Flexible Clinical Trial Design

Barnett International Conference on Adaptive Methods

March 4, 2004

Cyrus R. Mehta

President, Cytel Software Corporation

email: mehta@cytel.com – web: www.cytel.com – tel: 617-661-2011

Agenda

1. Group Sequential Design
   - Distribution Theory
   - Spending Functions and Stopping Boundaries
   - Early Stopping for Futility
   - Sample Size
2. Interim Monitoring
   - Deviation from number and timing of interim looks
   - Conditional Power
   - Repeated Confidence Intervals
   - Adjusted p-values
3. Sample Size Re-estimation
   - Information Based Design
   - Adaptive Design

Part I

Introduction and Statistical Principles of Flexible Clinical Trials

What are Flexible Clinical Trials

Traditional Design: fix the sample size in advance and only perform **one** efficacy analysis after all subjects have been enrolled and evaluated.

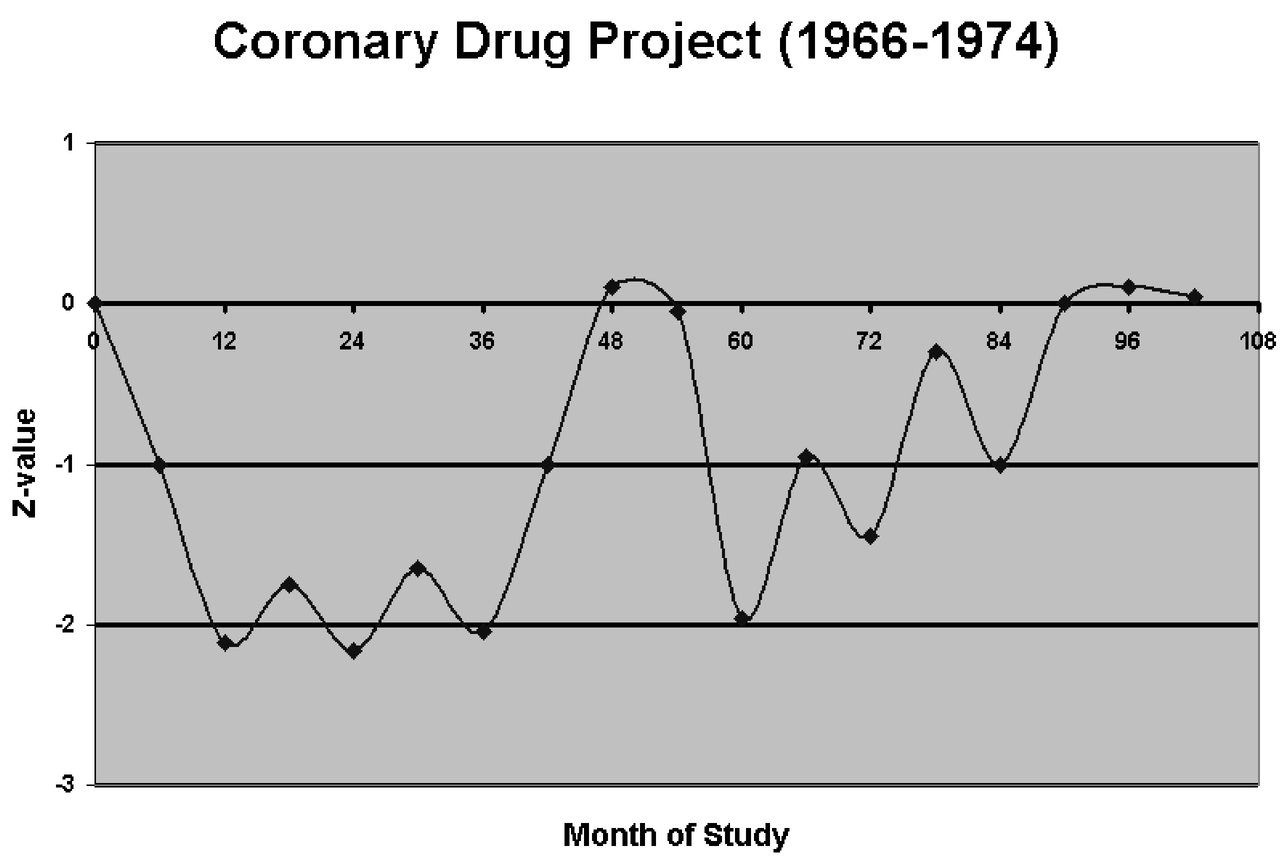
Flexible Design: monitor the accruing efficacy data at administratively convenient intervals and make important decisions concerning the future course of the study along the way.

Advantages of Flexible Clinical Trials

- **Early intimation of efficacy**; option to either terminate or prepare for an early regulatory submission.
- **Early intimation of inefficacy**; option to either terminate for futility, drop the ineffective arm, or divert key resources to more promising studies.
- **Verify design assumptions** (variance, effect size, covariates, etc) from accumulating data; option to revise the sample size to avoid an underpowered study.

What are the Statistical Problems?

If you take multiple looks you are much more likely to see spurious effects due to chance fluctuations in the data.



Appropriate Statistical Methods are Necessary

- **Unless appropriate statistical methods are used, multiple looks, premature termination, or boosting the sample size can affect the type-1 and type-2 errors**
- **The Good News: Appropriate statistical methods for interim monitoring do exist, are endorsed in the FDA Guidance Document ICH-E9, and validated software for implementing these methods is now available**

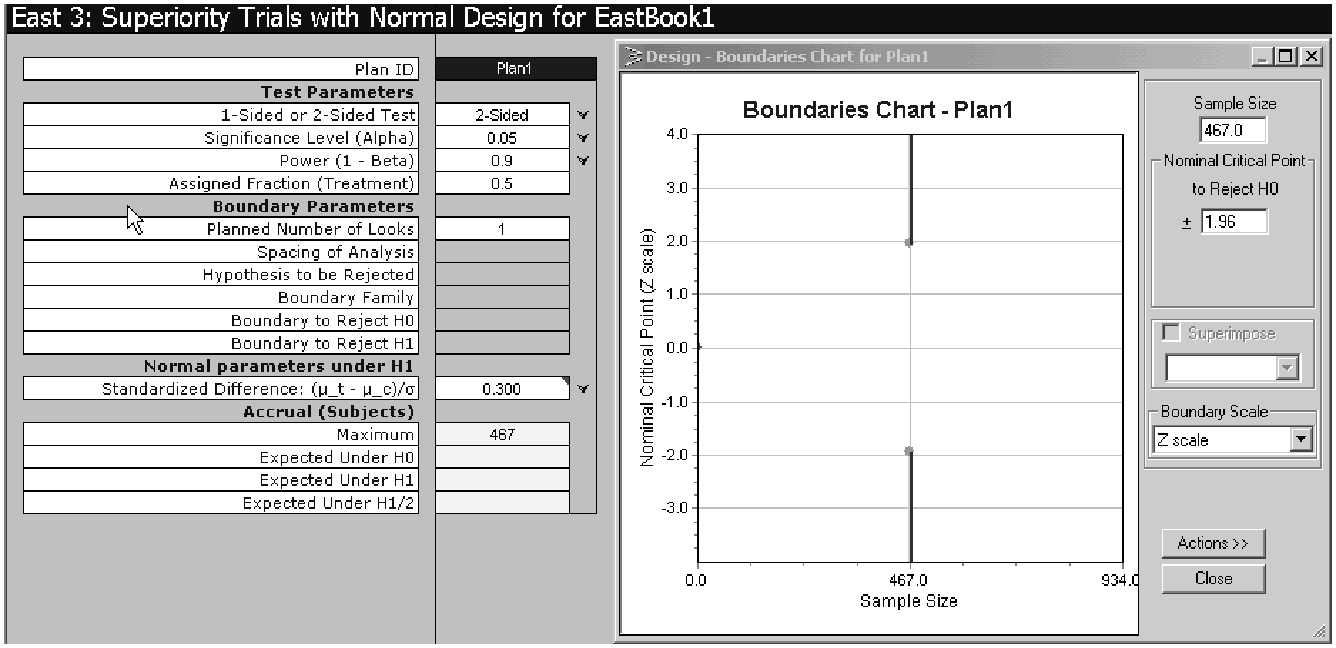
What are the Logistical Problems?

- Access to the interim results by trial investigators could bias the future conduct of the study
- For pivotal trials, the potential for bias is controlled through:
  - An external Data Monitoring Committee
  - An external statistical center preparing the interim analysis reports
- Future Possibility: One could consider controlling bias through direct links between EDC software and group sequential software

Cholesterol Reduction Example

- Placebo controlled efficacy trial with primary endpoint being reduction in total serum cholesterol over a 4-week period (Facey, 1992).
- A subject receives either Control drug (C) or Experimental drug (E).
- Design for 90% power to detect an average benefit of 60 mg/dl for drug E compared to drug C. Two-sided test with 5% significance level.
- Patient to patient variability assumed to be $\sigma^2 = 200^2$.

Static Single-Look Design



How interim monitoring can help

The static design makes many assumptions that cannot be corrected in mid-stream

- **Maybe $\delta << 60$. If so, we'd like to know early and cut our losses**
- **Maybe $\delta >> 60$. If so, we'd like to know early and possibly terminate for efficacy**
- **Maybe $\delta \approx 60$ but $\sigma > 200$. If so, we'd like to increase the sample size**
- **Maybe we should have actually designed for 90% power to detect $\delta = 30$, not $\delta = 60$**

Interim monitoring makes it possible to make early termination decisions and/or to implement mid-course corrections to the study design

Frequently Expressed Concerns

1. Will type-1 error be preserved?
2. Is there a penalty for taking interim looks?
   (a) stricter p-value requirements?
   (b) larger sample size requirement?
3. Will the sponsor be locked into a rigid schedule with respect to the number and timing of the interim looks?

Comparison of Single-Look and Multi-Look Designs

East 3: Superiority Trials with Normal Design for EastBook1

Plan ID	Plan1	Plan2	Plan3	Plan4
Test Parameters				
1-Sided or 2-Sided Test	2-Sided	2-Sided	2-Sided	2-Sided
Significance Level (Alpha)	0.05	0.05	0.05	0.05
Power (1 - Beta)	0.9	0.9	0.9	0.9
Assigned Fraction (Treatment)	0.5	0.5	0.5	0.5
Boundary Parameters				
Planned Number of Looks	1	3	3	3
Spacing of Analysis		Equal	Equal	Equal
Hypothesis to be Rejected		H0 Only	H0 Only	H0 Only
Boundary Family		SpF(Pub)	SpF(Pub)	SpF(Pub)
Boundary to Reject H0		LD(OF)	Gm(-2)	Gm(1)
Boundary to Reject H1				
Normal parameters under H1				
Standardized Difference: (μ_t - μ_c)/σ	0.300	0.300	0.300	0.300
Accrual (Subjects)				
Maximum	467	473	487	540
Expected Under H0		471	482	529
Expected Under H1		379	348	337
Expected Under H1/2		453	454	481

Statistical Principles

- Distribution Theory
- Spending Functions and Stopping Boundaries
- Sample Size Computation
- Flexible Interim Monitoring
- Conditional Power
- Repeated Confidence Intervals
- Adjusted p-values

Notation

Let δ denote the true (unknown) difference between the treatment and control groups. There are many different ways to define δ:

- difference of two means
- difference of two binomial probabilities
- log hazard ratio
- log odds ratio
- any general coefficient in a regression model

Types of Trials

- Superiority Trial – Two Sided: **We select a clinically meaningful superiority margin $\delta_1 > 0$. We are interested in testing**

$$H_0\colon\ \delta = 0 \text{ versus } H_1\colon\ |\delta| > 0$$

using a two-sided level-α test with power $1 - \beta$ to reject H_0 when $|\delta| = \delta_1$.

- Superiority Trial – One Sided: **We select a clinically meaningful superiority margin $\delta_1 > 0$. We are interested in testing**

$$H_0\colon\ \delta = 0 \text{ versus } H_1\colon\ \delta > 0$$

using a one-sided level-α test with power $1 - \beta$ to reject H_0 when $\delta = \delta_1$.

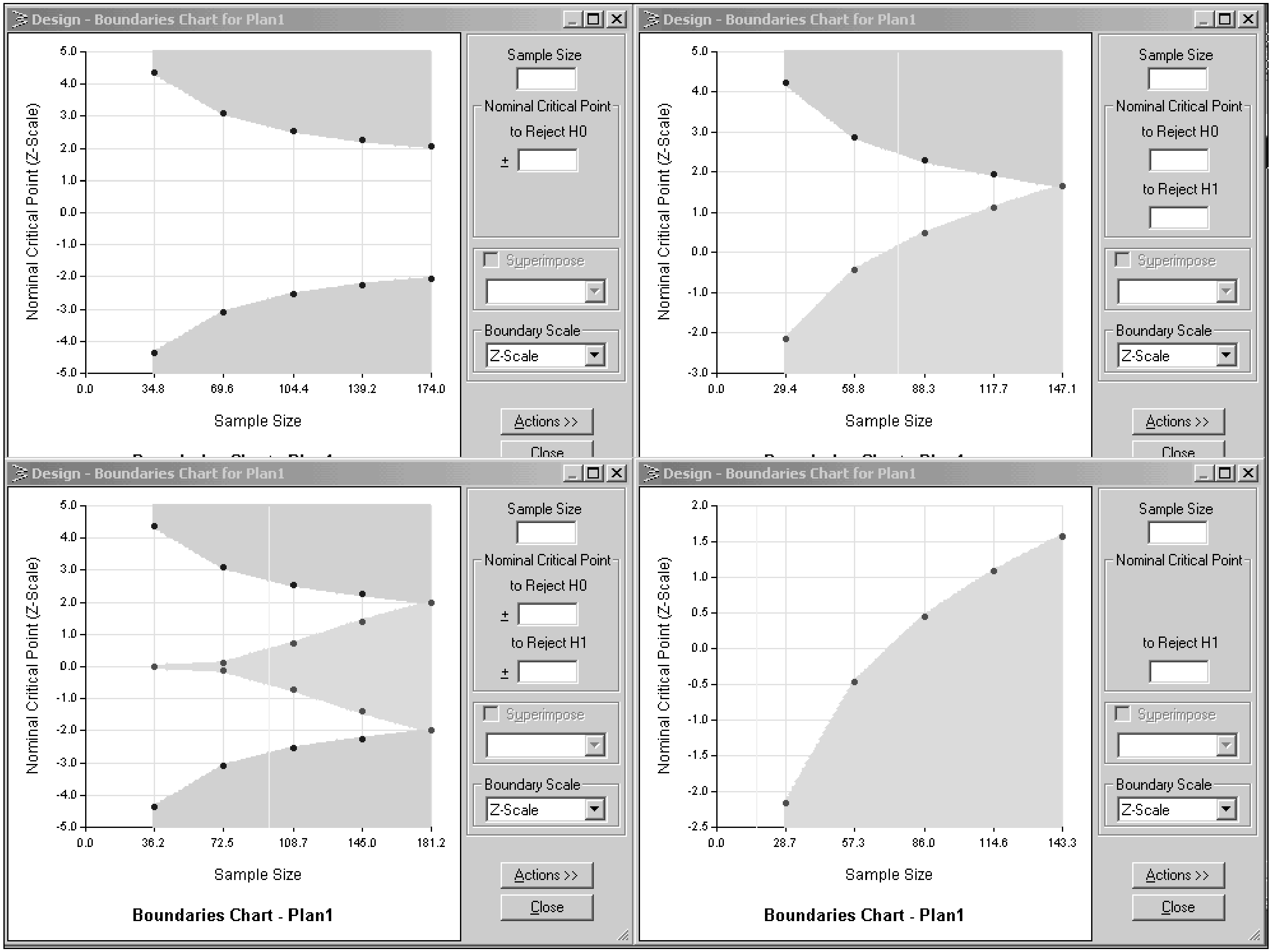
- Non-Inferiority Trial – One Sided: **We select a clinically meaningful non-inferiority margin $\delta_0 > 0$. We are interested in testing**

$$H_0\text{: } \delta = \delta_0 \text{ versus } H_1\text{: } \delta < \delta_0$$

using a one-sided level-α test with power $1 - \beta$ to reject H_0 when $\delta = 0$.

Note:

- **In this framework we are assuming that δ_1 (or δ_0) can be specified a priori. The goal is not to estimate δ_1, but rather to power the study so that if $\delta > \delta_1$, the trial will be positive with high probability.**
- **The interest in adaptive designs arises because sometimes it is difficult to specify δ_1 at the begining of a trial.**


Design - Boundaries Chart for Plan1
Nominal Critical Point (Z-Scale)
5.0 4.0 3.0 2.0 1.0 0.0 -1.0 -2.0 -3.0 -4.0 -5.0
0.0 34.8 69.6 104.4 139.2 174.0
Sample Size
Sample Size
Nominal Critical Point
to Reject H0
±
Superimpose
Boundary Scale
Z-Scale
Actions >>
Close
Design - Boundaries Chart for Plan1
Nominal Critical Point (Z-Scale)
5.0 4.0 3.0 2.0 1.0 0.0 -1.0 -2.0 -3.0
0.0 29.4 58.8 88.3 117.7 147.1
Sample Size
Sample Size
Nominal Critical Point
to Reject H0
to Reject H1
Superimpose
Boundary Scale
Z-Scale
Actions >>
Close
Design - Boundaries Chart for Plan1
Nominal Critical Point (Z-Scale)
5.0 4.0 3.0 2.0 1.0 0.0 -1.0 -2.0 -3.0 -4.0 -5.0
0.0 36.2 72.5 108.7 145.0 181.2
Sample Size
Boundaries Chart - Plan1
Sample Size
Nominal Critical Point
to Reject H0
±
to Reject H1
±
Superimpose
Boundary Scale
Z-Scale
Actions >>
Close
Design - Boundaries Chart for Plan1
Nominal Critical Point (Z-Scale)
2.0 1.5 1.0 0.5 0.0 -0.5 -1.0 -1.5 -2.0 -2.5
0.0 28.7 57.3 86.0 114.6 143.3
Sample Size
Boundaries Chart - Plan1
Sample Size
Nominal Critical Point
to Reject H1
Superimpose
Boundary Scale
Z-Scale
Actions >>
Close

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 Maximum Information

- For normal and binomial endpoints let

$$n_{\max} = \text{the maximum sample size required for the trial}$$

- For time-to-event endpoints let

$$d_{\max} = \text{ maximum number of events required for the trial}$$

- More generally, in terms of Fisher information, let

$$I_{\max} \approx [\text{se}(\hat{\delta}_{\max})]^{-2} = \text{maximum information to be collected}$$

That is, we will keep the trial open until the standard error of $\hat{\delta}$ is sufficiently small that its square inverse equals the desired information

The Information Fraction

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

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

- If K is intended to be the last look, we will often denote $I_{\max}$ by I_K, $n_{\max}$ by n_K and $d_{\max}$ by d_K
- 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_j we compute the Wald statistic

$$Z(t_j) = \frac{\hat{\delta}(t_j)}{\text{se}[\hat{\delta}(t_j)]}$$

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

Unified Distribution Theory

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

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

This general result is due to Scharfstein, Tsiatis and Robins (JASA, 1997), and Jennison and Turnbull (JASA, 1997).

Representation as a Process of Independent Increments

For $j = 1, 2, \ldots K$, define the score statistic

$$W(t_j) = \sqrt{t_j} Z(t_j)$$

1. $\{W(t_1), W(t_2), \ldots W(t_K)\}$ is multivariate normal.

2. $W(t_j) \sim N(\eta t_j, t_j)$, where $\eta = \delta\sqrt{I_{\max}}$.

3. For any $t_{j_1} < t_{j_2}$, cov$\{W(t_{j_1}), W(t_{j_2})\} = t_{j1}$.

This implies that $W(t_{j_1})$ and $W(t_{j_2}) - W(t_{j_1})$ are independent.

Note:

- **The drift, η, of the random variables $Z(t_j)$ and $W(t_j)$ is determined by δ and $I_{\max}$. Under H_0 the drift is zero.**
- **The independent increments structure enables us to compute group sequential stopping boundaries very efficiently by recursive integration instead of by multivariate normal integration. See, for example, Jennison and Turnbull (2000, chapter 19).**

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

$$|Z(t_j)| \geq c_j \ .$$

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

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

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

The α-Spending Function Approach

- 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 $c_1, c_2, \ldots c_K$:

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

$$\alpha(t_1) + P_0\{|Z(t_1)| < c_1, |Z(t_2)| \geq c_2\} = \alpha(t_2) \ ,$$

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

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

Why Type-1 Error is Preserved

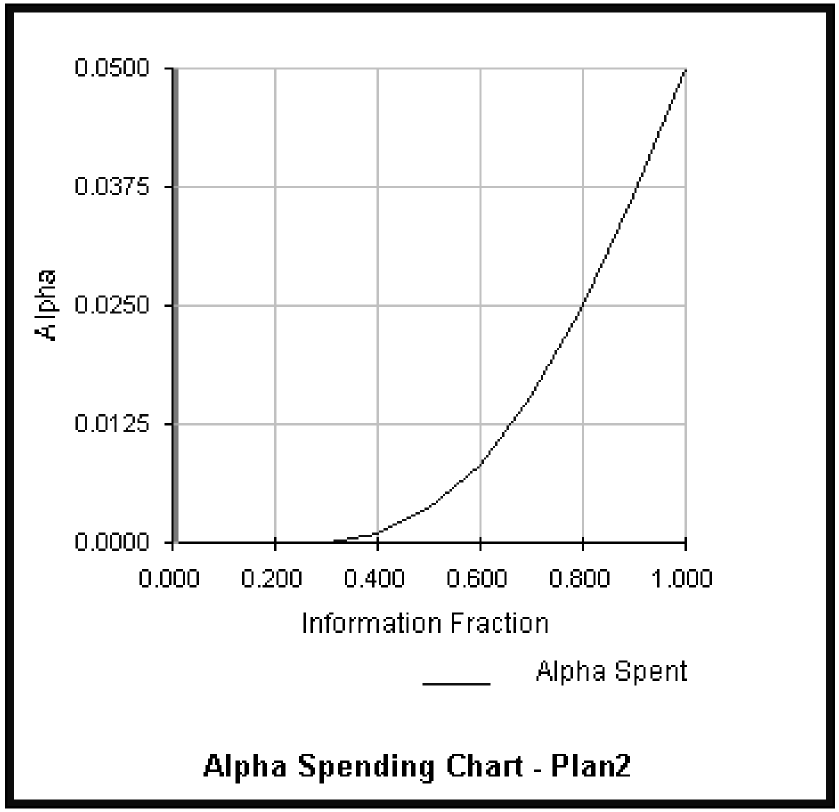
- The probability of ever crossing the boundary under the null hypothesis is

$$P_0\{Z(t_1) \geq c_1\} + P_0\{Z(t_1) < c_1, Z(t_2) \geq c_2\} + \dots$$
$$\dots + P_0\{Z(t_1) < c_1, \dots, Z(t_{K-1}) < c_{K-1}, Z(t_K) \geq c_K\}$$

- This simplifies to

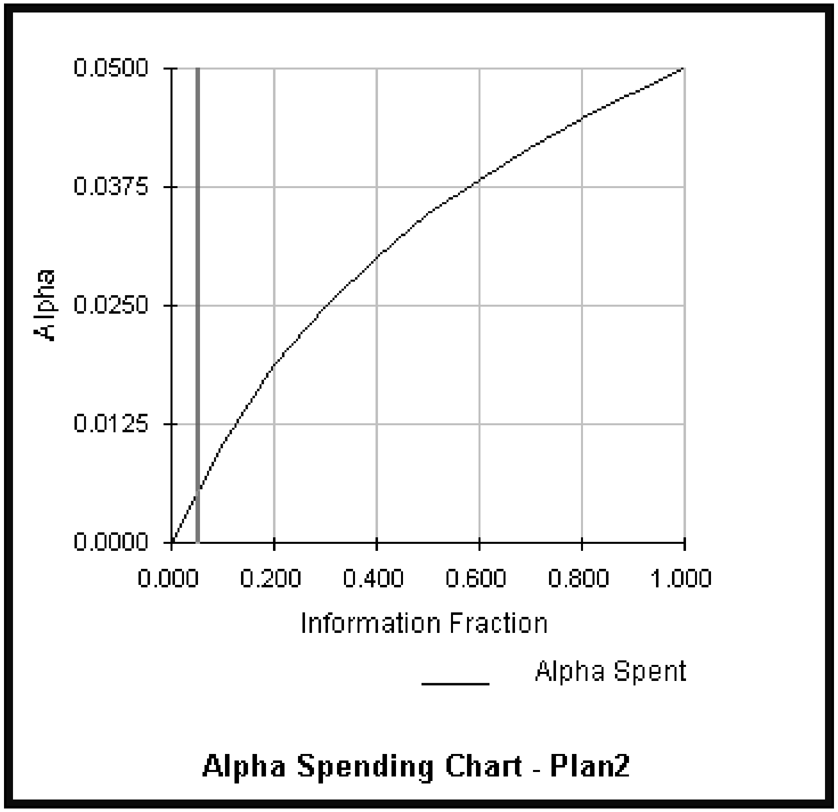
$$\alpha(t_1)+[\alpha(t_2)-\alpha(t_1)]+\dots+[\alpha(t_K)-\alpha(t_{K-1})] = \alpha(t_K) = \alpha$$

Lan-DeMets (OBF) α-Spending Function



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

Lan-DeMets (PK) α-Spending Function



$$\alpha(t) = \alpha \log\{1 + (e-1)t\} \ .$$

Flexibility of the α-Spending Function

- What if we deviate from the number and timing of the looks?
- The probability of crossing the boundary by taking looks at times $t'_1, t'_2, \ldots t'_{K'}$, where is $t'_j \neq t_j$ and $K' \neq K$ is

$$P_0\{Z(t'_1) \geq c'_1\} + P_0\{Z(t'_1) < c'_1, Z(t'_2 \geq c'_2)\} + \ldots + P_0\{Z(t'_1) < c'_1, \ldots, Z(t'_{K'-1}) < c'_{K'-1}, Z(t'_{K'}) \geq c'_{K'}\}$$

- We now read off the $\alpha(.)$ values at these new time points and thus obtain

$$\alpha(t'_1)+[\alpha(t'_2)-\alpha(t'_1)]+\ldots+[\alpha(t'_{K'})-\alpha(t'_{K'-1})] = \alpha(t'_{K'}) \leq \alpha$$

Warning: The distribution theory can be invalidated if $(t'_{j+1}, \ldots t'_{K'})$ or the choice of K' are determined adaptively, based on the observed values of $z_1, \ldots z_j$. However, extensive investigation by Lan and DeMets (1989), and Proschan, Follman and Waclawiw (1992) reveal that the impact on inflating the type-1 error of group sequential trials is slight.

Two Parametric Families of Spending Functions

Gamma Family Hwang IK, Shih WJ and DeCani JS (1990). Statistics in Medicine, 9, 1439-1445.

$$\alpha(t) = \alpha \frac{(1 - e^{-\gamma t})}{(1 - e^{-\gamma})}, \text{ where } \gamma \neq 0$$

Setting γ to -4 or -5 generates boundaries similar to O'Brien-Fleming, while setting γ to 1 generates boundaries similar to Pocock.

Rho Family Jennison and Turnbull (2000). Group Sequential Methods, Chapman and Hall/CRC, New York.

$$\alpha(t) = \alpha t^{\rho}, \text{ where } \rho > 0$$

Setting ρ to 2.5 or 3 generates boundaries similar to O'Brien-Fleming, while setting ρ to 0.75 or 1 generates boundaries similar to Pocock.

Computing the Maximum Information

Suppose the desired power to detect an effect size of δ is $1 - \beta$. The key is to compute the drift parameter that will achieve this power.

- Recall that $Z(t_j) \sim N(\eta\sqrt{t_j}, 1)$, where $\eta = \delta\sqrt{I_{\max}}$.
- Find the value of the drift parameter η that satisfies the equation

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

  Note that η depends on the the stopping boundary as well as β and K.

- **Having obtained η, solve for I_{max} from the relationship $\eta = \delta\sqrt{I_{\mathrm{max}}}$.**

- **Thus, under H_1: $\delta = \delta_1$, $I_{\mathrm{max}} = (\eta/\delta_1)^2$. This is the amount of information we need to achieve $1 - \beta$ power.**

Converting Information into Sample Size

Since $I \approx [\text{se}(\hat{\delta})]^{-2}$, and we require $I_{\max} = (\eta/\delta_1)^2$, we should go on admitting patients until the reciprocal of the variance estimate of $\hat{\delta}$ equals $(\eta/\delta_1)^2$. But what does this mean for sample size?

- For normal endpoints and balanced randomization,

$$[I_{\max}]^{-1} \approx \text{var}(\hat{\delta}_K) = \text{var}(\bar{X}_E - \bar{X}_C) = \frac{2\sigma^2}{(n_{\max}/2)} = \frac{4\sigma^2}{n_{\max}}$$

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

- For binomial endpoints and balanced randomization,

$$n_{\max} = 2[\pi_C(1 - \pi_C) + (\pi_c + \delta_1)(1 - \pi_c - \delta_1)]I_{\max} \; .$$

The Inflation Factor

- For any given α, β, K and spending function, we can compute (and store) the drift parameter η by recursive integration. Given η we can compute the maximum information $I_{\max} = (\eta/\delta)^2$.
- It is convenient to express $I_{\max} = (\eta/\delta_1)^2$ in the form

$$I_{\max} = \left[\frac{z_\alpha + z_\beta}{\delta_1}\right]^2 \times \left[\frac{\eta}{z_\alpha + z_\beta}\right]^2 = \left[\frac{z_\alpha + z_\beta}{\delta}\right]^2 \times \text{IF} ,$$

where

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

The information required for a fixed-sample study is multiplied by an "inflation factor". For example, for a Lan-DeMets (OBF) spending function, $\alpha = 0.05$, $1 - \beta = 0.9$, and $K = 5$, the inflation factor is IF $= 1.03$.

Inflation Factors for Pocock and O'Brien-Fleming Boundaries

$\alpha = 0.05$ (two-sided)					$\alpha = 0.01$ (two-sided)				
	Spending	Power $(1-\beta)$				Spending	Power $(1-\beta)$		
K	Function	0.80	0.90	0.95	K	Function	0.80	0.90	0.95
2	Pocock	1.11	1.10	1.09	2	Pocock	1.09	1.08	1.08
2	O-F	1.01	1.01	1.01	2	O-F	1.00	1.00	1.00
3	Pocock	1.17	1.15	1.14	3	Pocock	1.14	1.12	1.12
3	O-F	1.02	1.02	1.02	3	O-F	1.01	1.01	1.01
4	Pocock	1.20	1.18	1.17	4	Pocock	1.17	1.15	1.14
4	O-F	1.02	1.02	1.02	4	O-F	1.01	1.01	1.01
5	Pocock	1.23	1.21	1.19	5	Pocock	1.19	1.17	1.16
5	O-F	1.03	1.03	1.02	5	O-F	1.02	1.01	1.01

Stopping Early For Futility

- Sometimes we might wish to stop a trial early because the effect size observed at an interim analysis is too small to warrant continuation.
- In East we provide two ways to make this determination:
  - Informal – based on conditional power
  - Formal – based on futility stopping boundaries that are built into the design (see Pampallona and Tsiatis, 1994)
- We shall consider the informal conditional power approach later when we discuss interim monitoring.
- Formal futility boundaries are derived in such a way that the probability of crossing them under H_1 is equal to β.
- In contrast formal efficacy boundaries are derived in such a way that the probability of crossing them under H_0 is equal to α.

The β-Spending Function Approach

(Pampallona, Tsiatis and Kim, 2001)

- Just as we use an α-spending function to generate efficacy boundaries, we can also use a β-spending function to generate futility boundaries
- The probability of crossing the efficacy boundary under H_0 is α
- The probability of crossing the futility boundary under H_1 is β
- We force the two boundaries meet at the last look by appropriate choice of the drift parameter.
- Thereby we ensure that either the null hypothesis is rejected or the alternative hypothesis is rejected by the time the last look is taken.

Simultaneous Computation of Efficacy and Futility Boundaries and Drift Parameter

We may spend both α and β at each look and thereby obtain efficacy and futility boundaries, and the appropriate drift parameter (Pampallona and Tsiatis, 1994).

Step 1 **Fix a value for the drift parameter η.**

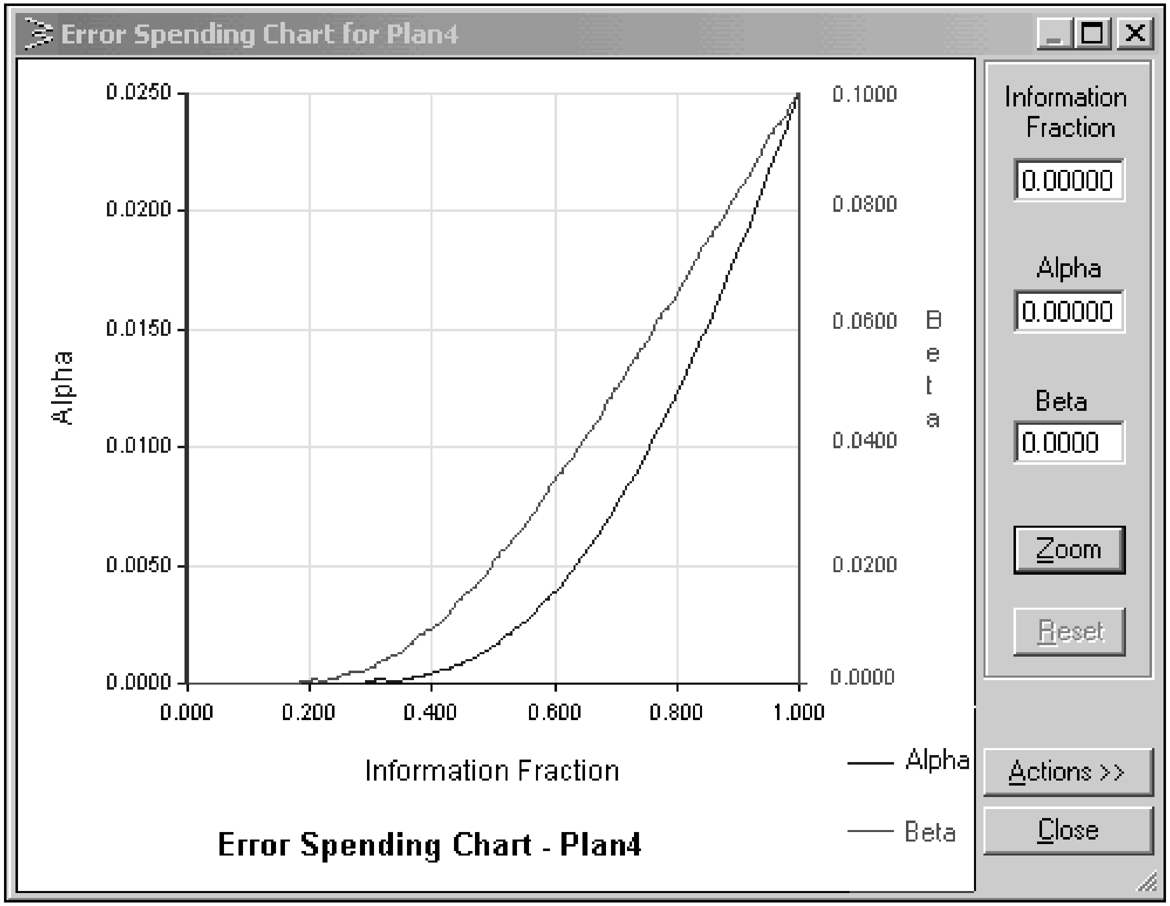
Step 2 **Find (l_1, u_1) such that $P_0\{Z(t_1) \geq u_1\} = \alpha(t_1)$, and $P_\eta\{Z(t_1) \leq l_1\} = \beta(t_1)$.**

Step 3 **Solve recursively for (l_j, u_j) for $j = 2, 3, \ldots K$, such that**

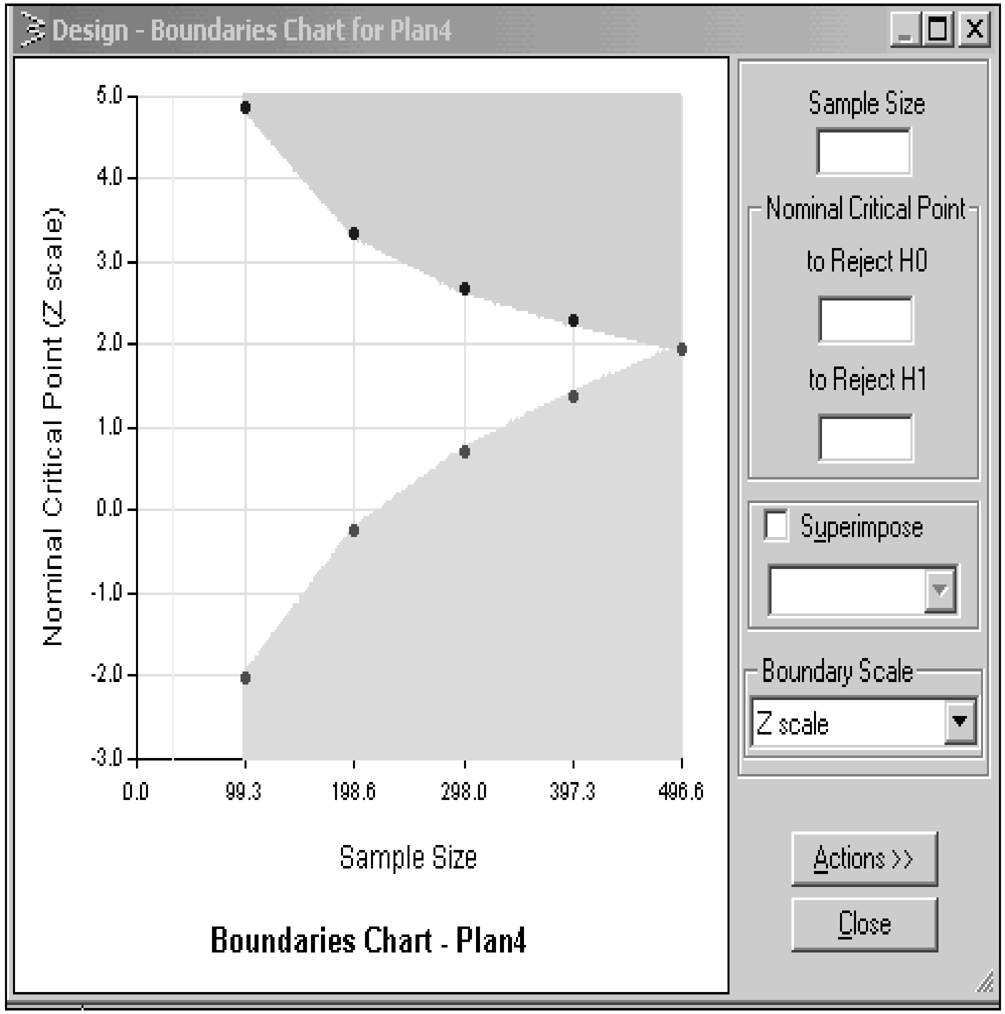
$$\begin{aligned} \alpha(t_{j-1}) + P_0\{l_1 < Z(t_1) < u_1, \cdots, l_{j-1} < Z(t_{j-1}) < u_{j-1}, Z(t_j) \geq u_j\} &= \alpha(t_j) \\ \beta(t_{j-1}) + P_\eta\{l_1 < Z(t_1) < u_1, \cdots, l_{j-1} < Z(t_{j-1}) < u_{j-1}, Z(t_j) \leq l_j\} &= \beta(t_j) \end{aligned}$$

Step 4 **Repeat Steps 1 through 3 with different values of η until $l_K = u_K$.**

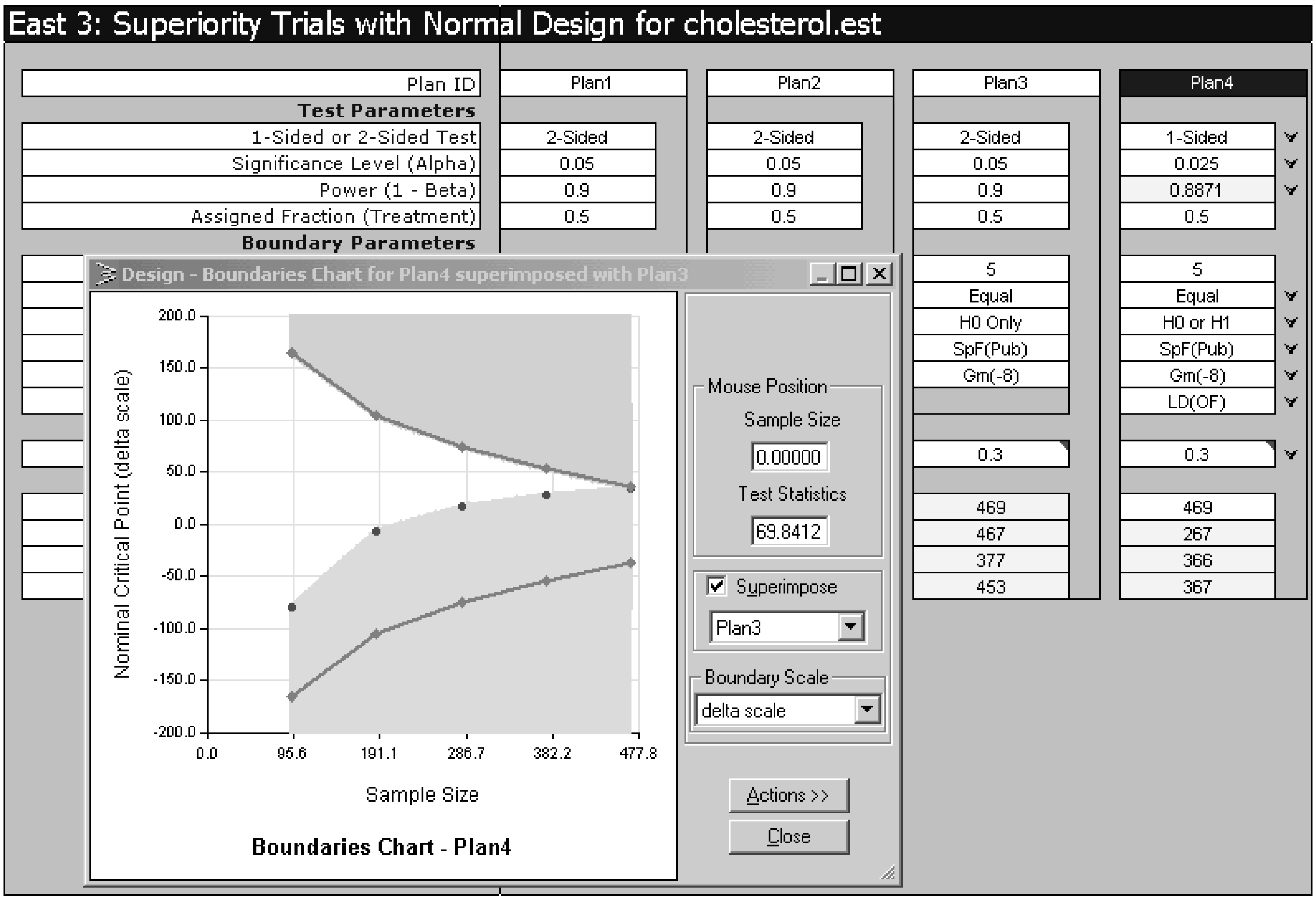
Lan-DeMets α and β Spending Functions



Corresponding Efficacy and Futility Boundaries



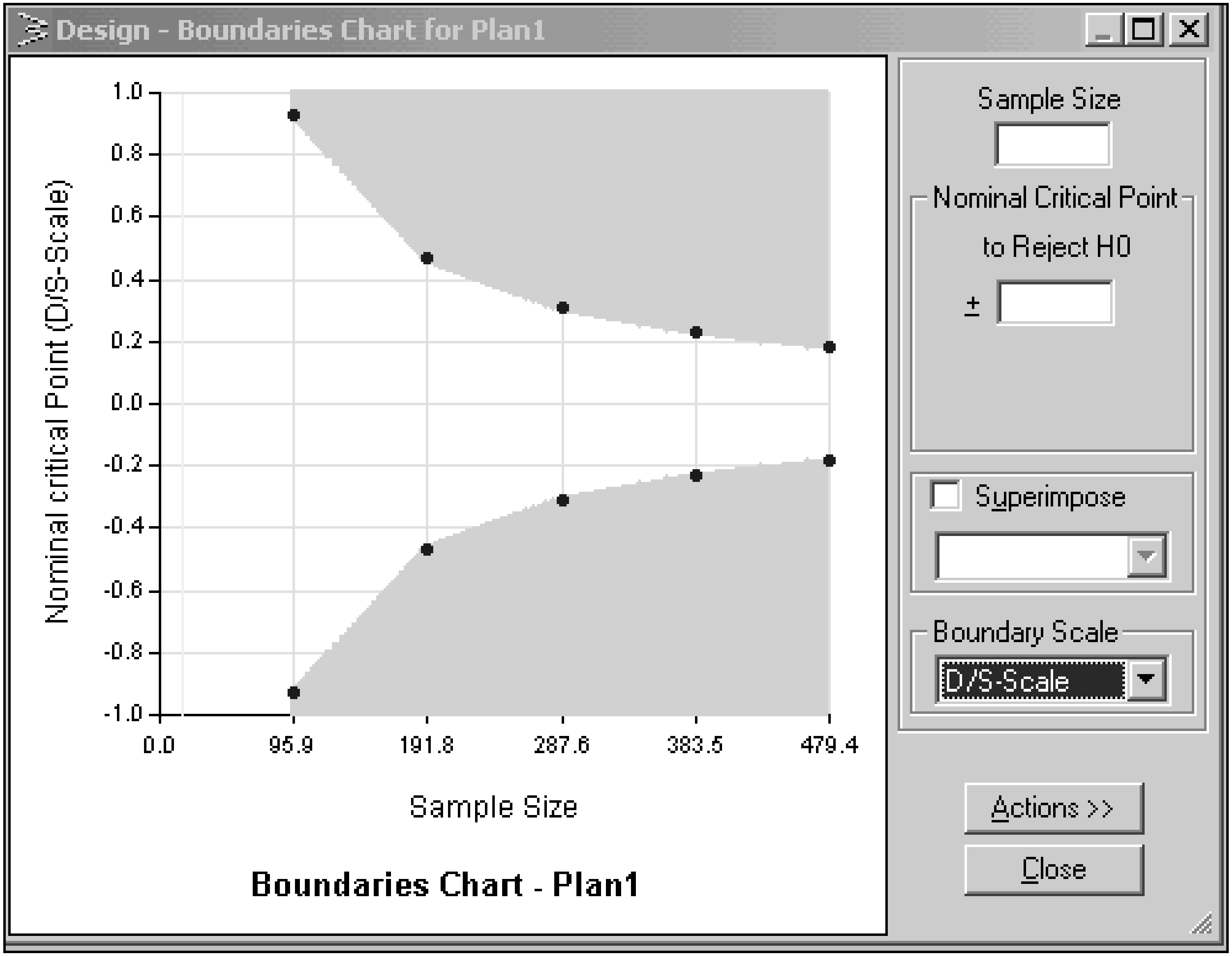
Cholesterol Study with Futility Boundaries



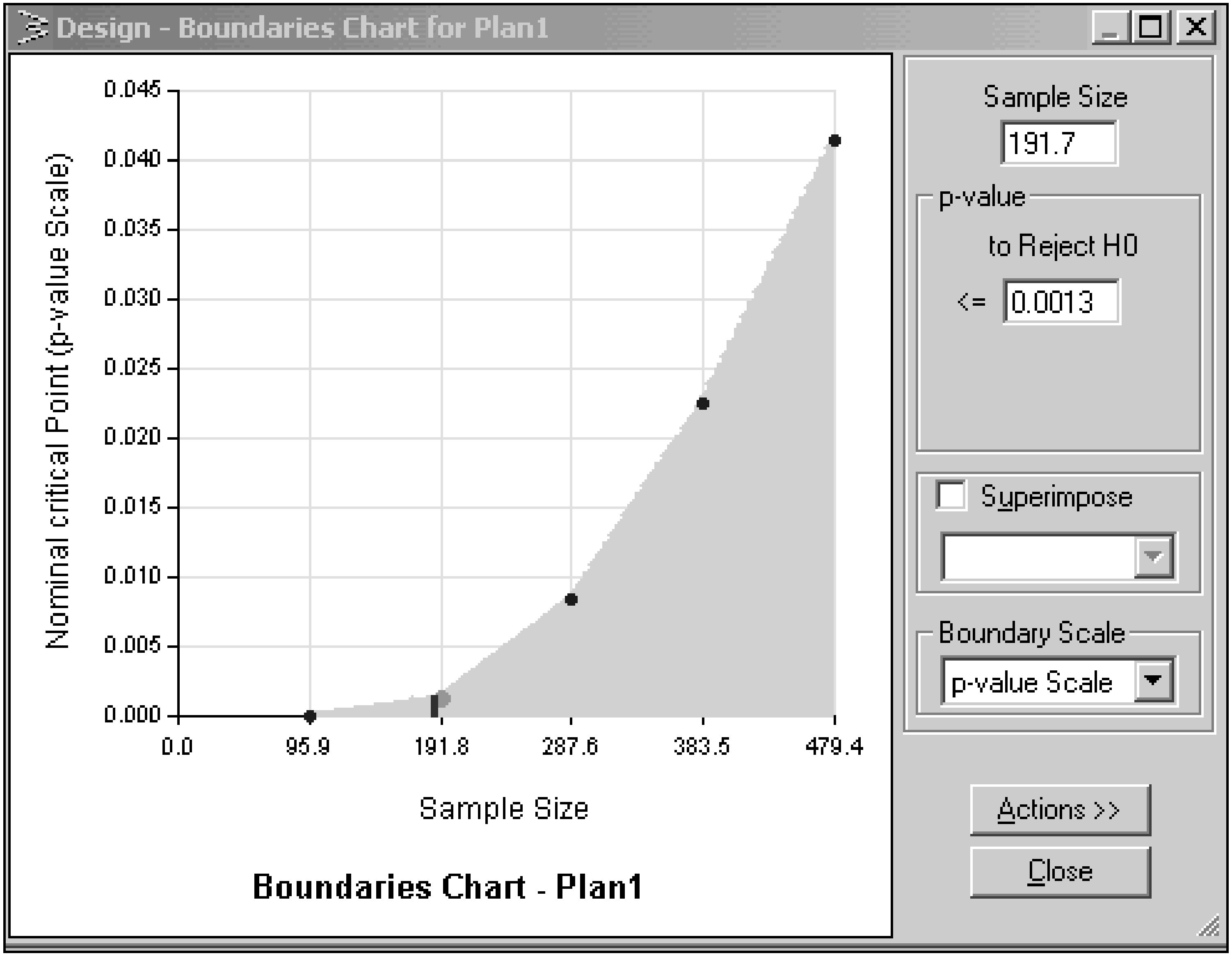
Scales for Displaying Stopping Boundaries

In EaSt-3 we display the stopping boundaries on five scales, all transformations of the Z-scale.

1. **The Z-Scale: Reject if $Z(t_j)$ crosses its stopping boundary.**
2. **The P-value Scale: Reject if $\Phi(Z(t_j))$ crosses its stopping boundary.**
3. **The δ Scale: Reject if $\hat{\hat{\delta}}$ crosses its stopping boundary.**
4. **The δ/σ Scale: Reject if $\hat{\hat{\delta}}/\hat{\sigma}$ crosses its stopping boundary.**
5. **The CP Scale: Reject if the conditional power at the current value of the test statistic is less than the corresponding boundary value.**



Design - Boundaries Chart for Plan1
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Nominal critical Point (D/S-Scale)
0.0
95.9
191.8
287.6
383.5
479.4
Sample Size
Boundaries Chart - Plan1
Sample Size
Nominal Critical Point
to Reject H0
±
Superimpose
Boundary Scale
D/S-Scale
Actions >>
Close



Design - Boundaries Chart for Plan1
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Nominal critical Point (p-value Scale)
0.0
95.9
191.8
287.6
383.5
479.4
Sample Size
Boundaries Chart - Plan1
Sample Size
191.7
p-value
to Reject H0
<=
0.0013
Superimpose
Boundary Scale
p-value Scale
Actions >>
Close

The Role of Simulation

- Useful for verifying that the distribution theory holds – especially for small samples
- Useful for communicating the properties of the design to non-statisticians on the team
- Useful for studying the properties of information based and adaptive designs

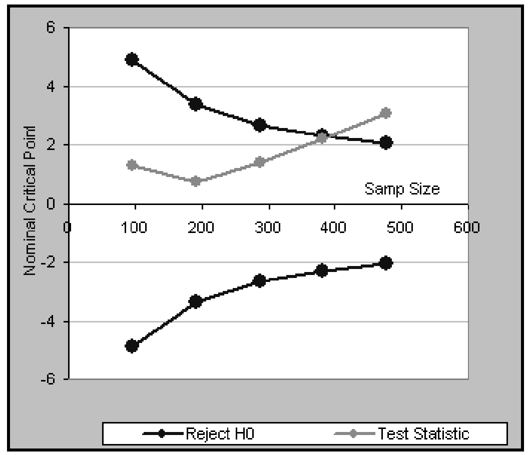
East 3: Simulation of Superiority Sequential Trials (Enhanced Simulation)

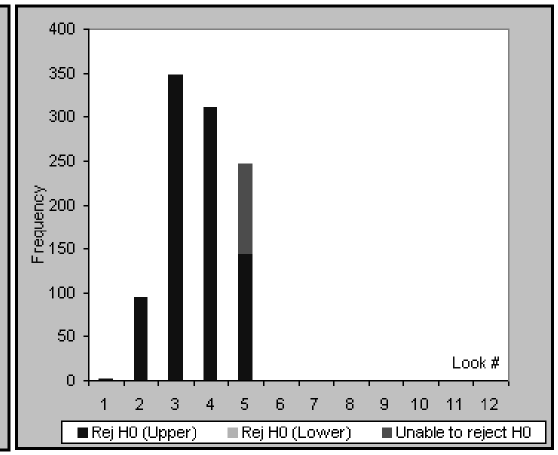
Design Parameters for Plan1	
1-Sided or 2-Sided Test	2-Sided
Significance Level(Alpha)	0.05
Power(1-Beta)	0.9
End Point	Normal
Standardized Difference	0.3
Difference in Two Means	0.6
Standard Deviation	2
Assigned Fraction	0.5
Planned # of Looks	5
Spacing of Analysis	Equal
Hypothesis to be rejected	H0 Only
Boundary Family	SpF(Pub)
Boundary to Reject H0	LD(OF)
Boundary to Reject H1	
Design Outputs	
Max. Sample Size (Nmax)	478
Max. Information (Imax)	29.8609
Simulation Parameters	
Simulation Scale	Sample Size
Difference in Two means	0.6
Standard Deviation	2
Number of Trials	1000
Refresh Every 'n' Trials, n =	100
Simulation Starting Seed	Clock

Run | Single Step | Reset | Stop | Help

Simulation Seed = 37787
Elapsed Time = 0:00:08

Simulation Boundary					Overall Simulation Results						
Look	Boundary			Simulated	Avg.	Avg.Sample	# Rejecting H0		# Unable to	Total Simulations	
#	Samp Size	H0-	H0+	Test Stat	Information	Size	Upper	Lower	reject H0	Count	%
1	95.55	-4.8771	4.8771	1.2887	6.68	96.00	1			1	0.10%
2	191.11	-3.3570	3.3570	0.7238	12.39	191.00	95			95	9.50%
3	286.66	-2.6803	2.6803	1.3995	18.26	287.00	348			348	34.80%
4	382.22	-2.2898	2.2898	2.2113	24.17	382.00	310			310	31.00%
5	477.77	-2.0310	2.0310	3.0885	29.68	478.00	143		103	246	24.60%
6											
7											
8											
9											
10											
11											
12											
				Total	22.33	354.13	897		103	1000	
				%			89.70%		10.30%		





(1) Simulate trials with large σ

(2) Simulate trials with small δ

(3) Simulate information based and adaptive methods to recover power without compromising the α

Part II

Interim Monitoring

Interim Monitoring of Cholesterol Study

We will monitor Plan 2 with 5 looks (N=478). Suppose we observe the following data at the first 3 interim monitoring time points.

Look (j)	N_j	t_j	$\hat{\delta}(t_j)$	se$[\hat{\delta}(t_j)]$
1	250	0.523	30	19.688
2	350	0.733	28	20.125
3	400	0.837	27.5	20.583

Neither the number nor spacing of the interim looks matches with the values proposed at the design stage. Monitor the trial on an interim monitoring worksheet.

The Interim Monitoring Worksheet

The interim monitoring worksheet allows you to:

- **Re-compute the boundaries to accomodate changes in the number and spacing of the interim analyses, including unplanned interim analyses**
- **Track the type-1 error spent at each look and ensure that it never exceeds α**
- **Compute conditional power at each look**
- **Compute repeated confidence intervals (Jennison and Turnbull, 1989) at each look**
- **Compute post-hoc power at each look**

Re-computing the Boundaries

Look 1 $t_1 = 0.523$ and $\alpha(t_1) = 0.004$. Find c_1 such that

$$P_0(|Z_1| \geq c_1) = 0.004 \ .$$

The re-computed boundary is $c_1 = \pm 2.887$.

Look 2 $t_2 = 0.733$ and $\alpha(t_2) = 0.018$. Find c_2 such that

$$0.004 + P_0(|Z_1| < 2.887, |Z_2| \geq c_2) = 0.018 \ .$$

The re-computed boundary is $c_2 = \pm 2.399$.

Look 3 $t_2 = 0.837$ and $\alpha(t_3) = 0.029$. Find c_3 such that

$$0.018 + P_0(|Z_1| < 2.887, |Z_2| < 2.399, |Z_3| \geq c_3) = 0.029 \ .$$

The re-computed boundary is $c_2 = \pm 2.27$.

Repeated Confidence Intervals

- Let δ be the unknown parameter. For $k = 1, 2, \ldots K$, let $Z_k = \hat{\delta}_k \sqrt{I_k} \equiv \hat{\delta}_k / \text{se}(\hat{\delta}_k)$ be the Wald statistic and let c_k be the corresponding stopping boundary. Then

$$P_0\left(\bigcap_{k=1}^{K} |Z_k| < c_k\right) = 1 - \alpha$$

- Now suppose that $\delta = \delta_0 \neq 0$. Then, for $k = 1, 2, \ldots K$, $Z_k - \delta_0 \sqrt{I_k}$ has a $N(0, 1)$ distribution and the same covariance structure as Z_k. Therefore, for any value of δ_0,

$$P_{\delta_0}\left(\bigcap_{k=1}^{K} |Z_k - \delta_0 \sqrt{I_k}| < c_k\right) = 1 - \alpha$$

- The Pivot Argument: The event $\{|Z_k - \delta_0 \sqrt{I_k}| < c_k\}$ occurs if and only if $\delta_0 \in (\hat{\delta}_k - c_k/\sqrt{I_k}, \hat{\delta}_k + c_k/\sqrt{I_k})$

- We have shown that the sequence of confidence intervals $(\hat{\delta}_k \pm c_k/\sqrt{I_k})$ for $k = 1, 2, \ldots K$ satisfies the property

  $$P_\delta(\hat{\delta}_k - c_k/\sqrt{I_k} < \delta < \hat{\delta}_k + c_k/\sqrt{I_k} \text{ for all } k = 1, 2, \ldots K) = 1 - \alpha$$

  i.e., all K intervals simultaneously cover δ with probability $1 - \alpha$.

- Therefore each individual interval of the form $(\hat{\delta}_k \pm c_k/\sqrt{I_k})$ also covers the unknown δ with confidence at least equal to $1 - \alpha$.

- This is a useful property for making valid inferences about δ if the study continues after a boundary has been crossed. For instance suppose the DMC continues to leave the study open and take further interim looks even though the boundary has been crossed.

- These RCI's are slightly conservative relative to the multiplicity-adjusted single confidence interval computed at the end of the study.

Computation of RCI's

The interval at look j is of the form $\hat{\delta}_j \pm c_j \times \text{se}(\hat{\delta}_j)$.

Look 1 $c_1 = 2.887$, $\hat{\delta}_1 = 30$, $\text{se}(\hat{\delta}_1) = 19.668$.
RCI = (-26.782, 86.782)

Look 2 $c_2 = 2.399$, $\hat{\delta}_2 = 28$, $\text{se}(\hat{\delta}_2) = 20.125$.
RCI = (-20.285, 76.285)

Look 3 $c_1 = 2.270$, $\hat{\delta}_3 = 27.5$, $\text{se}(\hat{\delta}_3) = 20.583$.
RCI = (-19.232, 74.232)

Conditional Power

- **Conditional power is the probability of declaring statistical significance if we make the next analysis the final one, conditional on the current value of the test statistic.**
- **Accurate computation of conditional power needs some care, however, because it is being done within the confines of a group sequential trial. Thus we must adjust the final stopping boundary and final sample size to compensate for the type-1 error being spent at each interim look.**
- **Suppose we have completed j looks, at information fractions $t_1, t_2, \ldots t_j$, and have not so far crossed a boundary. Let the next look be the last one. Suppose it is taken at information fraction t_L^* and the corresponding boundary for declaring statistical significance is c_L^*.**
- **We must choose t_L^* and c_L^* in such a way that the a priori unconditional type-1 error of of the trial is α and the a priori unconditional power (also called post-hoc power in East) is $1 - \beta$.**

(A) Ideal Last-Look Position and Final Boundary

The values of t_L^* and c_L^* are obtained by the following iterative three-step procedure.

1. **Select a value for t_L^* and find the corresponding last-look stopping boundary c_L^* by solving**

$$\alpha(t_j)+P_0\{|Z(t_1)| < c_1, |Z(t_2)| < c_2, \cdots |Z(t_j)| < c_j, |Z(t_L^*)| \geq c_L^*\} = \alpha$$

2. **Compute**

$$P_\eta\{|Z(t_1)| < c_1, |Z(t_2)| < c_2, \cdots |Z(t_j)| < c_j, |Z(t_L^*)| < c_L^*\} = \beta^*$$

3. **Iterate between step 1 and step 2 until $\beta^* = \beta$.**

(B) Computing the Conditional Power

- Suppose that the observed value of the test statistic at time t_j is $z(t_j)$. Then the conditional power for any drift parameter η is given by

$$\text{CP} = P_\eta\{|Z(t_L^*)| \geq c_L^* | z(t_j)|\} \ .$$

- Using the result that $W(t) = \sqrt{t}Z(t)$ is a Brownian process with independent increments the conditional power can be simplified to

$$\text{CP} = 1 - \Phi\left(\frac{u_L^* - w(t_j) - \eta(t_L^* - t_j)}{\sqrt{t_L^* - t_j}}\right) + \Phi\left(\frac{-u_L^* - w(t_j) - \eta(t_L^* - t_j)}{\sqrt{t_L^* - t_j}}\right)$$

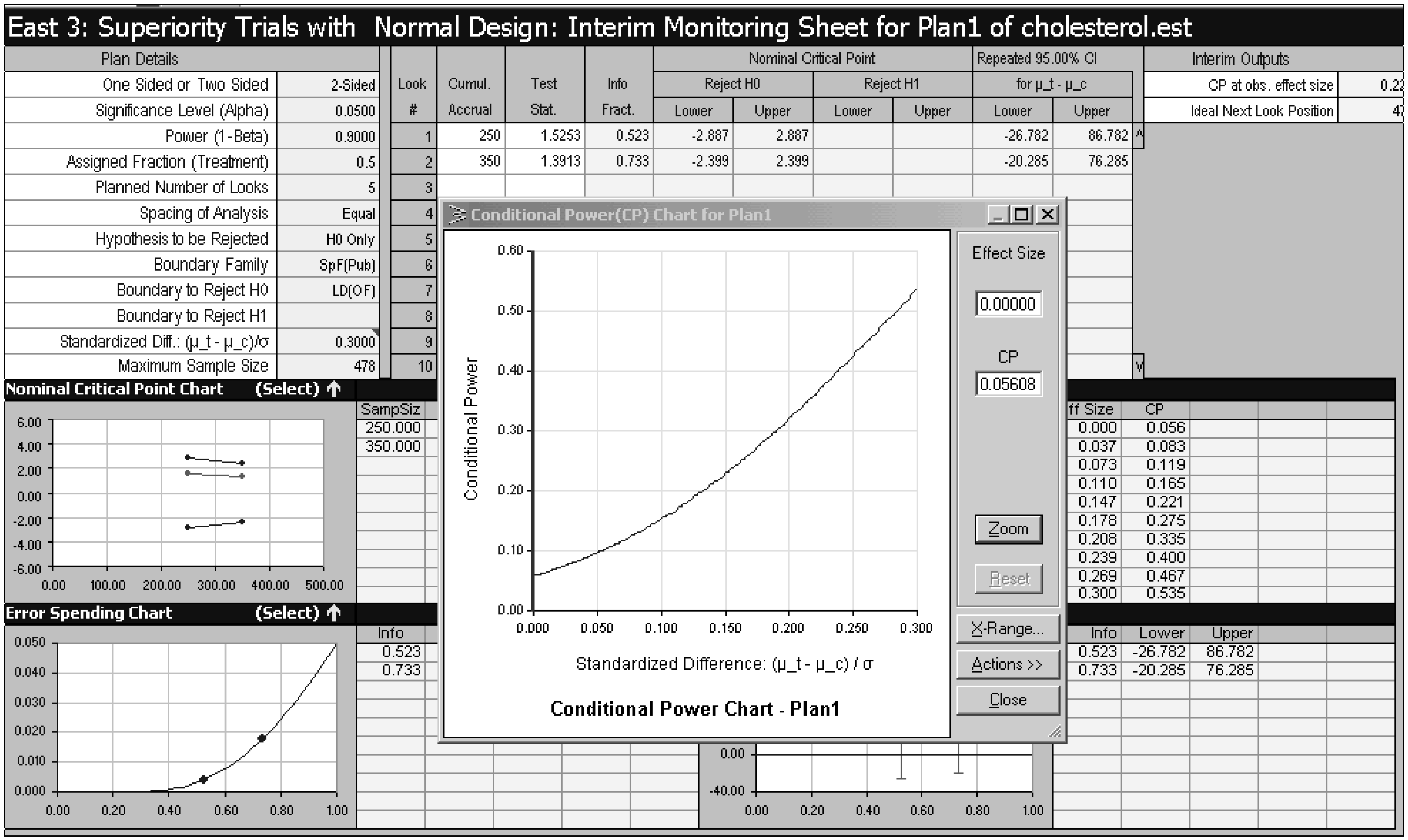
where $u_L^* = \sqrt{t_L^*}c_L^*$ and $w(t_j) = \sqrt{t_j}z(t_j)$.

(C) Cholesterol Example: Conditional Power at the End of Look 2

- The first look was taken at $N_1 = 250$ and the second look at $N_2 = 350$.
- East shows that at the end of look 2, the ideal next look position is $N_3 = 474$. This means that full 90% unconditional power would be achieved if the study were designed for 3 looks, at sample sizes of 250, 350 and 474, respectively.
- The final stopping boundary can be obtained by entering the third look at $N_3 = 474$ into the East interim monitoring worksheet. It is seen to be 2.009.
- At the second look, $z_2 = 1.3913$. Thus the conditional power is

$$\text{CP} = P_{\delta/\sigma}(Z_3 \geq 2.009 | z_2 = 1.3913)$$

- The conditional power can be plotted against different values of δ/σ. At look 2 the estimate is $\hat{\delta}/\hat{\sigma} = 0.127$ so that the corresponding CP is 0.224.


East 3: Superiority Trials with Normal Design: Interim Monitoring Sheet for Plan1 of cholesterol.est
Conditional Power(CP) Chart for Plan1
Conditional Power
Standardized Difference: (μ_t - μ_c) / σ
Conditional Power Chart - Plan1
Effect Size
0.00000
CP
0.05608
Zoom
Reset
X-Range...
Actions >>
Close
Nominal Critical Point Chart
Error Spending Chart

Adjusted Inference at the Last Look

Take a fourth look at the data, at $N = 480$. Now the maximum information needed to attain 90% power has been exceeded. Therefore East will spend all the remaining α and make this the last look. Suppose $\hat{\delta} = 35$ and $\mathrm{se}(\hat{\delta}) = 23.41$.

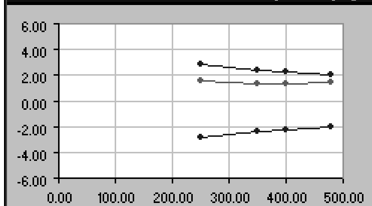
East 3: Superiority Trials with Normal Design: Interim Monitoring Sheet for Plan1 of cholesterol.est

Plan Details	
One Sided or Two Sided	2-Sided
Significance Level (Alpha)	0.0500
Power (1-Beta)	0.9000
Assigned Fraction (Treatment)	0.5
Planned Number of Looks	5
Spacing of Analysis	Equal
Hypothesis to be Rejected	H0 Only
Boundary Family	SpF(Pub)
Boundary to Reject H0	LD(OF)
Boundary to Reject H1	
Standardized Diff.: (μ_t - μ_c)/σ	0.3000
Maximum Sample Size	478

Look #	Cumul. Accrual	Test Stat.	Info Fract.	Nominal Critical Point: Reject H0 Lower	Reject H0 Upper	Reject H1 Lower	Reject H1 Upper	Repeated 95.00% CI for μ_t - μ_c: Lower	Upper
1	250	1.5253	0.523	-2.887	2.887			-26.782	86.782
2	350	1.3913	0.733	-2.399	2.399			-20.285	76.285
3	400	1.3361	0.837	-2.270	2.270			-19.232	74.232
4	480	1.495	1.005	-2.050	2.050			-11.098	70.898
5									
6									
7									
8									
9									
10									

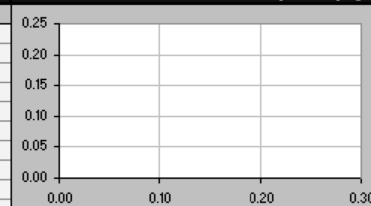
Final Outputs at Look #4	
Adjusted p-value	0.137
Adj. Pt. Est. for μ_t - μ_c	27.198
Adj. 95.00% CI for μ_t - μ_c	
Lower Confidence Bound	-8.677
Upper Confidence Bound	63.083
Post-Hoc Power	0.900

Nominal Critical Point Chart (Select)



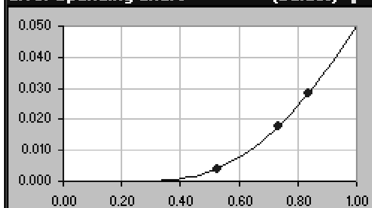
SampSiz	H0-	H0+	H1-	H1+
250.000	-2.887	2.887		
350.000	-2.399	2.399		
400.000	-2.270	2.270		
480.000	-2.050	2.050		

Conditional Power Chart (Select)



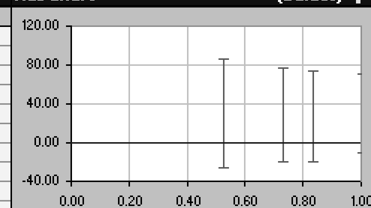
Eff Size	CP

Error Spending Chart (Select)



Info	Alpha	Beta
0.523	0.004	
0.733	0.018	
0.837	0.029	
1.005	0.050	

RCI Chart (Select)



Info	Lower	Upper
0.523	-26.782	86.782
0.733	-20.285	76.285
0.837	-19.232	74.232
1.005	-11.098	70.898

Stage-Wise Ordering of Sample Space

(Tsiatis, Rosner and Mehta, 1984; Kim and DeMets, 1987)

- East displays the adjusted p-value (0.137), point estimate(27.198 and 95% confidence interval (-8.677, 63.083) for δ.
- To obtain these estimates we order the sample space in a stage-wise manner. There are two simple rules for the ordering.
  1. At the last look, K, values of $|z|$ that are larger than $|Z_K|$ are more extreme.
  2. The earlier a boundary is crossed, the more extreme is the outcome.
- Therefore the rejection region for this example is

$$\mathcal{E}^* = \{|Z_4| \geq 1.495\} \cup \{|Z_3| \geq 2.27\} \cup \{|Z_2| \geq 2.399\} \cup \{|Z_1| \geq 2.887\}$$

- **The adjusted p-value is**

$$p^* = P_0\{\mathcal{E}^*\}.$$

- **Let η be the drift parameter. The $100 \times (1-2\nu)$ confidence interval for η is (η^L, η^U) where**

$$\begin{aligned}\eta^U &= \sup\{\eta : P_\eta\{\mathcal{E}^*\} \le 1-\nu\} \ , \\ \eta^L &= \inf\{\eta : P_\eta\{\mathcal{E}^*\} \ge \nu\} \ .\end{aligned}$$

The relationship that links η to the effect size, δ, is $\eta = \delta\sqrt{I_{\max}}$.

- **The median unbiased estimate (MUE) for η, is the value of $\tilde{\eta}$ that satisfies**

$$P_{\tilde{\eta}}\{\mathcal{E}^*\} = 0.5 \ .$$

Part III

Sample Size Re-estimation

Designs that Permit Sample Size Re-Estimation

- We consider two types of designs that permit sample size re-estimation based on interim results:
  - Information based designs
  - Adaptive designs
- Information based designs make adjustments to the sample size using revised estimates of nuisance parameters (like variance, control response rate, or covariate effects) obtained from the interim data
- Adaptive designs make adjustments to the sample size based on a revised assessment of the primary effect size parameter. The revised assessment could be based on external data or internal results available at the interim analysis

Information Based Designs

Example 1: Binomial Endpoint Trial.

Consider the PRISM trial (Steve Snappin, DIA Workshop, Sept 12, 2003) sponsored by Merck

- **Patients with acute coronary syndrome randomized to Heparin alone (control arm), Tirofiban alone (monotherapy arm), or Heparin + Tirofiban (Combination therapy arm)**
- **Composite endpoint (refractory ischemia, MI or death) evaluated 7 days from randomization**
- **Perform two 2-sided tests (control versus monotherapy; control versus combo therapy) each at $\alpha = 0.025$**
- **Design for 80% power to detect a 25% drop in the event rate of the treatment arm relative to the control arm**
- **Take one interim look half way through the trial**

Event Rate for Heparin is Uncertain

- The Heparin (control) arm is expected to have an event rate of 30%. But this is just an initial estimate based on early studies. The sample size is sensitive to this estimate.

- Let π_c be the event rate for the control arm and π_e be the response rate for the experimental arm. We require 80% power to detect a 25% drop in the event rate. In other words,we would like to detect an improvment of

$$\pi_e = 0.75\pi_c$$

  with 80% probability

- Possible values of π_c and π_e under the alternative hypothesis are:

π_c	$(\pi_e = 0.75 \times \pi_c)$
0.30	0.225
0.20	0.15
0.10	0.075

Sensitivity of Sample Size to π_c

East 3: Superiority Trials with Binomial Design for EastBook1

Plan ID	Plan1	Plan2	Plan3
Test Parameters			
1-Sided or 2-Sided Test	2-Sided	2-Sided	2-Sided
Significance Level (Alpha)	0.025	0.025	0.025
Power (1 - Beta)	0.8	0.8	0.8
Assigned Fraction (Treatment)	0.5	0.5	0.5
Boundary Parameters			
Planned Number of Looks	2	2	2
Spacing of Analysis	Equal	Equal	Equal
Hypothesis to be Rejected	H0 Only	H0 Only	H0 Only
Boundary Family	SpF(Pub)	SpF(Pub)	SpF(Pub)
Boundary to Reject H0	Gm(-8)	Gm(-8)	Gm(-8)
Boundary to Reject H1			
Binomial parameters under H1			
Proportion Response (Control: π_c)	0.3	0.2	0.1
Proportion Response (Treatment: π_t)	0.225	0.15	0.075
Accrual (Subjects)			
Maximum	1300	2188	4851
Expected Under H0	1300	2187	4850
Expected Under H1	1240	2087	4628
Expected Under H1/2	1295	2179	4832

Maximum Information vs Maximum Sample Size

How can we preserve the desired power in the presence of unknown nuisance parameters?

- **We first figure out how much information we need to achieve the desired power at the alternative hypothesis $\delta = \delta_1$**

$$I_{\max} = \left[\frac{z_\alpha + z_\beta}{\delta_1}\right]^2 \times \left[\frac{\eta}{z_\alpha + z_\beta}\right]^2 = \left[\frac{z_\alpha + z_\beta}{\delta_1}\right]^2 \times \mathsf{IF}(\alpha, \beta, \Delta, K)$$

- **Once we know how much information is required, we can use**

$$I_{\max} \approx \frac{1}{\mathsf{var}(\hat{\delta}_K)}$$

  to see how small the var$(\hat{\delta}_K)$ has to be so as to obtain the desired amount of information.

- **We then recruit just the right number of patients that will make the reciprocal of the variance estimate of $\hat{\delta}$ sufficiently small.**

Binomial Endpoints

The sample size will depend on how the effect size δ is defined

1. **Difference of Proportions:** $\delta = \pi_e - \pi_c$

$$\begin{aligned} \hat{\delta} &= \hat{\pi}_e - \hat{\pi}_c \\ \text{var}(\hat{\delta}_K) &= \frac{\hat{\pi}_e(1-\hat{\pi}_e)}{n_{\max}/2} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_{\max}/2} = [I_{\max}]^{-1} \\ n_{\max} &= 2[\hat{\pi}_e(1-\hat{\pi}_e) + \hat{\pi}_c(1-\hat{\pi}_c)]I_{\max} \end{aligned}$$

As an initial estimate, at the design stage, we would use

$$n_{\max} = 2[(\pi_c + \delta_1)(1 - \pi_c - \delta_1) + (\pi_c)(1 - \pi_c)]I_{\max}$$

2. Log Ratio of Proportions: $\delta = \ln(\pi_e/\pi_c)$

$$\hat{\delta} = \ln\left(\frac{\hat{\pi}_e}{\hat{\pi}_c}\right)$$

$$\text{var}(\hat{\delta}_K) = \frac{\hat{\pi}_c^{-1}[1 + \exp(-\hat{\delta}_K) - 2\hat{\pi}_c]}{n_{\max}/2} = [I_{\max}]^{-1}$$

$$n_{\max} = 2\hat{\pi}_c^{-1}[1 + \exp(-\hat{\delta}_K) - 2\hat{\pi}_c]I_{\max}$$

As an initial estimate, at the design stage, we would use

$$n_{\max} = 2\pi_c^{-1}[1 + \exp(-\delta_1) - 2\pi_c]I_{\max}$$

(where $\delta_1 = \ln(0.75) = -0.29$ for the PRISM trial)

3. Log Odds Ratio: $\delta = \ln\{[(\pi_e)(1-\pi_c)]/[(\pi_c)(1-\pi_e)]\}$

$$\hat{\delta} = \ln\{\frac{\hat{\pi}_e(1-\hat{\pi}_c)}{\hat{\pi}_c(1-\hat{\pi}_e)}\}$$

$$\text{var}(\hat{\delta}_K) = \frac{2[\hat{\pi}_c^{-1}(1-\hat{\pi}_c)^{-1} + \hat{\pi}_e^{-1}(1-\hat{\pi}_e)^{-1}]}{n_{\max}/2} = [I_{\max}]^{-1}$$

$$n_{\max} = 2[\hat{\pi}_c^{-1}(1-\hat{\pi}_c)^{-1} + \hat{\pi}_e^{-1}(1-\hat{\pi}_e)^{-1}]I_{\max}$$

As an initial estimate, at the design stage, we would use

$$n_{\max} = 2[\pi_c^{-1}(1-\pi_c)^{-1} + (\pi_c+\delta_1)^{-1}(1-\pi_c-\delta_1)^{-1}]I_{\max}$$

Normal Endpoints

Difference of Means: $\delta = \mu_e - \mu_c$

$$\hat{\delta} = \bar{X}_e - \bar{X}_c$$

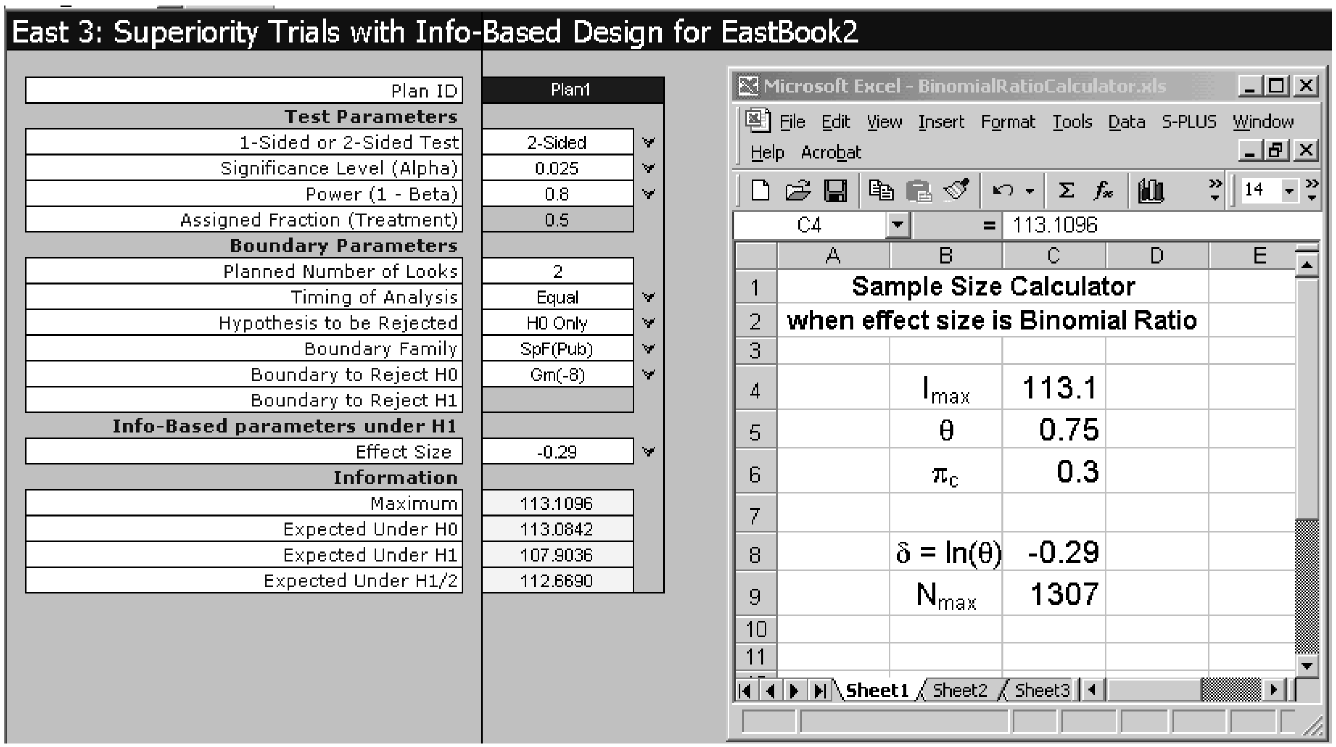
$$\text{var}(\hat{\delta}_K) = \frac{\hat{\sigma}^2}{n_{\max}/2} + \frac{\hat{\sigma}^2}{n_{\max}/2} = [I_{\max}]^{-1}$$

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

As an initial estimate, at the design stage, we would use

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

Keep $I_{\max}$ fixed. Be flexible about $n_{\max}$



Sample Size Requirements for PRISM

Heparin Rate	Maximum Information	Maximum Sample Size
30%	113.1	1307
25%	113.1	1659
20%	113.1	2187
15%	113.1	3067
10%	113.1	4826

Interim Analysis of PRISM

- **The observed rates were $\hat{\pi}_c = 59/351$ (16.8%) for Heparin and $\hat{\pi}_e = 39/336$ (11.6%) for Heparin plus Tirofiban; $\hat{\delta} = -0.37$**

- **Current information is**

$$I = [\text{var}(\hat{\delta})]^{-1} = \frac{\hat{\pi}_c^{-1}[1 + \exp(-\hat{\delta}) - 2\hat{\pi}_c]}{(n_c + n_e)/2} = 27.3$$

- **Current information fraction is $t = 27.3/113.1 = 0.24$.**

- **Current test statistic is $Z = \hat{\delta}\sqrt{I} = -1.93$.**

The stopping boundary is -4.058, so the trial continues. The new estimate of sample size is obtained by solving:

$$\boldsymbol{N/N_{\max} = I/I_{\max}}$$

so that

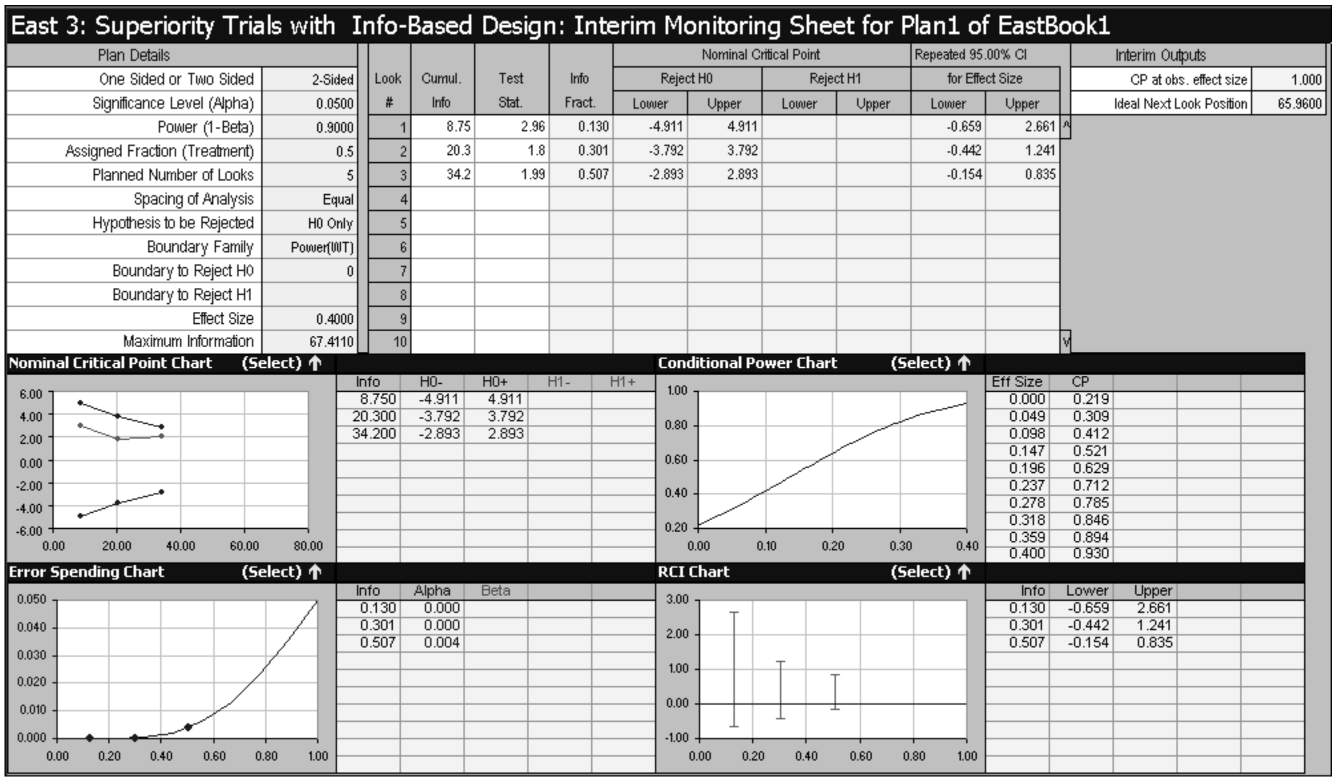
$$\boldsymbol{N_{\max} = N \times I_{\max}/I = 687 \times 113.1/27.3 = 2844}$$

Example 2: Normal Endpoint Trial

Normal data; $\delta = 0.4$; 5-looks; two-sided O'Brien-Fleming; $\alpha = 0.05$ and 90% power. If we design the above study for maximum information rather than maximum sample size then: $I_{\max} = 67.4$; initial estimate of $\sigma^2 = 1$; initial estimate of $n_{\max} = 270$.

Look	n	s_p^2	$\hat{\delta}$	var($\hat{\delta}$)	I	Z	$\frac{n}{n_{\max}}$	$\frac{I}{I_{\max}}$
1	56	1.6	1.0	0.114	8.75	2.96	0.21	0.13
2	112	1.38	0.4	0.049	20.3	1.8	0.41	0.30
3	200	1.46	0.35	0.029	34.2	1.99	0.74$^{(\dagger)}$	0.51$^{(\dagger)}$
4	276	1.44	0.34	0.021	47.9	2.35	0.72$^{(\dagger)}$	0.71$^{(\dagger)}$
5	392	1.48	0.34	0.015	65.5	2.75	1.0$^{(\dagger)}$	0.97$^{(\dagger)}$
($\dagger$) Increase $n_{\max}$ to $n \times (I_{\max}/I)$ = 200/.51 = 392								

Third Interim Look at Cholesterol Data



East 3: Superiority Trials with Info-Based Design: Interim Monitoring Sheet for Plan1 of EastBook1

Plan Details	
One Sided or Two Sided	2-Sided
Significance Level (Alpha)	0.0500
Power (1-Beta)	0.9000
Assigned Fraction (Treatment)	0.5
Planned Number of Looks	5
Spacing of Analysis	Equal
Hypothesis to be Rejected	H0 Only
Boundary Family	Power(WT)
Boundary to Reject H0	0
Boundary to Reject H1	
Effect Size	0.4000
Maximum Information	67.4110

Look #	Cumul. Info	Test Stat.	Info Fract.	Nominal Critical Point Reject H0 Lower	Reject H0 Upper	Reject H1 Lower	Reject H1 Upper	Repeated 95.00% CI for Effect Size Lower	Upper
1	8.75	2.96	0.130	-4.911	4.911			-0.659	2.661
2	20.3	1.8	0.301	-3.792	3.792			-0.442	1.241
3	34.2	1.99	0.507	-2.893	2.893			-0.154	0.835
4									
5									
6									
7									
8									
9									
10									

Interim Outputs	
CP at obs. effect size	1.000
Ideal Next Look Position	65.9600

Nominal Critical Point Chart (Select)

Info	H0-	H0+	H1-	H1+
8.750	-4.911	4.911		
20.300	-3.792	3.792		
34.200	-2.893	2.893		

Conditional Power Chart (Select)

Eff Size	CP
0.000	0.219
0.049	0.309
0.098	0.412
0.147	0.521
0.196	0.629
0.237	0.712
0.278	0.785
0.318	0.846
0.359	0.894
0.400	0.930

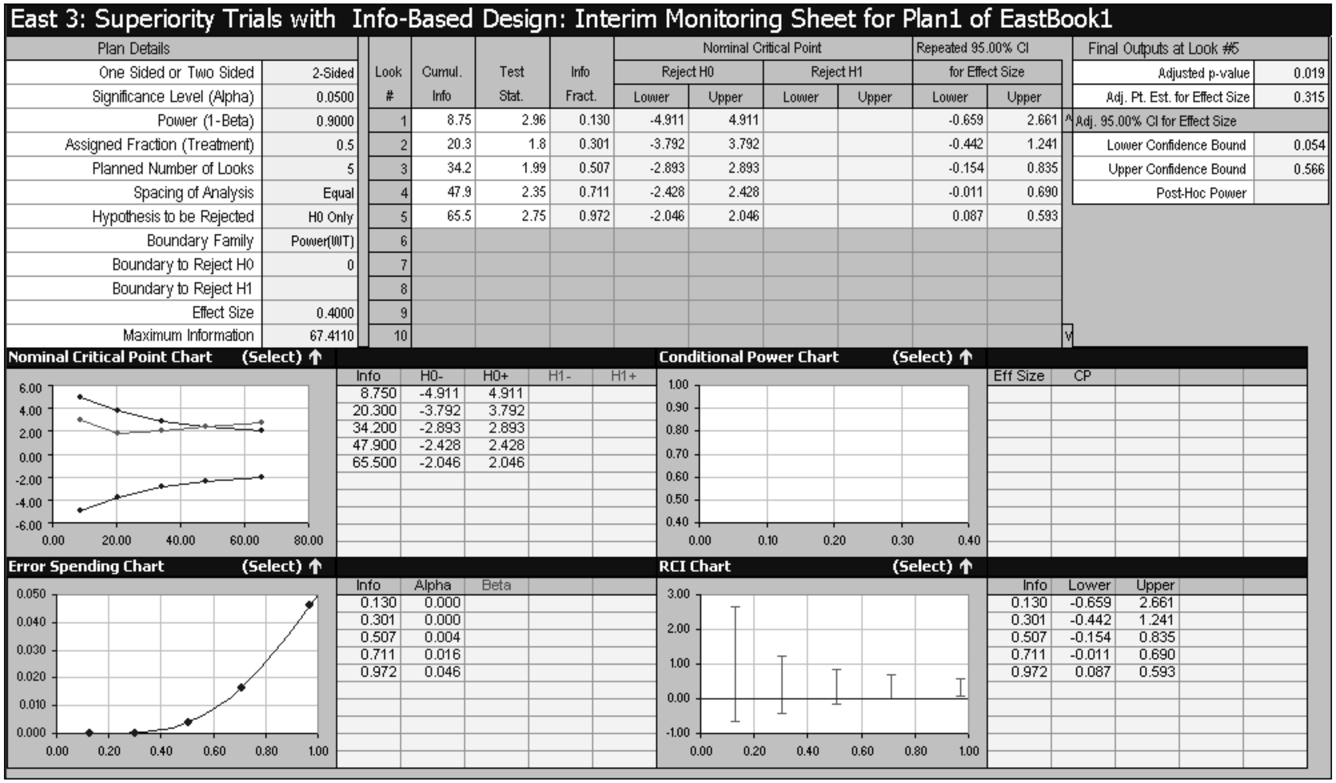
Error Spending Chart (Select)

Info	Alpha	Beta
0.130	0.000	
0.301	0.000	
0.507	0.004	

RCI Chart (Select)

Info	Lower	Upper
0.130	-0.659	2.661
0.301	-0.442	1.241
0.507	-0.154	0.835

Final Look at Cholesterol Data



East 3: Superiority Trials with Info-Based Design: Interim Monitoring Sheet for Plan1 of EastBook1

Plan Details	
One Sided or Two Sided	2-Sided
Significance Level (Alpha)	0.0500
Power (1-Beta)	0.9000
Assigned Fraction (Treatment)	0.5
Planned Number of Looks	5
Spacing of Analysis	Equal
Hypothesis to be Rejected	H0 Only
Boundary Family	Power(WT)
Boundary to Reject H0	0
Boundary to Reject H1	
Effect Size	0.4000
Maximum Information	67.4110

Look #	Cumul. Info	Test Stat.	Info Fract.	Nominal Critical Point: Reject H0 Lower	Reject H0 Upper	Reject H1 Lower	Reject H1 Upper	Repeated 95.00% CI for Effect Size: Lower	Upper
1	8.75	2.96	0.130	-4.911	4.911			-0.659	2.661
2	20.3	1.8	0.301	-3.792	3.792			-0.442	1.241
3	34.2	1.99	0.507	-2.893	2.893			-0.154	0.835
4	47.9	2.35	0.711	-2.428	2.428			-0.011	0.690
5	65.5	2.75	0.972	-2.046	2.046			0.087	0.593
6									
7									
8									
9									
10									

Final Outputs at Look #5	
Adjusted p-value	0.019
Adj. Pt. Est. for Effect Size	0.315
Adj. 95.00% CI for Effect Size	
Lower Confidence Bound	0.054
Upper Confidence Bound	0.566
Post-Hoc Power	

Nominal Critical Point Chart (Select)

Info	H0-	H0+	H1-	H1+
8.750	-4.911	4.911		
20.300	-3.792	3.792		
34.200	-2.893	2.893		
47.900	-2.428	2.428		
65.500	-2.046	2.046		

Conditional Power Chart (Select)

Eff Size	CP

Error Spending Chart (Select)

Info	Alpha	Beta
0.130	0.000	
0.301	0.000	
0.507	0.004	
0.711	0.016	
0.972	0.046	

RCI Chart (Select)

Info	Lower	Upper
0.130	-0.659	2.661
0.301	-0.442	1.241
0.507	-0.154	0.835
0.711	-0.011	0.690
0.972	0.087	0.593

Example 3: Cox Model with Covariates

- The data are generated by the model

$$\lambda(t|Z_0, Z_1) = \lambda_0(t)e^{\delta Z_0 + \eta Z_1} \; .$$

- We wish to test $H_0 : \; \delta = 0$ in the presence of several nuisance parameters, η.

- Set up stopping boundaries for $\hat{\delta}/\text{se}(\hat{\delta})$, where $\hat{\delta}$ is the maximum partial likelihood estimate of δ, and $\text{se}(\hat{\delta})$ is its standard error.

- Enroll subjects into the trial and follow them until either a boundary is crossed or

$$[\text{se}(\hat{\delta})]^{-2} \geq \left[\frac{z_{\alpha/2} + z_{\beta}}{\delta_1}\right]^2 \times \text{IF}$$

Example 4: Random Effects Longitudinal Model

- Let $Y_{ijk} = \alpha_{ik} + \gamma_{ik} t_{ijk} + \epsilon_{ijk}$, where

$$\begin{bmatrix} \alpha_{ik} \\ \gamma_{ik} \end{bmatrix} \sim N\left(\begin{bmatrix} A_k \\ G_k \end{bmatrix}, \begin{bmatrix} \sigma^2_{A_k} & \sigma_{A_k,G_k} \\ \sigma_{A_k,G_k} & \sigma^2_{G_k} \end{bmatrix} \right)$$

  and

$$\epsilon_{ijk} \sim N(0, \sigma^2_{\epsilon_k}) \ ,$$

$$k = 1, 2, \ j = 1, \ldots, n_{ik}, \ i = 1, \ldots, m_k \ .$$

- To test H_0: $\delta = G_1 = G_2 = 0$, enroll subjects until

$$[\text{se}(\hat{\delta})]^{-2} \geq \left[\frac{z_{\alpha/2} + z_\beta}{\delta_1} \right]^2 \ .$$

Comments on Information Based Design

- **With information based design you only need to specify the effect size at the design stage. Moreover the effect size may be specified on any suitable metric, such as difference of proportions, ratio of proportions, odds ratio, hazard ratio, or difference of means**

- **Whereas the required value of $n_{\max}$ depends on nuisance parameters (σ^2, π_c, covariate values), the required value of $I_{\max}$ does not. $I_{\max}$ only depends on the pre-specified values of δ, α, β, and the spending function.**

- **But if we want to translate information into sample size, we do need to provide initial estimates of the nuisance parameters. These estimates can be revised at each interim look**

- **Information based designs are already in use for survival studies where we keep the study open until $d_{\max}$ events are obtained rather than until $n_{\max}$ patients are enrolled.**

Comments on Information Based Monitoring

- **You must calculate the information at each interim monitoring time point and monitor on the basis of $(I_j/I_{\max})$ instead of $(n_j/n_{\max})$.**
- **Each time you monitor the data you have the opportunity to review your initial assumptions about the nuisance parameters and revise them.**
- **Based on these revised estimates you can re-estimate the sample size.**
- **Thus, maximum information remains the same but sample size can be changed over the course of the study.**

- **It should be possible in the future to use electronic data capture technology to monitor continuously and stop the trial when the right amount of information is obtained. One could estimate nuisance parameters along the way, compute the implied value of $n_{\max}$, and produce a report of how much longer the trial might be expected to last based on the current accrual rates. Early stopping rules could also be built into this automated monitoring scheme. This idea could also be extended to regression models with covariates.**

Adaptive Design

We motivate this idea with an example from Cui, Hung and Wang (Biometrics, 1999).

- Phase III clinical trial for prevention of myocardial infarct in patients undergoing coronary artery bypass graft surgery
- Design parameters:
  - Placebo MI rate, $\pi_c = 22\%$. Treatment MI rate, $\pi_t = 11\%$. Looking for an improvement of $\delta = \pi_t - \pi_c = -11\%$
  - One sided test with $\alpha = 0.025$ and 95% power to detect $\delta = -11\%$
  - One interim look at midcourse
- Required up-front commitment: 579 patients (both arms)

But when the interim look was taken:

- The drop in MI was only half the design specification. $\hat{\pi}_c \approx 22\%$, and $\hat{\pi}_t \approx 16.5\%$; i.e., $\hat{\delta} \approx -6.5\%$, whereas the study was powered for $\delta = -11\%$
- Sponsor was concerned that if these estimates accurately reflected the true parameters, the trial would only have 39% power
- At the time there was no valid procedure to increase the sample size without inflating the type-1 error.
- Eventually the trial failed to show statistical significance

Could this trial have been saved?

Yes! An adaptive approach might have saved this trial

- **Basic idea of adaptive trials:**
  - **increase the sample size if the observed data suggest loss of power**
  - **at the same time make appropriate adjustments to preserve the type-1 error**
- **Extensive Literature: Bauer and Kohne (1994), Proschan and Hunsberger (1995), Cui, Hung and Wang (1999), Shen and Fisher (1999), Lemacher and Wassmer (1999), Liu and Chi (2001), Lan and Trost (1997), Tsiatis and Mehta (2003).**

The CHW adaptive method

- Two-stage trial is planned for n_1 observations/arm at stage 1 and $n_{\max}$ observations/arm overall.
- At stage 1 compute increase sample size so as to recover:

Unconditional Power

$$n^*_{\max} = \left[\frac{\delta}{\hat{\delta}}\right]^2 n_{\max}$$

Conditional Power

$$1-\Phi\left[\frac{u_2\sqrt{n^*_{\max}} - z_1\sqrt{n_1} - \sqrt{(n_{\max}-n_1)(n^*_{\max}-n_1)}\hat{\delta}/\hat{\sigma}}{\sqrt{n^*_{\max}-n_1}}\right]$$

where $\hat{\sigma} = \sqrt{2\hat{\pi}_t(1-\hat{\pi}_t) + 2\hat{\pi}_t(1-\hat{\pi}_t)}$

Test Statistic at Final Look

- At the final stage the usual test statistic is $Z = \hat{\delta}/\text{se}(\hat{\delta})$. If we do not adapt we can re-write Z as

$$Z = \sqrt{\left(\frac{n_1}{n_{\max}}\right)} \sum_{i=1}^{n_1} \frac{(X_{ti} - X_{ci})}{\hat{\sigma}\sqrt{n_1}} + \sqrt{\left(\frac{n_{\max} - n_1}{n_{\max}}\right)} \sum_{i=n_1+1}^{n_{\max}} \frac{(X_{ti} - X_{ci})}{\hat{\sigma}\sqrt{n_{\max} - n_1}}$$

- If we increase the sample size from $n_{\max}$ to $n^*_{\max}$, the statistic becomes

$$Z^* = \sqrt{\left(\frac{n_1}{n^*_{\max}}\right)} \sum_{i=1}^{n_1} \frac{(X_{ti} - X_{ci})}{\hat{\sigma}\sqrt{n_1}} + \sqrt{\left(\frac{n^*_{\max} - n_1}{n^*_{\max}}\right)} \sum_{i=n_1+1}^{n^*_{\max}} \frac{(X_{ti} - X_{ci})}{\hat{\sigma}\sqrt{n^*_{\max} - n_1}}$$

Downweight the final test statistic

Cui, Hung and Wang (1999) propose that instead of using

$$Z^* = \sqrt{\left(\frac{n_1}{n^*_{\max}}\right)} \sum_{i=1}^{n_1} \frac{(X_{ti} - X_{ci})}{\hat{\sigma}\sqrt{n_1}} + \sqrt{\left(\frac{n^*_{\max} - n_1}{n^*_{\max}}\right)} \sum_{i=n_1+1}^{n^*_{\max}} \frac{(X_{ti} - X_{ci})}{\hat{\sigma}\sqrt{n^*_{\max} - n_1}}$$

we should use

$$Z^{**} = \sqrt{\frac{n_1}{n_{\max}}} \sum_{i=1}^{n_1} \frac{(X_{it} - X_{ic})}{\sqrt{\hat{\sigma}}\sqrt{n_1}} + \sqrt{\frac{n_{\max} - n_1}{n_{\max}}} \sum_{i=n_1+1}^{n^*_{\max}} \frac{(X_{it} - X_{ic})}{\sqrt{\hat{\sigma}}\sqrt{(n^*_{\max} - n_1)}}$$

- Under H_0, the distribution of Z^{**} is identical to that of Z. So type-1 error is preserved
- But the contribution of the stage 1 patients counts more than the contribution of the stage 2 patients.

Alternative Methods

Adjust the Final Boundary In this method, the usual unweighted test statistic is used, but the final stopping boundary is changed from u_2 to u_2^* so as to preserve the type-1 error. We solve the following equation for u_2^*

$$1-\Phi(u_1)+\int_{l_1}^{u_1}\left\{1-\Phi\left[u_2^*\sqrt{1+\frac{n_1}{(n^*_{\max}-n_1)}}-z_1\sqrt{\frac{n_1}{(n^*_{\max}-n_1)}}\right]\right\}\phi(z_1)dz_1=\alpha$$

Combination of P-values This method utilizes the fact that under the null hypothesis the p-values are uniformly distributed between 0 and 1. The p-values from the first and second stages can be combined in a pre-specified way to so as to preserve the type-1 error

Two-look design with no early stopping but possible adaptation at look 1

East 3: Superiority Trials with Binomial Design fo	
Plan ID	Plan1
Test Parameters	
1-Sided or 2-Sided Test	1-Sided
Significance Level (Alpha)	0.025
Power (1 - Beta)	0.95
Assigned Fraction (Treatment)	0.5
Boundary Parameters	
Planned Number of Looks	2
Spacing of Analysis	Equal
Hypothesis to be Rejected	H0 Only
Boundary Family	SpF(Pub)
Boundary to Reject H0	Gm(-12)
Boundary to Reject H1	
Binomial parameters under H1	
Proportion Response (Control: π_c)	0.22
Proportion Response (Treatment: π_t)	0.11
Accrual (Subjects)	
Maximum	579
Expected Under H0	579
Expected Under H1	550
Expected Under H1/2	577

Simulation Results: $\pi_t = 16.5\%$. No adaptation at look 1.

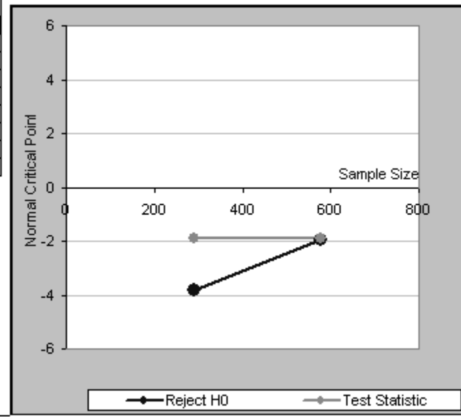
East 3: Simulation of Superiority Sequential Trials (Cui, Hung, Wang Adaptive Method)

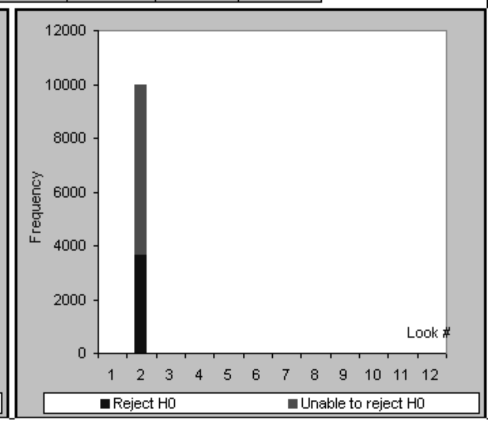
Design Parameters for Plan1-	
1-Sided or 2-Sided Test	1-Sided
Significance Level(Alpha)	0.025
Power(1-Beta)	0.95
End Point	Binomial
Proportion Response(Control)	0.22
Proportion Response(Treatment)	0.11
Assigned Fraction	0.5
Planned # of Looks	2
Spacing of Analysis	Equal
Hypothesis to be rejected	H0 Only
Boundary Family	SpF(Pub)
Boundary to Reject H0	Gm(-12)
Boundary to Reject H1	
Design Outputs	
Max. Sample Size (Nmax)	579
Max. Information (Imax)	1074.2115
Simulation Parameters	
Proportion Response(Control)	0.22
Proportion Response(Treatment)	0.165
Use Nmax Till 'L' Looks, L =	1
Crit. to Rescale Nmax (f_0)	1
Max. Usable Sample Size	539
Number of Trials	10000
Refresh Every 'n' Trials, n =	1000
Simulation Starting Seed	Clock

Run | Single Step | Reset | Stop | Help

Simulation Seed =
Elapsed Time = 0:00:05

Simulation Boundary				Overall Simulation Results					
Look	Boundary		Simulated	Avg.	Avg.Sample	# Rejecting	# Unable to	Total simulations	
#	Info	H0	Test Stat	Information	Size	H0	reject H0	Count	%
1	289.44	-3.8388	-1.8832	448.65	289.00	23		23	0.23%
2	578.87	-1.9601	-1.8751	887.11	547.78	3635	6342	9977	99.77%
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
			Total	886.11	547.18	3658	6342	10000	
			%			36.58%	63.42%		





Simulation Results: $\pi_t = 16.5\%$. With adaptation at look 1.

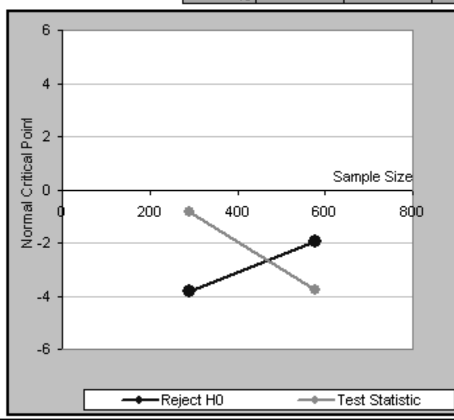
East 3: Simulation of Superiority Sequential Trials (Cui, Hung, Wang Adaptive Method)

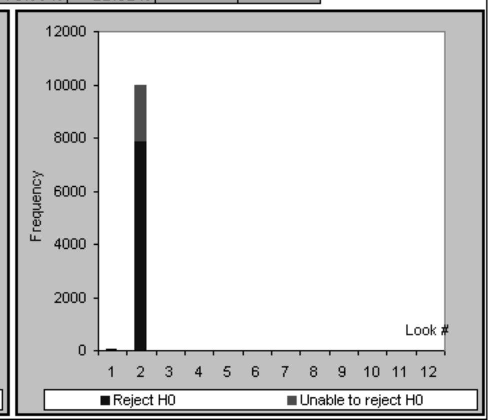
Design Parameters for Plan1-	
1-Sided or 2-Sided Test	1-Sided
Significance Level(Alpha)	0.025
Power(1-Beta)	0.95
End Point	Binomial
Proportion Response(Control)	0.22
Proportion Response(Treatment)	0.11
Assigned Fraction	0.5
Planned # of Looks	2
Spacing of Analysis	Equal
Hypothesis to be rejected	H0 Only
Boundary Family	SpF(Pub)
Boundary to Reject H0	Gm(-12)
Boundary to Reject H1	
Design Outputs	
Max. Sample Size (Nmax)	579
Max. Information (Imax)	1074.2115
Simulation Parameters	
Proportion Response(Control)	0.22
Proportion Response(Treatment)	0.165
Use Nmax Till 'L' Looks, L =	1
Crit. to Rescale Nmax (f_0)	1
Max. Usable Sample Size	3000
Number of Trials	10000
Refresh Every 'n' Trials, n =	1000
Simulation Starting Seed	Clock

Run | Single Step | Reset | Stop | Help

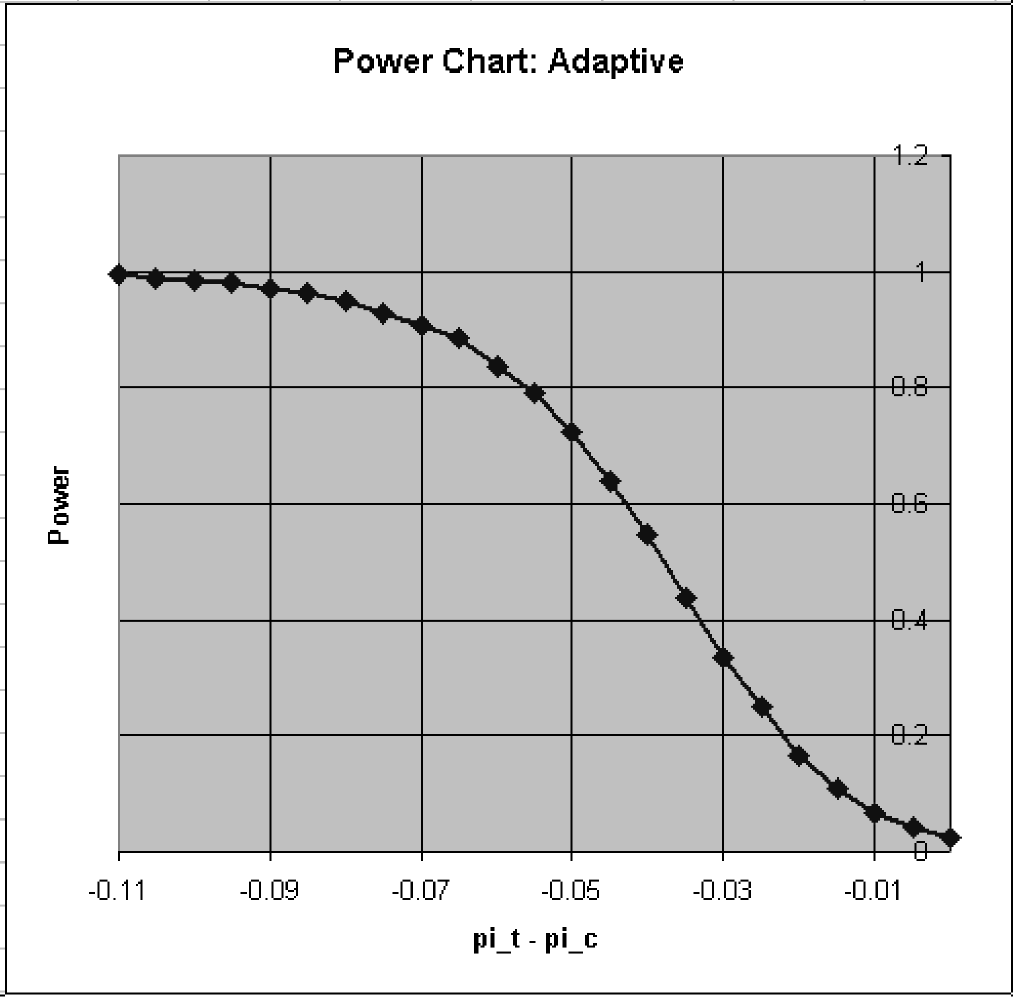
Simulation Seed =
Elapsed Time = 0:00:13

Simulation Boundary				Overall Simulation Results					
Look	Boundary		Simulated	Avg.	Avg.Sample	# Rejecting	# Unable to	Total simulations	
#	Info	H0	Test Stat	Information	Size	H0	reject H0	Count	%
1	289.44	-3.8388	-0.8229	466.16	289.00	36		36	0.36%
2	578.87	-1.9601	-3.7923	2841.62	1763.35	7863	2101	9964	99.64%
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
			Total	2833.07	1758.04	7899	2101	10000	
			%			78.99%	21.01%		

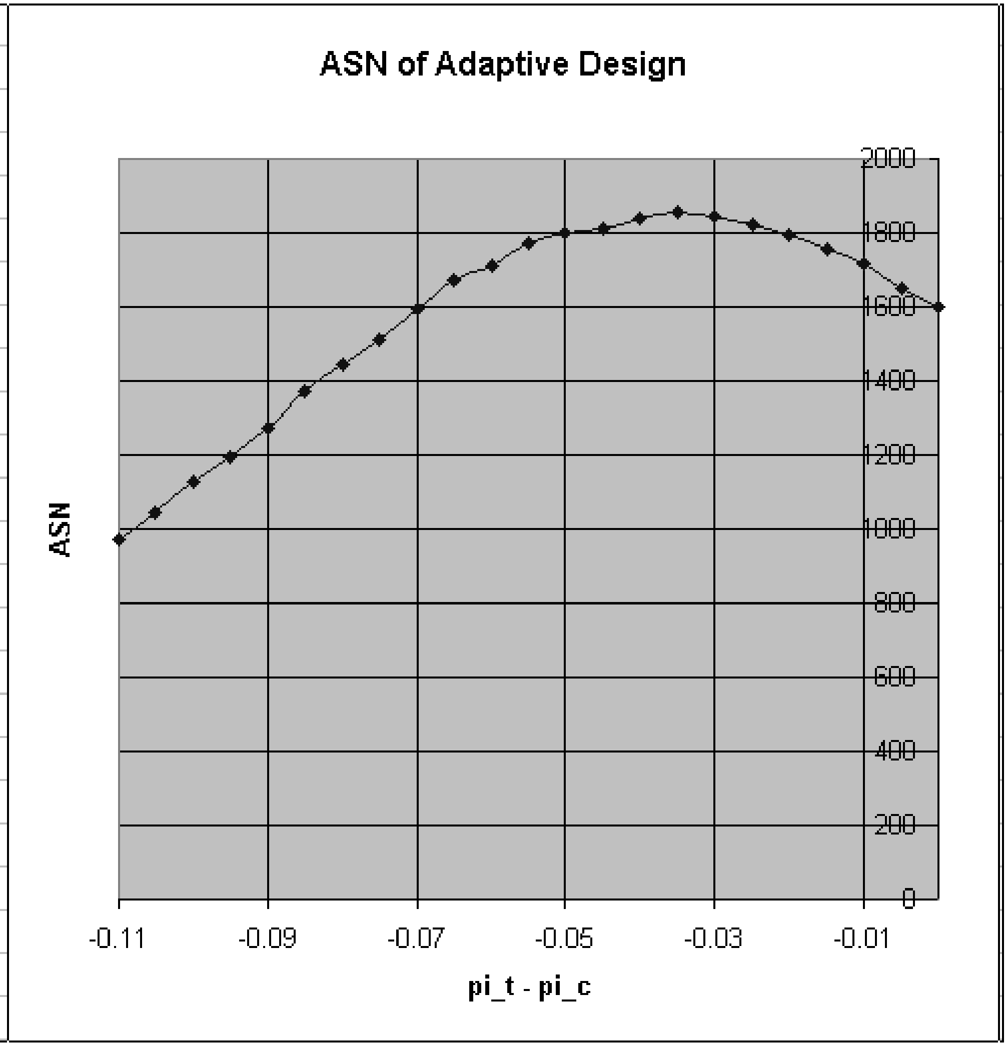




pi_t-pi_c	power
-0.11	0.994
-0.105	0.989
-0.1	0.986
-0.095	0.982
-0.09	0.972
-0.085	0.963
-0.08	0.948
-0.075	0.927
-0.07	0.907
-0.065	0.886
-0.06	0.835
-0.055	0.791
-0.05	0.723
-0.045	0.638
-0.04	0.546
-0.035	0.437
-0.03	0.334
-0.025	0.252
-0.02	0.166
-0.015	0.111
-0.01	0.067
-0.005	0.041
0	0.025



pi_t-pi_c	asn
-0.11	973
-0.105	1045
-0.1	1126
-0.095	1194
-0.09	1274
-0.085	1373
-0.08	1445
-0.075	1509
-0.07	1593
-0.065	1671
-0.06	1709
-0.055	1771
-0.05	1798
-0.045	1812
-0.04	1839
-0.035	1857
-0.03	1847
-0.025	1823
-0.02	1797
-0.015	1757
-0.01	1718
-0.005	1649
0	1601



Compare the Adaptive and Group Sequential Approaches

Adaptive Approach: Start small, then ask for more.

- We designed the trial for 579 patients, to detect a drop in the event rate of 11%.
- But mid-way through the trial, when we saw that the drop was only 6.5% we were willing to increase the sample size.

Group Sequential Approach: Ask for more up front, then cut back.

- Start out by designing a trial that has good power to detect a drop of 6.5%.
- But have high probability of early detection and termination if actual drop is 11%.
- By judicious selection of sample size, number of looks, spacing of looks and stopping boundaries, one can improve on the adaptive approach uniformly. (See general theorem by Tsiatis and Mehta, Biometrika, 2003)

How to search for the dominating group sequential design

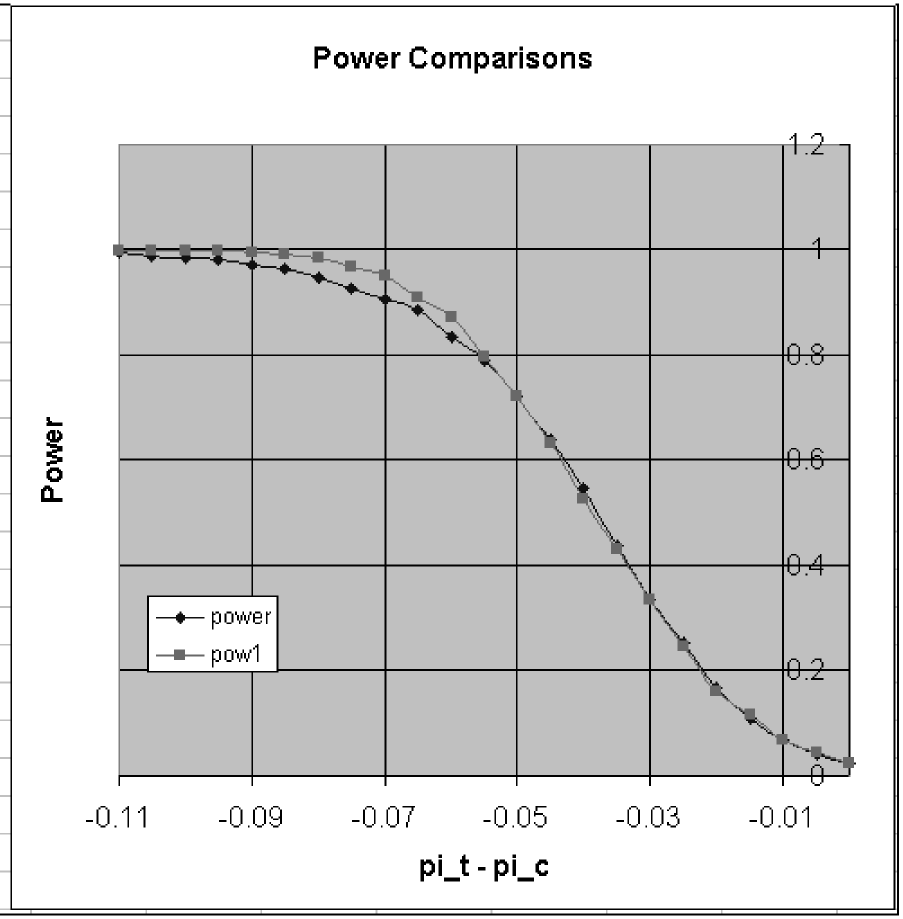
Some rules of thumb:

1. **Start with $N_{\max}$ approximately equal to the highest ASN of the adaptive design. Then increase of decrease to move the power curve up or down.**
2. **Select both an efficacy and a futility boundary.**
3. **To decrease ASN and power, add more interim looks.**
4. **To decrease ASN and power, increase the concavity of the spending function.**
5. **For added sensitivity, change spacing of the looks**

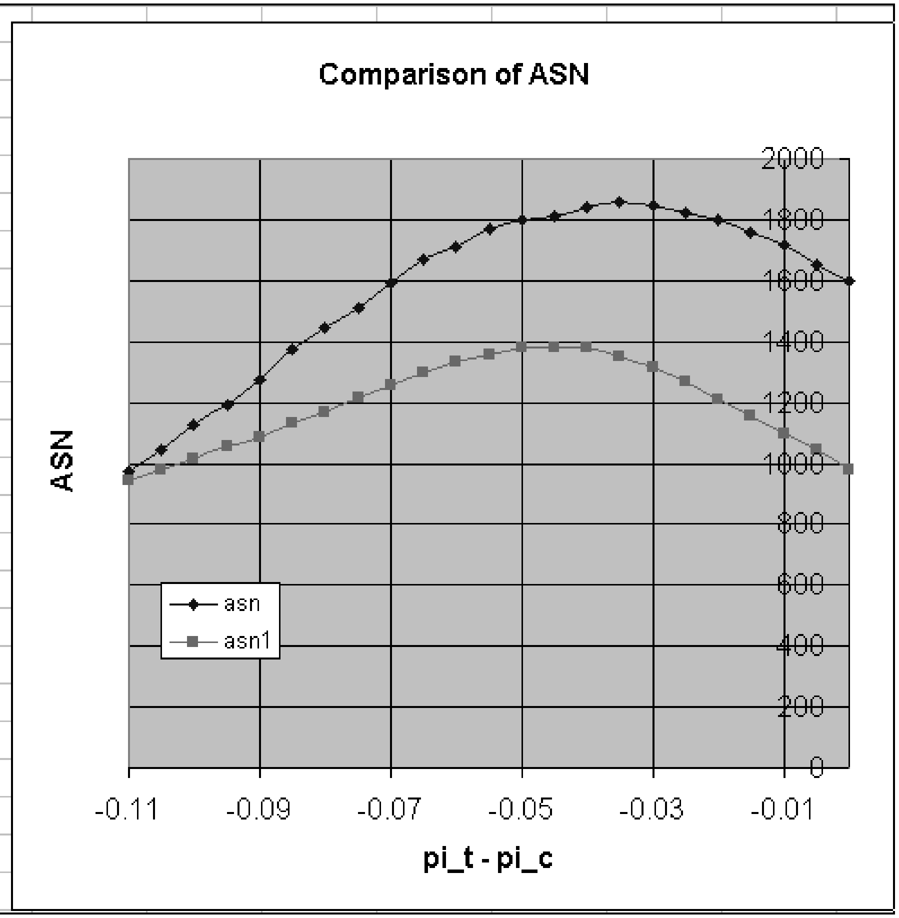
East 3: Superiority Trials with Binomial Design for EastBook1

Plan ID	Plan1	Plan2	Plan3
Test Parameters			
1-Sided or 2-Sided Test	1-Sided	1-Sided	1-Sided
Significance Level (Alpha)	0.025	0.025	0.025
Power (1 - Beta)	0.95	0.95	0.7544
Assigned Fraction (Treatment)	0.5	0.5	0.5
Boundary Parameters			
Planned Number of Looks	2	2	3
Spacing of Analysis	Equal	Equal	Unequal
Hypothesis to be Rejected	H0 Only	H0 Only	H0 or H1
Boundary Family	SpF(Pub)	SpF(Pub)	SpF(Pub)
Boundary to Reject H0	Gm(-12)	Gm(-12)	Gm(0)
Boundary to Reject H1			Gm(0)
Binomial parameters under H1			
Proportion Response (Control: π_c)	0.22	0.22	0.22
Proportion Response (Treatment: π_t)	0.11	0.165	0.165
Accrual (Subjects)			
Maximum	579	2658	1700
Expected Under H0	579	2658	954
Expected Under H1	550	2527	1204
Expected Under H1/2	577	2651	1180

pi_t-pi_c	power	pow1
-0.11	0.994	0.999
-0.105	0.989	0.999
-0.1	0.986	0.999
-0.095	0.982	0.998
-0.09	0.972	0.996
-0.085	0.963	0.991
-0.08	0.948	0.986
-0.075	0.927	0.968
-0.07	0.907	0.949
-0.065	0.886	0.909
-0.06	0.835	0.873
-0.055	0.791	0.796
-0.05	0.723	0.723
-0.045	0.638	0.631
-0.04	0.546	0.527
-0.035	0.437	0.431
-0.03	0.334	0.334
-0.025	0.252	0.246
-0.02	0.166	0.162
-0.015	0.111	0.115
-0.01	0.067	0.068
-0.005	0.041	0.044
0	0.025	0.025



pi_t-pi_c	asn	asn1
-0.11	973	942
-0.105	1045	979
-0.1	1126	1013
-0.095	1194	1054
-0.09	1274	1088
-0.085	1373	1130
-0.08	1445	1167
-0.075	1509	1214
-0.07	1593	1255
-0.065	1671	1295
-0.06	1709	1336
-0.055	1771	1358
-0.05	1798	1380
-0.045	1812	1382
-0.04	1839	1379
-0.035	1857	1352
-0.03	1847	1315
-0.025	1823	1270
-0.02	1797	1211
-0.015	1757	1156
-0.01	1718	1098
-0.005	1649	1042
0	1601	980



Does the flexibility of adaptive designs offset their loss of efficiency?

There has been a great deal of recent interest in adaptive designs. Although in theory they can be dominated by standard group sequential trials there are several reasons for giving them serious consideration

Reason 1: The clinically meaningful effect size can change

- At the design stage you felt that the clinically meaningful drop in MI is 11%. You designed the study for 90% power to detect this difference
- But after the trial starts, possibly because results from other trials become available, you feel that even a drop in MI of 6.5% is worth detecting. Now you have the flexibility to re-design the study for 90% conditional power to detect this difference
- We need published discussions, with specific examples of this type

Reason 2: Investigators need to see some data before finalizing the design

Sometimes the pilot study data and phase II data are inadequate for specifying a clinically meaningful effect size. Investigators want to see some data from the actual study before finalizing where they want to power it.

Reason 3: Budgetary Considerations

Investigators might find it easier to request a small budget at the design stage and ask for supplemental funding after seeing the interim data.

- But this leads to an interesting question. Could you request supplemental funding for increasing the sample size without unblinding the interim results to the trial sponsor?
- Again, we need published discussions, with specific examples of how the actual logistics of running an adaptive trial would be handled. Who would make the decision to adjust the sample size? Who would know about the decision? What is the potential for bias?

What's in Store for the Future?

- The pressure to speed up trial, reduce their costs and get new products to the market place will only increase
- Flexible methods for design and monitoring will be key factors for responding to this pressure
- There will be a merger of two forces, electronic data capture technology and group sequential technology, to provide an integrated solution to the logistical and statistical complexities of monitoring trials in flexible ways without biasing their conclusions.