Adaptive Sample Size Re-estimation: Design and Inference

Sarah Emerson and Greg Levin

Outline

Considerations in Designing Clinical Trials

- Goal: determine efficacy of a treatment (or difference between treatments)
- One- or two-sided hypothesis test based on a statistic of interest, chosen to be scientifically/clinically relevant

Considerations in Designing Clinical Trials

- Scientific:
  - Answer clinical question of interest with useful estimates and intervals
  - Evaluate mechanistic questions
- Ethical:
  - Quickly identify treatments that cause harm
  - Get effective treatments to patients quickly
  - Release patients from a less promising trial so that they might participate in other trials
- Financial:
  - Patient costs expensive; limit number of patients required
  - Long duration increases operating costs
  - Bringing a good drug to market sooner allows an earlier profit, advantage in competition

Example Study Setting

- Goal: Determine efficacy of new/experimental treatment (A) relative to standard of care/placebo control (B).
- Protocol:
  - Accrue n subjects
  - Randomize at 1:r ratio to treatment A or B (we will consider r = 1, so 1:1 randomization)
    - $n_A = \frac{n}{1+r} = \frac{n}{2}$ = Number of subjects receiving treatment A
    - $n_B = \frac{rn}{1+r} = \frac{n}{2}$ = Number of subjects receiving treatment B
  - Measure outcomes
    - X_{Ai} for subject i receiving experimental treatment A
    - X_{Bi} for subject i receiving control treatment B.
    - For now, we assume outcomes are immediately available for all subjects.

Example Study Setting

- Population parameters:
  - $\mu_A = E[X_{Ai}]$ (unknown)
  - $\mu_B = E[X_{Bi}]$ (unknown)
  - $\sigma^2 = \text{Var}[X_{Ai}] = \text{Var}[X_{Bi}]$ (common variance, assumed known for now, and taken to be $\sigma^2 = 1$)
- Parameter of interest: $\theta = \mu_A - \mu_B$
  - Null hypothesis $H_0 : \theta = \theta_0 = 0$ (no difference in mean treatment effect)
  - Alternative hypothesis $H_0 : \theta = \theta_A = 0.46$ (experimental treatment mean is larger, indicating superiority over control)

Example Study Setting

- Sample statistic notation:
  - $\bar{X}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} X_{Ai}$ (sample mean of treatment A group)
  - $\bar{X}_B = \frac{1}{n_B} \sum_{i=1}^{n_B} X_{Bi}$ (sample mean of treatment B group)
- Note:

$$\begin{aligned} \text{Var}[\bar{X}_A - \bar{X}_B] &= \sigma^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right) \\ &= 1 \left(\frac{2}{n} + \frac{2}{n} \right) \quad \text{if } n_A = n_B = \frac{n}{2} \text{ and } \sigma^2 = 1 \\ &= \frac{4}{n} \end{aligned}$$

Possible Designs

Fixed Sample

- Gather $n = 290$ subjects randomized at 1:1 ratio to treatments A and B ($n_A = n_B = 145$)
- Measure outcomes X_{Ai} or X_{Bi} for each subject.
- Compute two-sample z-statistic:

$$z(\theta_0) = \frac{\overline{X}_A - \overline{X}_B - \theta_0}{\sqrt{\sigma^2(\frac{1}{n_A} + \frac{1}{n_B})}} = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{4\sigma^2/n}} = \frac{\sqrt{n}}{2}\left(\overline{X}_A - \overline{X}_B\right)$$

- Reject H_0 if $z(\theta_0) > z_\alpha = \Phi^{-1}(1-\alpha) = 1.96$
- Equivalently, reject H_0 if $(\overline{X}_A - \overline{X}_B) > \frac{1.96}{\sqrt{n}/2} = 0.2298$

Significance Level $= 0.025$

Power $= 0.975$ at Design Alternative $\theta_A = 0.46$

Sequential Trials

- Ethical and financial issues in clinical trials may be improved by performing multiple interim analyses during the trial
- Maintain control of the significance level and the power at the design alternative by adjusting the decision criteria at each analysis
- Allowing early stopping of the trial at interim analyses typically reduces the expected trial duration and number of subjects required

Possible Designs

Fixed Sample

- Gather $n = 290$ subjects randomized at 1:1 ratio to treatments A and B ($n_A = n_B = 145$)
- Measure outcomes X_{Ai} or X_{Bi} for each subject.
- Compute two-sample z-statistic:

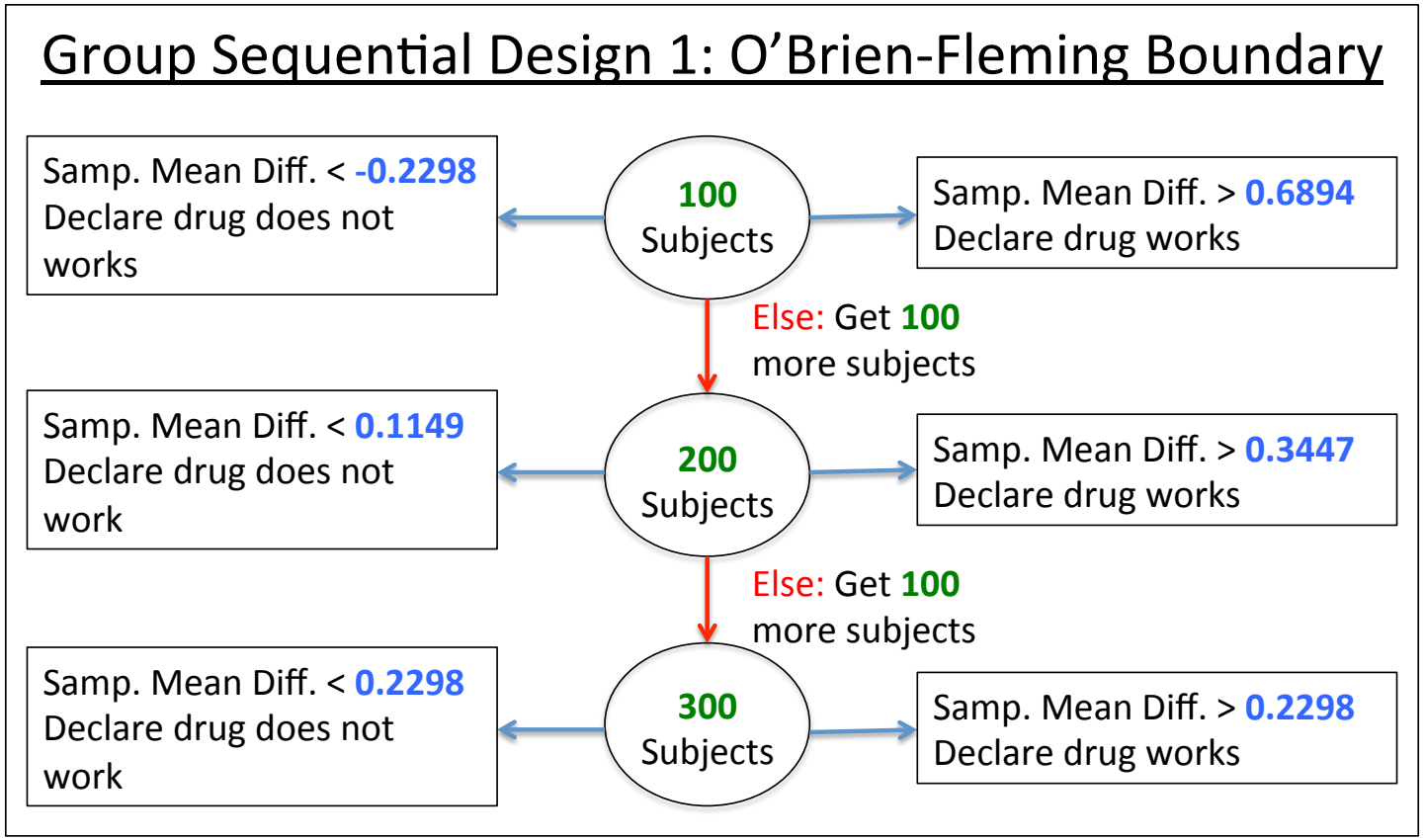
$$z(\theta_0) = \frac{\overline{X}_A - \overline{X}_B - \theta_0}{\sqrt{\sigma^2(\frac{1}{n_A} + \frac{1}{n_B})}} = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{4\sigma^2/n}} = \frac{\sqrt{n}}{2}\left(\overline{X}_A - \overline{X}_B\right)$$

- Reject H_0 if $z(\theta_0) > z_\alpha = \Phi^{-1}(1-\alpha) = 1.96$
- Equivalently, reject H_0 if $(\overline{X}_A - \overline{X}_B) > \frac{1.96}{\sqrt{n}/2} = 0.2298$

Significance Level $= 0.025$

Power $= 0.975$ at Design Alternative $\theta_A = 0.46$

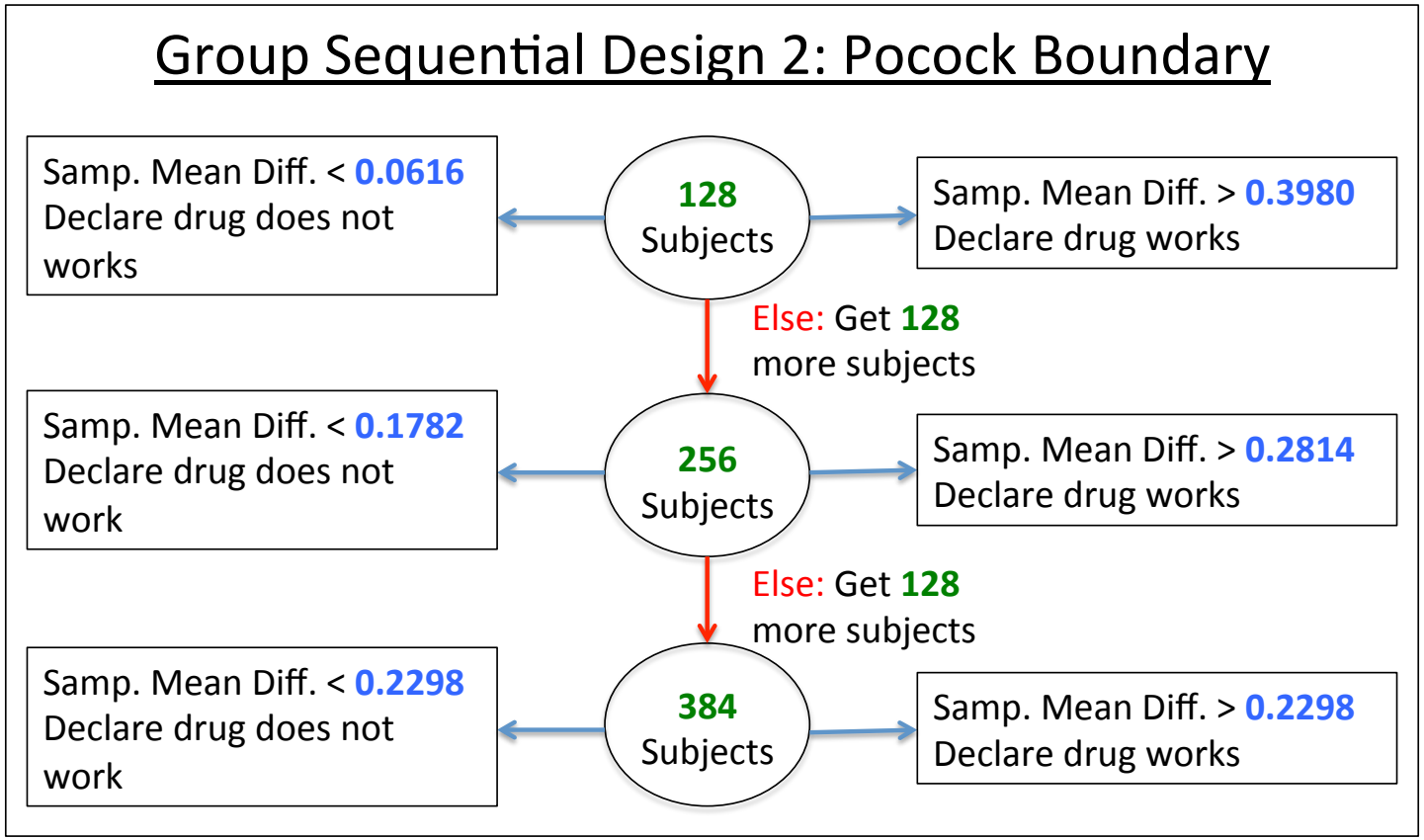
Possible Designs



Significance Level $= 0.025$

Power $= 0.975$ at Design Alternative $\theta_A = 0.46$

Possible Designs



Significance Level $= 0.025$

Power $= 0.975$ at Design Alternative $\theta_A = 0.46$

Group Sequential Trials: Definition

A group sequential design is defined by a **Stopping Rule** consisting of:

1. **Analysis Times:** A set of J analysis times $n_1, n_2, \ldots, n_J$ defined in terms of the amount of *statistical information* accumulated
   - n_j = total number of subjects or number of events (across all arms) observed up to the jth analysis.
   - n_{Aj}, n_{Bj} = number of subjects on arm A or B, respectively, observed up to the jth analysis. $n_{Aj} + n_{Bj} = n_j$.

Group Sequential Trials: Definition

Incremental Analysis Times/Sample Sizes:

- $n_j^* = n_j - n_{j-1}$ = *incremental* number of subjects/events (across all arms) added between $(j-1)$st and jth analyses
- $n_{Aj}^* = n_{Aj} - n_{A(j-1)}$, $n_{Bj}^* = n_{Bj} - n_{B(j-1)}$ = *incremental* number of subjects/events added on each arm

Group Sequential Trials: Definition

2. **Test Statistic:** A test statistic T_j calculated from the data accumulated so far at each analysis time $j = 1, 2, \ldots J$. Examples:
   - Partial Sum/Partial Sum Difference:

$$S_j = \sum_{i=1}^{n_{Aj}} X_{Ai} - \sum_{i=1}^{n_{Bj}} X_{Bi}$$

   - MLE:

$$\begin{aligned} \hat{\theta}_j &= \frac{1}{n_{Aj}} \sum_{i=1}^{n_{Aj}} X_{Ai} - \frac{1}{n_{Bj}} \sum_{i=1}^{n_{Bj}} X_{Bi} \\ &= \bar{X}_{Aj} - \bar{X}_{Bj} \end{aligned}$$

Group Sequential Trials: Definition

- z-statistic:

$$Z_j = \frac{\hat{\theta}_j - \theta_0}{\sqrt{\frac{\sigma^2}{n_{Aj}} + \frac{\sigma^2}{n_{Bj}}}} = \frac{\hat{\theta}_j - \theta_0}{\sigma\sqrt{\frac{1}{n_{Aj}} + \frac{1}{n_{Bj}}}}$$

- Fixed-sample p-value:

$$P_j = 1 - \Phi(Z_j)$$

Group Sequential Trials: Definition

Incremental Test Statistics:

- Incremental Partial Sum/Partial Sum Difference:

$$S_j^* = \sum_{i=n_{A(j-1)}+1}^{n_{Aj}} X_{Ai} - \sum_{i=n_{B(j-1)}+1}^{n_{Bj}} X_{Bi}$$

- Incremental MLE:

$$\begin{aligned}\hat{\theta}_j^* &= \frac{1}{n_{Aj}^*} \sum_{n_{A(j-1)}+1}^{n_{Aj}} X_{Ai} - \frac{1}{n_{Bj}^*} \sum_{i=n_{B(j-1)}+1}^{n_{Bj}} X_{Bi} \\ &= \bar{X}_{Aj}^* - \bar{X}_{Bj}^*\end{aligned}$$

Group Sequential Trials: Definition

- Incremental z-statistic:

$$Z_j^* = \frac{\hat{\theta}_j^* - \theta_0}{\sqrt{\frac{\sigma^2}{n_{Aj}^*} + \frac{\sigma^2}{n_{Bj}^*}}} = \frac{\hat{\theta}_j^* - \theta_0}{\sigma\sqrt{\frac{1}{n_{Aj}^*} + \frac{1}{n_{Bj}^*}}}$$

- Incremental Fixed-sample p-value:

$$P_j^* = 1 - \Phi(Z_j^*)$$

Group Sequential Trials: Definition

3. **Stopping Boundary:** A set of boundary values $a_j \leq b_j \leq c_j \leq d_j$ for each analysis time $j = 1, 2, \ldots, J$
   - Decision rule:

$T_j \geq d_j$	Stop trial at j^{th} analysis and accept upper hypothesis
$c_j < T_j < d_j$	Continue trial
$b_j \leq T_j \leq c_j$	Stop trial at j^{th} analysis and accept null hypothesis (two-sided test)
$a_j < T_j < b_j$	Continue trial
$T_j \leq a_j$	Stop trial at j^{th} analysis and accept lower hypothesis

Group Sequential Trials: Definition

- The regions $\mathcal{C}_j = (a_j, b_j) \cup (c_j, d_j)$ are called the *continuation regions* at analysis j.
  - ⋆ If the test statistic belongs to this interval or set of intervals, the trial is continued beyond analysis j.
- The complement of the continuation regions $\mathcal{S}_j = \mathcal{C}_j'$ are called the *stopping regions* at analysis j.
  - ⋆ If the test statistic belongs to this interval or set of intervals, the trial is stopped at analysis j.

Group Sequential Trials: Definition

- (a_j, b_j, c_j, d_j) for $j = 1, \ldots, J$ must be chosen to obtain desired significance level α:

$$P_{\theta_0}(\text{Reject } H_0 \text{ at any } j = 1, \ldots, J) = \alpha$$

- $a_J = b_J$ and $c_J = d_J$ to guarantee that a decision is made by the final analysis
- For a one-sided design, $b_j = c_j$ for all j.

Stopping Boundary Specification

- Great flexibility in choice of boundary
  - Error spending designs
  - Unified family of group sequential designs (Kittelson and Emerson 1999), includes
    - ★ O'Brien-Fleming
    - ★ Pocock
    - ★ Wang and Tsiatis
    - ★ ...and others

    as special cases.

Group Sequential Trials: Sufficient Statistic

When a group sequential trial is stopped, the sufficient statistic is (M, S_M) (or $(M, \hat{\theta}_M)$) where

- M is analysis time at which trial stops, $M \in \{1, 2, \ldots, J\}$; $M = j$ if the trial stops at the jth analysis.
- S_M is the observed partial sum/partial sum difference when the trial stops.
- $\hat{\theta}_M$ is the observed MLE when the trial stops.

This statistic (M, S_M) or $(M, \hat{\theta}_M)$ may be abbreviated as (M, S) or $(M, \hat{\theta})$.

Possible Designs

Fixed Sample

- Gather $n = 290$ subjects randomized at 1:1 ratio to treatments A and B ($n_A = n_B = 145$)
- Measure outcomes X_{Ai} or X_{Bi} for each subject.
- Compute two-sample z-statistic:

$$z(\theta_0) = \frac{\overline{X}_A - \overline{X}_B - \theta_0}{\sqrt{\sigma^2(\frac{1}{n_A} + \frac{1}{n_B})}} = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{4\sigma^2/n}} = \frac{\sqrt{n}}{2}\left(\overline{X}_A - \overline{X}_B\right)$$

- Reject H_0 if $z(\theta_0) > z_\alpha = \Phi^{-1}(1-\alpha) = 1.96$
- Equivalently, reject H_0 if $(\overline{X}_A - \overline{X}_B) > \frac{1.96}{\sqrt{n}/2} = 0.2298$

Significance Level $= 0.025$

Power $= 0.975$ at Design Alternative $\theta_A = 0.46$

Possible Designs

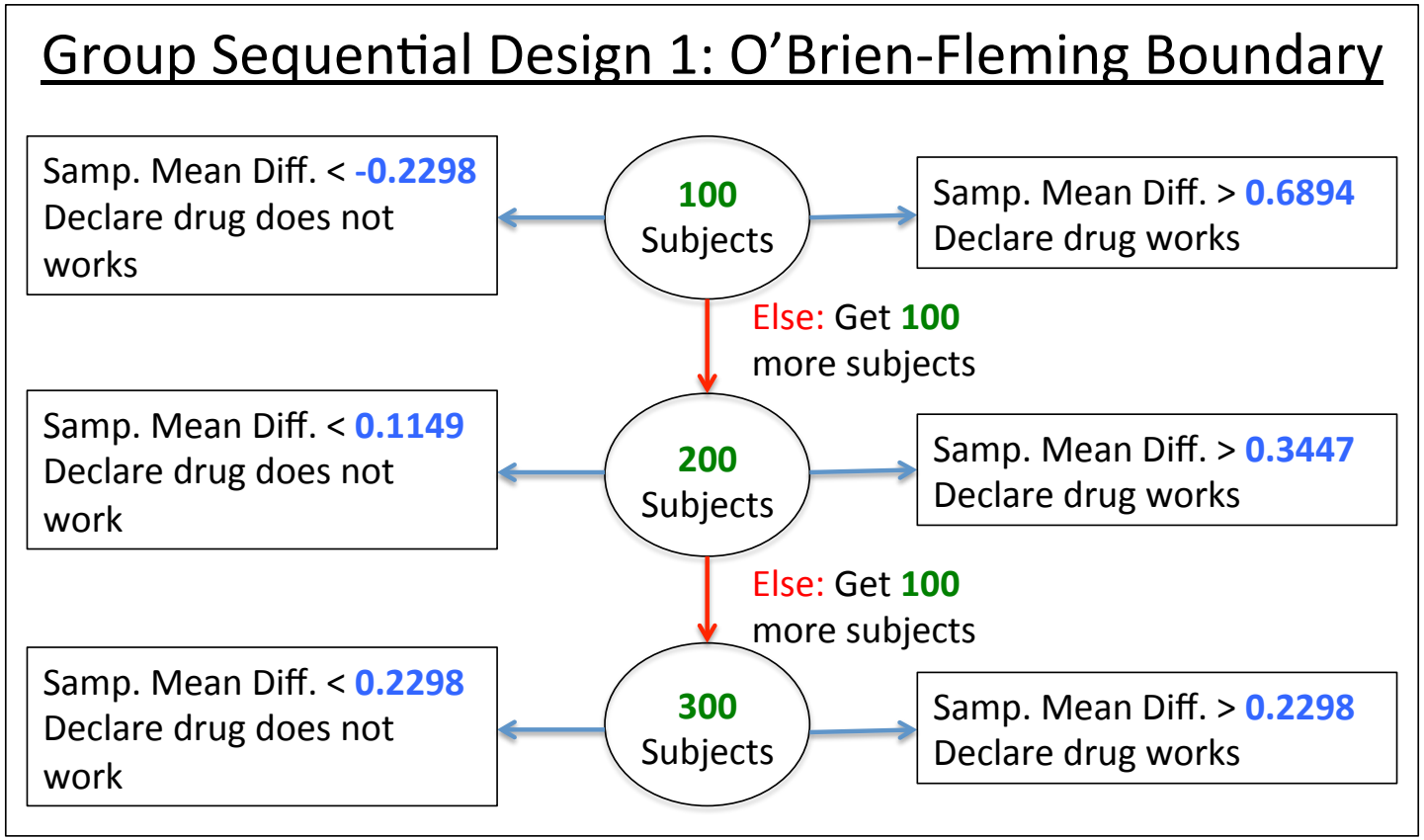
Fixed Sample Design

Number of Analyses: $J = 1$

Test Statistic: $T_j = \hat{\theta}_j =$ Sample Mean

j	n_j	a_j	b_j	c_j	d_j
1	290	0.2298	0.2298	0.2298	0.2298

Possible Designs



Significance Level $= 0.025$

Power $= 0.975$ at Design Alternative $\theta_A = 0.46$

Possible Designs

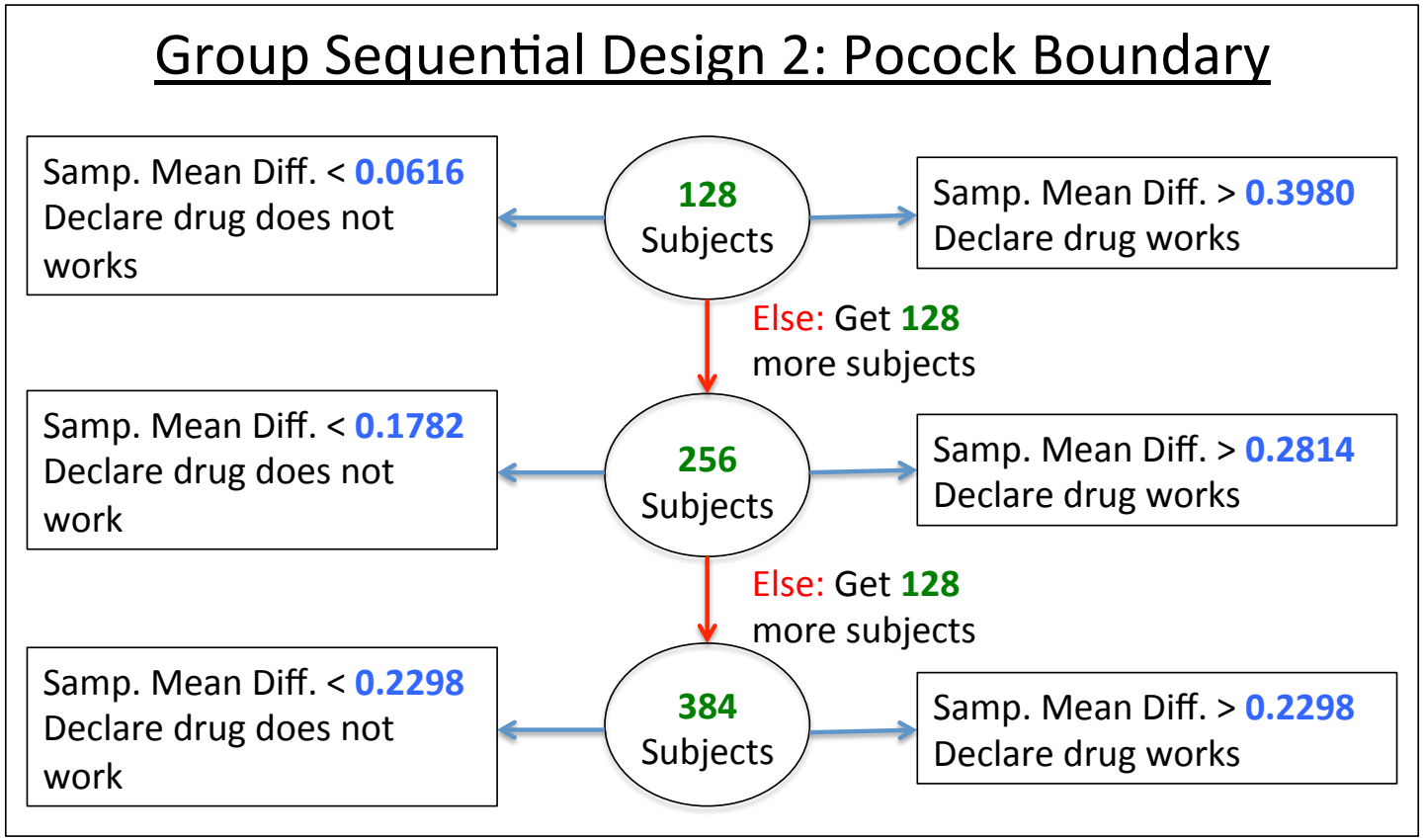
O'Brien-Fleming Group Sequential Design

Number of Analyses: $J = 3$

Test Statistic: $T_j = \hat{\theta}_j =$ Sample Mean

j	n_j	a_j	b_j	c_j	d_j
1	100	-0.2298	0.2298	0.2298	0.6894
2	200	0.1149	0.2298	0.2298	0.3447
3	300	0.2298	0.2298	0.2298	0.2298

Possible Designs



Significance Level $= 0.025$

Power $= 0.975$ at Design Alternative $\theta_A = 0.46$

Possible Designs

Pocock Group Sequential Design

Number of Analyses: $J = 3$

Test Statistic: $T_j = \hat{\theta}_j =$ Sample Mean

j	n_j	a_j	b_j	c_j	d_j
1	128	0.0616	0.2298	0.2298	0.3980
2	256	0.1782	0.2298	0.2298	0.2814
3	384	0.2298	0.2298	0.2298	0.2298

Boundary Scales

- The same stopping boundary can be represented on many different test-statistic scales, including partial sum difference, sample mean, z-statistic, p-value, etc.

Clinical Trial Optimality Criteria: Comparing Sequential Designs

- In comparing different types of sequential designs, we must select criteria that we wish to constrain or optimize. Possibilities include:
  - ⋆ Maximal possible sample size n_J
  - For range of parameter values θ:
    - ⋆ Power $P_\theta(\text{Reject } H_0)$
    - ⋆ Average sample size (ASN)
    - ⋆ Probability of using more than q subjects
    - ⋆ Median sample size (or any other quantile)
- All but the first of these criteria require knowledge of the sampling distribution of the sufficient statistic (M, S_M) given a parameter value θ.

Clinical Trial Optimality Criteria

Maximal Sample Size

- The maximal sample size for a sequential design is just n_J: the largest sample size at which an analysis may possibly be performed.

Fixed	O'Brien-Fleming	Pocock
291	300	384

Sequential Trial Sampling Density

- For the remaining optimality criteria, the sampling density for the statistic (M, S) is required.
- We will use S_j as the test statistic T_j for ease of discussion; recall that the stopping boundary can be equivalently expressed on many scales.
- For simplicity we will assume $n^*_{Aj} = n^*_{Bj} = \frac{1}{2} n^*_j$ for all j. Analogous formulae for different randomization ratios may be extended from this case.
- We use the fact that

$$S_j = S^*_1 + S^*_2 + \ldots + S^*_j \quad \text{and} \quad S^*_j \sim N\left(\frac{n^*_j}{2}\theta, n^*_j\sigma^2\right)$$

Sequential Trial Sampling Density

To obtain the sampling density at an observed value $(M = j, S = s)$, we have to consider the possible paths that could reach this point.

- If $j = 1$:
  - The test statistic S_1 must have been in the stopping region $\mathcal{S}_j \Leftrightarrow S_1 \notin \mathcal{C}_1$.
  - The value of the test statistic S_1 is $S_1 = s$.
- If $j > 1$:
  - At all analyses $\ell = 1, 2, \ldots, j-1$, the test statistic S_ℓ must have been in the continuation region $\mathcal{C}_\ell$
  - At analysis j the test statistic S_j must have been in the stopping region $\mathcal{S}_j \Leftrightarrow S_j \notin \mathcal{C}_j$.
  - The value of the test statistic S_j is $S_j = s$.

Sequential Trial Sampling Density

Following Armitage et al. (1969), the density of $(M = j, S = s)$ is

$$p_{M,S}(j, s;\, \theta) = \begin{cases} f_{M,S}(j, s;\, \theta) & \text{if } s \in \mathcal{S}_j \\ 0 & \text{otherwise} \end{cases}$$

where the (sub)density $f_{M,S}(j,\, s;\, \theta)$ is recursively defined as

$$f_{M,S}(1, s;\, \theta) \;=\; \frac{1}{\sigma\sqrt{n_1}}\, \phi\left(\frac{s - n_1\, \theta/2}{\sigma\sqrt{n_1}}\right)$$

$$f_{M,S}(j, s;\, \theta) \;=\; \int_{\mathcal{C}_{j-1}} \frac{1}{\sigma\sqrt{n_j^*}}\, \phi\left(\frac{s - u - n_j^*\, \theta/2}{\sigma\sqrt{n_j^*}}\right)\, f_{M,S}(j-1,\, u,\, ;\theta)\, du$$

for $j = 2, \ldots, J$

Sequential Trial Stopping Probabilities

- Using the density $p_{M,S}(j, s; \theta)$, analysis time stopping probabilities may be obtained as

$$\text{Total:} \quad P_\theta(M = j \text{ Total}) = \int_{\mathcal{S}_j} p_{M,S}(j, u)\, du$$

$$\text{Upper:} \quad P_\theta(M = j, \text{ Upper}) = \int_{u \geq d_j} p_{M,S}(j, u)\, du$$

$$\text{Null:} \quad P_\theta(M = j, \text{ Null}) = \int_{b_j \leq u \leq c_j} p_{M,S}(j, u)\, du$$

$$\text{Lower:} \quad P_\theta(M = j, \text{ Lower}) = \int_{u \leq a_j} p_{M,S}(j, u)\, du$$

Clinical Trial Optimality Criteria

Power

- Here we consider the upper power for a one-sided test of a greater alternative.
- Using the total analysis time stopping probabilities $P_\theta(M = j, \text{ Total})$, the power may be obtained as

$$\text{Power}(\theta) = 1 - \beta(\theta) = \sum_{j=1}^{J} P_\theta(M = j, \text{ Upper})$$

Clinical Trial Optimality Criteria

Power

- For example, under the design alternative $\theta = 0.4596$, the O'Brien-Fleming design has the following upper stopping probabilities:

j	N_j	$P_{\theta=0}(M = j,\ \text{Upper})$
1	100	0.1253
2	200	0.6670
3	300	0.1827

The power when $\theta = 0.4596$ is therefore

$$\begin{aligned}\text{Power}(\theta = 0.4596) &= 0.1253 + 0.6670 + 0.1827 \\ &= 0.975\end{aligned}$$

Clinical Trial Optimality Criteria

Average Sample Size (ASN)

- Using the total analysis time stopping probabilities $P_\theta(M = j,\ \text{Total})$, the average sample size may be obtained as

$$\text{ASN}(\theta) = \sum_{j=1}^{J} P_\theta(M = j,\ \text{Total}) n_j$$

Clinical Trial Optimality Criteria

Average Sample Size (ASN)

- For example, under the null hypothesis $\theta = 0$ the O'Brien-Fleming design has the following total stopping probabilities:

j	n_j	$P_{\theta=0}(M = j, \text{ Total})$
1	100	0.1256
2	200	0.6742
3	300	0.2002

The average sample size (ASN) when $\theta = 0$ is therefore

$$\begin{aligned} \text{ASN}(\theta = 0) &= 100(0.1256) + 200(0.6742) + 300(0.2002) \\ &= 207.4663 \end{aligned}$$

Clinical Trial Optimality Criteria

Probability of More Than q Subjects

- Using the total analysis time stopping probabilities $P_\theta(M = j, \text{ Total})$, the probability of using more than q subjects may be obtained as

$$P_\theta(n_M \geq q) = \sum_{j:n_j>q} P_\theta(M = j, \text{ Total})$$

Clinical Trial Optimality Criteria

Percentile of Sample Size Distribution

- Using the total analysis time stopping probabilities $P_\theta(M = j,\ \text{Total})$, the pth percentile of the sample size distribution may be obtained as

$$q_p(\theta) = \min\{n_j : \sum_{\ell=1}^{j} P_\theta(M = \ell,\ \text{Total}) \geq p\}$$

Group Sequential Design Inference Goals

- Trial designed to decide between $H_0 : \theta = \theta_0$ and $H_A : \theta = \theta_A > \theta_0$ with significance level α and power $1 - \beta$ when $\theta = \theta_A$.
- Decision at end of trial:
  - Reject $H_0 : \theta = \theta_0$
  - Fail to reject $H_0 : \theta = \theta_0$ ('Accept' H_0)
- Almost always want more information than just this binary decision:
  - How large is the effect?
  - How confident are we in the estimated effect?

Group Sequential Design Inference Goals

- Hypothesis Testing: Decide between $H_0 : \theta = \theta_0$ and $H_A : \theta > \theta_0$.
  - Design constructed to test particular value of θ_0 at desired level α, with desired power $1 - \beta$ to detect a particular $\theta_A > \theta_0$.
  - We may want to perform a test of a different null hypothesis at the conclusion of the test.
- Point Estimates: Estimates of θ satisfying various optimality criteria. (How large is the effect?)
- Confidence Intervals: Interval estimates $\mathcal{C}_{1-\alpha}$ of θ satisfying $P_\theta(\theta \in \mathcal{C}_{1-\alpha}) = 1 - \alpha$ (How confident are we in the estimated effect?)

Obtaining p-values for tests of general null hypotheses

- First consider fixed sample inference: $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ where μ is unknown but σ^2 is known.
- $H_0 : \mu = 0$ vs. $H_A : \mu = 1$
- Recall the interpretation of a p-value for a fixed sample test of $H_0 : \theta = \theta_0'$ when the observed statistic is $X = x$:

$$p_{\theta_0} = \text{Probability of a more 'extreme' result than } X = x \text{ when } \theta = \theta_0'$$

Obtaining p-values for tests of general null hypotheses

- Need to decide which of the possible sample results (outcomes) at the end of the trial are more 'extreme': More convincing for the alternative/less convincing for the null.
  - Larger values of $\bar{X}_n$ are more convincing for H_A and less convincing for H_0.
  - e.g., $\bar{X}_n = 0.7$ is stronger evidence for the alternative/against the null than $\bar{X}_n = 0.3$.

Obtaining p-values for tests of general null hypotheses

- This concept of a p-value may be used in the group sequential design setting to obtain tests of null hypotheses other than the design null hypothesis θ_0.
- Ordering of the sample space (outcome space): Define an ordering or partial ordering of all possible outcomes (M, S) to specify which results will be considered more extreme under the null/stronger evidence for the alternative.
- Unlike in fixed sample inference (at least in normal setting), no obvious ordering exists since sufficient statistic is bivariate (outcome space is 2-dimensional)

Group Sequential Design Inference Approaches

- Suppose, as in earlier example, testing $H_0 : \theta = \theta_0 = 0$ vs. $H_A : \theta = \theta_A = 0.4596$.
- Recall O'Brien-Fleming design with $\alpha = 0.025$, power $1 - \beta = 0.975$:

Number of Analyses: $J = 3$

Test Statistic: $T_j = \hat{\theta}_j =$ Sample Mean

j	n_j	a_j	b_j	c_j	d_j
1	100	-0.2298	0.2298	0.2298	0.6894
2	200	0.1149	0.2298	0.2298	0.3447
3	300	0.2298	0.2298	0.2298	0.2298

Group Sequential Design Inference Approaches

- Consider two possible outcomes:
  - Outcome 1: $(M_1 = 1, \hat{\theta}_1 = 0.7)$
  - Outcome 2: $(M_2 = 3, \hat{\theta}_2 = 0.8)$
- Which of these outcomes would you consider stronger evidence for the alternative, $\theta = 0.45$/weaker evidence for the null $\theta = 0$?

Sample Mean Ordering

Sample Mean Ordering:

- Outcomes are ordered according to the value of the MLE $\hat{\theta}_M = \hat{\theta}$.
- Consider two outcomes
  - Outcome 1: $(M = j_1, \hat{\theta} = t_1)$
  - Outcome 2: $(M = j_2, \hat{\theta} = t_2)$

Outcome 1 would be considered more extreme under the Sample Mean ordering as follows:

$$(j_1, t_1) \succ_{SM} (j_2, t_2) \text{ if } t_1 > t_2$$

Analysis Time Ordering

Analysis Time Ordering:

- Outcomes are ordered according to
  1. Stopping time M
  2. MLE $\hat{\theta}$
- Consider two outcomes:
  - Outcome 1: $(M = j_1, \hat{\theta} = t_1)$
  - Outcome 2: $(M = j_2, \hat{\theta} = t_2)$

Outcome 1 would be considered more extreme under the Analysis Time ordering as follows:

$$(j_1, t_1) \succ_{AT} (j_2, t_2) \text{ if } \begin{cases} j_1 < j_2 & \text{and } t_1 \geq d_{j_1} \\ j_1 > j_2 & \text{and } t_2 \leq a_{j_2} \\ j_1 = j_2 & \text{and } t_1 > t_2 \end{cases}$$

Likelihood Ratio Ordering

(Signed) Likelihood Ratio Ordering:

- Outcomes are ordered according to signed likelihood ratio test statistic for hypothesized θ_0'
- Consider two outcomes:
  - Outcome 1: $(M = j_1, \hat{\theta} = t_1)$
  - Outcome 2: $(M = j_2, \hat{\theta} = t_2)$

Outcome 1 would be considered more extreme under the Likelihood Ratio ordering as follows:

$$(j_1, t_1) \succ_{AT} (j_2, t_2) \text{ if}$$

$$\text{sign}(t_1 - \theta_0') \frac{p_{M,T}(j_1, t_1; \theta = t_1)}{p_{M,T}(j_1, t_1; \theta = \theta_0')} > \text{sign}(t_2 - \theta_0') \frac{p_{M,T}(j_2, t_2; \theta = t_2)}{p_{M,T}(j_2, t_2; \theta = \theta_0')},$$

$$\text{i.e., if} \quad \sqrt{n_{j_2}}(t_2 - \theta_0') > \sqrt{n_{j_1}}(t_1 - \theta_0')$$

Outcome Space Orderings

- For both the Sample Mean ordering and the Analysis Time ordering, the ordering does not depend upon the null hypothesis being tested.
- In contrast, note that the Likelihood Ratio ordering depends on the value of θ_0' being tested, and therefore may order the outcome space differently for different θ_0' values.

Outcome Space Orderings

- It can be shown that the Sample Mean and Analysis Time orderings produce *stochastically ordered* distributions of the outcomes under the proposed ordering:

$$P_\theta\left((M,\hat{\theta}) \succ (j,t)\right) \text{ is an increasing function of } \theta$$

  for both $\succ_{SM}$ and $\succ_{AT}$ orderings.
- In contrast, stochastic ordering has not been proven for the Likelihood Ratio ordering.

Confidence Intervals from p-values

- Construct one-sided p-values $p_1(\theta_0)$ for test of $H_0 : \theta = \theta_0$ vs. $H_A : \theta > \theta_0$ using chosen ordering of sample space.
- Obtain two-sided p-values as $p(\theta_0) = 2 * \min(p_1(\theta_0), 1 - p_1(\theta_0))$
- Construct confidence intervals using hypothesis test/confidence interval duality:

$$\mathcal{C} = \{\theta_0 : p(\theta_0) > \alpha\}$$

Alternative Confidence Interval Approach: Repeated Confidence Intervals

Repeated Confidence Intervals (Jennison and Turnbull, 1989):

- Invert a level α two-sided group sequential test at each stage $j = 1, \ldots, J$ to obtain intervals $\mathcal{I}_j$ such that

$$P_\theta \left(\theta \in \mathcal{I}_j \text{ for all } j = 1, \ldots, J \right) = 1 - \alpha$$

- $\mathcal{I}_j$ is the set of all values of θ_0' for which a group sequential test of $H_0 : \theta = \theta_0'$ would not reject at stage j.

Alternative Confidence Interval Approach: Repeated Confidence Intervals

- $\mathcal{I}_j$ can be rephrased in terms of the test statistic $T_j(\theta_0')$ which depends upon the null hypothesis value.
- The group sequential stopping rule can be expressed as

$$\text{Reject } H_0 : \theta = \theta_0' \text{ if } T_j(\theta_0') < a_j \text{ or } T_j(\theta_0') > d_j.$$

- Thus we have

$$\mathcal{I}_j = \left\{ \theta_0' : a_j \leq T_j(\theta_0') \leq d_j \right\}$$

Alternative Confidence Interval Approach: Repeated Confidence Intervals

- Consider the normal setting (with no mean variance relationship)
- Let $\{J, n_j, T_j, (a_j, b_j, c_j, d_j)$ for $j = 1, \ldots, J\}$ be a level α group sequential test of $H_0 : \theta = 0$ vs. $H_A : \theta \neq 0$
- Consider the boundary scale $T_j = \hat{\theta}_j$
- The interval $\mathcal{I}_j$ at stage j is

$$\mathcal{I}_j = \left\{\theta_0' : a_j \leq \hat{\theta}_j - \theta_0' \leq d_j\right\}$$

- The repeated confidence interval for θ is therefore

$$\left\{\theta_0' : a_j \leq \hat{\theta}_j - \theta_0' \leq d_j \text{ for all } j = 1, \ldots, J\right\}$$

Point Estimation

- Given that we observe an outcome $(M, \hat{\theta}) = (j, t)$, we would like to provide a point estimate for the parameter θ.
- Several options have been proposed:
  - Maximum Likelihood Estimator
  - Bias-Adjusted Mean
  - Median-Unbiased Estimator
  - (And several others)

Point Estimation

- Maximum Likelihood Estimate:

$$\begin{aligned}\hat{\theta}_{\text{MLE}} &= \hat{\theta} \\ &= \bar{X}_A - \bar{X}_B\end{aligned}$$

- The MLE is typically a biased estimate of θ:

$$E_\theta[\hat{\theta}] \neq \theta$$

- For example, when $\theta = 0$, the expected value of the difference in sample means when the trial stops is

$$E_{\theta=0}[\hat{\theta}] = -0.033 \neq 0$$

Point Estimation

- Bias-adjusted Mean: $\hat{\theta}_{\text{BAM}}$ is the value of θ' satisfying

$$\mathsf{E}\left[\hat{\theta}; \theta'\right] = t;$$

that is, the value of the parameter for which the observed statistic is the expected value under that parameter value.

Point Estimation

- Median-unbiased Mean: $\hat{\theta}_{\text{MUE}}$ is the value of θ' satisfying

$$P\left((M,S) \succ (m,s); \theta'\right) = 0.5;$$

that is, the value of the parameter for which the observed statistic would be the median of the sampling distribution under that parameter value.

- Note that this estimator depends on the ordering of the outcome space.

Inference Optimality Criteria: Confidence Intervals

- How should we decide which method of CI construction is better?
  - Coverage probability close to nominal level $1 - \alpha$
  - Confidence interval width: narrow intervals preferred, more efficient
  - Convexity: Does the method produce a true interval?
  - Agreement with design hypothesis test decision
  - Agreement with a reasonable point estimate

Confidence Interval Optimality Criteria

Coverage Probability:

- Let $\mathcal{C}_{1-\alpha}$ be a nominal $(1-\alpha)100\%$ confidence interval for the parameter θ.
- Recall that the confidence interval is random: when the experiment is repeated, we will obtain different limits for the interval.
- The coverage probability is $P(\theta \in \mathcal{C}_{1-\alpha})$.
- This should be the target level $(1-\alpha)100\%$ by construction, but it is important to assess whether that is actually being achieved.

Confidence Interval Optimality Criteria

Interval Width:

- Methods that produce shorter/narrower intervals with the same coverage probability are preferred: more precision about the estimate of θ.
- In fixed sample setting with known variance, the width is constant for a given sample size.
- In contrast, in the group sequential setting interval width is random and its distribution depends upon the true value of θ.
- Interval length may be compared on basis of
  - Average width
  - Median/other quantile of width
  - Probability that the width exceeds some given size

Confidence Interval Optimality Criteria

Convexity:

- Are the confidence regions true intervals?
- If $\theta_1 \in \mathcal{C}$ and $\theta_2 \in \mathcal{C}$, then we would like to have all parameter values θ^* between θ_1 and θ_2 also in $\mathcal{C}$.
- That is, we want

$$\beta\theta_1 + (1-\beta)\theta_2 \in \mathcal{C}$$

for any $\beta \in (0, 1)$.

Confidence Interval Optimality Criteria

Agreement with Decision:

- If the study is stopped for efficacy then we would prefer that θ_0 not be in the $(1-\alpha)100\%$ confidence region, where α is level for which the stopping boundaries were designed.
  - That is, if the design null hypothesis $H_0 : \theta = \theta_0$ is rejected by the level α by the stopping boundary, $\theta_0 \notin \mathcal{C}$.
- If the study is stopped for futility then the design alternative at which the design has power $1-\beta = 1-\alpha$ should not be in the confidence region.
  - That is, if the design null hypothesis $H_0 : \theta = \theta_0$ is accepted by the stopping boundary that has power $1-\alpha$ to detect the alternative $\theta = \theta_A$, then $\theta_A \notin \mathcal{C}$

Confidence Interval Optimality Criteria

Agreement with Point Estimate:

- If $\tilde{\theta}$ is an estimate of the parameter θ, and $\mathcal{C}$ is a confidence interval for θ, it is preferable to have $\tilde{\theta} \in \mathcal{C}$.
- This is an optimality criterion for both the estimate and the interval
- It is more important to have agreement with well-behaved estimators like the Bias-adjusted Mean than with poorer estimators like the MLE.
- Some reasonable confidence intervals may not contain the MLE with non-negligible probability, which is more acceptable due to the bias of the MLE.

Inference Optimality Criteria: Point Estimates

- How should we decide which method of point estimate construction is better/which to use?
  - Bias
  - Mean-squared Error
  - Agreement with reasonable confidence interval
  - Agreement with design hypothesis test decision
  - (Consistency)

Point Estimates Optimality Criteria

Bias:

- Is the expected value of the estimator equal to the true parameter value?
- The bias of an estimator $\tilde{\theta}$ for the parameter θ is

$$B(\theta;\theta) = \mathrm{E}(\tilde{\theta}) - \theta$$

- An estimator $\tilde{\theta}$ is unbiased if

$$E(\tilde{\theta}) = \theta$$

- Low or zero bias is desirable, other properties being equal.

Point Estimates Optimality Criteria

Mean-squared Error:

- What is the expected squared distance between the estimator and the true parameter value?
- The mean-squared error of an estimator $\tilde{\theta}$ for the parameter θ is

$$\text{MSE}(\tilde{\theta};\theta) = \text{E}\left[(\tilde{\theta}-\theta)^2\right]$$

- It can be shown that

$$\text{MSE}(\tilde{\theta};\theta) = \left[B(\tilde{\theta};\theta)\right]^2 + \text{Var}[\tilde{\theta}]$$

- Small mean-squared error is desirable.

Point Estimates Optimality Criteria

Agreement with Confidence Interval:

- If $\tilde{\theta}$ is an estimate of the parameter θ, and $\mathcal{C}$ is a confidence interval for θ, it is preferable to have $\tilde{\theta} \in \mathcal{C}$.
- This is an optimality criterion for both the estimate and the interval
- It is more important to have agreement with well-behaved confidence intervals (i.e. those that are narrower, form true intervals, etc.)

Point Estimates Optimality Criteria

Agreement with Decision:

- Is it possible that the estimate be in the null hypothesis region of the parameter space, but the decision based on the boundary is to reject the null?
- That is, if $\tilde{\theta} \leq \theta_0$ we do not want to reject $H_0 : \theta = \theta_0$ in favor of a greater alternative.

Point Estimates Optimality Criteria

Consistency

- Does the estimator converge (in probability) to the true value as the sample size increases to infinity?
- This property is less emphasized for sequential designs, as we are primarily interested in the sample size for which the study is planned.
- (You may, nevertheless, encounter papers where consistency of estimators in a group sequential design setting is considered.)

Types of Design Adaptation

- Many possible ways to adaptively modify future analysis plan at an interim analysis. Examples:
  - Sample size re-estimation
  - Adaptive randomization
  - Dropping inferior treatment groups
  - Change of endpoint
  - Change of hypothesis

Types of Design Adaptation

- The common theme among adaptive designs is the use of an interim effect estimate to adjust the plans for future analyses.
- Here we focus solely on adaptive sample size and stopping boundary modification based on interim effect estimate.
- Note that modifying sample sizes due to updated information on ancillary statistics/information growth is not considered in this setting, and does not require as careful attention to protecting Type I error rate.

Adaptive Sample Size and Stopping Boundary Modification

- The type of design adaptation we consider here involves using an interim estimate of effect size to modify the future analyses.
- Modification may affect any or all of the following components of the future analysis plan:
  - Number of future analyses
  - Timing/sample size for future analyses
  - Stopping boundary/critical value(s) for future analyses

Why Adapt?

- Proposed Benefits of Adaptive Sample Size and Stopping Boundary modification:
  - Re-power study to detect smaller/larger effect size if interim estimate indicates a value substantially different from design hypotheses
  - Increased flexibility in accrual decisions, justification for sample size
  - Possibly improve efficiency
  - Potential cost reduction, particularly in time-to-event setting

Considerations in Adapting Sample Size/Stopping Boundary

- If adaptation is performed without careful adjustment of stopping boundary, Type I error can be greatly inflated.
- Proschan and Hunsberger (1995):
  - Two-stage design: n_1 in first stage
  - Interim effect size estimate at first stage used to choose n_2 for second stage
  - Depending upon how n_2 chosen, Type I error probability can more than double: $0.05 \rightarrow 0.1146$
  - Even Bonferroni correction would not fix this inflation of Type I error rate.

Some Proposed Adaptation Rules

- Proschan and Hunsberger (conditional error)
- Lehmacher and Wassmer 1999 (reweighted statistic)
- Cui, Hung, Wang 1999 (reweighted statistic)
- Muller and Schafer 2001 (conditional error)
- Brannath, Posch, Bauer 2002 (recursive combination tests/conditional error)
- Gao, Ware, Mehta 2008 (conditional error, sample size guided by conditional power)
- Mehta and Pocock 2010 (conditional error, sample size guided by conditional power)
- More general: Any path of group sequential designs chosen to have correct rejection rate under null

Some Proposed Adaptation Rules

- Adaptation rules in literature can be categorized into three general approaches:
  - Reweighting the test statistic: using the same stopping boundary (critical values) with different sample sizes
  - Conditional error preservation: using possibly different stopping boundary (critical values) and different sample sizes
  - General pre-specified design such that overall type I error rate is controlled
- We will see that these are listed in order of increasing flexibility.

Adaptation Rules: Reweighted Statistic/Combining p-values

- The first type of adaptation rule starts with a group sequential design:
  - $\{J, n_j, T_j, (a_j, b_j, c_j, d_j)$ for $j = 1, \ldots, J\}$
  - Let T_j be either the z-statistic Z_j or the fixed sample p-value P_j.
  - Incremental test statistics Z_j^* and P_j^*, computed only from the data acquired in the jth group.
- At some interim analysis $h\,(1 \leq h < J)$, the future incremental sample sizes may be modified:
  - $n_j^* \rightarrow \tilde{n}_j^*$ for $j = h+1, h+2, \ldots, J$
  - For notational conveniene, we let $\tilde{n}_j^* = n_j^*$ for $j = 1, \ldots, h$.
  - Let $\tilde{T}_j^*$ be the incremental test statistic computed using the new sample size for the jth stage, $\tilde{n}_j^*$.

Adaptation Rules: Reweighted Statistic/Combining p-values

- First consider the normal mean setting with $n_{Aj} = n_{Bj} = \frac{n_j}{2}$, so

$$Z_j = \frac{(\hat{\theta}_j - \theta_0)\sqrt{n_j}}{2\sigma}$$

$$Z_j^* = \frac{(\hat{\theta}_j^* - \theta_0)\sqrt{n_j^*}}{2\sigma}$$

- Note that we can write

$$\hat{\theta}_j = \frac{\frac{n_1^*}{2}\hat{\theta}_1^* + \frac{n_2^*}{2}\hat{\theta}_2{}^* + \ldots + \frac{n_j^*}{2}\hat{\theta}_j^*}{\frac{n_1^*}{2} + \frac{n_2^*}{2} + \ldots + \frac{n_j^*}{2}} = \frac{\sum_{\ell=1}^{j} n_\ell^* \hat{\theta}_\ell^*}{\sum_{\ell=1}^{j} n_\ell}$$

Adaptation Rules: Reweighted Statistic/Combining p-values

- Therefore, we can decompose Z_j in terms of the incremental Z_ℓ^* as

$$\begin{aligned} Z_j &= \frac{\sqrt{n_j}}{2\sigma}\left(\frac{\sum_{\ell=1}^{j} n_\ell^* \hat{\theta}_\ell^*}{\sum_{\ell=1}^{j} n_\ell} - \theta_0\right) \\ &= \frac{\sqrt{n_j}}{2\sigma}\left(\frac{\sum_{\ell=1}^{j} n_\ell^* (\hat{\theta}_\ell^* - \theta_0)}{n_j}\right) \\ &= \frac{\sum_{\ell=1}^{j} \sqrt{n_\ell^*} Z_\ell^*}{\sqrt{n_j^*}} \end{aligned}$$

Adaptation Rules: Reweighted Statistic/Combining p-values

- If the incremental group sizes are modified, we still have

$$\tilde{Z}_j^* = \frac{(\tilde{\hat{\theta}}_j^* - \theta_0)\sqrt{\tilde{n}_j^*}}{2\sigma} \sim N(0,1) \quad \text{under } H_0 : \theta = \theta_0$$

- Note, however, that if the sample sizes $\tilde{n}_j$ and the incremental sample sizes $\tilde{n}_j^*$ depend on interim effect estimates, we do not have $\tilde{n}_j$ independent of Z_ℓ^* for $\ell \neq j$.

Adaptation Rules: Reweighted Statistic/Combining p-values

- Therefore, the statistic

$$\tilde{Z}_j = \frac{\sum_{\ell=1}^{j} \sqrt{\tilde{n}_\ell^*}\, \tilde{Z}_\ell^*}{\sqrt{\tilde{n}_j^*}}$$

may not be $N(0,1)$ under H_0, as it is no longer a standardized sum of independent normal random variables.

Adaptation Rules: Reweighted Statistic/Combining p-values

- If instead we use pre-specified weights (variances) w_ℓ for each Z_ℓ^* in computing the test statistic, we do obtain a standard normal random variable under H_0:

$$Y_j = \frac{\sum_{\ell=1}^{j} \sqrt{w_\ell} \tilde{Z}_\ell^*}{\sqrt{\sum_{\ell=1}^{j} w_\ell}}$$

  since $\sqrt{w_\ell} Z_\ell^* \sim N(0, w_\ell)$ so $\sum_{\ell=1}^{j} \sqrt{w_\ell} Z_\ell^* \sim N(0, \sum_{\ell=1}^{j} w_\ell)$
- A natural choice for the weights w_ℓ is the originally planned sample sizes

$$w_\ell = n_\ell^*$$

Adaptation Rules: Reweighted Statistic/Combining p-values

- The statistic Y_j is compared to the originally planned stopping boundary critical values for the jth stage.
- This procedure has the same Type I error rate as the original group sequential design.

Adaptation Rules: Reweighted Statistic/Combining p-values

- The reweighting approach can also be expressed, more generally, as an approach of combining p-values.
- Setting
  - A total of J potential analyses are allowed.
  - The data gathered in each stage is independent of all other stages (independent increments)
  - Incremental p-values P_j^* for each stage are exact (or at least near-exact) in the sense that

$$P_{H_0}(P_j^* \leq u) \approx u \quad \text{for all } u \in [0,1]$$

Adaptation Rules: Reweighted Statistic/Combining p-values

- Then the following test statistic may be compared to a level α stopping boundary with J analyses on the Z-scale (reject H_0 for large $Q_j \Leftrightarrow$ greater alternative).

$$Q_j = \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \Phi(1 - P_j^*)$$

- This approach protects the type I error rate at level α, no matter what incremental sample sizes n_j^* are used for each stage.
- Incremental sample sizes may be modified at any time
- Number of possible future analyses may *not* be changed

Adaptation Rules: Conditional Error/Modified Critical Value

- The second type of adaptation rule also starts with a group sequential design:
  - $\{J, n_j, T_j, (a_j, b_j, c_j, d_j) \text{ for } j = 1, \ldots, J\}$
  - At some interim analysis $h\,(1 \leq h < J)$, the entire future sampling plan and stopping boundary may be modified:
    - $n_j^* \rightarrow \tilde{n}_j^*$ for $j = h+1, h+2, \ldots, J$
    - $(a_j, b_j, c_j, d_j) \rightarrow (\tilde{a}_j, \tilde{b}_j, \tilde{c}_j, \tilde{d}_j)$ for $j = h+1, h+2, \ldots, \tilde{J}$
  - Note that the entire **sample path**: (sample size, boundary, and number of future analyses) may change.

Adaptation Rules: Conditional Error/Modified Critical Value

- The **conditional rejection rate**
  - at a specified true value of the parameter θ,
  - given the current test statistic value/estimate of effect size $\hat{\theta}_h = t_h$, and
  - using a particular future sampling path k: $\{\tilde{J}^{(k)}, \tilde{n}_j^{(k)}, \tilde{T}_j^{(k)}, (\tilde{a}_j^{(k)}, \tilde{b}_j^{(k)}, \tilde{c}_j^{(k)}, \tilde{d}_j^{(k)}) \text{ for } j = h+1, \ldots, \tilde{J}^{(k)}\}$

  is given by

$$CP_\theta(\text{Sampling Path } k | \hat{\theta}_h = t_h) =$$

$$P_\theta(\text{Reject } H_0 \text{ at any } j = h+1, \ldots, \tilde{J}^k \text{ using sampling path } k | \hat{\theta}_h = t_h)$$

Adaptation Rules: Conditional Error/Modified Critical Value

- **Conditional type I error rate** is the conditional rejection rate under the null hypothesis $H_0 : \theta = \theta_0$
- If we constrain our future sampling paths to match the original conditional type I error rate, i.e. ensure that

$$CP_{\theta_0}(\text{Original Sampling Path } k = 0 | \hat{\theta}_h = t_h) =$$
$$CP_{\theta_0}(\text{Sampling Path } k | \hat{\theta}_h = t_h)$$

then the overall type I error rate of the adaptive design is controlled at the original level α.

Adaptation Rules: Conditional Error/Modified Critical Value

- Popular methods of conditional error adaptation:
  - Consider adaptation at the next to last stage $h = J - 1$.
  - Modification of final sample size only; no increase in number of future analyses is considered.
  - To maintain conditional type I error rate, the critical values $(\tilde{a}_J, \tilde{b}_J, \tilde{c}_J, \tilde{d}_J)$ must be adjusted based on the new final sample size $\tilde{n}_J = n_{J-1} + \tilde{n}^*_J$
  - Typically, a one-sided design is considered, so $\tilde{a}_J = \tilde{d}_J$ and therefore a single critical value $\tilde{a}_J(\tilde{n}_J)$ must be solved for, in terms of the new incremental sample size $\tilde{n}^*_J$.

Adaptation Rules: Conditional Error/Modified Critical Value

- Gao, Ware, and Mehta 2008 provide formulae for the critical value $\tilde{a}(\tilde{n}_J^*)$ given an observed test statistic $Z_{J-1} = z_{J-1}$ and a new incremental sample size $\tilde{n}_J^*$:

$$\tilde{a}_J(\tilde{n}_J^*) = \frac{1}{\sqrt{\tilde{n}_J}} \left[\frac{\sqrt{\tilde{n}_J^*}}{\sqrt{n_J^*}} \left(a_J \sqrt{n_J} - z_h \sqrt{n_{J-1}} \right) + z_h \sqrt{n_{J-1}} \right]$$

- It can be shown that this is equivalent to reweighting the z-statistic and using the original critical value a_J:
  - Changing the statistic, keeping the boundary $\Leftrightarrow$ Keeping the statistic, changing the boundary

Adaptation Rules: Conditional Error/Modified Critical Value

- Contrary to the statement in Gao, Ware, and Mehta 2008:

  > *"The equivalence of the three methods demonstrates that the sample size re-estimation method of Cui, Hung, and Wang is valid and does not truly down-weight any portion of the data."*

  this equivalence instead demonstrates that modifying the critical value based on a new sample size *is the same as down-weighting some of the data* and is therefore likely to be an inefficient approach.

Example Adaptation Rules: Conditional Error/Modified Critical Value

- Adaptive final sample size may be chosen according to any desired criteria
- Popular choice of adaptive final sample size is to chose $\tilde{n}_J$ to attain a desired level of **conditional power**, where the conditional power is evaluated using the current effect size estimate $\hat{\theta}_h$ as the true parameter.

Adaptation Rules: Conditional Error/Modified Critical Value

- Given desired conditional power $1 - \beta$, we find $\tilde{n}_J^*$ such that

$$P_{\hat{\theta}_L}\left(\tilde{Z}_J > \tilde{a}_J(\tilde{n}_J^*)\right) = 1 - \beta$$

  where $\tilde{Z}_J$ is the cumulative z-statistic using $\tilde{n}_J = n_{J-1} + \tilde{n}_J^*$ observations.
- Since $\tilde{a}_J(\tilde{n}_J^*)$ is a function of $\tilde{n}_J^*$, this expression can be solved for the desired value $\tilde{n}_J^*$.

Adaptation Rules: Conditional Error/Modified Critical Value

- Gao, Ware, and Mehta 2008 also provide formulae for the new final sample size $\tilde{n}_J$ needed to obtain conditional power of $1-\beta$, given that $Z_{J-1} = z_{J-1}$:

$$\tilde{n}_J = \frac{n_{J-1}}{z_{J-1}^2}\left[\frac{(a_J\sqrt{n_J} - z_{J-1}\sqrt{n_{J-1}})}{\sqrt{n_J^*}} + z_\beta\right]^2 + n_{J-1}$$

Adaptation Rules: Adaptive Switching between Sampling Path

- The third, most general type of adaptation rule can be thought of as adaptively switching between different group sequential designs (different **sampling paths**).
- Starting with a group sequential design:
  - $\{J^{(0)}, n_j^{(0)}, T_j^{(0)}, (a_j^{(0)}, b_j^{(0)}, c_j^{(0)}, d_j^{(0)})$ for $j = 1, \ldots, J^{(0)}\}$
- At some interim analysis $h\,(1 \leq h < J)$, adaptively select one of r possible future sampling paths.
- Valid adaptive design controlling overall type I error rate as long as:
  - Total probability under $H_0 : \theta = \theta_0$ of rejecting H_0 is constrained to be $\leq \alpha$
  - Exactly one future sampling path is selected.

Adaptation Rules: Adaptive Switching between Sampling Path

Details:

- At interim analysis $h\,(1 \leq h < J)$, the continuation region for $T_h^{(0)}$ is partitioned into r disjoint continuation sets $\mathcal{C}_h^{(k)}$, for $k = 1, \ldots, r$.
- If $T_h^{(0)} \in \mathcal{C}_h^{(k)}$, then the future stopping boundary will be the kth future group sequential sampling path:
  - $\{J^{(k)}, n_j^{(k)}, T_j^{(k)}, (a_j^{(k)}, b_j^{(k)}, c_j^{(k)}, d_j^{(k)})$ for $j = h+1, \ldots, J^{(k)}\}$, for $k = 1, \ldots, r$
- Let K be the random variable denoting which path is chosen, $K \in \{1, \ldots, r\}$: $K = k$ if $T_h^{(0)} \in \mathcal{C}_h^{(k)}$.

Adaptation Rules: Adaptive Switching between Sampling Path

- Compared to the previous two approaches (reweighting and preserving conditional type I error rates), the adaptive switching approach is more flexible.
- Control of the unconditional type I error rate may be accomplished without constraining the conditional type I error rates.

Adaptation Rules: Adaptive Switching between Sampling Path

- Adaptive designs as proposed in Gao, Ware, Mehta 2008 (GWM) and others can be represented in this adaptive switching framework.
- Since
  - sample sizes must be discrete, and
  - there is almost always (always) a maximal possible sample size set

  any adaptive rule can be regarded as switching between a finite number of future group sequential sampling paths.

Adaptation Rules: Adaptive Switching between Sampling Path

- In practice, we have found that the performance of an adaptive rule with a large number r of different possible group sequential sampling paths is not much different from an adaptive rule with a small number of possible sampling paths.
- e.g. A discretized GWM design with just $r = 4$ different possible group sequential sampling paths has practically identical performance to one with $r = 100$ different sampling paths.

Adaptive Switching Designs using Standard Group Sequential Software

- Any pre-specified adaptive design can be represented and calculated using standard group sequential software that allows:
  - Arbitrarily spaced analyses
  - Constrained and partially constrained boundary searches
  - Numerical integration to find stopping probabilities and stopping densities
- Basic idea:
  - Specify each possible group sequential sampling path after analysis time h as a different group sequential design with first analysis at time n_h.