# Introduction to Clinical Trials - Day 2

## Session 6 - Group Sequential Monitoring

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Trial monitoring

## Elements and motivation for trial monitoring

- Motivation: Many trials have been stopped early:
  - Physician health study showed that aspirin reduces the risk of cardiovascular death.
  - A phase III study of tamoxifen for prevention of breast cancer among women at risk for breast cancer showed a reduction in breast cancer incidence.
  - A phase III study of anti-arrhythmia drugs for prevention of death in people with cardiac arrhythmia stopped due to excess deaths with the anti-arrhythmia drugs.
  - A phase III study of folic acid supplements for prevention of neural tube defects.
  - Women's Health Initiative: Hormones cause heart disease.

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Trial monitoring



## Elements and motivation for trial monitoring

- What is trial monitoring?
  - Monitoring for quality control; for example,
    - Patient accrual.
    - Data quality/completeness.
    - Unanticipated adverse events.
  - Monitoring study endpoints(s); for example,
    - Treatment benefits.
    - Toxicity differences.
  - Good quality control should be part of every study to ensure that the study achieves its goals.
  - Monitoring study endpoints is not applicable in every study, and requires special statistical methods to avoid increased statistical errors.

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Trial monitoring

## Elements and motivation for trial monitoring

- Reasons to monitor study 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).

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Trial monitoring

## Elements and motivation for trial monitoring

- If not done properly, monitoring of endpoints can lead to biased results:
  - Data driven analyses cause bias:
    - Analyzing study results because they look good leads to an overestimate of treatment benefits.
  - Publication or presentation of 'preliminary results' can affect:
    - Ability to accrue subjects.
    - Type of subjects that are referred and accrued.
    - Treatment of patients not in the study.
  - Failure to design for interim analyses can lead to hasty decisions. Decisions made 'in the heat of the moment' are subject to:
    - Inadequate consideration of trade-offs between competing endpoints (toxicity versus benefit).
    - External pressures from study investigators or sponsors.
    - Lack of objectivity by study monitors.

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

## Elements and motivation for trial monitoring

- Thus,
  - Monitoring of study endpoints is often required for ethical reasons.
  - Monitoring of study endpoints must carefully planned as part of study design to:
    - Avoid bias
    - Assure careful decisions
    - Maintain desired statistical properties

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Elements and motivation for trial monitoring

## Key elements of monitoring

- How are trials monitored?
  - Investigator knowledge of interim results can lead to biased results:
    - Negative results may lead to loss of enthusiasm.
    - Positive interim results may lead to inappropriate early publication.
    - Either result may cause changes in the types of subjects who are recruited into the trial.
  - “Data Safety and Monitoring Boards (DSMB)" are used to avoid biased decisions:
    - DSMB members are *independent* of the study investigators
    - The DSMB reviews unblinded data in the midst of a trial to:

      1. Assure the trial is safe to continue.
      2. Make decisions about early termination based on the statistical monitoring plan (“group-sequential clinical trial design").

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Elements and motivation for trial monitoring



## Key elements of monitoring

The trial monitoring plan is typically pre-specified in two documents:

- DSMB charter:
  - Defines scope of trial monitoring
  - Defines DSMB responsibilities
  - Defines sponsor responsibilities
  - Pre-specifies monitoring plans and decisions (reasons for stopping)
- Interim Statistical Analysis Plan (ISAP):
  - Defines monitoring endpoint(s)
  - Pre-specifies analysis timing, decision criteria, and rationale
  - Pre-specifies methods for implementation (changes to analysis timing)
  - Pre-specifies adjustments to statistical inference about treatment effects

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Elements and motivation for trial monitoring



## Key elements of monitoring

- Typical content for DSMB charter:
  - Trial synopsis; for example:
    - Summary of design
    - Eligibility/exclusions
    - Statistical design and sample size
  - DSMB organization
    - Composition and selection of members
  - Responsibilities of DSMB
    - What will be monitored (accrual, QC, safety, endpoints?)
  - Responsibilities of sponsor
    - Providing open/closed reports; data summaries
  - Committee meetings:
    - Open session; closed session; executive session
  - Communication
    - Open report; closed report to be provided to DSMB
    - Responsibility for meeting minutes (open and closed minutes)
    - Process for DSMB recommendations

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Elements and motivation for trial monitoring



## Key elements of monitoring

- Typical content for ISAP:
  - Safety monitoring plan (if there are formal safety interim analyses)
    - Decision rules for formal safety analyses
    - Evaluation of decision rules (power, expected sample size, stopping probability)
    - Methods for modifying rules (changes in timing of analyses)
    - Methods for inference (bias adjusted inference)
  - Monitoring plan for primary endpoint(s)
    - Decision rules and reasons for early termination (e.g., efficacy, futility, equivalence, harm)
    - Evaluation of decision rules (power, expected sample size, stopping probability)
    - Methods for modifying rules (changes in timing of analyses)
    - Methods for inference (bias adjusted inference)
  - Data handling and responsibilities for analysis

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

## Statistical framework for trial monitoring: Statistical design of the fixed-sample trial

- The interim statistical analysis plan is based on the fixed sample design
  - Primary endpoint
  - Probability model
  - Functional
  - Contrast
  - Statistical hypotheses
  - Statistical standards for decisions (interval estimate)

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

## Statistical framework for trial monitoring: Statistical design of the fixed-sample trial

- The statistical decision criteria are referenced to the trial's design hypotheses. For example:
  - One-sided superiority test (assume small $\theta$ favors new treatment):

$$\begin{aligned} \text{Null:} \quad & \theta \geq \theta_{\emptyset} \\ \text{Alternative:} \quad & \theta \leq \theta_{+} \end{aligned}$$

with $\theta_{+} < \theta_{\emptyset}$, and $\theta_{+}$ is chosen to represent the smallest difference that is clinically important.

  - Two-sided (equivalence) test:

$$\begin{aligned} \text{Null:} \quad & \theta = \theta_{\emptyset} \\ \text{Lower Alternative:} \quad & \theta \leq \theta_{-} \\ \text{Upper Alternative:} \quad & \theta \geq \theta_{+} \end{aligned}$$

with $\theta_{-} < \theta_{\emptyset} < \theta_{+}$. $\theta_{-}$ and $\theta_{+}$ denote the smallest important differences.



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

## Statistical framework for trial monitoring: Selecting decision criteria

- A decision to stop needs to consider what has or has not been ruled out. For example
  - One-sided superiority test (assume small $\theta$ favors new treatment):
    - Stop for superiority when any harm ($\theta \geq \theta_{\emptyset}$) has been ruled out.
    - Stop for futility when important benefits ($\theta \leq \theta_{+}$) have been ruled out.
  - Two-sided (equivalence) test:
    - Stop for treatment *A* better than treatment *B* when inferiority of *A* ($\theta \leq \theta_{\emptyset}$) has been ruled out.
    - Stop for treatment *B* better than treatment *A* when inferiority of *B* ($\theta \geq \theta_{\emptyset}$) has been ruled out.
    - Stop for equivalence when important differences (either $\theta \geq \theta_{+}$ or $\theta \leq \theta_{-}$ ) have been ruled out.
- The hypotheses that have been ruled in/out are given by the interval estimate.



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Overview of group sequential designs

## Statistical framework for trial monitoring: Group sequential designs (superiority trial)

- Suppose that the trial is planned for $j = 1, ..., J$ interim analyses.
- Let $\hat{\theta}_j$ denote the estimated treatment effect at the $j$th analysis.
- Consider stopping criteria $a_j < d_j$ with:

$$\begin{aligned} \hat{\theta}_j \leq a_j &\Rightarrow \text{Decide new treatment is superior} \\ \hat{\theta}_j \geq d_j &\Rightarrow \text{Decide new treatment is not superior} \\ a_j < \hat{\theta}_j < d_j &\Rightarrow \text{Continue trial} \end{aligned}$$

  Set $a_J = d_J$ so that the trial stops by the $J$th analysis.

- How should we choose these critical values?

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Statistical framework for trial monitoring

## Inadequacy of Fixed Sample Methods

- Suppose we simply ignore the fact that we are repeatedly testing our hypothesis
- We can quickly see the impact of this via simulation
  - Let $X_i \sim_{\text{iid}} \mathcal{N}(\theta, \sigma^2)$
  - $j = 1, ..., 4$ equally spaced analyses at 25, 50, 75, and 100 observations
  - Test statistic after $n_j$ observations have been accrued

$$\bar{X}_{n_j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_i$$

  - Test $H_0 : \theta = 0$ with level $\alpha = .05$
- Fixed sample methods (2-sided test): Reject $H_0$ first time

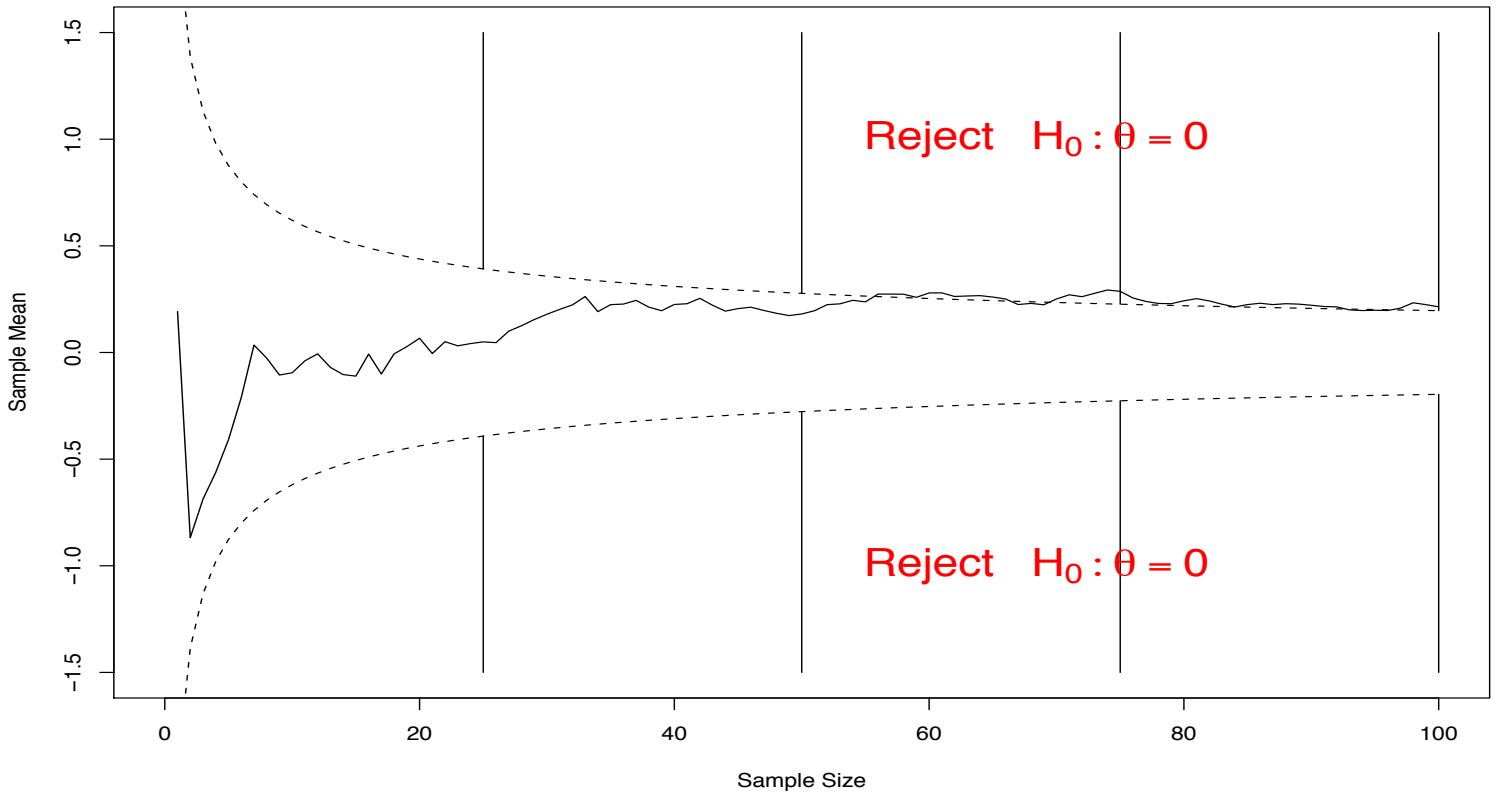
$$|\bar{X}_{n_j}| > z_{1-\alpha/2} \frac{\sigma}{\sqrt{n_j}}, \qquad j = 1, 2, 3, 4$$

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Statistical framework for trial monitoring

## Inadequacy of Fixed Sample Methods : Simulation

- Consider the sample path of the statistic for a single simulated trial

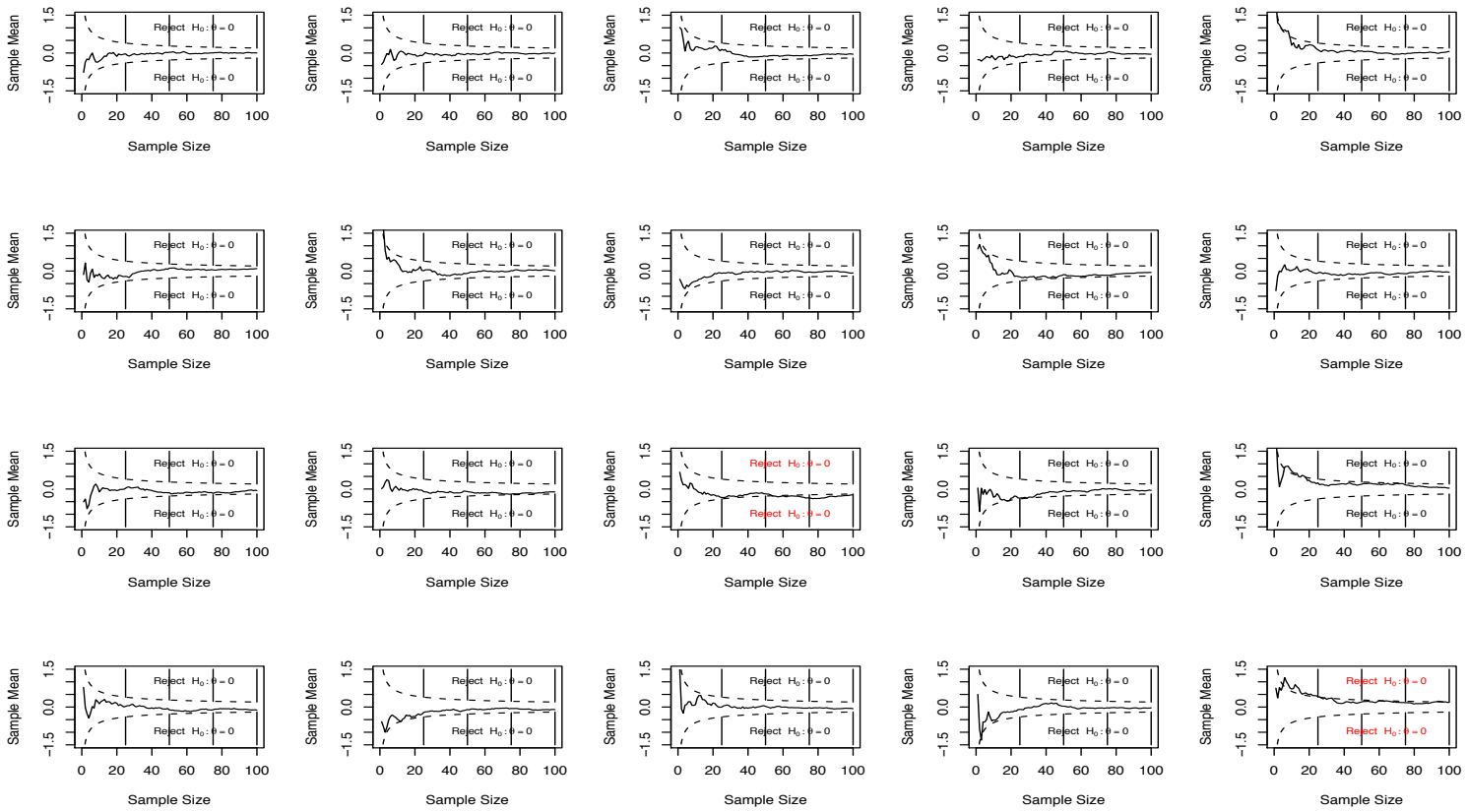



Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Statistical framework for trial monitoring

## Inadequacy of Fixed Sample Methods : Simulation

- Consider the sample path of the statistic for 20 randomly sampled trials






Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Statistical framework for trial monitoring

## Inadequacy of Fixed Sample Methods : Simulation

- Simulated type I error rate using fixed sample methods
- Based on 100,000 simulations

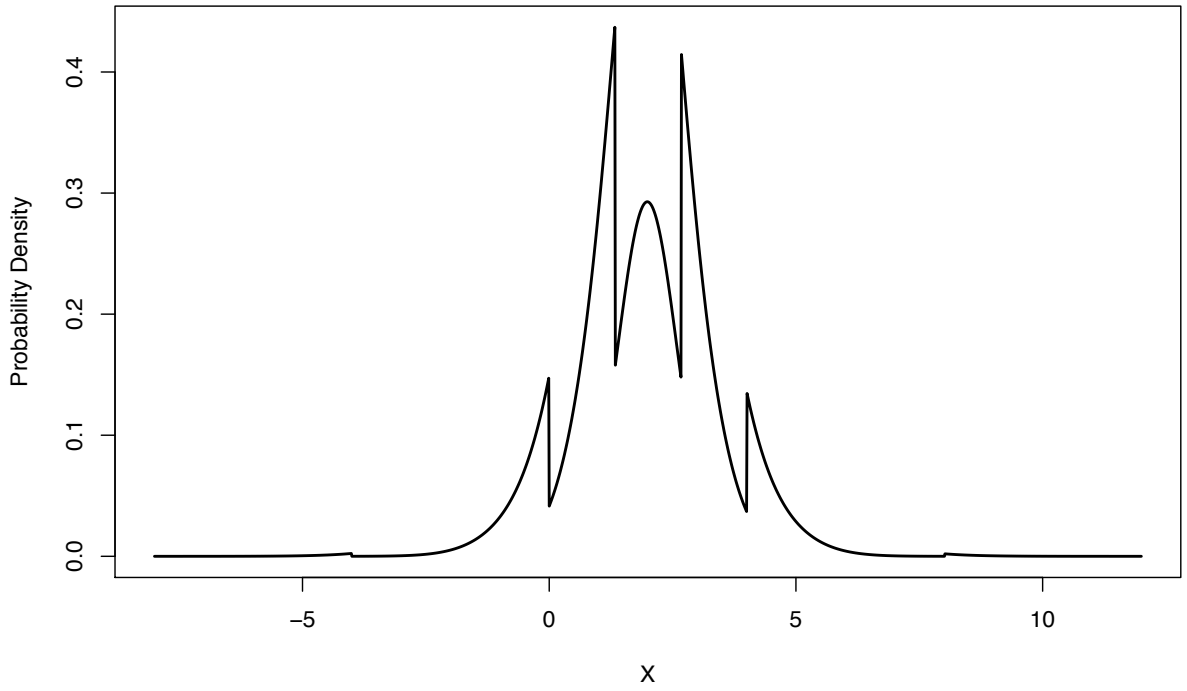
| Significant at | Proportion Significant | Number Significant | Proportion Significant |
|---|---|---|---|
| Analysis 1 | 0.05075 | Exactly 1 | 0.07753 |
| Analysis 2 | 0.04978 | Exactly 2 | 0.02975 |
| Analysis 3 | 0.05029 | Exactly 3 | 0.01439 |
| Analysis 4 | 0.05154 | All 4 | 0.00554 |
| | | Any | 0.12721 |

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Interim analyses require special methods

## Sampling density for sequentially-monitored test statistic

- The filtering due to interim analyses creates non-standard sampling densities as the basis for inference.
- Sampling density depends on the stopping rule.
- In order to correct the type 1 error rate, we must be able to compute the density of the statistic that accounts for the possibility of stopping at interim analyses






Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Sampling density for sequentially sampled test statistic

- Let $C_j$ denote the continuation set at the $j$th interim analysis.
- Let $(M, S)$ denote the bivariate statistic 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.
- The sampling density for the observation $(M = m, S = s)$ is:

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

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

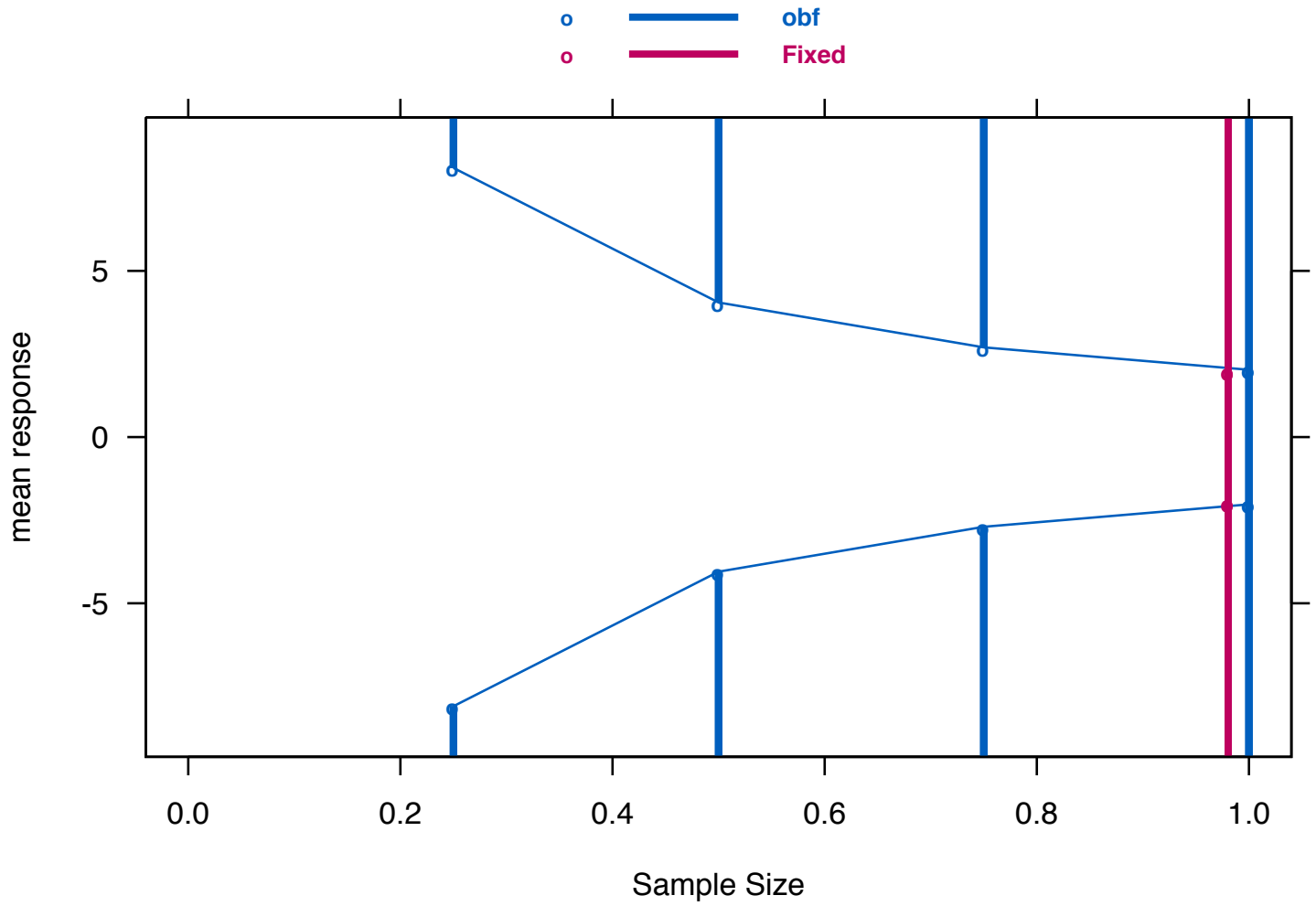
$$\begin{aligned} f(1, s; \theta) &= \frac{1}{\sqrt{n_1 V}} \phi\left(\frac{s - n_1\theta}{\sqrt{n_1 V}}\right) \\ f(j, s; \theta) &= \int_{C_{(j-1)}} \frac{1}{\sqrt{n_j V}} \phi\left(\frac{s - u - n_j\theta}{\sqrt{n_j V}}\right) f(j-1, u; \theta)\, du, \\ & \quad j = 2, \ldots, m \end{aligned}$$

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Types of group sequential designs

## Example: O'Brien-Fleming (OBF) 2-sided design

- Using the correct sampling density, we can choose boundary values that maintain experiment wise Type I error






Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Types of group sequential designs



## Example: O'Brien-Fleming (OBF) 2-sided design

- Simulated type I error rate using fixed sample methods
- Based on 100,000 simulations

| Significant at | Proportion Significant | Number Significant | Proportion Significant |
|---|---|---|---|
| Analysis 1 | 0.00006 | Exactly 1 | 0.03610 |
| Analysis 2 | 0.00409 | Exactly 2 | 0.01198 |
| Analysis 3 | 0.01910 | Exactly 3 | 0.00210 |
| Analysis 4 | 0.04315 | All 4 | 0.00001 |
| | | Any | 0.05019 |

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

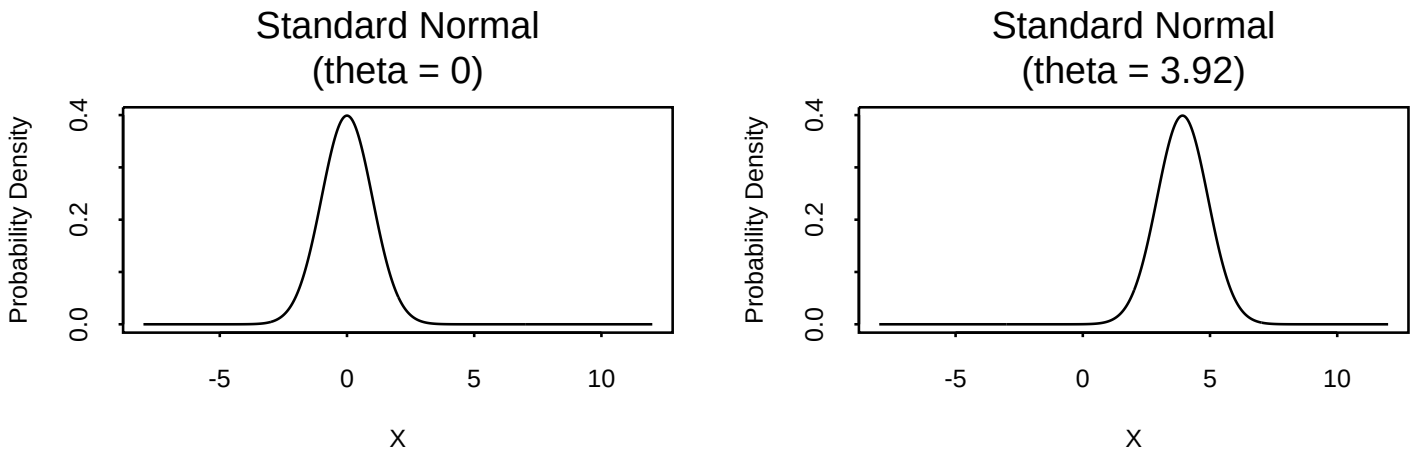
Types of group sequential designs

Example: Sepsis trial

# Types of group sequential designs

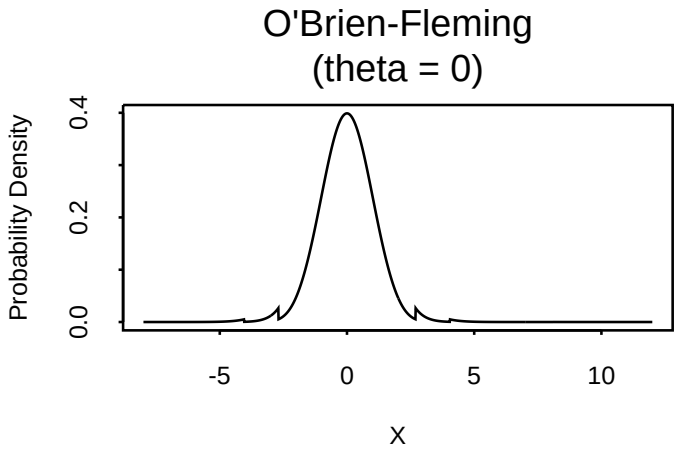
## Example: O'Brien-Fleming (OBF) 2-sided design

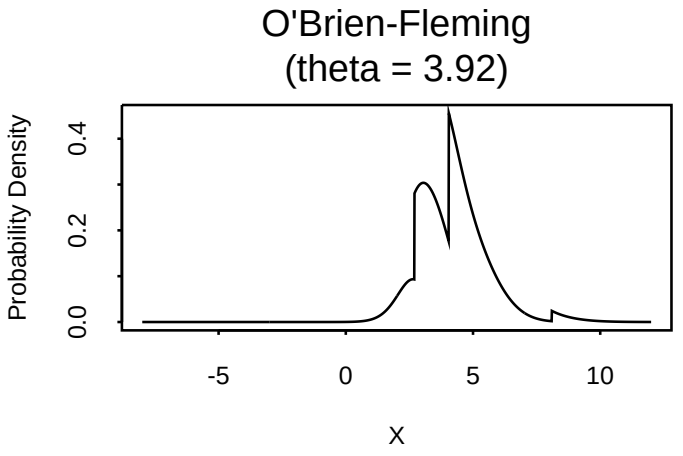
- Sampling density for OBF boundaries with $\theta = 0$ and $\theta = 3.92$ (corresponding Normal sampling density for comparison):









# Types of group sequential designs

## Boundary shape functions

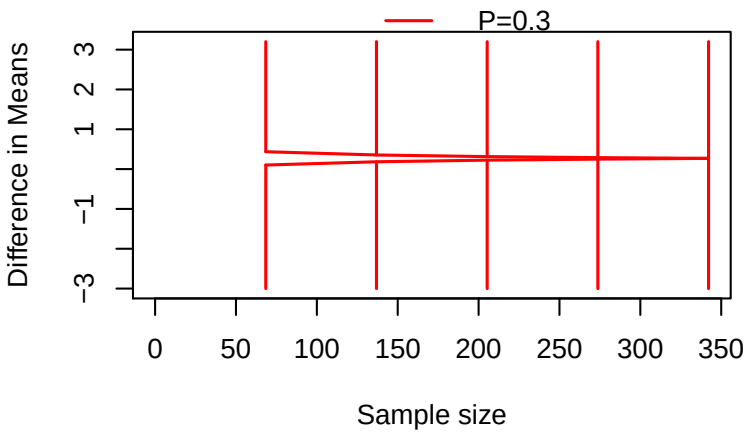
- There are an infinite number of stopping boundaries to choose from that will maintain a given family-wise error
  - They will differ in required sample size and power
- Kittelson and Emerson (1999) described a “unified family" of designs that are parameterized by three parameters ($A$, $R$, and $P$)
- Parameterization of boundary shape function includes many previously described approaches
  - Wang & Tsiatis boundary shape functions:
    - $A = 0$, $R = 0$, and $P > 0$
    - $P = 0.5$ : Pocock (1977)
    - $P = 1.0$ : O’Brien-Fleming (1979)
  - Triangular Test boundary shape functions (Whitehead):
    - $A = 1$, $R = 0$, and $P = 1$
  - Sequential Conditional Probability Ratio Test (Xiong):
    - $R = 0.5$, and $P = 0.5$

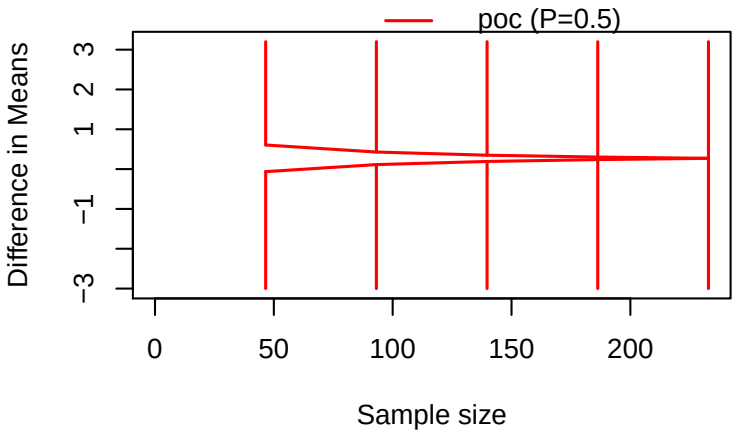
Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

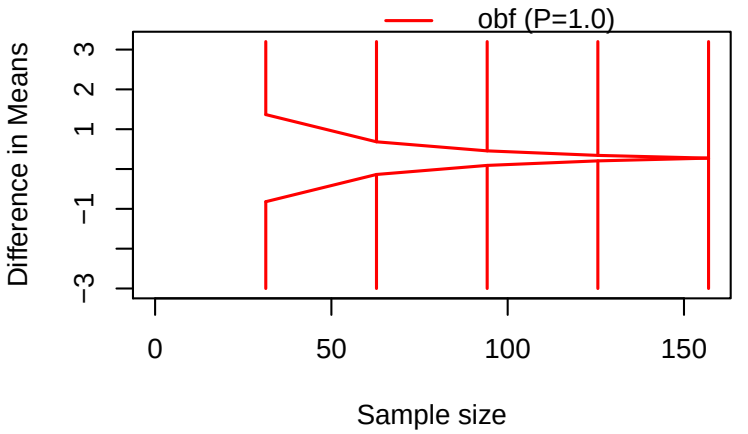
# Types of group sequential designs

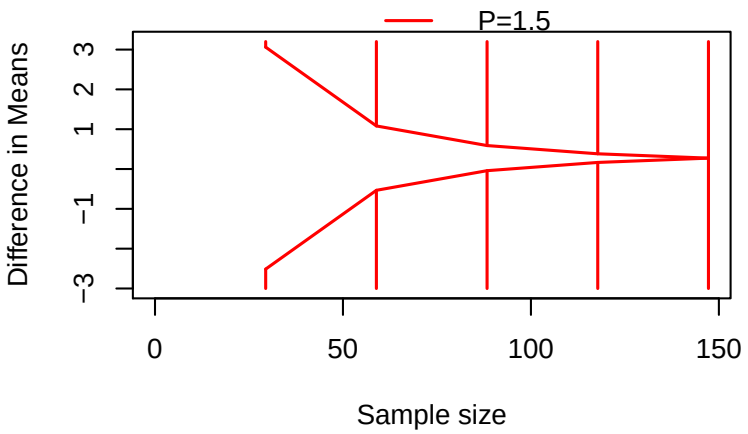
## Boundary shape functions

- Consider differing choices of $P$



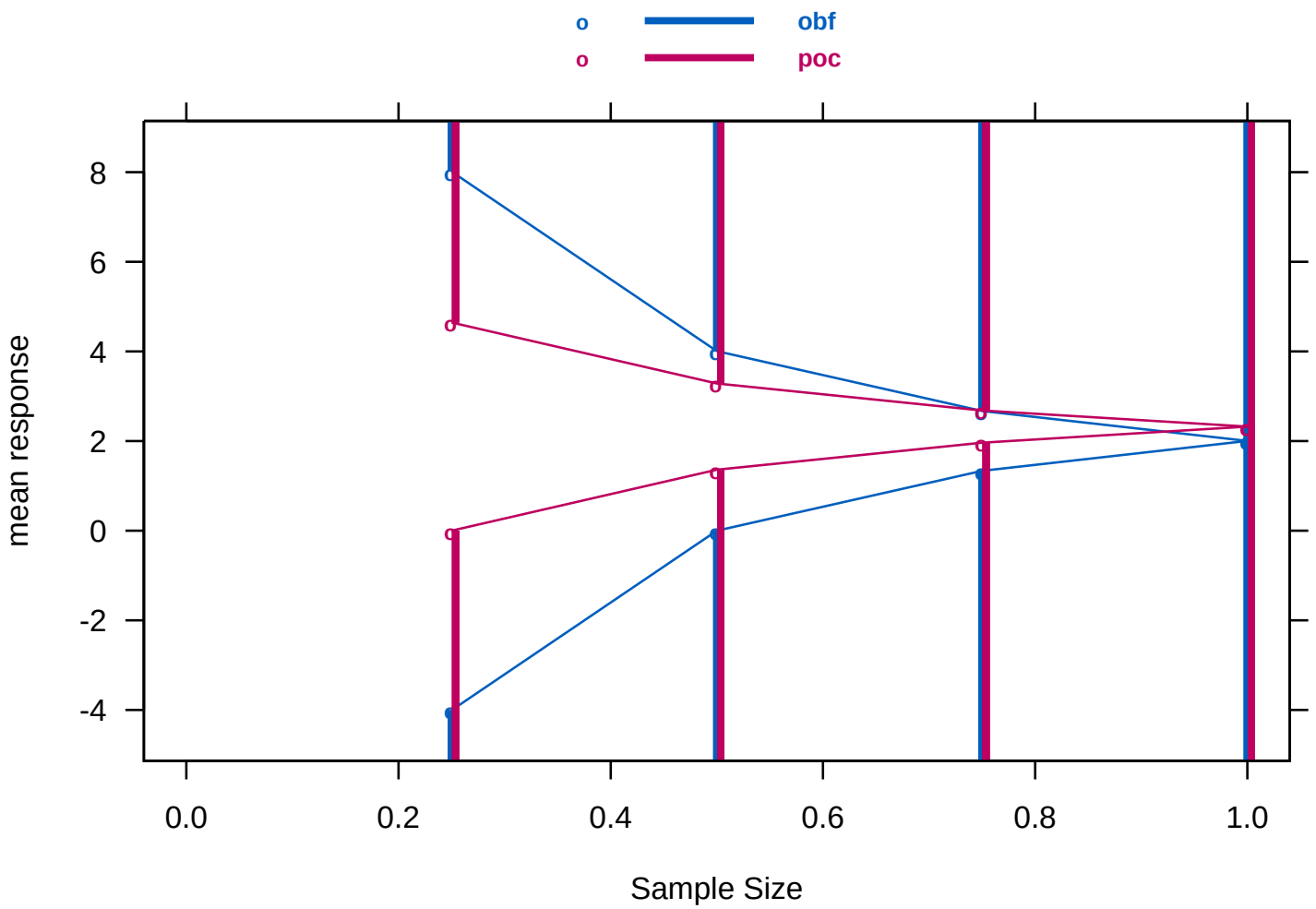






Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: OBF (P=1) versus Pocock (P=0.5) 1-sided designs






Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Types of group sequential designs



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

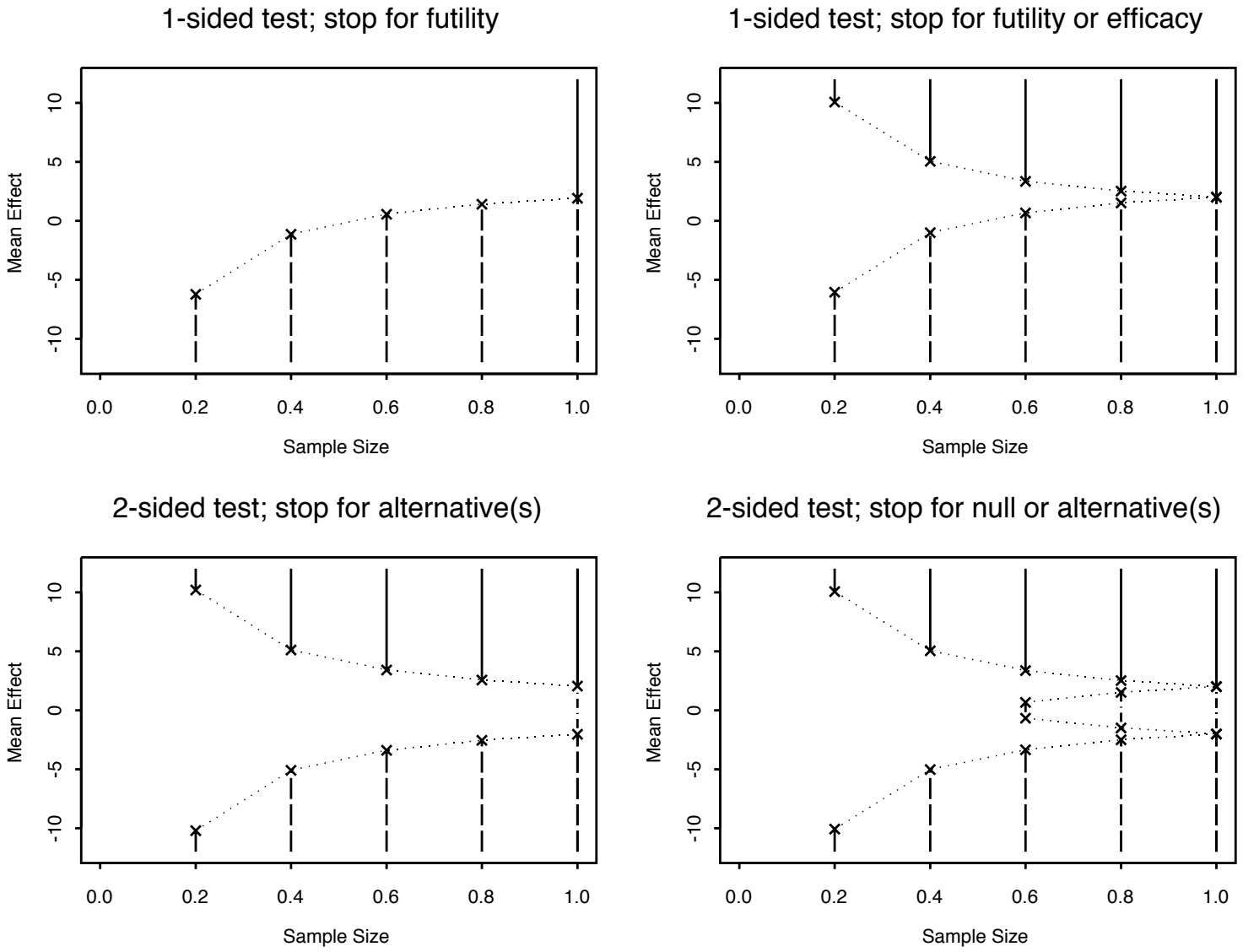
Types of group sequential designs

Example: Sepsis trial

## Group sequential designs can be formulated for various hypotheses

- Four design categories:
  - 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)

# Four general design categories






Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Types of group sequential designs



## So how should we choose a stoping rule?

- Consider appropriate type of hypothesis to test
- Maintain statistical design criteria of the fixed sample trial:
  - Type I error rate of $\alpha = 0.025$ (one-sided test) or $\alpha = 0.05$ (two-sided test).
  - Maintain maximal sample size (with potential loss of power)
  - Maintain power (with larger maximal sample size)
- Other considerations when selecting critical values:
  - Number of interim analyses
  - Timing of interim analyses
  - Degree of early conservatism
  - Characteristics of the sample size distribution:
    - Expected sample size (Average Sample Number; ASN)
    - Quantiles of the sample size distribution
    - Maximal sample size
    - Stopping probabilities at each of the interim analyses

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Interim analyses require special methods



## Characteristics of the group sequential sampling density

- Density is not shift invariant
- Jump discontinuities
- Requires numerical integration
- Sequential testing introduces bias:

| | $E(\hat{\theta})$ | |
|---|---|---|
| $\theta$ | OBF | Pocock |
| 0.00 | -0.29 | -0.48 |
| 1.96 | 1.95 | 1.82 |
| 3.92 | 4.21 | 4.38 |

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Case Study : Sepsis Trial



## Background

- Critically ill patients often get overwhelming bacterial infection (sepsis), after which mortality is high
- Gram negative sepsis is often characterized by production of endotoxin, which is thought to be the cause of much of the ill effects of gram negative sepsis
- Hypothesis: Administering antibody to endotoxin may decrease morbidity and mortality
- Two previous randomized clinical trials showed a slight benefit
- There were no safety concerns at the inception of the trial

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Case Study : Sepsis Trial



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

## Definition of Treatment

- Single administration of antibody to endotoxin within 24 hours of diagnosis of sepsis
- Reductions in dose not applicable
- Ancillary treatments unrestricted

# Case Study : Sepsis Trial



## Defining the target population

- Patients in ICU with newly diagnosed sepsis
- Infected with gram negative organisms
  - culture proven
  - gram stain

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Case Study : Sepsis Trial



## Defining the Comparison Group

- Need to ensure scientific credibility for regulatory approval
- Crossover designs impossible
- Ultimate decision:
  - Single comparison group treated with placebo
    - Not interested in studying dose response
    - No similar current therapy (still ethical to use placebo)
  - Randomized
    - Allow for causal inference
    - No blocking

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Case Study : Sepsis Trial



## Defining the Outcomes of Interest

- **Goals**:
  - Primary: Increase survival
    - Long term (always best)
    - Short term (many other processes may intervene)
  - Secondary: Decrease morbidity
- **Refinement of the primary endpoint**
  - Possible primary endpoints
    - Time to death
    - Mortality rate at a fixed point in time
    - Time alive out of ICU during fixed period of time

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Case Study : Sepsis Trial



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

## Refinement of the primary endpoint

Option 1: Time to death (censored continuous data)

- Trial is likely to have early censoring due to logistical constraints of the trauma centers
- Such early censoring might place emphasis on clinically meaningless improvements in very short term survival
  - eg. We may be detecting differences in 1 day survival even though there is no difference in survival at 10 days

# Case Study : Sepsis Trial



## Refinement of the primary endpoint

Option 2: Mortality rate at a fixed point in time (binary data)

- Allows for choice of a *scientifically* relevant time frame
  - Treatment is a single administration; short half-life
- Allows for choice of a *clinically* relevant time frame
  - Avoids sensitivity to improvements lasting only short periods of time

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Case Study : Sepsis Trial



Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

## Refinement of the primary endpoint

Option 3: Time alive out of the ICU during a fixed period of time (continuous data)

- Incorporates morbidity endpoints
- Addresses patient quality of life
- May be sensitive to clinically meaningless improvements depending upon the time frame chosen

# Case Study : Sepsis Trial



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

## Refinement of the primary endpoint

Final Choice: Mortality rate at a fixed point in time (binary data)

- Sponsor proposed 14 day mortality
- FDA countered with a suggestion of 28 day mortality

# Case Study : Sepsis Trial



## Method of analysis

- Test for differences in binomial proportions
  - Ease of interpretation
  - 28 day mortality not a rare event
  - 1:1 correspondence with tests of odds ratio (for known baseline event rates)
- No adjustment for covariates
- Statistical information dictated by mean variance relationship of Bernoulli random variables:
  - Let $Y_{ki}$ denote binary response (mortality at 28 days) for $i$-th subject in group $k$, $k = 0, 1$
  - $Y_{ki} \sim \mathcal{B}(1, \theta_k)$
  - $\mathrm{E}[Y_{ki}] = \theta_k$ and $\mathrm{Var}[Y_{ki}] = \theta_k(1 - \theta_k)$

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Case Study : Sepsis Trial



## Definition of statistical hypotheses

### Null hypothesis

- No difference in mortality between groups
- Estimated baseline rate
  - 28 day mortality: 30%
  - (needed in this case to estimate variability)

### Alternative hypothesis

- One-sided test for decreased mortality
- Targeted 28 day mortality rate in antibody arm: 25%
  - 5% absolute difference in mortality

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Case Study : Sepsis 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

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Case Study : Sepsis Trial



## 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 5% absolute difference in mortality
  - Statistical variability based on mortality rate of 30% in placebo arm

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Case Study : Sepsis Trial



## Determination of sample size

- General sample size formula:
  - $\delta$ = standardized alternative
  - $\Delta$ = difference between null and alternative treatment effects
  - $V$ = variability of a single sampling unit
  - $n$ = number of sampling units

$$n = \frac{\delta^2 V}{\Delta^2}$$

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Case Study : Sepsis Trial

## Determination of sample size

- Parameter values in the present case:
  - $\delta = (z_{1-\alpha} + z_{\beta})$ with $\alpha = 0.025$ and $\beta = 0.80$
  - $\Delta = \theta_{1,H_1} - \theta_{0,H_1} = -0.05$
  - $V = \theta_{1,H_1}(1 - \theta_{1,H_1}) + \theta_{0,H_1}(1 - \theta_{0,H_1}) =$ $.25 \times .75 + .3 \times .7 = .3975$
  - $n$ = sample size per arm

$$n = \frac{\delta^2 V}{\Delta^2} = \frac{(1.96 + .841)^2 \times .3975}{(-.05)^2} = 1247.97 \rightarrow 1248$$

Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Case Study : Sepsis Trial

## Resulting Fixed sample design

- Problem: Sponsor was concerned that 2496 ($2 \times 1248$) patients would be logistically infeasible and wanted to consider a design with 1700 patients
- Operating characteristics with $N$=1700:
  - Critical value : -0.0424
  - 64% power for alternative of 5% absolute difference; 90% power for alternative of 7% absolute difference; Corresponding p-value : 0.025
  - 95% confidence interval : (-0.085, 0)
  - Interpretation: Smallest magnitude of (observed) effect which would result in a significant result is a 4.24% decrease in mortality on the treatment arm with corresponding CI ( -0.085, 0).



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Example: Sepsis Trial



## Addition of interim analyses

- FDA requires an interim safety analysis
- DSMB considers 4 interim analyses to stop for harm or futility using an O'Brien-Fleming stopping rule

```
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 = 0.9021)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N=  425)     -Inf   0.0883
    Time 2 (N=  850)     -Inf   0.0019
    Time 3 (N= 1275)     -Inf  -0.0269
    Time 4 (N= 1700)  -0.0413  -0.0413
```

Elements of Trial Monitoring

Group Sequential Designs

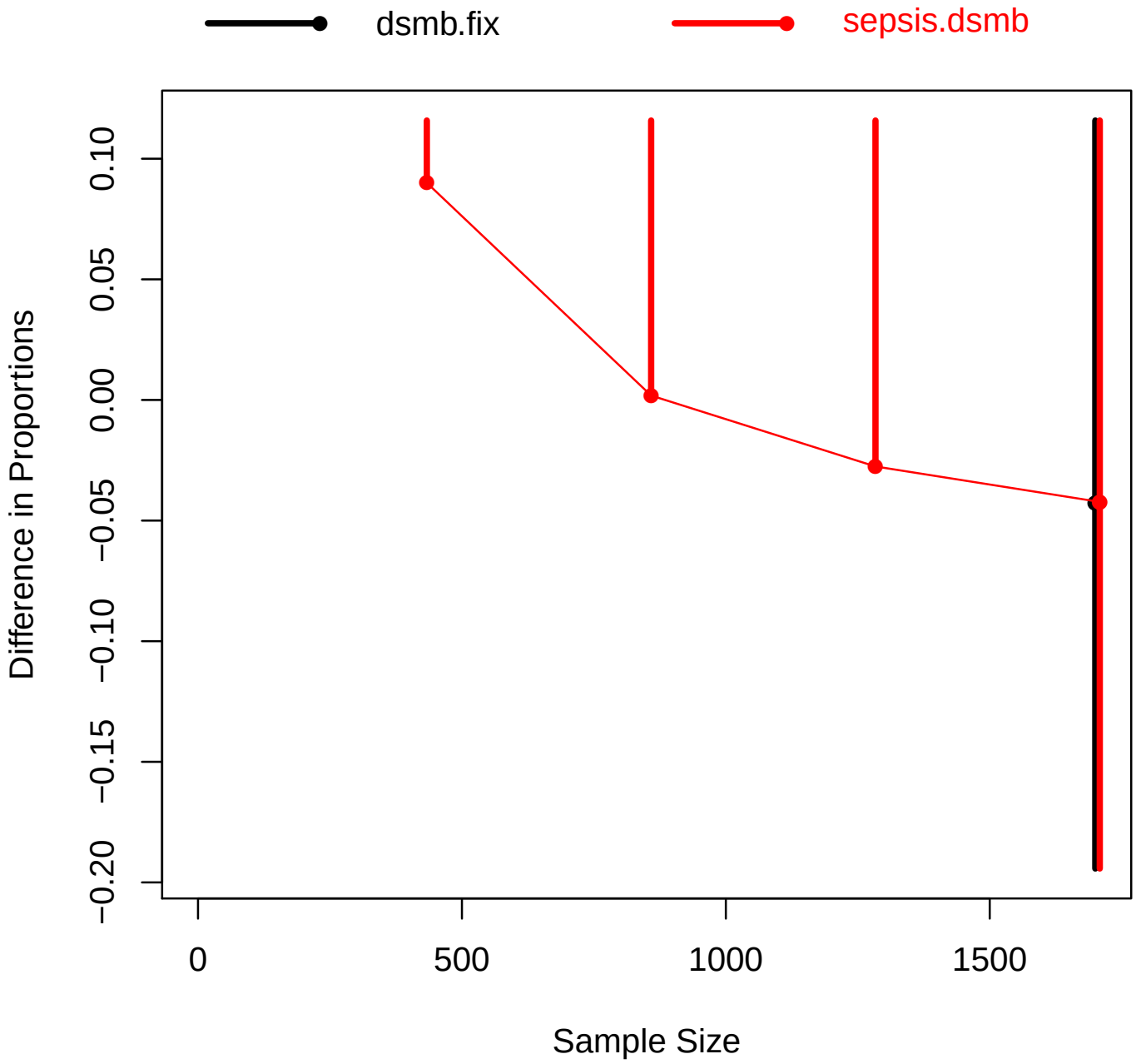
Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Example: Sepsis Trial

- Stopping boundaries






Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial



## Addition of interim analyses

- Sponsor and DSMB would also like to consider stopping for efficacy
- Consider an O'Brien-Fleming boundary for both efficacy and futility

```
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 = 0.8947)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N=  425)  -0.1710   0.0855
    Time 2 (N=  850)  -0.0855   0.0000
    Time 3 (N= 1275)  -0.0570  -0.0285
    Time 4 (N= 1700)  -0.0427  -0.0427
```

Elements of Trial Monitoring

Group Sequential Designs

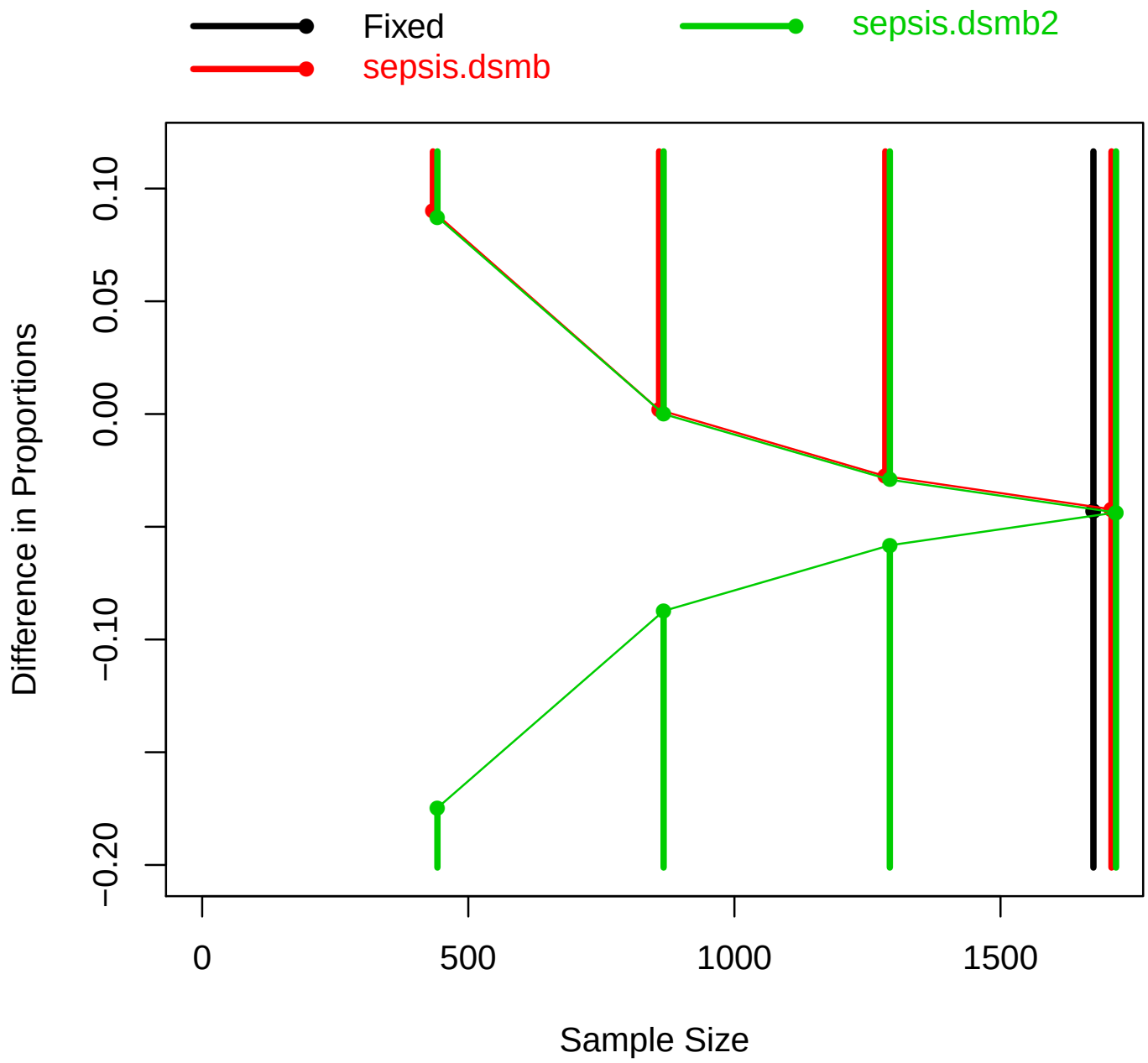
Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Example: Sepsis Trial

- Stopping boundaries






Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

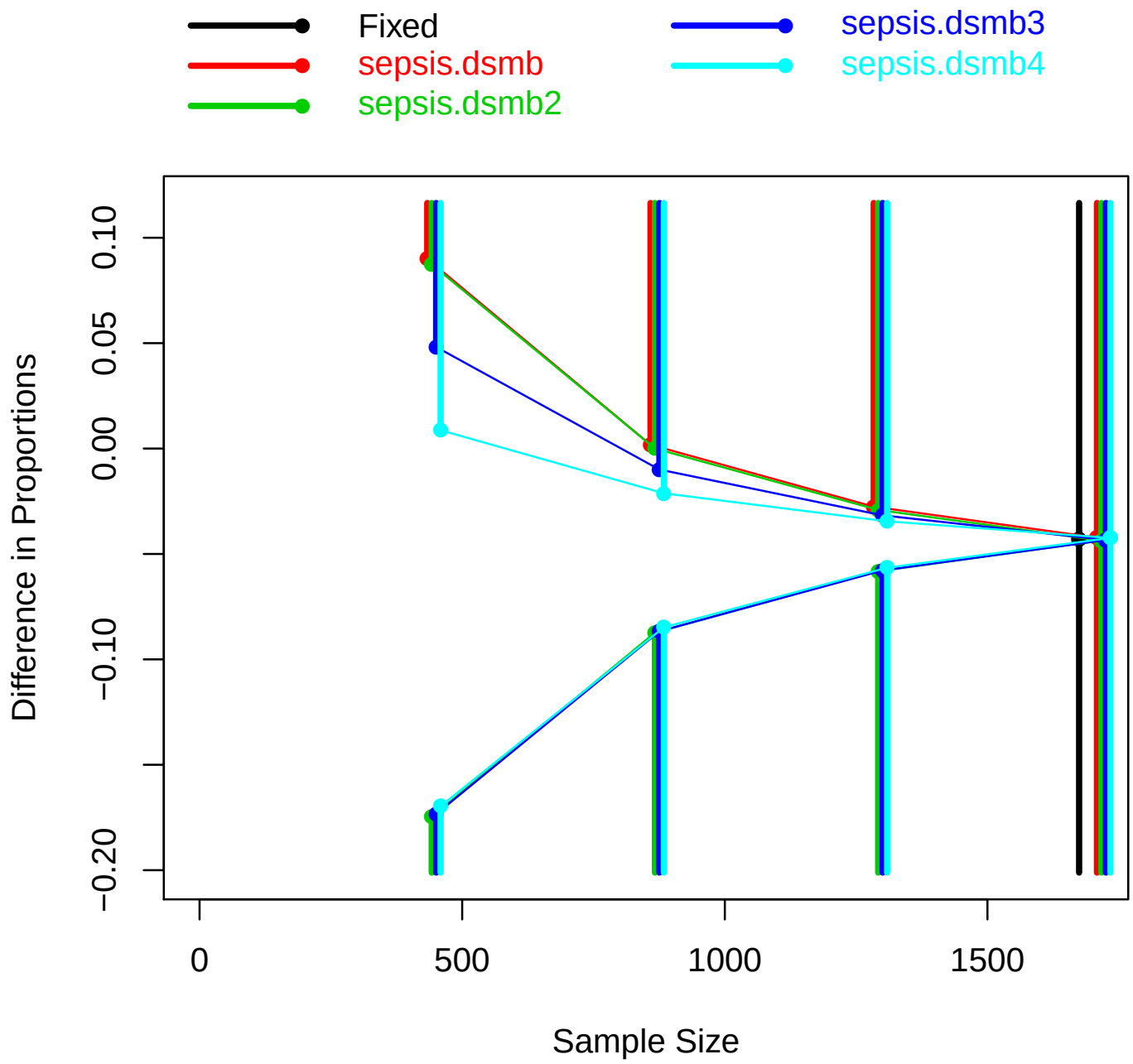
Example: Sepsis trial

## Addition of interim analyses

- DSMB sought a design with less early conservatism for futility
- Sponsor considered a Pocock futility bound and something between an O'Brien-Fleming and Pocock design

# Example: Sepsis Trial

- Stopping boundaries






Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Example: Sepsis Trial



## Choosing a boundary

- In order to choose between the considered designs, need to consider operating characteristics
  - Point estimates of treatment effect at boundary decisions
  - Confidence intervals resulting from decisions on the boundary
  - Statistical power as a function of treatment effect
  - Sample size distribution as a function of treatment effect

Elements of Trial Monitoring

Group Sequential Designs

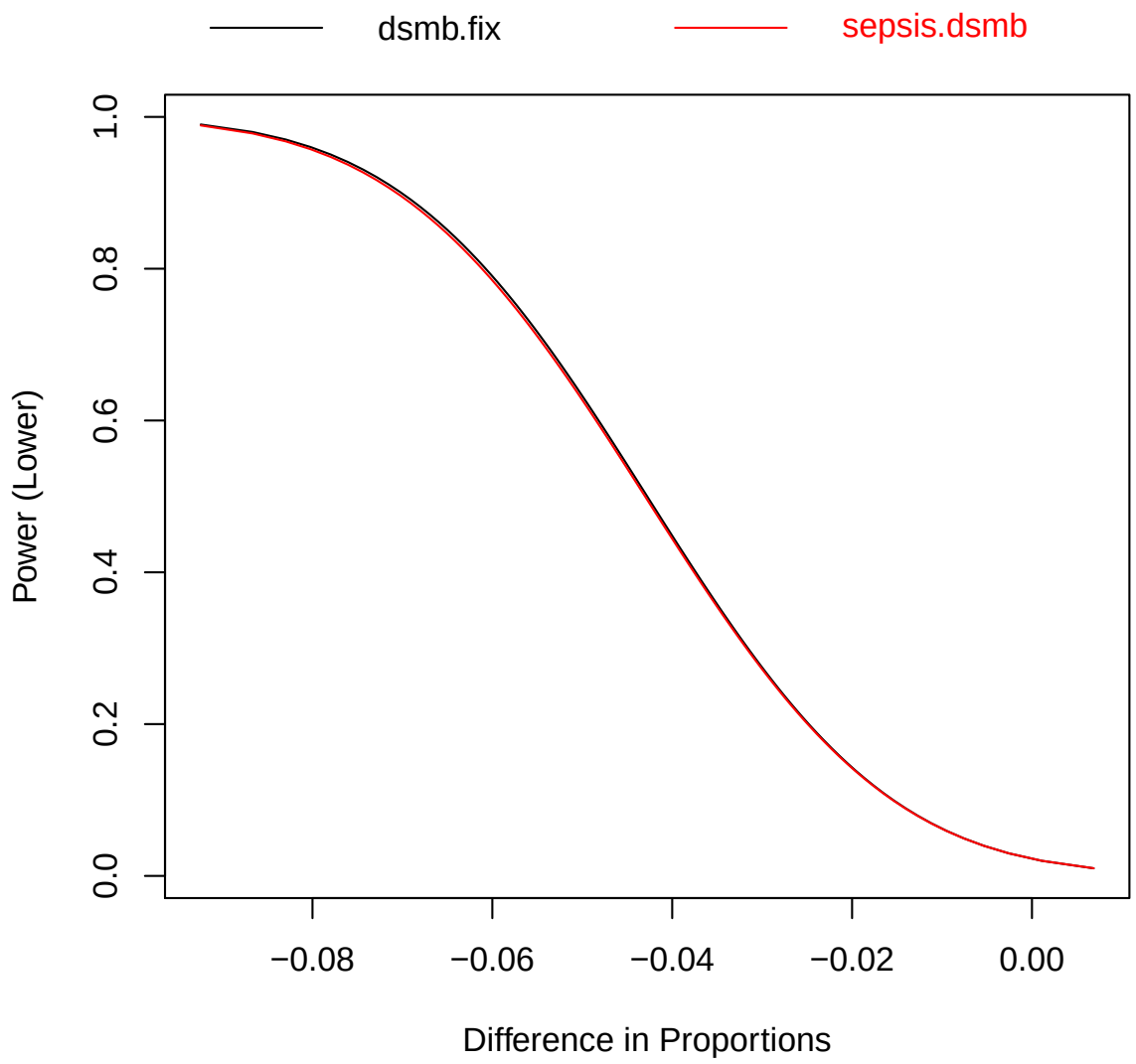
Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

# Example: Sepsis Trial

- Comparing power (adding futility-only stopping):

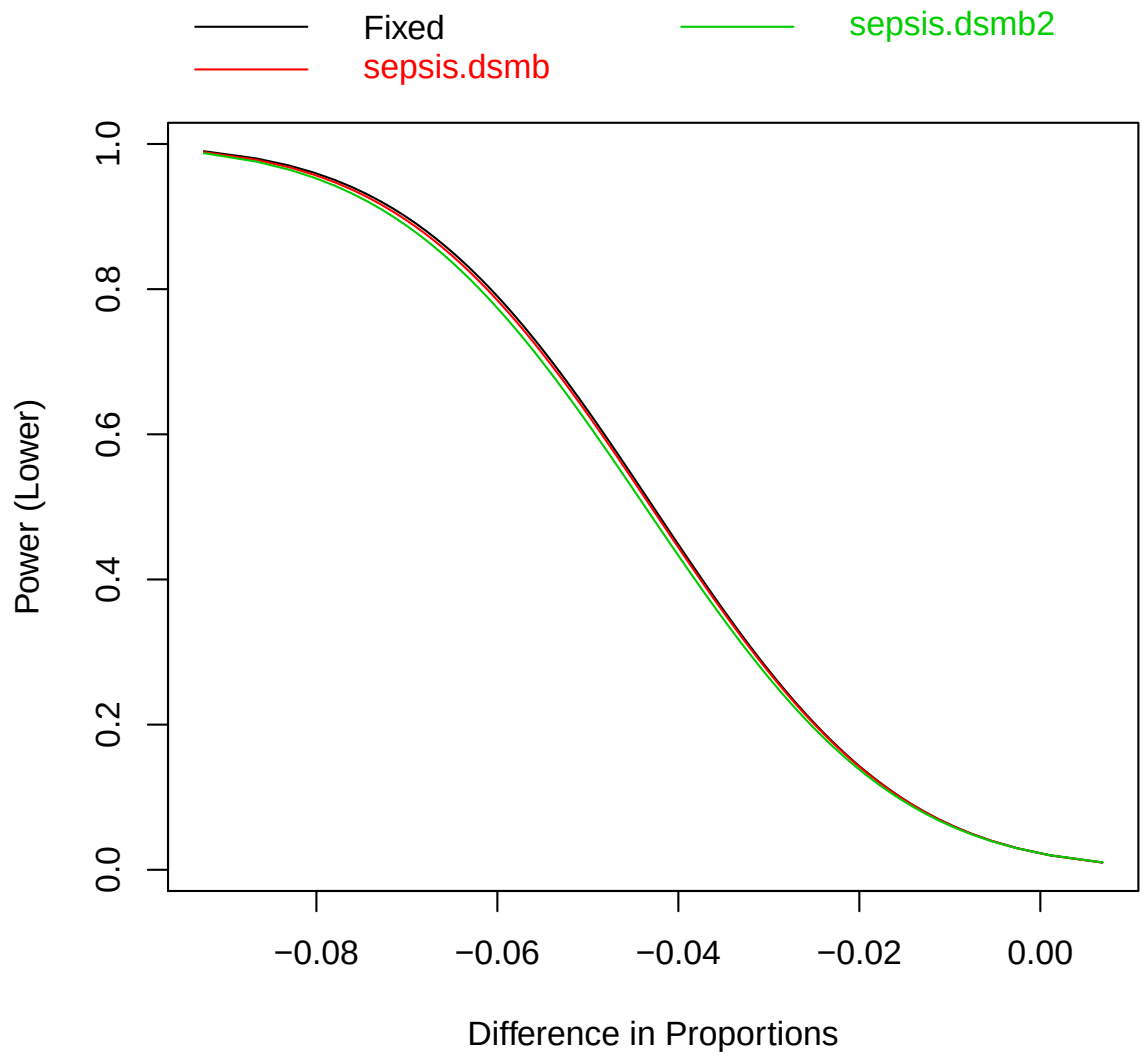





Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial

- Comparing power (adding futility and efficacy stopping):

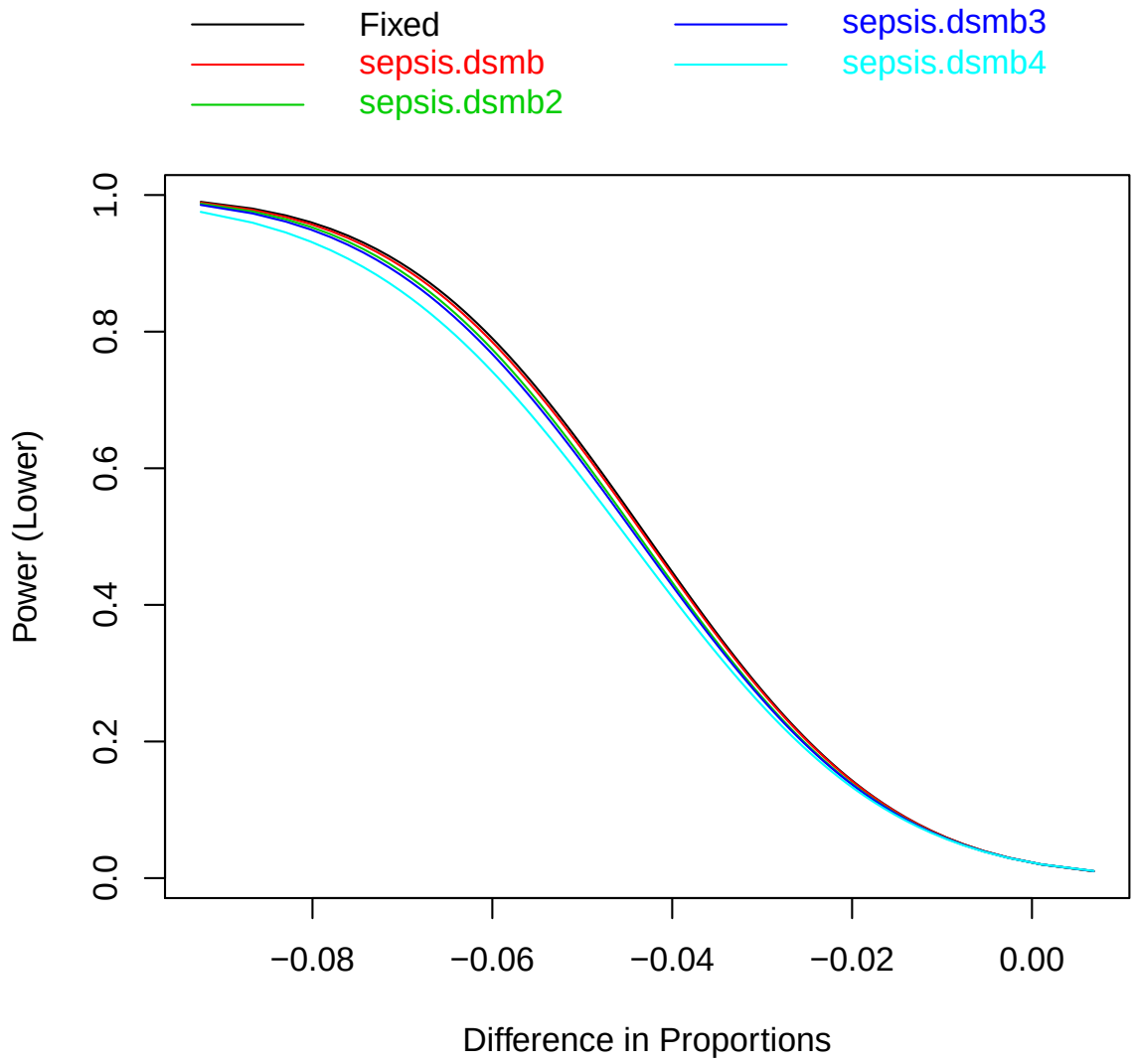





Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial

- Comparing power (effect of conservatism):

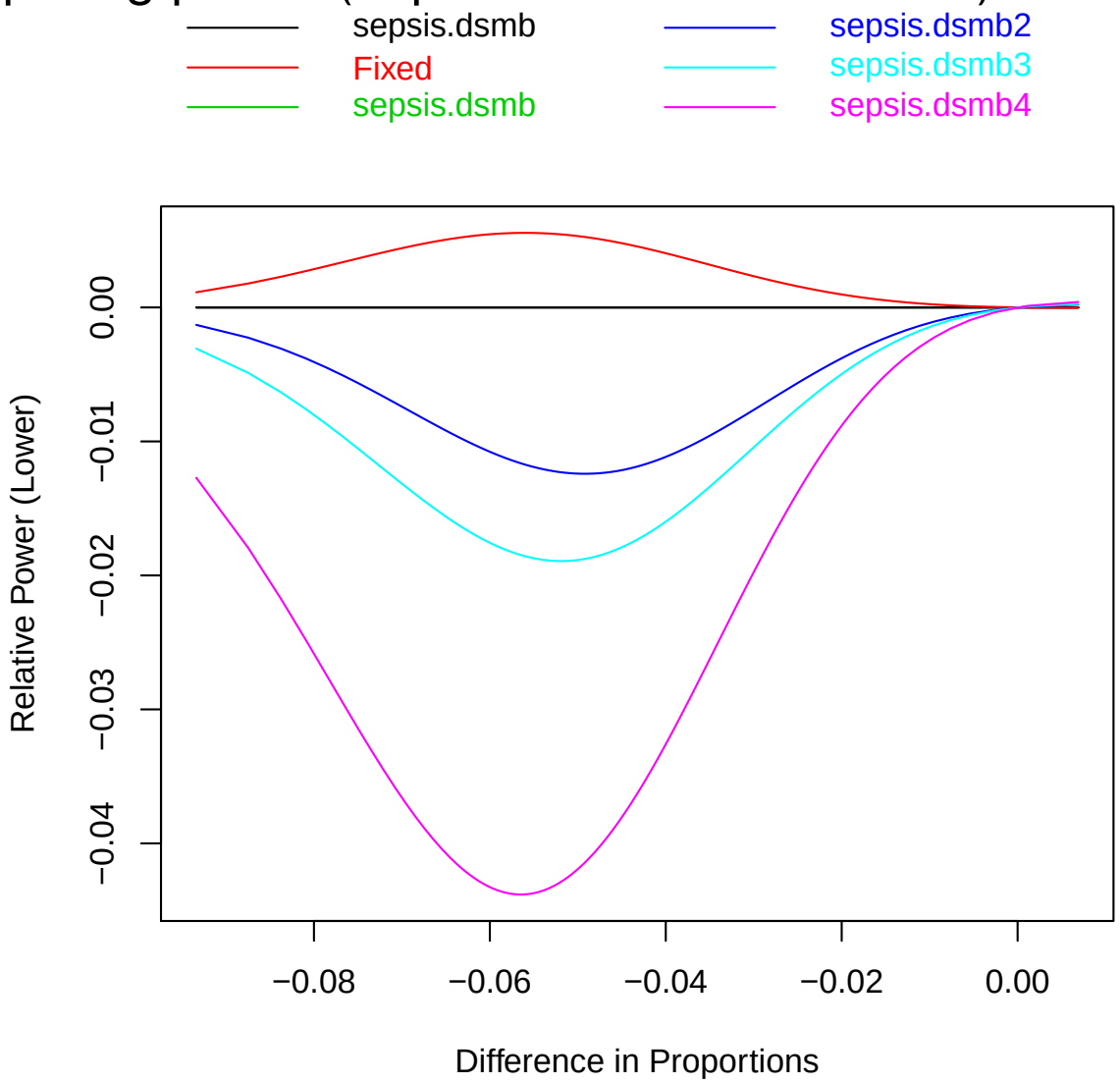





Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial

- Comparing power (sepsis.dsmb as reference):






Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

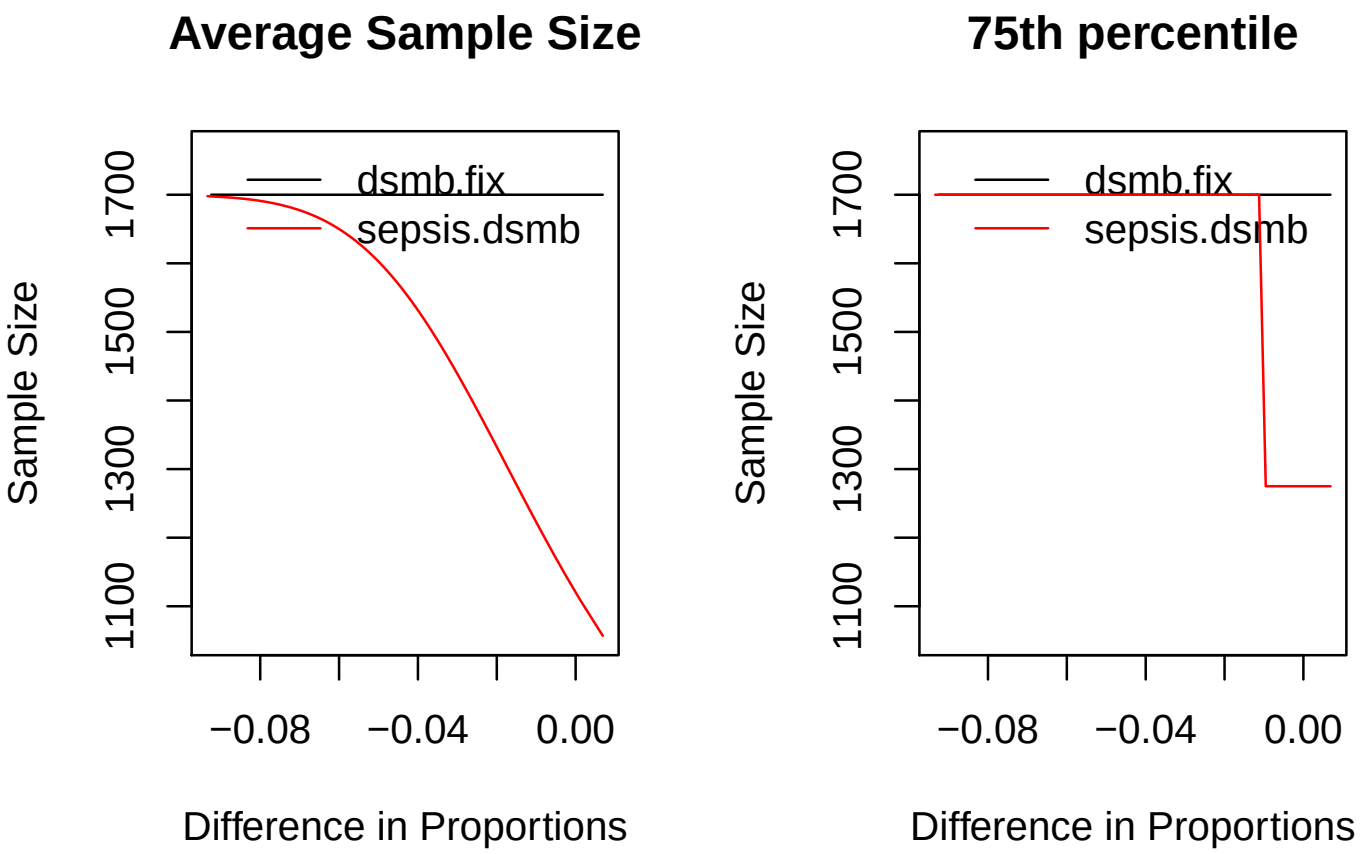
Types of group sequential designs

Example: Sepsis trial

# Example: Sepsis Trial



- Comparing expected sample size (ASN): adding futility-only stopping:

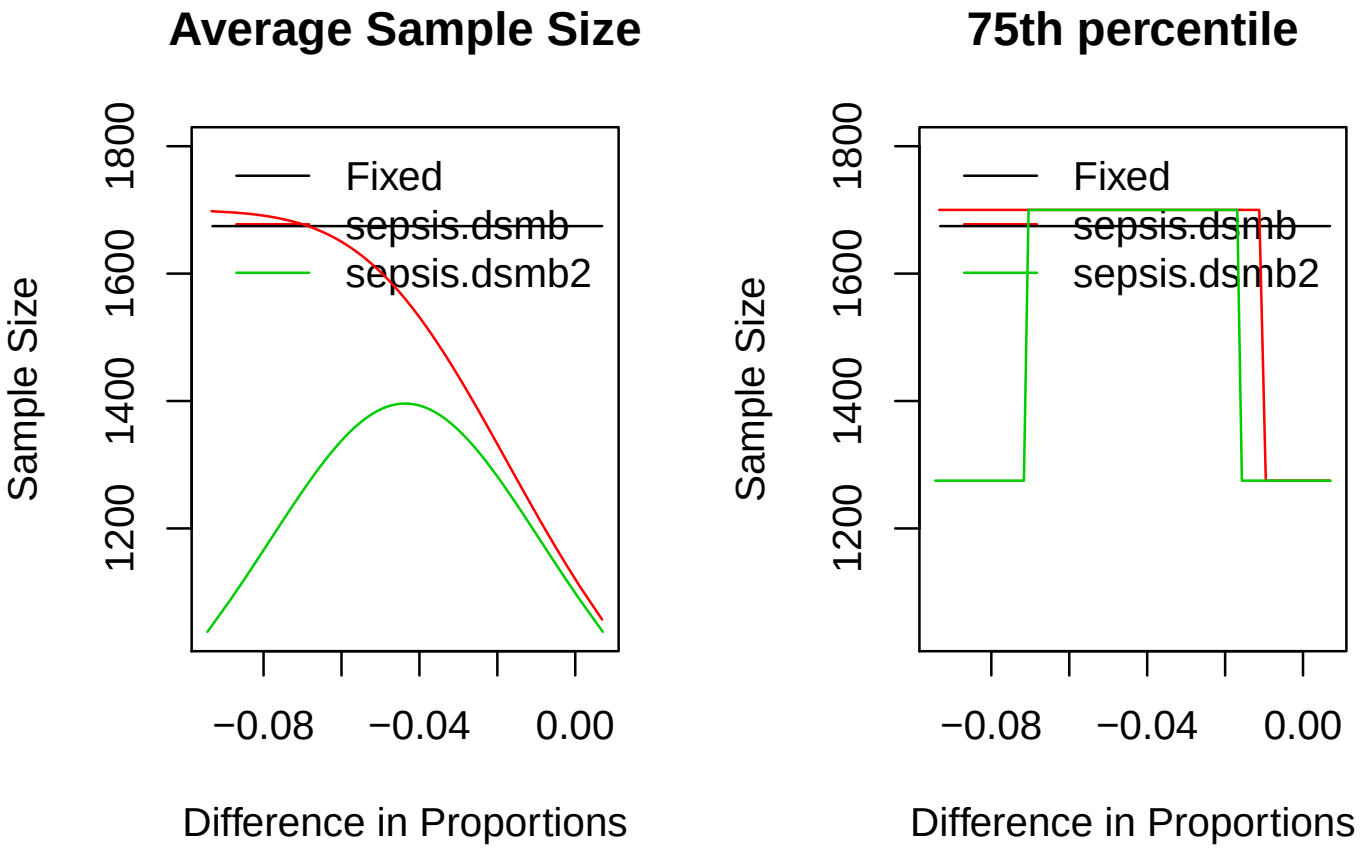



Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial



- Comparing expected sample size (ASN): futility and efficacy stopping:

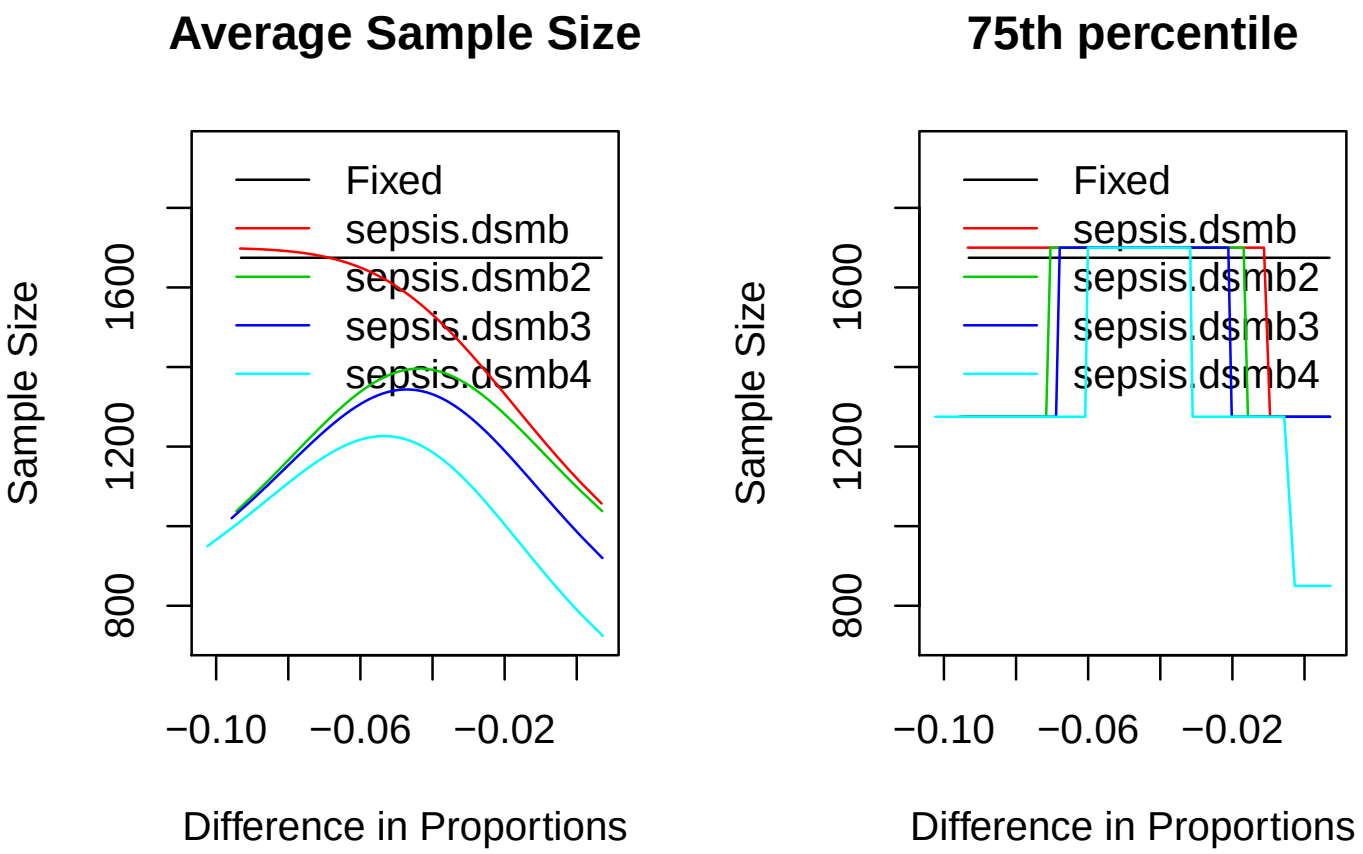



Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial



- Comparing expected sample size (ASN): early conservatism:




Elements of Trial Monitoring
Group Sequential Designs
Statistical framework for trial monitoring
Types of group sequential designs
Example: Sepsis trial

# Example: Sepsis Trial



Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial

## General behavior of interim analyses

- Decreasing early conservatism gave smaller ASN for unimportant benefits.
- Decreasing early conservatism also reduces power for efficacy.

# Example: Sepsis Trial



## General behavior of interim analyses

- 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.

Elements of Trial Monitoring

Group Sequential Designs

Statistical framework for trial monitoring

Types of group sequential designs

Example: Sepsis trial