   0.6 139.24      0.9354
   0.8 151.43      0.3319
   1.0 114.51      0.0250

Stopping Probabilities:
 Theta Time 1 Time 2 Time 3 Time 4
   0.6 0.0049 0.3339 0.4757 0.1855
   0.8 0.0286 0.2174 0.3891 0.3649
   1.0 0.1308 0.4939 0.2830 0.0923
```