Symposium on Group Sequential Inference
New Jersey Chapter of the ASA
June 4, 2002

Cyrus R. Mehta
President, Cytel Software Corporation
and
Adjunct Professor of Biostatistics, Harvard University

Outline of Workshop

- Motivating Example
- Classical Group Sequential Design
- Infomation Based Design
- Adaptive Design
- Inefficiency of Adaptive Design

Cholesterol Reduction Study (Facey, 1992)

Placebo controlled efficacy trial investigating a new treatment for hyperchloesterolemia, with primary endpoint being reduction in total serum cholesterol over a 4-week period.

- Normally distributed data with $\sigma^2 = 1$ for each subject.
- Each subject receives either Control drug (C) or Experimental drug (E). Let $\delta = \mu_C - \mu_E$.
- We wish to test the null hypothesis $H_0 : \delta = 0$ versus the one-sided alternative hypothesis $H_1 : \delta > 0$ at level $\alpha = 0.05$.
- We want 90% power to detect $\delta = 0.615$.
- We intend to monitor the data 5 times and stop early for benefit or futility; i.e., either to reject or accept the null hypothesis.

Part I: Classical Group Sequential Design

- We believe that $\sigma^2 = 1$ is an accurate estimate and do not make any provision for altering the sample size in case we are wrong.
- We don't mind losing power if $\delta < 0.615$. (Possibly because any value of δ smaller that 0.615 won't be clinically meaningful.)

Measures of Information

- Monitor the data K times at calendar times $\tau_1, \tau_2, \ldots \tau_K$
- For normal and binomial endpoints let

$$n_j = \text{ sample size at calendar time } \tau_j$$

- For time-to-event endpoints let

$$d_j = \text{ number of events at calendar time } \tau_j$$

- More generally, in terms of Fisher information, let

$$I_j = \text{ information at calendar time } \tau_j \approx [\text{se}(\hat{\delta}_j)]^{-2}$$

where $\hat{\delta}_j$ is an efficient estimate of δ at calendar time τ_j. Notice that n_j and d_j are special cases of I_j

The Information Fraction

- Define the "information fraction", t_j, at calendar time τ_j:

$$t_j = \begin{cases} \frac{n_j}{n_K} & \text{for normal and binomial} \\ \frac{d_j}{d_K} & \text{for time-to-event} \\ \frac{I_j}{I_K} \approx \frac{[\text{se}(\hat{\delta}_j)]^{-2}}{[\text{se}(\hat{\delta}_K)]^{-2}} & \text{in general} \end{cases}$$

- Since K is the last look and information is assumed to increase with each successive look, we will often denote I_K by $I_{\max}$, n_K by $n_{\max}$, and d_K by $d_{\max}$.
- We may regard the information fraction t, $0 \leq t \leq 1$, as the internal time axis of the clinical trial.

The Test Statistic

At any interim monitoring time t we compute the Wald statistic

$$T(t) = \frac{\hat{\delta}(t)}{\text{se}(\hat{\delta}(t))}$$

where $\hat{\delta}(t)$ is an efficient estimator for δ using all the data available to us up to time t, and $\text{se}(\hat{\delta}(t))$ is the estimated standard error of $\hat{\delta}(t)$.

Unified Distribution Theory

Provided $\hat{\delta}(t_j)$ is an efficient estimate of δ, the asymptotic joint distribution of the sequence of test statistics $\{T(t_1), T(t_2), \ldots T(t_K)\}$ has the following properties regardless of the underlying model generating the data:

1. $\{T(t_1), T(t_2), \ldots T(t_K)\}$ is multivariate normal.
2. $T(t_j) \sim N(\eta\sqrt{t_j}, 1)$, where $\eta = \delta\sqrt{I_{\max}}$ is known as the drift parameter .
3. For any $t_{j_1} < t_{j_2}$, $\text{cov}\{T(t_{j_1}), T(t_{j_2})\} = \sqrt{\frac{I_{j_1}}{I_{j_2}}}$.

This general result, due to Scharfstein, Tsiatis and Robins (JASA, 1997), enables us to compute boundary crossing probabilities very efficiently (by recursive integration) for any design.

Obtaining Stopping Boundaries

- Monitor the data K times at the interim monitoring time points $\{t_1, t_2, \ldots t_K\}$.
- Let $\{c_1, c_2, \ldots c_K\}$ be the corresponding stopping boundaries.
- Stop the trial and reject H_0 at the first t_j such that

$$T(t_j) \geq c_j \ .$$

- We must select the c_j's so as to preserve the type-1 error.

$$1 - P_0\{\bigcap_{j=1}^{K} T(t_j) < c_j\} = \alpha \ .$$

Many c_j's satisfy this condition. How shall we choose among them?

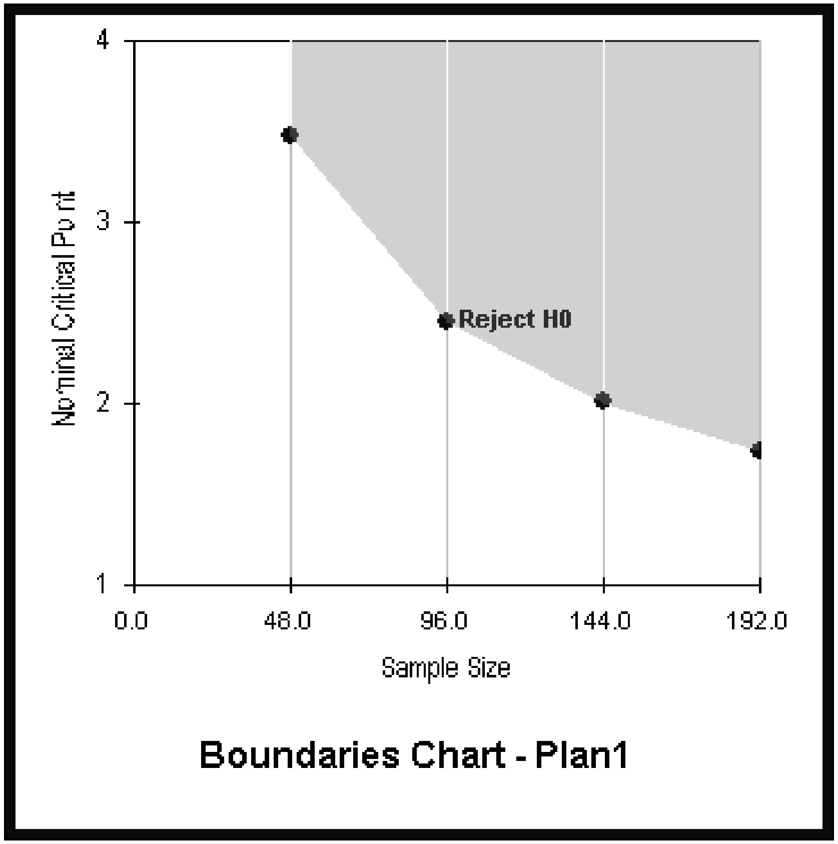
Method 1: Wang-Tsiatis Power Boundaries

Use the Wang-Tsiatis (1987) family of power stopping boundaries

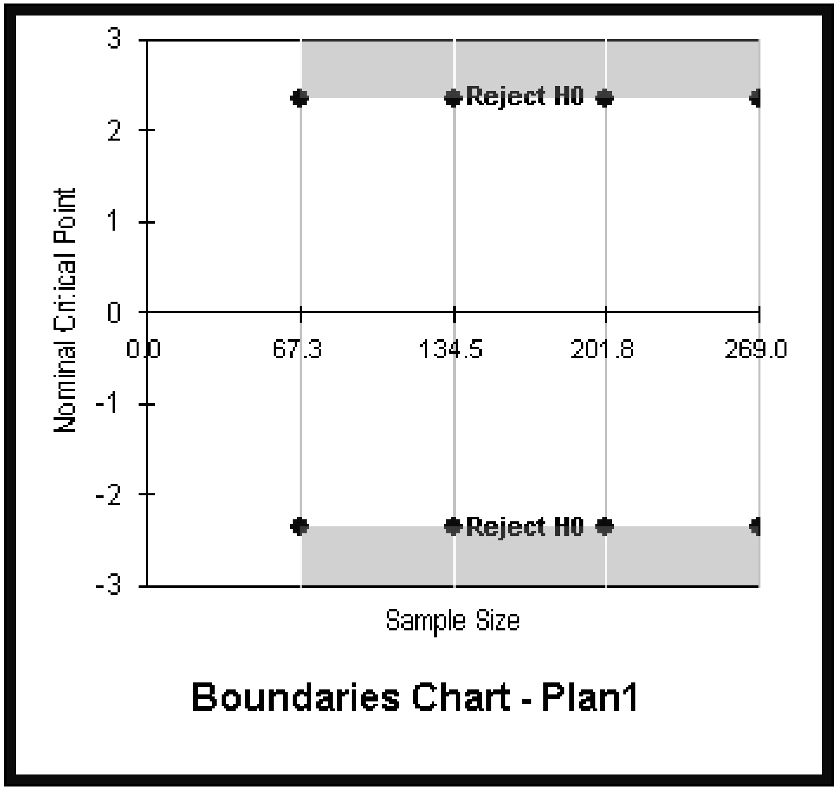
$$c_j = C(\Delta, \alpha, K) j^{\Delta - 0.5}$$

Δ is a shape parameter that puts some structure on the c_j's.

One-Sided O'Brien Fleming Boundary ($\Delta = 0$)



Two-Sided Pocock Boundary ($\Delta = 0.5$)



Method 2: Invert an α-Spending Function

- Specify a monotone increasing function of t for $t \in [0,1]$ with $\alpha(0) = 0$, $\alpha(1) = \alpha$. Lan and DeMets (1983) have proposed

$$\alpha(t) = 4 - 4\Phi\left(\frac{z_{\alpha/4}}{\sqrt{t}}\right) \ .$$

  but any other montone function could be used also.

- Solve recursively for $c_1, c_2, \ldots c_K$:

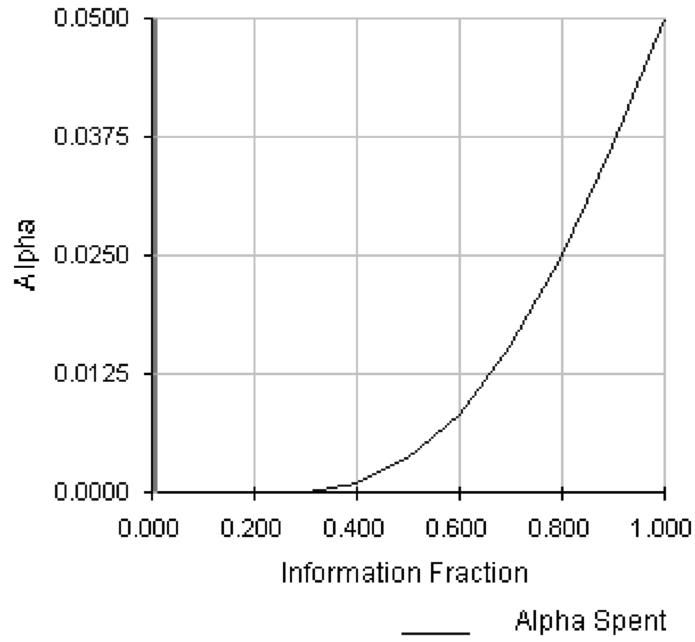
$$P_0\{T(t_1) \geq c_1\} = \alpha(t_1) \ ,$$

$$P_0\{T(t_1) < c_1, T(t_2) \geq c_2\} = \alpha(t_2) - \alpha(t_1) \ ,$$

  and for $j = 3, \ldots K$,

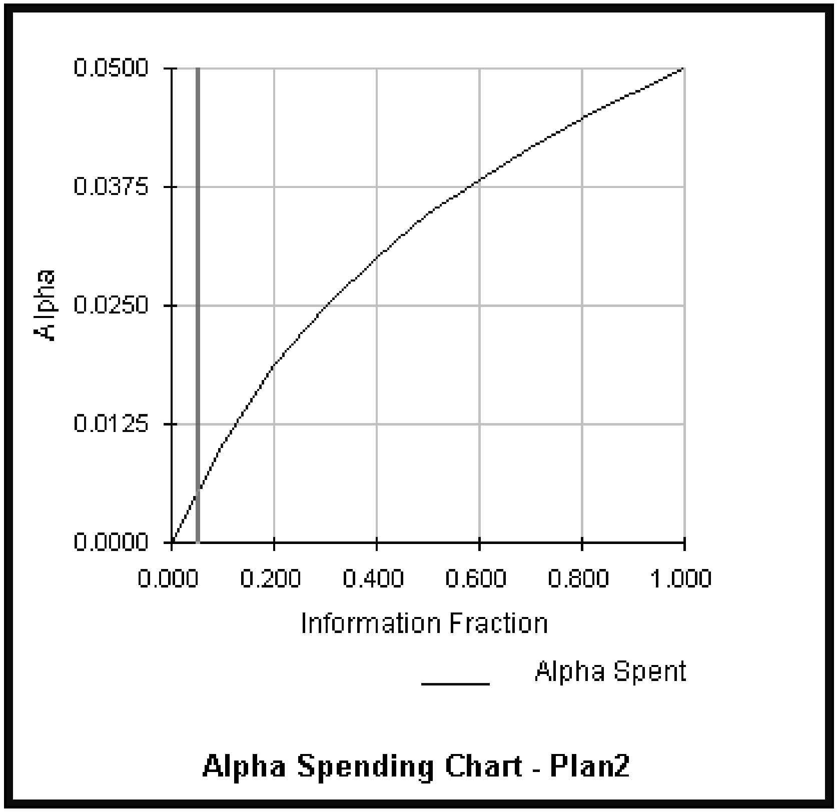
$$P_0\{T(t_1) < c_1, \ldots, T(t_{j-1}) < c_{j-1}, T(t_j) \geq c_j\} = \alpha(t_j) - \alpha(t_{j-1}) \ .$$

α-Spending Function Yielding O'Brien-Fleming Boundary

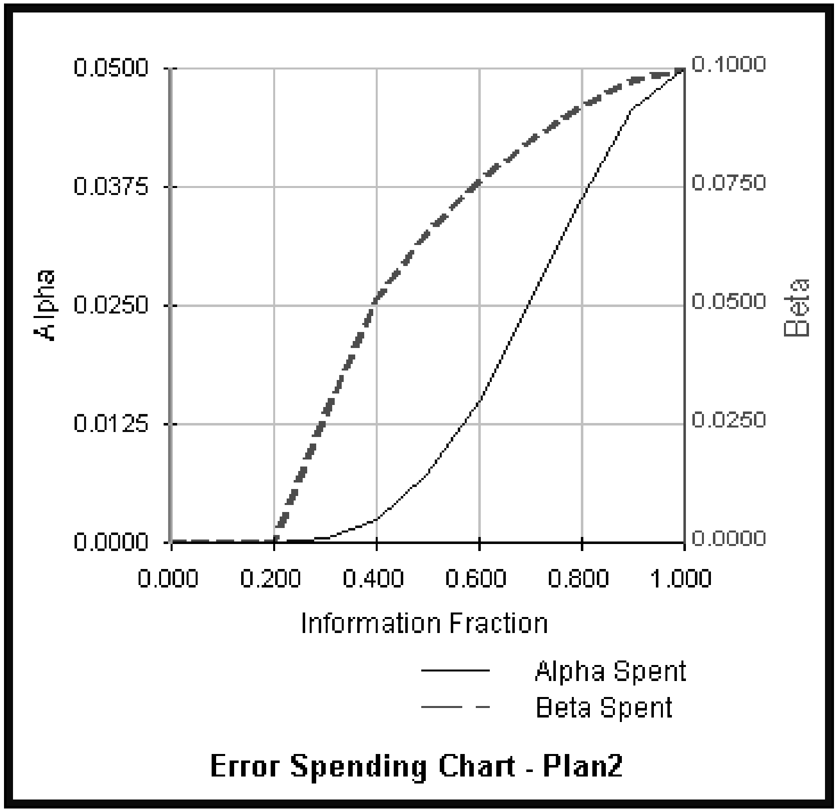


Alpha Spending Chart - Plan2

α-Spending Function Yielding Pocock Boundary



α- and β-Spending Functions (where β is the type-2 error)



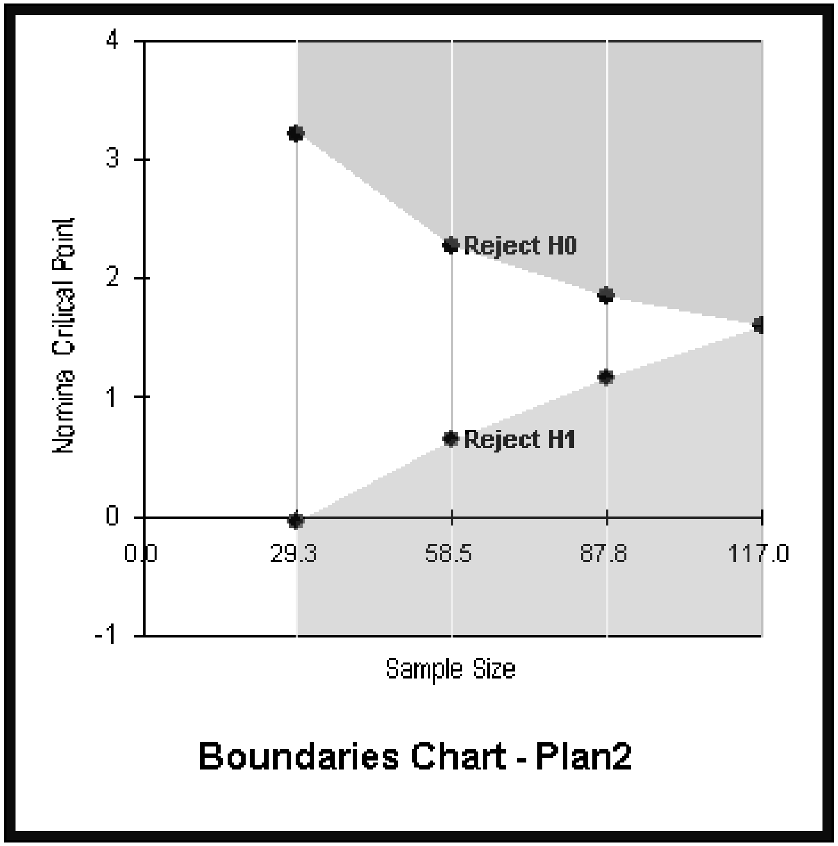
Simultaneous Inversion of α and β Spending Functions (Pampallona and Tsiatis, 1994)

We obtain benefit and futility stopping boundaries $\{(l_i, u_i) : \ i = 1, 2, \ldots K\}$ by recursively satisfying the following pairs of equations for $j = 1, 2, \ldots K$.

$$P_0\{l_1 < T(t_1) < u_1, \cdots, l_{j-1} < T(t_{j-1}) < u_{j-1}, T(t_j) \geq u_j\} = \alpha(t_j) - \alpha(t_{j-1}) \ ,$$

$$P_1\{l_1 < T(t_1) < u_1, \cdots, l_{j-1} < T(t_{j-1}) < u_{j-1}, T(t_j) \geq l_j\} = \beta(t_j) - \beta(t_{j-1}) \ .$$

Benefit and Futility Boundaries Corresponding to the α- and β-Spending Functions



Computing the Sample Size

Suppose the desired power to detect an effect size of δ is $1 - \beta$.

- Recall that $T(t_j) \sim N(\eta\sqrt{t_j}, 1)$.
- Find the value of the drift parameter η that satisfies the equation

$$1 - P_\eta\{\bigcap_{j=1}^{K} |T(t_j)| < c_j\} = 1 - \beta$$

- Given η and δ, solve for $I_{\max}$ from the relationship $\eta = \delta\sqrt{I_{\max}}$.

The Inflation Factor

- Thus we should go on admitting patients until

$$I_{\max} \approx [\text{se}(\hat{\delta}_K)]^{-2} = [\frac{\eta}{\delta}]^2$$

- It is convenient to express $I_{\max}$ in the form

$$I_{\max} = \left[\frac{z_\alpha + z_\beta}{\delta}\right]^2 \times \left[\frac{\eta}{z_{\alpha/2} + z_\beta}\right]^2 = \left[\frac{z_\alpha + z_\beta}{\delta}\right]^2 \times \text{IF} \ ,$$

where

$$\text{IF} = \left[\frac{\eta}{z_{\alpha/2} + z_\beta}\right]^2 \ .$$

$I_{\max}$ is the information required for a corresponding fixed-sample study multiplied by an "inflation factor". For example, for $\Delta = 0$, $\alpha = 0.05$, $1 - \beta = 0.9$, and $K = 5$, IF $= 1.03$ for an O'Brien-Fleming H_0-only boundary.

Converting Information into Sample Size

Recall that $I_{\max} = [\text{se}(\hat{\delta}_K)]^{-2}$. Thus it is easy to convert maximum information into maximum sample size or maximum number of events.

- For normal endpoints and balanced randomization,

$$n_{\max} = 4\sigma^2 I_{\max} \ .$$

- For binomial endpoints and balanced randomization,

$$n_{\max} = 2[\pi_C(1-\pi_C) + \pi_E(1-\pi_E)]I_{\max} \ .$$

- For time-to-event endpoints and balanced randomization,

$$d_{\max} = 4I_{\max} \ .$$

Demonstration of the above design principles in EaSt

Part 2: Information Based Design

- Recall that we first computed $I_{\max}$ and then used our knowledge of σ^2 to convert it into $n_{\max}$.
- But we could be wrong about σ^2. So why not design the study on the basis of $I_{\max}$ alone?

Need for Sample Size Flexibility

- One-sided test; 5% significance and 90% power to detect $\delta = 0.615$; 4 looks with early stopping for H_0 or H_1 with O'Brien-Fleming boundaries. The initial estimate of variance is $\sigma^2 = 1$

$$I_{\max} = \left[\frac{z_\alpha + z_\beta}{\delta}\right]^2 \times \mathrm{IF}(\alpha, \beta, K, \Delta) = [\frac{2.93}{0.615}]^2 \times 1.09 = 24.8$$

- Since $n_{\max} = 4\sigma^2 I_{\max}$:
  - If $\sigma = 1$, we require $n_{\max} = 101$ subjects
  - If $\sigma = 1.2$, we require $n_{\max} = 145$ subjects
  - If $\sigma = 1.3$, we require $n_{\max} = 171$ subjects

Build Sample Size Flexibility into the Protocol

- If we specify 101 patients in the protocol and $\sigma = 1$, the study will have 90% power. But if $\sigma = 1.2$, the power will drop to 79%. If $\sigma = 1.3$, the power will drop fruther to 73%.
- So, rather than specify $n_{\max} = 100$, it is safer to specify the $I_{\max} = 24.8$ in the protocol. That way we are protected even if σ^2 was underestimated.
- We can estimate σ^2 at one or more interim monitoring time-points and can adjust the sample size if necessary to preserve the power of the study.
- By judicious choice of the stopping boundary, we can be very conservative about early stopping, but monitor primarily to improve the design.

Demonstration of Information Based Design in EaSt.

Part 3: Adaptive Design

- The study was designed for 90% power to detect $\delta = 0.615$
- Suppose that half-way through, you take an interim look and observe $\hat{\delta} = 0.4$. If $\hat{\delta}$ is an accurate estimate of δ, you only have 54% power, not 90% power.
  - Can you rescue this study by increasing the sample size?
  - Should you rescue this study by increasing the sample size? i.e., shouldn't you rather have designed **up-front** for 90% power to detect 0.4?

Adaptive Strategies that Preserves α

- Adaptive strategies to increase the sample size have been proposed by Cui, Hung and Wang (1999), Shen and Fisher (1999), Lemacher and Wassmer (1999), and many others.
- All these methods preserve the type-1 error by modifying the test statistic after the sample size has been adapted.

Example: Method of Cui, Hung and Wang, 1999

- **Original Design:** Treatment vs. placebo; $\alpha = 0.05$, power $= 90\%$; normal data; 5 equal looks; $\sigma^2 = 1$; one-sided test of $\delta = 0$ vs $\delta = 0.615$. Based on these input parameters, maximum sample size is $n_{\max} = 101$ with $H_0 \& H_1$ Pampallona-Tsiatis boundaries, and no adaptation.

- **Adaptive Design:** Suppose you adapt **after** seeing interim results at look $L = 2$ because $\hat{\delta}$ is much smaller than 0.615. New maximum sample size is

$$m_{\max} = \min\left\{ n_{\max} \times \left(\frac{0.615}{\hat{\delta}}\right)^2, \text{ Practical Upper Limit} \right\}$$

- Remaining $5 - L$ looks are spaced equally with $(m_{\max} - Ln_{\max}/5)/(5 - L)$ new subjects per look.

- It is important to recognize that $m_{\max}$ is a random variable whose value depends on the value of $\hat{\delta}$ that will be observed at the end of look L.

- Therefore, a priori, the remaining $5 - L$ interim monitoring time points could lie anywhere in the range $Ln_{\max}/5$ to $m_{\max}$.

Modify the Test Statistic

- Consider look $j > L$. Let n_j be the original sample size and $\tilde{n}_j$ be the new (increased) sample size after adapting.
- Let
$$Z_i = (X_{Ci} - X_{Ei})/\sqrt{2}$$
denote the standardized difference in responses for the ith patient pair. Replace
$$T_j = \frac{Z_1 + Z_2 + \cdots + Z_{\tilde{n}_j/2}}{\sqrt{\tilde{n}_j/2}}$$
by
$$\tilde{T}_j = \sqrt{\frac{n_L}{n_j}} \sum_{l=1}^{n_L/2} \frac{Z_l}{\sqrt{n_L/2}} + \sqrt{1 - \frac{n_L}{n_j}} \sum_{l=n_L/2+1}^{\tilde{n}_j/2} \frac{Z_l}{\sqrt{(\tilde{n}_j - n_L)/2}}$$
for $j = L+1, L+2, \ldots 4$.

- Cui, Hung and Wang (1999) have shown that the modified statistic will preserve the type-1 error.
- Also, because it permits the sample size to be increased, it will boost up the power if in fact $\delta < 0.615$.
- There is, however, a price to be paid. The modified test statistic is not a sufficient statistic for δ. It "down-weights" the contributions to the estimate of δ that are made after look L. Therefore it loses efficiency and can be improved.

Theorem (Tsiatis and Mehta, 2000): Any adaptive test, no matter how obtained, can be improved upon by a likelihood ratio test that utilizes the same acceptance and rejection regions as the adaptive test.

- When $\delta > 0$ the likelihood ratio test has a higher probability of crossing the rejection boundary than the adaptive test at every value of $t \in [0, 1]$.
- When $\delta > 0$ the likelihood ratio test has a lower probability of crossing the acceptance boundary than the adaptive test at every value of $t \in [0, 1]$.
- When $\delta < 0$ the corresponding opposite conclusions hold.

Demonstrate the adaptive design in EaSt. Show that it is inefficient relative to the classical design.

Obtaining the Acceptance and Rejection Regions of any Adaptive Test

- Any level-α adaptive test is completely described by:
  - the maximum sample size, $m_{\max}$;
  - a rejection spending function, $\alpha(t)$, that specifies the probability, under the null, of rejecting H_0 by information time t;
  - an acceptence spending function, $\theta(t)$, that specifies the probability, under the null, of accepting H_0 by information time t;
- Suppose the rejecting and acceptance spending functions have discrete mass at time points $t_1, t_2, \ldots t_K$. We can extract the acceptance and rejection regions of the adaptive design by finding the stopping boundaries $\{(l_i, u_i) : \ i = 1, 2, \ldots K\}$. To do this we must solve the following pair of equations recursively, for $j = 1, 2, \ldots K$.

$$P_0\{l_1 < T(t_1) < u_1, \cdots, l_{j-1} < T(t_{j-1}) < u_{j-1}, T(t_j) \geq u_j\} = \alpha(t_j) - \alpha(t_{j-1}) \, ,$$

$$P_0\{l_1 < T(t_1) < u_1, \cdots, l_{j-1} < T(t_{j-1}) < u_{j-1}, T(t_j) \geq l_j\} = \theta(t_j) - \theta(t_{j-1}) \, .$$

Comments on Adaptive Design

- No matter what adaptive design is adopted, it is possible to improve on it with the classical (non-adaptive) group sequential test using stopping boundaries induced by the adaptive test's own spending functions.
- The improvements can be substantial!
- That being the case, a better strategy might be to design up-front with the classical group sequential test for the smaller δ but choose optimal boundaries (e.g., Jennison, 1987, Eales and Jennison, 1992) that are tailored to stop the trial very early if the larger δ is true.