4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time
4. Designing the Trial Monitoring Plan

4.1 Elements of Trial Monitoring
4.2 Interim analyses: group sequential trials
4.3 Design of group sequential trials
4.4 General characteristics of group sequential trials
   (a) Unified family (a general formulation for stopping rules)
   (b) Evaluating and comparing design characteristics
   (c) Case study: ROCKET-AF
4.5 Inference following group sequential testing
4.6 Implementing the trial monitoring plan
4.7 Other considerations in trial monitoring
   (a) Stopping for futility
   (b) Monitoring multiple endpoints
   (c) Adaptive designs
Monitoring group sequential trials

Operating characteristics we considered at the design stage

1. Standard for evidence and efficiency of designs
   - Type I error
   - Power at various alternatives
   - Average sample number (ASN) / stopping probabilities

2. Point estimates of treatment effect corresponding to boundary decisions in favor of
   - Efficacy – Futility – Harm

3. Frequentist/Bayesian/Likelihood inference on the boundaries

4. Conditional futility/reversal of decision corresponding to boundary decisions

All dependent on the sampling density of the test statistic...
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

Monitoring group sequential trials

RECALL: Group sequential sampling density

- Consider independent observations $X_1, \ldots, X_{nJ}$ with $E[X_i] = \theta$, $i = 1, \ldots, n_J$

- Interested in testing $H_0 : \theta = \theta_0$ based upon a maximum of $J$ analyses

- Let $S_j$ denote the test statistic computed at interim analysis $j$ using observations $1, \ldots, n_j$, and suppose that $S_j \sim N(\theta V_j, V_j)$, $j = 1, \ldots, J$

- At each analysis we partition the outcome space for statistic $S_j$ into stopping set $S_j$ and continuation set $C_j$
  
  - If $S_j \in S_j$, the trial is stopped.
  - Otherwise, $S_j \in C_j$ and the study continues to gather additional observations.
Monitoring group sequential trials

RECALL: Group sequential sampling density

- Under an independent increments covariance structure, the sampling density of the bivariate group sequential statistic $(M, S_M)$, where $M = \min\{j : S_j \notin C_j\}$ is given by

$$p(m, s; \theta) = \begin{cases} f(m, s; \theta) & s \notin C_m \\ 0 & \text{otherwise} \end{cases},$$

where the function $f(j, s; \theta)$ is given recursively by,

$$f(1, s; \theta) = \frac{1}{\sqrt{V_1}} \phi \left( \frac{s - \theta \sqrt{V_1}}{\sqrt{V_1}} \right)$$

$$f(j, s; \theta) = \int_{C_{j-1}} \sqrt{V_j} \phi \left( \frac{s - u - \sqrt{V_j}}{\sqrt{V_j}} \right) f(j - 1, u; \theta) du, j = 2, \ldots, m$$

with $v_j = V_j - V_{j-1}$ and $\phi(x) = \frac{\exp \left( -x^2 / 2 \right)}{\sqrt{2\pi}}$. 
Monitoring group sequential trials

Operating characteristics condition upon exact timing

- When $S_j$ represents the score statistic resulting from a parametric probability model, $\text{Var}[S_j] = V_j = \mathcal{I}_j$ is Fisher Information

- The group sequential density (and hence all of the previously mentioned operating characteristics) will depend upon the timing of analyses as measured by the information accrued

- Most commonly, we carry out *maximal information trials*
  - Specify the maximum information that will be entertained
    - Usually in order to guarantee a specified power at a clinically relevant alternative
  - Interim analyses are then planned according to the proportion of the maximal sample size that has been accrued to the trial ($\Pi_j \equiv V_j / V_J$)
### Operating characteristics condition upon exact timing

- During the conduct of a study the timing of analyses may change because:
  - Monitoring scheduled by calendar time
  - Slow (or fast) accrual
  - External causes (should not be influenced by study results)
  - Statistical information from a sampling unit may be different than originally estimated
    - Variance of measurements
    - Baseline event rates (binary outcomes)
    - Censoring and survival distributions (weighted survival statistics)

- Consequences of these changes can include
  - Change in nominal type I error rate from originally planned design
  - Change in power from originally planned design
Monitoring group sequential trials

Example: Stopping rule chosen at design

- Test of normal mean:
  - $H_0 : \theta \leq 0.0$
  - $H_1 : \theta \geq 0.5$

- One-sided symmetric test
  - Size .025, Power .975
  - Four equally spaced analyses
  - Pocock (1977) boundary relationships
4.6 Implementing the trial monitoring plan

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

Monitoring group sequential trials

Example: Stopping rule chosen at design

```r
> dsn <- seqDesign( prob.model="normal", arms=1, null.hypothesis=0,
+                  alt.hypothesis=0.5, test.type="greater", variance=4,
+                  power=0.975, P=0.5, nbr.analyses=4, early.stopping="both" )

> dsn

PROBABILITY MODEL and HYPOTHESES:
  Theta is mean response
  One-sided hypothesis test of a greater alternative:
    Null hypothesis : Theta <= 0.0 (size = 0.025)
    Alternative hypothesis : Theta >= 0.5 (power = 0.975)
  (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale

Futility Efficacy

<table>
<thead>
<tr>
<th>Time</th>
<th>N=</th>
<th>0.0000</th>
<th>0.5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>86.31</td>
<td>0.0000</td>
<td>0.5000</td>
</tr>
<tr>
<td>Time 2</td>
<td>172.62</td>
<td>0.1464</td>
<td>0.3536</td>
</tr>
<tr>
<td>Time 3</td>
<td>258.92</td>
<td>0.2113</td>
<td>0.2887</td>
</tr>
<tr>
<td>Time 4</td>
<td>345.23</td>
<td>0.2500</td>
<td>0.2500</td>
</tr>
</tbody>
</table>
```
Monitoring group sequential trials

Analyses after 40%, 60%, 80%, 100% (maintain power)

```r
> dsn.late.power <- update(dsn, sample.size=c(.4,.6,.8,1) )
> dsn.late.power

PROBABILITY MODEL and HYPOTHESES:
Theta is mean response
One-sided hypothesis test of a greater alternative:
  Null hypothesis : Theta <= 0.0  (size = 0.025)
  Alternative hypothesis : Theta >= 0.5  (power = 0.975)
(Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
  Futility Efficacy
  Time 1 (N= 131.97)  0.1047  0.3953
  Time 2 (N= 197.95)  0.1773  0.3227
  Time 3 (N= 263.93)  0.2205  0.2795
  Time 4 (N= 329.91)  0.2500  0.2500
```
Monitoring group sequential trials

Analyses after 40%, 60%, 80%, 100% (maintain max sample size)

```r
> dsn.late.n <- update(dsn,
    sample.size=c(.4,.6,.8,1)*max(dsn$parameters$sample.size),
    alt.hypothesis="calculate" )

> dsn.late.n

PROBABILITY MODEL and HYPOTHESES:
 Theta is mean response
 One-sided hypothesis test of a greater alternative:
    Null hypothesis : Theta <= 0.0000  (size = 0.025)
    Alternative hypothesis : Theta >= 0.4888  (power = 0.975)
(Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
    Futility Efficacy
    Time 1  (N= 138.09)  0.1024  0.3864
    Time 2  (N= 207.14)  0.1733  0.3155
    Time 3  (N= 276.19)  0.2155  0.2732
    Time 4  (N= 345.23)  0.2444  0.2444
```
Changes in the number of analyses

- During the conduct of a study, the number of analyses may also be different from design stage
  - Monitoring scheduled by calendar time
  - Slow (or fast) accrual
  - External causes (should not be influenced by study results)

- This will also result in changes to design operating characteristics
Monitoring group sequential trials

Example: Stopping rule chosen at design (cont’d)

> dsn <- seqDesign( prob.model="normal", arms=1, null.hypothesis=0, +                        alt.hypothesis=0.5, test.type="greater", variance=4, +                        power=0.975, P=0.5, nbr.analyses=4, early.stopping="both" )

> dsn

PROBABILITY MODEL and HYPOTHESES:
   Theta is mean response
   One-sided hypothesis test of a greater alternative:
      Null hypothesis : Theta <= 0.0 (size = 0.025)
      Alternative hypothesis : Theta >= 0.5 (power = 0.975)
(Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
   Futility Efficacy

   Time 1 (N= 86.31)  0.0000  0.5000
   Time 2 (N= 172.62)  0.1464  0.3536
   Time 3 (N= 258.92)  0.2113  0.2887
   Time 4 (N= 345.23)  0.2500  0.2500
Monitoring group sequential trials

Analyses after 20%, 40%, 60%, 80%, 100% (maintain power)

```r
> dsn.5.power <- update(dsn, sample.size=c(.2,.4,.6,.8,1) )

> dsn.5.power

PROBABILITY MODEL and HYPOTHESES:
  Theta is mean response
  One-sided hypothesis test of a greater alternative:
    Null hypothesis : Theta <= 0.0   (size = 0.025)
    Alternative hypothesis : Theta >= 0.5   (power = 0.975)
  (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale

<table>
<thead>
<tr>
<th>Time</th>
<th>(N)</th>
<th>Futility</th>
<th>Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72.10</td>
<td>-0.0590</td>
<td>0.5590</td>
</tr>
<tr>
<td>2</td>
<td>144.20</td>
<td>0.1047</td>
<td>0.3953</td>
</tr>
<tr>
<td>3</td>
<td>216.31</td>
<td>0.1773</td>
<td>0.3227</td>
</tr>
<tr>
<td>4</td>
<td>288.41</td>
<td>0.2205</td>
<td>0.2795</td>
</tr>
<tr>
<td>5</td>
<td>360.51</td>
<td>0.2500</td>
<td>0.2500</td>
</tr>
</tbody>
</table>
Monitoring group sequential trials

Analyses after 20%, 40%, 60%, 80%, 100% (maintain max sample size)

```r
> dsn.5.n <- update(dsn,
     sample.size=c(.2,.4,.6,.8,1)*max(dsn$parameters$sample.size),
     alt.hypothesis="calculate"
)

> dsn.5.n

PROBABILITY MODEL and HYPOTHESES:
  Theta is mean response
  One-sided hypothesis test of a greater alternative:
    Null hypothesis : Theta <= 0.0000  (size = 0.025)
    Alternative hypothesis : Theta >= 0.5109  (power = 0.975)
(Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                      Futility  Efficacy
    Time 1 (N= 69.05)  -0.0603  0.5713
    Time 2 (N= 138.09)  0.1070  0.4039
    Time 3 (N= 207.14)  0.1811  0.3298
    Time 4 (N= 276.19)  0.2253  0.2856
    Time 5 (N= 345.23)  0.2555  0.2555
```
4.6 Implementing the trial monitoring plan

**Result of changing schedule of analyses**

- **Summary for Pocock boundary relationships**

<table>
<thead>
<tr>
<th>Analysis Times</th>
<th>Alt</th>
<th>Max N</th>
<th>Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25, .50, .75, 1.00</td>
<td>.500</td>
<td>345.23</td>
<td>.2500</td>
</tr>
<tr>
<td>.40, .60, .80, 1.00</td>
<td>.500</td>
<td>329.91</td>
<td>.2500</td>
</tr>
<tr>
<td>.40, .60, .80, 1.00</td>
<td>.489</td>
<td>345.23</td>
<td>.2444</td>
</tr>
<tr>
<td>.20, .40, .60, .80, 1.00</td>
<td>.500</td>
<td>360.51</td>
<td>.2500</td>
</tr>
<tr>
<td>.20, .40, .60, .80, 1.00</td>
<td>.511</td>
<td>345.23</td>
<td>.2555</td>
</tr>
</tbody>
</table>
# Monitoring group sequential trials

## Result of changing schedule of analyses

### Summary for O’Brien-Fleming boundary relationships

<table>
<thead>
<tr>
<th>Analysis Times</th>
<th>Alt</th>
<th>Max N</th>
<th>Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25, .50, .75, 1.00</td>
<td>.500</td>
<td>256.83</td>
<td>.2500</td>
</tr>
<tr>
<td>.40, .60, .80, 1.00</td>
<td>.500</td>
<td>259.44</td>
<td>.2500</td>
</tr>
<tr>
<td>.40, .60, .80, 1.00</td>
<td>.503</td>
<td>256.83</td>
<td>.2513</td>
</tr>
<tr>
<td>.20, .40, .60, .80, 1.00</td>
<td>.500</td>
<td>259.45</td>
<td>.2500</td>
</tr>
<tr>
<td>.20, .40, .60, .80, 1.00</td>
<td>.503</td>
<td>256.83</td>
<td>.2513</td>
</tr>
</tbody>
</table>
Constrained Boundaries Example

Constrained O’Brien-Fleming Design

- It is often desirable to modify a stopping rule at the design stage to maintain a particular set of boundary constraints.
- For example, an O’Brien-Fleming stopping rule is known for extreme conservatism at early analysis.
  - One-sided level .025 test of a normal mean with four equally spaced analyses.
  - Stopping at first analysis for efficacy requires a fixed sample P-value of less than .0001.

```
> obf <- seqDesign( prob.model="normal", arms=1, null.hypothesis=0, 
+     alt.hypothesis=0.5, test.type="greater", variance=4, 
+     power=0.975, P=1, nbr.analyses=4, early.stopping="both" )
```
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

---

Constrained Boundaries Example

**Constrained O’Brien-Fleming Design**

```r
> obf

PROBABILITY MODEL and HYPOTHESES:

Theta is mean response

One-sided hypothesis test of a greater alternative:

Null hypothesis : Theta <= 0.0  (size = 0.025)

Alternative hypothesis : Theta >= 0.5  (power = 0.975)

(Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale

<table>
<thead>
<tr>
<th>Time</th>
<th>Futility</th>
<th>Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.5000</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.5000</td>
</tr>
<tr>
<td>3</td>
<td>0.1667</td>
<td>0.3333</td>
</tr>
<tr>
<td>4</td>
<td>0.2500</td>
<td>0.2500</td>
</tr>
</tbody>
</table>

> seqBoundary(obf, scale="P")

STOPPING BOUNDARIES: Fixed Sample P-value scale

<table>
<thead>
<tr>
<th>Time</th>
<th>Futility</th>
<th>Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9774</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.5000</td>
<td>0.0023</td>
</tr>
<tr>
<td>3</td>
<td>0.1237</td>
<td>0.0104</td>
</tr>
<tr>
<td>4</td>
<td>0.0226</td>
<td>0.0226</td>
</tr>
</tbody>
</table>
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring
Error Spending Functions
Constrained Boundaries
Issues When Monitoring a Trial

Constrained Boundaries Example

Constrained O’Brien-Fleming Design

- Some sponsor’s wish for the operating characteristics of an O’Brien-Fleming design but desire a slightly less conservative first boundary

- One possibility is to constrain the O’Brien-Fleming design at the first analysis so that the efficacy bound corresponds to a P-value of 0.0005

- In order to maintain the overall type I error rate, the value of G must be re-computed using this constraint

- This can be done using an `exact.constraint`:

```r
> bnd.const <- as.seqBoundary( cbind(matrix(NA,nrow=4,ncol=3), 
                                 c(.0005,rep(NA,3))), scale="P" )
> bnd.const
STopping boundaries: fixed sample p-value scale
   a  b  c  d
time 1 NA NA NA 5e-04
time 2 NA NA NA NA
time 3 NA NA NA NA
time 4 NA NA NA NA
```
4.6 Implementing the trial monitoring plan

**Background**

Example: Constrained OBF design

**Flexible Trial Monitoring**

**Error Spending Functions**

**Constrained Boundaries**

**Issues When Monitoring a Trial**

Estimation of statistical information

Measuring study time

---

**Constrained Boundaries Example**

**Constrained O’Brien-Fleming Design**

```r
> obf.const <- update( obf, exact.constraint=bnd.const )
> obf.const

PROBABILITY MODEL and HYPOTHESES:
Theta is mean response
   One-sided hypothesis test of a greater alternative:
      Null hypothesis : Theta <= 0.0  (size = 0.025)
      Alternative hypothesis : Theta >= 0.5  (power = 0.975)

STOPPING BOUNDARIES: Sample Mean scale
   Futility  Efficacy
     Time 1 (N= 64.31) -0.4990  0.8207
     Time 2 (N= 128.61) 0.0005  0.5005
     Time 3 (N= 192.92) 0.1670  0.3337
     Time 4 (N= 257.23) 0.2502  0.2502

> seqBoundary(obf.const, scale="P")

STOPPING BOUNDARIES: Fixed Sample P-value scale
   Futility  Efficacy
     Time 1 (N= 64.31) 0.9773  0.0005
     Time 2 (N= 128.61) 0.4989  0.0023
     Time 3 (N= 192.92) 0.1231  0.0102
     Time 4 (N= 257.23) 0.0224  0.0224
```
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

Constrained Boundaries Example

Constrained O’Brien-Fleming Design

▶ Comparison of stopping boundaries (sample mean scale)
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

Constrained Boundaries Example

Constrained O’Brien-Fleming Design

- Comparison of statistical power
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

Constrained Boundaries Example

Constrained O’Brien-Fleming Design

▶ Comparison of statistical power
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time
Monitoring group sequential trials

Result of changing schedule of analyses

- As previously noted, during the conduct of a study the timing of analyses may change because:
  - Monitoring scheduled by calendar time
  - Slow (or fast) accrual
  - External causes (should not be influenced by study results)
  - Statistical information from a sampling unit may be different than originally estimated
    - Variance of measurements
    - Baseline event rates (binary outcomes)
    - Censoring and survival distributions (weighted survival statistics)
4.6 Implementing the trial monitoring plan

Background
Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information
Measuring study time

Monitoring group sequential trials

Result of changing schedule of analyses

- Need methods that allow flexibility in determining number and timing of analyses

- Should maintain some (but not, in general, all) desired operating characteristics, e.g.:
  - Type I error
  - Type II error
  - Maximal sample size
  - Futility properties
  - Bayesian properties
Monitoring group sequential trials

Popular methods for flexible implementation of group sequential boundaries


Monitoring group sequential trials

Common features

- Stopping rule specified at design stage parameterizes the boundary for some statistic (boundary scale)
  - Error spending family (Lan & Demets, 1983) → proportion of type I error spent
  - Unified family (Emerson & Kittelson, 1999) → point estimate (MLE)

- At the first interim analysis, parametric form is used to compute the boundary for actual time on study

- At successive analyses, the boundaries are recomputed accounting for the exact boundaries used at previously conducted analyses

- Maximal sample size estimates may be updated to maintain power
  - For binary outcomes, generally use pooled estimate of event rates to withhold treatment effect from study sponsor
Error spending functions

Implementing error spending functions

- **Error spending** (also known as \( \alpha \)-spending) allow flexible implementation by pre-specifying a rate at which the type I error will be “spent” at each interim analysis; specifically:
  
  - Let \( \alpha \) denote the type I error probability for the trial.
  - Use the group sequential sampling density to calculate the stopping probabilities (\( \alpha_j \)) over the prior interim analyses.
  - Let \( \alpha_j \) denote the probability of rejecting the null hypothesis at the \( j \)th interim analysis (then \( \alpha = \sum_j \alpha_j \)).
  - **Error spending function**: Let \( \alpha(\Pi) \) denote a function that constrains the probability of rejecting the null hypothesis at or before \( 100 \times \Pi \% \) of the total information; that is:

\[
\alpha(\Pi) = \frac{1}{\alpha} \sum_{j: \Pi_j < \Pi} \alpha_j
\]

Thus, \( \alpha(\Pi) \) is the proportion of the total type I error that has been “spent” when there is \( \Pi \) information in the trial.
### Implementing error spending functions

<table>
<thead>
<tr>
<th>Example of error spending functions:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant spending:</strong> $\alpha(\Pi) = \Pi$</td>
</tr>
<tr>
<td><strong>Power family:</strong> $\alpha(\Pi) = \Pi^P, P &gt; 1$</td>
</tr>
<tr>
<td><strong>Approximate O’Brien-Fleming:</strong> $\alpha(\Pi) = \Phi\left(\frac{Z_{\alpha/2}}{\sqrt{\Pi}}\right)$</td>
</tr>
<tr>
<td><strong>Approximate Pocock:</strong> $\alpha(\Pi) = ln[1 + (e - 1)\Pi]$</td>
</tr>
<tr>
<td><strong>Hwang, Shih, Decani, 1990:</strong> $\alpha(\Pi) = \frac{1 - e^{-\gamma\Pi}}{1 - e^{-\gamma}}, \gamma \neq 0$</td>
</tr>
</tbody>
</table>

Where $\Phi()$ is the standard normal cdf.
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

---

Error spending functions

Implementing error spending functions - Sepsis trial

- Consider the sepsis trial introduced earlier (see lctSec4-3) and suppose we wish to conduct a trial with four equally spaced analyses utilizing an O’Brien-Fleming stopping rule
  - One-sided type I error .025
  - N=1700 maximal patients

```r
> sepsis.fix <- seqDesign(prob.model="proportions", arms=2,
  size=.025, power="calculate",
  null.hypothesis= c(.30, .30),
  alt.hypothesis=c(0.25,0.30),
  sample.size=1700, test.type="less")

> #****** pre-trial monitoring plan
> sepsis.obf <- update(sepsis.fix,nbr.analyses=4,P=1)
> sepsis.obf

STOPPING BOUNDARIES: Sample Mean scale
  Efficacy Futility
  Time 1 (N= 425)  -0.1733  0.0866
  Time 2 (N= 850)  -0.0866  0.0000
  Time 3 (N= 1275) -0.0578  -0.0289
  Time 4 (N= 1700) -0.0433  -0.0433
```
Implementing error spending functions - Sepsis trial

- Pre-trial analysis timing in terms of information:
  - Recall $V = 0.25 \times 0.75 + 0.3 \times 0.7$
  - Pre-trial planned information:
    \[
    I = \frac{N_j/2}{V} = \frac{850}{0.3975} = 2138.4
    \]

- Pre-trial plan for analysis timing:
  \[
  \begin{array}{ccc}
  \Pi_j & N_j & \text{Information: } \frac{N_j}{2V} \\
  0.25 & 425 & 534.6 \\
  0.50 & 850 & 1069.2 \\
  0.75 & 1275 & 1603.8 \\
  1.00 & 1700 & 2138.4 \\
  \end{array}
  \]
Error spending functions

Implementing error spending functions - Sepsis trial

▶ Suppose the first interim analysis was conducted after data on 520 subjects (263 on the antibody arm, 257 on the placebo arm)

▶ Further suppose that 52 deaths were observed on the antibody arm and 65 deaths were observed on the placebo arm

\[
\hat{\theta}_1 = \frac{52}{263}, \quad \hat{\theta}_0 = \frac{65}{257}
\]

▶ Observed information at first interim analysis:

\[
\hat{S}_1 = \frac{\hat{\theta}_1 (1 - \hat{\theta}_1)}{263} + \frac{\hat{\theta}_0 (1 - \hat{\theta}_0)}{257} = 0.0013384
\]

\[
\frac{1}{\hat{S}_1} = 747.2
\]

\[
\Pi = 747.2/2138.4 = 0.34942
\]

Thus, we estimate that the first interim analysis has occurred at 34.9% of the planned total information.
Implementing error spending functions - Sepsis trial

- Pre-trial error-spending function:
  - Use `seqOC(sepsis.obf,theta=0)` to get the lower stopping probabilities at the interim analyses. These are the values of $\alpha_j$. The pretrial error-spending function, $\alpha(\Pi)$ has values at $\Pi_j$ defined by equation (1).

<table>
<thead>
<tr>
<th>$\Pi_j$</th>
<th>$a_j$</th>
<th>Stopping Prob ($\alpha_j$)</th>
<th>Cumulative type I error</th>
<th>Error spending function $\alpha(\Pi_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>-0.1733</td>
<td>0.00003</td>
<td>0.00003</td>
<td>0.00123</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.0866</td>
<td>0.00229</td>
<td>0.00232</td>
<td>0.09274</td>
</tr>
<tr>
<td>0.75</td>
<td>-0.0578</td>
<td>0.00886</td>
<td>0.01176</td>
<td>0.44703</td>
</tr>
<tr>
<td>1.00</td>
<td>-0.0433</td>
<td>0.01382</td>
<td>0.02500</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

- To get values of $\alpha(\Pi)$ for $\Pi \neq \Pi_j$ we can either:
  - Use an error-spending function that approximates the pre-trial plan
  - Use linear interpolation
### Error spending functions

#### Implementing error spending functions - Sepsis trial

- Using linear interpolation to find the critical value at 34.9% of total information:

\[
\alpha(0.349) = \alpha(0.25) + [\alpha(0.50) - \alpha(0.25)] \frac{0.349 - 0.25}{0.50 - 0.25}
\]

\[
= 0.00003 + 0.00229 \times \frac{0.099}{0.25}
\]

\[
= 0.00091872
\]

- Because this is the first interim analysis, we can calculate the revised value for \(a_1\) directly from the normal density:

\[
\frac{a_1}{\sqrt{\hat{S}_1}} = \Phi^{-1}(0.00091872)
\]

\[
= -3.1153
\]

Thus, \(a_1 = -3.1938\sqrt{0.0013384} = -0.11397\), and so we would continue because \(\hat{\theta}^{(1)} = -0.0552 > -0.11397\).
4.6 Implementing the trial monitoring plan

Background
Example: Constrained OBF design
Flexible Trial Monitoring

Error Spending Functions
Constrained Boundaries
Issues When Monitoring a Trial
Estimation of statistical information
Measuring study time

Error spending functions

Implementing error spending functions

Notes:

- At subsequent interim analyses we would repeat this process, but would need to account for the decision criteria used at earlier interim analyses to determine how much error should be spent and what the critical value should be.

- We can develop analogous stopping criteria for the futility ($d_j$) boundary using a $\beta$-spending function.

- I am not illustrating the above points because:
  - Error-spending scales do not directly elucidate the scientific/clinical aspects of the stopping criteria.
  - Error-spending scales do not directly address changes in the estimated standard deviation at subsequent interim analyses.

- (Note: any scale can be expressed on the sample mean scale, so you can (and should) consider the inference on the boundary when evaluating error-spending decision criteria.)
Error spending functions

Implementing error spending functions

- Error spending families have been implemented in RCTdesign
  - To get the error spending function from an existing design:
    ```r
    > update(sepsis.obf, display.scale="E")
    ```
  - To design a monitoring plan in the error spending scale:
    ```r
    > update(sepsis.obf, design.scale="E", P=-1, display.scale="E")
    > update(sepsis.obf, design.scale="E", P=-1, display.scale="X")
    ```
  - This implements the power family of error spending functions described above: \( \alpha(\Pi) = \Pi^P \times \alpha \)
Constrained Boundaries

Constrained boundaries

- Constrained boundaries allow the same flexibility as error spending functions, but are constructed in the scale of the estimated treatment effects (or any scale desired).

Overview:

- Calculate the estimated information at the interim analysis as a proportion of the total information.
- Calculate a revised group sequential design:
  - Use the values of $a_\ell$ and $d_\ell$ that were actually used at earlier interim analyses ($\ell < j$).
  - Calculate the new future values for $a_\ell$ and $d_\ell$ for $\ell \geq j$ using the original boundary shape function.
  - Find the value of $G$ that maintains the desired operating characteristics.
  - (Implemented in the function seqMonitor).
Constrained boundaries - Sepsis example

▶ As above, suppose the first interim analysis was conducted after data on 520 subjects (263 on the antibody arm, 257 on the placebo arm)

▶ Further suppose that 52 deaths were observed on the antibody arm and 65 deaths were observed on the placebo arm with:

\[
\hat{\theta}_1 = \frac{52}{263} \quad \hat{\theta}_0 = \frac{65}{257}
\]

Which represents 34.9% of the planned total information

▶ How should we adjust the pre-trial plan to accommodate this change?

* Refer to R-code in SepsisMonitor.R.
4.6 Implementing the trial monitoring plan

Background
Example: Constrained OBF design
Flexible Trial Monitoring
Error Spending Functions

Constrained Boundaries - Sepsis example

First interim analysis

Construct a new design with interim analyses at the actual information at the first interim analysis:

```r
> thetal.1 <- 52/263
> theta0.1 <- 65/257
> sepsis.IA1 <- update(sepsis.obf,
+   sample.size=c(520,850,1275,1700), ratio=c(263, 257),
+   variance=c(thetal.1*(1-thetal.1),theta0.1*(1-theta0.1)))
> sepsis.IA1$bound

STOPPING BOUNDARIES: Sample Mean scale

<table>
<thead>
<tr>
<th>Time</th>
<th>(N)</th>
<th>Efficacy</th>
<th>Futility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>520</td>
<td>-0.1325</td>
<td>0.0515</td>
</tr>
<tr>
<td>2</td>
<td>850</td>
<td>-0.0811</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>1275</td>
<td>-0.0541</td>
<td>-0.0270</td>
</tr>
<tr>
<td>4</td>
<td>1700</td>
<td>-0.0405</td>
<td>-0.0405</td>
</tr>
</tbody>
</table>
```

Notice that the new information timing is specified by updating *both* the sample size and the variance.

The trial should continue because the observed result $\hat{\theta}_{ia1} = \hat{\theta}_1 - \hat{\theta}_0 = -0.0552$ is in the continuation set.
4.6 Implementing the trial monitoring plan

Background

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

Constrained Boundaries - Sepsis example

First interim analysis

▶ Compare `sepsis.obf` and `sepsis.IA1`

```r
> seqPlotBoundary(sepsis.obf,sepsis.IA1, fixed=F)
> points(520,theta1.1-theta0.1,pch=4)
```

![Graph showing sample size vs. difference in proportions for sepsis.obf and sepsis.IA1](image)
Constrained Boundaries - Sepsis example

Second interim analysis

- Suppose the second analysis occurs after 1098 subjects with the observed result:
  - $\hat{\theta}_1 = \frac{120}{545}$ and $\hat{\theta}_0 = \frac{153}{553}$
  - So $\hat{\theta}_{ia2} = 0.220 - 0.277 = -0.565$

- Construction new decision criteria at for the second interim analysis with the following constraints:
  - Use the same critical values that were used at the first interim analysis $(a_1, d_1) = (-0.1325, 0.0515)$.
    (Accomplished using constrained boundaries)

- The information time of the second interim analysis is correct:
  - Sample size = 1098, with ratio = (545, 553)
  - Variance = $\hat{\theta}_{12}(1 - \hat{\theta}_{12}) + \hat{\theta}_{02}(1 - \hat{\theta}_{02})$
Constrained Boundaries - Sepsis example

Second interim analysis

- New design using constrained boundaries:
  
  ```
  > theta1.2 <- 120/545
  > theta0.2 <- 153/553
  
  > bnd <- rbind(sepsis.IA1$bound[1,],NA,NA)
  > bnd <- as.seqBoundary(bnd, scale="X" )
  
  > sepsis.IA2 <- update(sepsis.IA1,sample.size=c(520,1098,1700),
  + variance=c(theta1.2*(1-theta1.2),theta0.2*(1-theta0.2)),
  + ratio=c(545,553),exact.constraint=bnd)
  > sepsis.IA2$bound
  
  STOPPING BOUNDARIES: Sample Mean scale
  Efficacy Futility
  
  Time 1 (N= 520)  -0.1325  0.0515
  Time 2 (N= 1098) -0.0644 -0.0188
  Time 3 (N= 1700) -0.0416 -0.0416
  ```

- Time 1 decision criteria are same as `sepsis.IA1`.
- Observed result is in the continuation set:
  
  \[-0.0644 < \hat{\theta}_{ia2} < -0.0188\] so trial continues.
4.6 Implementing the trial monitoring plan

**Background**

Example: Constrained OBF design

**Flexible Trial Monitoring**

**Error Spending Functions**

**Constrained Boundaries**

**Issues When Monitoring a Trial**

Estimation of statistical information

Measuring study time

---

**Constrained Boundaries - Sepsis example**

**Second interim analysis**

- Compare `sepsis.IA2` with earlier designs:

```r
> seqPlotBoundary(sepsis.IA2, sepsis.IA1, sepsis.obf, fixed=F, col=(3:1))
> points(520, theta1.1-theta0.1, pch=4)
> points(1098, theta1.2-theta0.2, pch=4)
```

![Graph comparing sepsis.IA2 with earlier designs](image)
### Third interim analysis

- Suppose the second analysis occurs after 1098 subjects with the observed result:
  - $\hat{\theta}_1 = \frac{120}{545}$ and $\hat{\theta}_0 = \frac{153}{553}$
  - So $\hat{\theta}_{ia2} = 0.220 - 0.277 = -0.565$

- Construction new decision criteria at for the second interim analysis with the following constraints:
  - Use the same critical values that were used at the first interim analysis $(a_1, d_1) = (-0.1325, 0.0515)$.
    (Accomplished using constrained boundaries)

- The information time of the second interim analysis is correct:
  - Sample size = 1098, with ratio = (545, 553)
  - Variance = $\hat{\theta}_{12}(1 - \hat{\theta}_{12}) + \hat{\theta}_{02}(1 - \hat{\theta}_{02})$
4.6 Implementing the trial monitoring plan

Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

---

Constrained Boundaries - Sepsis example

**Third interim analysis**

- New design using constrained boundaries:
  
```r
> theta1.2 <- 120/545
> theta0.2 <- 153/553

> bnd <- rbind(sepsis.IA1$bound[1,],NA,NA)
> bnd <- as.seqBoundary(bnd, scale="X")

> sepsis.IA2 <- update(sepsis.IA1,sample.size=c(520,1098,1700),
+ variance=c(theta1.2*(1-theta1.2),theta0.2*(1-theta0.2)),
+ ratio=c(545,553),exact.constraint=bnd)
> sepsis.IA2$bound

STOPPING BOUNDARIES: Sample Mean scale

<table>
<thead>
<tr>
<th>Time</th>
<th>N</th>
<th>Efficacy</th>
<th>Futility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>520</td>
<td>-0.1325</td>
<td>0.0515</td>
</tr>
<tr>
<td>Time 2</td>
<td>1098</td>
<td>-0.0644</td>
<td>-0.0188</td>
</tr>
<tr>
<td>Time 3</td>
<td>1700</td>
<td>-0.0416</td>
<td>-0.0416</td>
</tr>
</tbody>
</table>
```

- Time 1 decision criteria are same as `sepsis.IA1`.
- Observed result is in the continuation set: 
  
\[-0.0644 < \hat{\theta}_{ia2} < -0.0188\] so trial continues.
Constrained Boundaries - Sepsis example

Third interim analysis

▶ Compare sepsis.IA3 with earlier designs:

```r
> seqPlotBoundary(sepsis.IA3,sepsis.IA2,sepsis.IA1,sepsis.obf,fixed=F,col=(4:1))
> points(520,theta1.1-theta0.1,pch=4)
> points(1098,theta1.2-theta0.2,pch=4)
> points(1700,theta1.3-theta0.3,pch=4)
```

```
   0  500  1000  1500
-0.20 -0.15 -0.10 -0.05  0.00  0.05  0.10
Sample Size
Difference in Proportions
● sepsis.IA3
● sepsis.IA2
● sepsis.IA1
● sepsis.obf
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
●
```

Sample Size vs Difference in Proportions
4.6 Implementing the trial monitoring plan

Background
Example: Constrained OBF design
Flexible Trial Monitoring
Error Spending Functions

Constrained Boundaries

Automating the process using the `seqMonitor` function

▶ The `seqMonitor` function:
  ▶ Analyzes the data at the interim analysis using the preceding design.
  ▶ Recalculates the information timing.
  ▶ Constrains on the boundaries from the preceding designs.
  ▶ Reports bias-adjusted inference at trial termination

▶ Sepsis example:

```r
> IA1 <- seqMonitor(sepsis.obf, response=Y.1, treatment=tx.1,
+   future.analyses=c(850,1275,1700))
> plot(IA1)

> IA2 <- seqMonitor(IA1, response=Y.2, treatment=tx.2,
+   future.analyses=1700)
> plot(IA2)

> IA3 <- seqMonitor(IA2, response=Y.3, treatment=tx.3)
> plot(IA3)
```
Choice of design characteristics to maintain during implementation

---

**It may not be possible to maintain all pre-trial design properties**

- The properties of the pre-trial group sequential design are retained during implementation:

  - Boundary shape function (e.g., Pocock or O’Brien-Fleming shapes) is retained.

  - Properties are recalculated as specified in the pre-trial design:

    * Calculate the power for a fixed alternative
    * Calculate the alternative for fixed power
    * Calculate maximal sample size for fixed alternative and power
    * (Getting it right can require careful thought)
Estimation of Statistical Information

Design stage vs. implementation stage

► At time of study design
  ▶ Sample size (power, alternative) calculations based on specifying statistical information available from each sampling unit

► During conduct of study
  ▶ Statistical information from a sampling unit may be different than originally estimated
    ▶ Variance of measurements
    ▶ Baseline event rates
    ▶ (Altered sampling distribution for treatment levels)
Estimation of Statistical Information

Computation of sample size

- Sample size formulas used in group sequential test design

\[ N = \frac{\delta_1^2 V}{(\Delta_1 - \Delta_0)^2} \]

- \( N \): maximal number of sampling units
- \( \delta_1 \): alternative for which a standardized form of a level \( \alpha \) test has power \( \beta \)
- \( 1/V \): statistical information contributed by each sampling unit
Estimation of Statistical Information

Computing sample size

- Sample size formulas used in group sequential test design are completely analogous to those used in fixed sample studies.

\[ N = \frac{\delta_1^2 V}{(\Delta_1 - \Delta_0)^2} \]

- In a fixed sample two arm test of an (approximately) normal mean we have:
  \[ \delta_1 = z_{1-\alpha/2} + z_\beta \]
  \[ V = 2\sigma^2 \]
4.6 Implementing the trial monitoring plan

Background
Example: Constrained OBF design

Flexible Trial Monitoring

Error Spending Functions

Constrained Boundaries

Issues When Monitoring a Trial

Estimation of statistical information

Measuring study time

---

**Incorrect estimates of information at design stage**

- Effect of using incorrect estimates of statistical information at the design stage
  - Using the specified sample size, the design alternative will not be detected with the desired power
  - Using the specified sample size, the alternative detected with the desired power will not be the design alternative
  - In order to detect the design alternative with the desired power, a different sample size is needed
Estimation of Statistical Information

Maintaining maximal sample size or power

- If maximal sample size is maintained, the study discriminates between null hypothesis and an alternative measured in units of statistical information

\[ N = \frac{\delta_1^2 V}{(\Delta_1 - \Delta_0)^2} = \frac{\delta_1^2}{\left(\frac{(\Delta_1 - \Delta_0)^2}{V}\right)} \]

- If statistical power is maintained, the study sample size is measured in units of statistical information

\[ \frac{N}{V} = \frac{\delta_1^2}{(\Delta_1 - \Delta_0)^2} \]
Estimation of Statistical Information

### Measuring study time

- Flexible methods compute boundaries at an interim analysis according to study time at that analysis.

- Study time can be measured by:
  - Proportion of planned number of subjects accrued (maintains maximal sample size)
  - Proportion of planned statistical information accrued (maintains statistical power)
  - (Calendar time— not really advised)
Estimation of Statistical Information

**Measuring study time**

- In either case, we must decide how we will deal with estimates of statistical information at each analysis when constraining boundaries.

- Statistical information in clinical trials typically has two parts:
  - $V =$ variability associated with a single sampling unit
  - The distribution of sampled levels of treatment

- In many clinical trials, the dependence on the distribution of treatment levels across analyses is only on the sample size $N$.
Estimation of Statistical Information

Possible approaches

- At each analysis estimate the statistical information available, and use that estimate at all future analyses
  - Theoretically, this can result in estimates of negative information gained between analyses

- At each analysis use the sample size with the current best estimate of $\sqrt{V}$
  - The 1:1 correspondence between boundary scales (see Session 3) is broken at previously conducted analyses
## Estimation of Statistical Information

### Possible approaches

- **In RCT design**, all probability models have statistical information directly proportional to sample size for block randomized experiments, thus we chose to update $V$ at all analyses using the current best estimate.

- Other statistical packages (PEST, EaSt) constrain boundaries using the estimate of statistical information available at the previous analyses.

- There is no clear best approach.
Estimation of Statistical Information

Possible approaches

- Overall, I think it makes more sense to use the best estimate of the variance of an observation when estimating a sampling distribution.

- This avoids the possibility of negative information, but allows the conflicting results described above.