Adaptive Designs for Confirmatory Clinical Trials

Society for Clinical Trials, Boston

May 19, 2013

Cyrus R. Mehta, Ph.D
President, Cytel Inc.

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

Part 1: Case Studies in East of Adaptive Sample Size Reestimation

Two Adaptive Designs

Simulation based design and monitoring of two late-stage adaptive clinical trials using the East-Adapt and East-SurvAdapt software

- **Normal endpoint example: Negative Symptoms Schizophrenia**
- **Survival endpoint example: Acute Myeloid Leukemia**

Motivation for Mid-Course Sample Size Correction in Pivotal Trials

We don't know what δ and σ to power the study for

- **Prior experience limited to small pilot studies**
- **Improved standard of care dilutes treatment effect**
- **Powering for smallest clinically important effect expensive**
- **Easier to obtain milestone-based financing for trials involving novel compounds**

If only σ is unknown, blinded SSR is recommended by FDA

Why not design for the smallest clinically meaningful treatment effect?

Large effects are uncommon, but designing for very small clinically meaningful effects requires huge up-front investments that management will not approve. A strategy of staged investment is more practical

Unreliability of Pilot Studies: **Most large treatment effects emerge from small studies, and when additional trials are performed, the effect sizes typically become much smaller. Well-validated large effects are uncommon and pertain to nonfatal outcomes.**

Pereira et. al., JAMA. 2012; 308(16): 1676-1684

Milestone-Driven Investment: **Sunesis Pharmaceuticals to Implement One-Time Sample Size Increase to Phase 3 VALOR Trial in AML. DSMB Recommends Increase Following Single, Pre-Planned Interim Efficacy and Safety Analysis of VALOR; Enrollment Completion Expected in 2013. DSMB Recommendation Triggers $25.0 Million Investment in Sunesis from Royalty Pharma.**

Press Release, September 11, 2012. Sunesis Pharma, South San Francisco

Case 1: Negative Symptoms Schizophrenia

- New drug versus placebo for treatment of negative symptoms schizophrenia
- Primary endpoint is the change in negative symptoms assessment (NSA) at week 26 relative to the baseline assessment
- Based on the limited information available sponsor powers the trial to detect a 2-point improvement ($\delta = 2$ with $\sigma = 7.5$) with respect to NSA
- Sponsor would like some insurance against power loss in case $\delta = 1.6$, the smallest clinically important effect

Fixed Sample Designs

	Des1	Des2
Design Type	MN-2S-DI	MN-2S-DI
Test Parameters		
Trial Type	Superiority	Superiority
No. of Looks	1	1
Test Type	1-Sided	1-Sided
Specified α	0.025	0.025
Specified Power	0.8	0.8
Attained Power	0.8	0.8
Model Parameters		
Input Method	Individual Means	Individual Means
Diff. in Means (δ = μt - μc)	2	1.6
Mean Control (μc)	0	0
Mean Treatment (μt1)	2	1.6
Std. Deviation (σ)	7.5	7.5
Test Statistic	Normal	Normal
Allocation Ratio (nt/nc)	1	1
Sample Size		
Maximum	442	690

Sponsor is willing to invest the resources for 442 subjects up-front. But would like to see some data from the trial before investing the resources for 690 subjects

Table 1: Operating Characteristics of Plan1 and Plan2

	Plan1		Plan2	
δ	Sample Size	Power	Sample Size	Power
1.6	442	61%	690	80%
1.7	442	66%	690	85%
1.8	442	71%	690	88%
1.9	442	76%	690	91%
2.0	442	80%	690	94%

- Trial is adequately powered if $\delta \geq 2$
- Trial is underpowered if $\delta = 1.6$, an effect that is still clinically meaningful

Group Sequential Option

- Design for $\delta = 1.6$ but with an early stopping boundary
- Perform interim analysis after observing 50% of the total information
- Use a conservative spending function to ensure that positive results based on premature termination will be compelling and alter medical practice (Pocock, 2005)

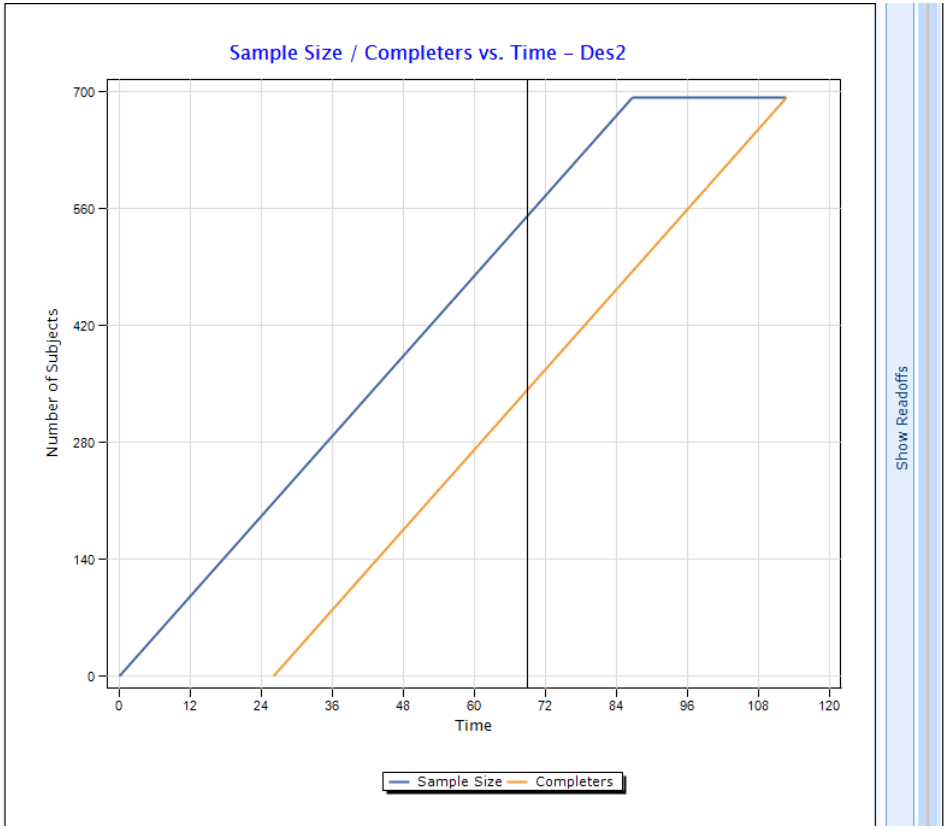
Problems: There will be overruns due to the 26-week delay in response

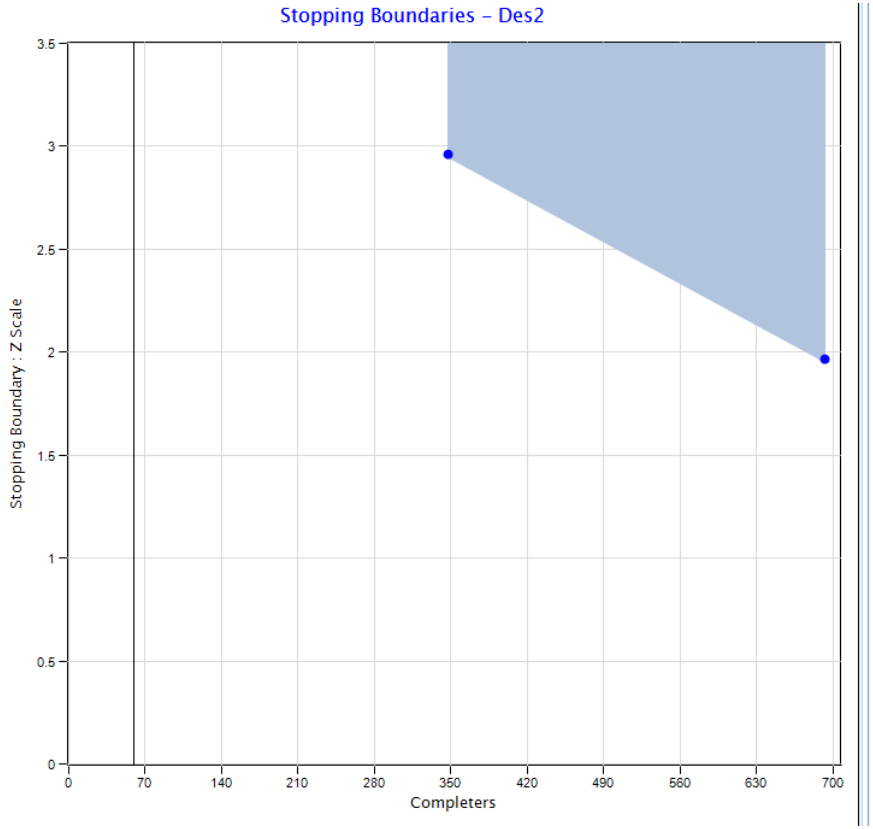
The Overruns Problem

- Suppose enrollment rate is 8/week. Then there will be $8 \times 26 = 208$ overruns (i.e., enrolees who have not yet become completers) at the time of early termination
- These overruns will offset some of the sample size savings gained by the early termination
- In addition, there is the risk that these overruns might reverse the early stopping decision when they become completers

Conclusion: Group sequential designs are useful for acute settings where the endpoint is observed quickly and for survival settings where the endpoint is event-driven, and study duration rather than sample size is the primary consideration

Illustration of Overruns Problem





Show Readoffs

By the time 347 patients have completed treatment (i.e., 50% of total information has arrived), 555 patients will already have enrolled. Sample size saving is thus rather limited

Table 2: Operating Characteristics of Plan3 (Group Sequential) and Plan2 (Fixed Sample)

	Plan3 (Group Sequential)				Plan2 (Fixed Sample)	
	Probability of	Expected Sample Size				
δ	Early Stopping	No Overruns	With Overruns	Power	SampSiz	Power
1.6	17%	635	670	80%	690	80%
1.7	20%	624	666	84%	690	85%
1.8	23%	611	661	88%	690	88%
1.9	27%	598	655	91%	690	91%
2.0	32%	583	649	94%	690	94%

- The power is identical for the two plans
- However, after accounting for the overruns the group sequential design offers only marginal savings over the fixed sample design
- Might improve performance with a more agressive spending function, but that would require an even larger up-front commitment

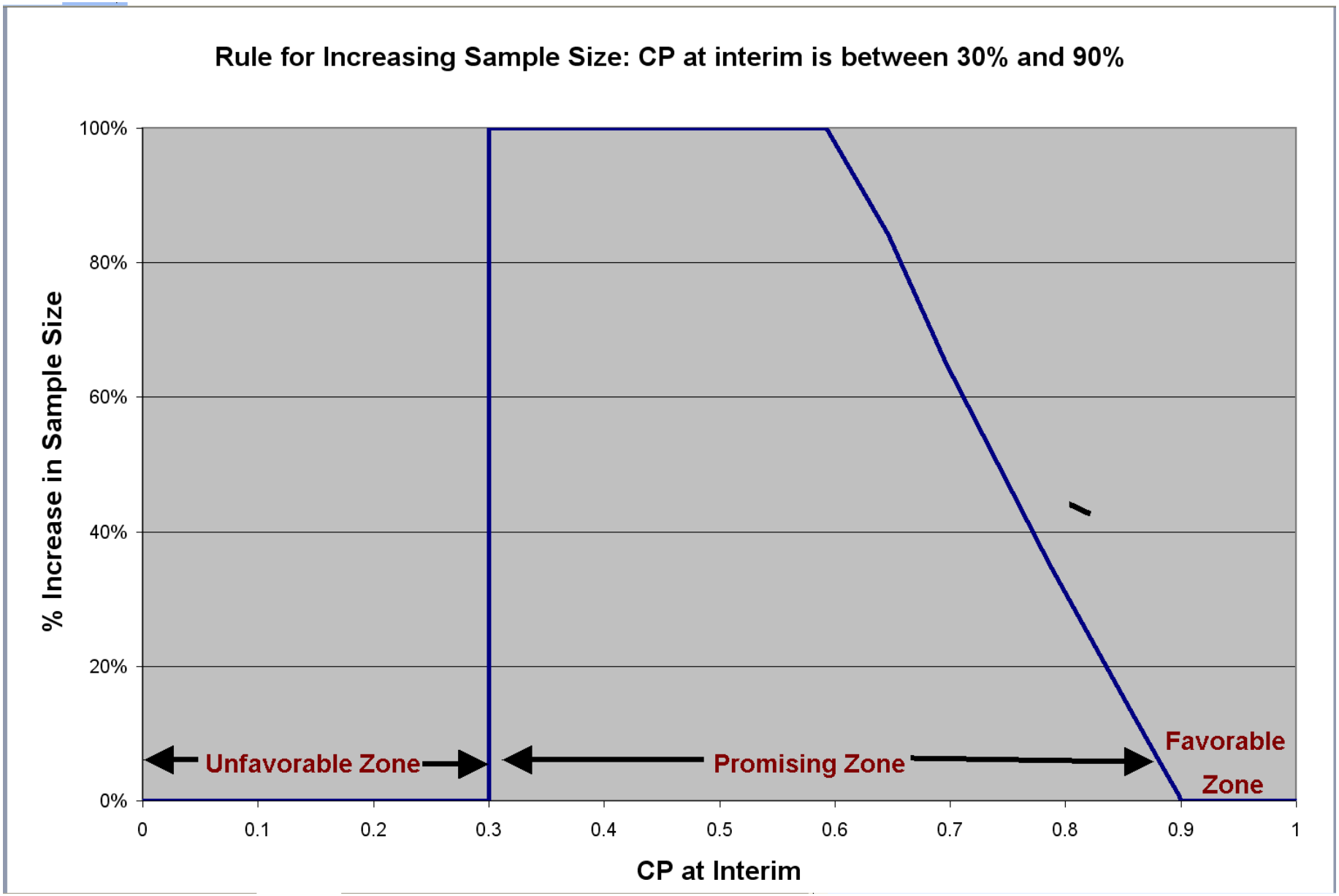
Problem: The 208 patient overrun could reverse early efficacy conclusion

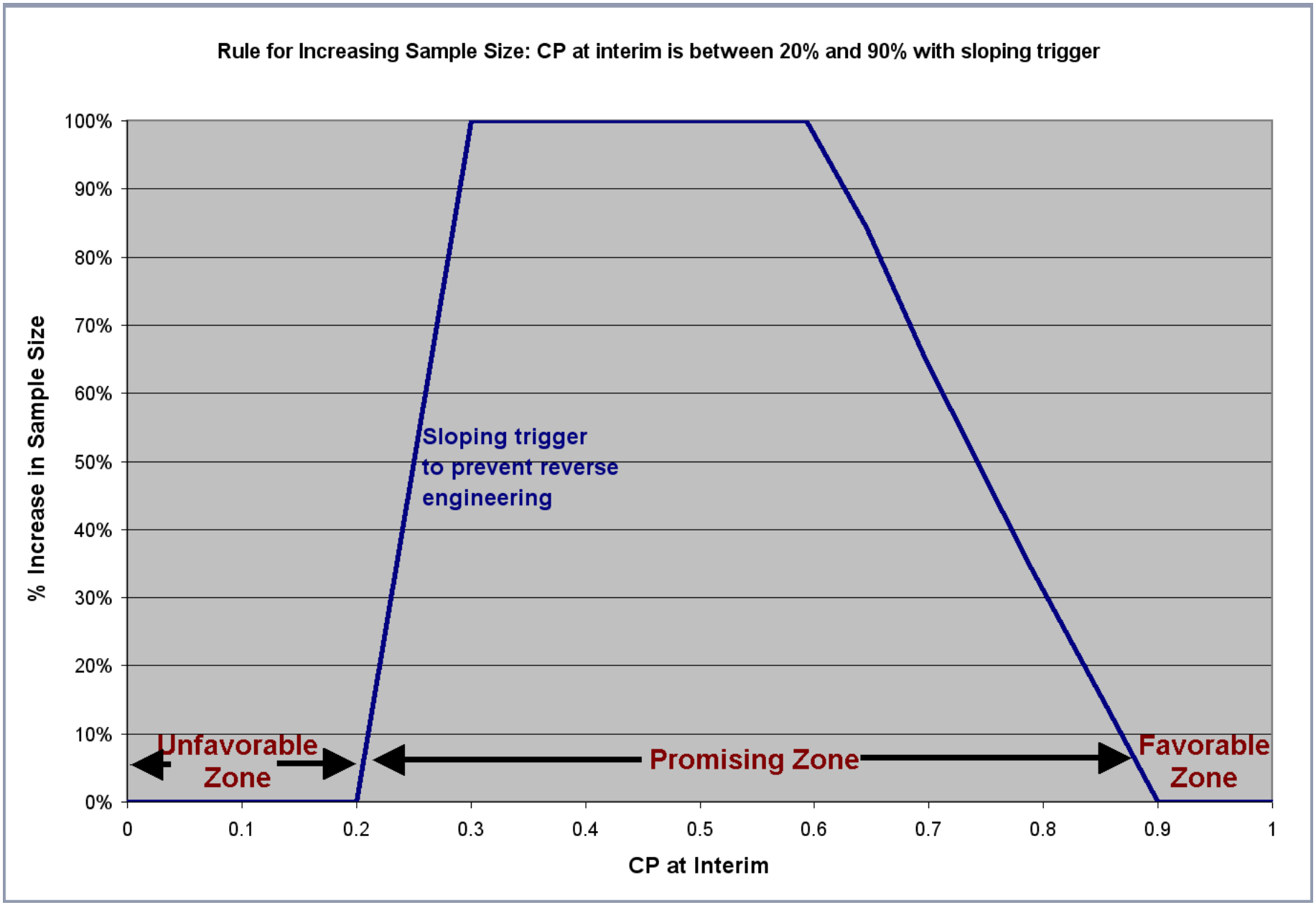
Adaptive Strategy

1. Only commit 442 subjects up-front
2. Perform an interim analysis after seeing data on 208 completers
3. Increase the sample size, and hence the conditional power, if the interim results fall in a promising zone

Criterion for Increasing Sample Size

- Partition interim outcome into unfavorable, promising and favorable zones, based on conditional power
- Only increase sample size if interim outcome is promising
- Specify a target conditional power (also known as "shape parameter") and calculate sample size needed to achieve the target
- Set minimum and maximum limits on sample size change


Rule for Increasing Sample Size: CP at interim is between 30% and 90%
% Increase in Sample Size
100%
80%
60%
40%
20%
0%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
CP at Interim
Unfavorable Zone
Promising Zone
Favorable Zone


Rule for Increasing Sample Size: CP at interim is between 20% and 90% with sloping trigger
% Increase in Sample Size
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
CP at Interim
Sloping trigger
to prevent reverse
engineering
Unfavorable
Zone
Promising Zone
Favorable
Zone

Simulation Based Adaptive Design

Simulation Parameters	
Difference of Means	1.6000
Standard Deviation	7.5000
Use Nmax Till 'L' Looks, L =	1
Criterion for Adapting: Min. CP	0.3000
Max. CP	0.8000
Min. Usable Sample Size	442
Max. Usable Sample Size	884
Desired Conditional Power (CP)	0.8000
Number of Trials	100000
Refresh Every 'n' Trials, n =	10000
Simulation Starting Seed	Clock

- **Use the EastAdapt simulation worksheet to evaluate operating characteristics of the adaptive design**
- **Experiment with the above criteria if operating characteristics until satisfactory design is obtained**

Overall Simulation Results					
Avg. Info.	Avg. Sample Size	# Rejecting H0	# Unable to Reject H0	Total Simulations	
				Count	%
0.92	208.00	7		7	0.01%
2.26	508.87	66970	33023	99993	99.99%
2.26	508.85	66977	33023	100000	
		66.98%	33.02%		
Simulation Results for Adapted Trial Only					
3.09	695.77	21862	4488	26350	100.00%
3.09	695.77	21862	4488	26350	
		82.97%	17.03%		

Simulation Results by Zone

Zone	Avg. Sample Size	Simulations Rejecting H0		Simulations not Rejecting H0		Total Simulations	
		Count	%	Count	%	Count	%
Unfavorable: CP < 0.300	442.00	9076	28.00%	23344	72.00%	32420	32.42%
Promising: 0.300 <= CP < 0.800	695.28	21898	82.94%	4503	17.06%	26401	26.40%
Favorable: CP >= 0.800	441.96	36003	87.43%	5176	12.57%	41179	41.18%

Simulation Results

- Unconditional power is 67% with average sample size 509
- But conditional of falling in promising zone, power is 83% with average sample size 696.
- Sponsor only has to commit the additional subjects if the interim results are promising, and would then be happy (presumably) to do so
- This design has 26% chance of falling in promising zone and 41% chance of falling in favorable zone, both highly desirable interim outcomes

Unconditional Power

Value of δ	Fixed Sample		Adaptive	
	Power	Expected SampleSize	Power	Expected Sample Size
1.6	61%	442	67%	510
1.7	66%	442	72%	508
1.8	71%	442	76%	505
1.9	76%	442	81%	502
2.0	80%	442	84%	498

Power Conditional on Zone

δ	Interim Outcome	Probability of Interim Outcome	Power Conditional on Interim Outcome		Expected Sample Size	
			Fixed	Adaptive	Fixed	Adaptive
1.6	Unfavorable	33%	27%	27%	442	442
	Promising	26%	61%	83%	442	696
	Favorable	41%	87%	87%	442	442
1.7	Unfavorable	29%	32%	32%	442	442
	Promising	26%	65%	86%	442	694
	Favorable	45%	90%	90%	442	442
1.8	Unfavorable	26%	36%	36%	442	442
	Promising	26%	69%	89%	442	690
	Favorable	48%	91%	91%	442	442
1.9	Unfavorable	23%	41%	41%	442	442
	Promising	25%	72%	91%	442	689
	Favorable	52%	93%	93%	442	442
2.0	Unfavorable	20%	44%	44%	442	442
	Promising	24%	76%	93%	442	685
	Favorable	56%	95%	95%	442	442

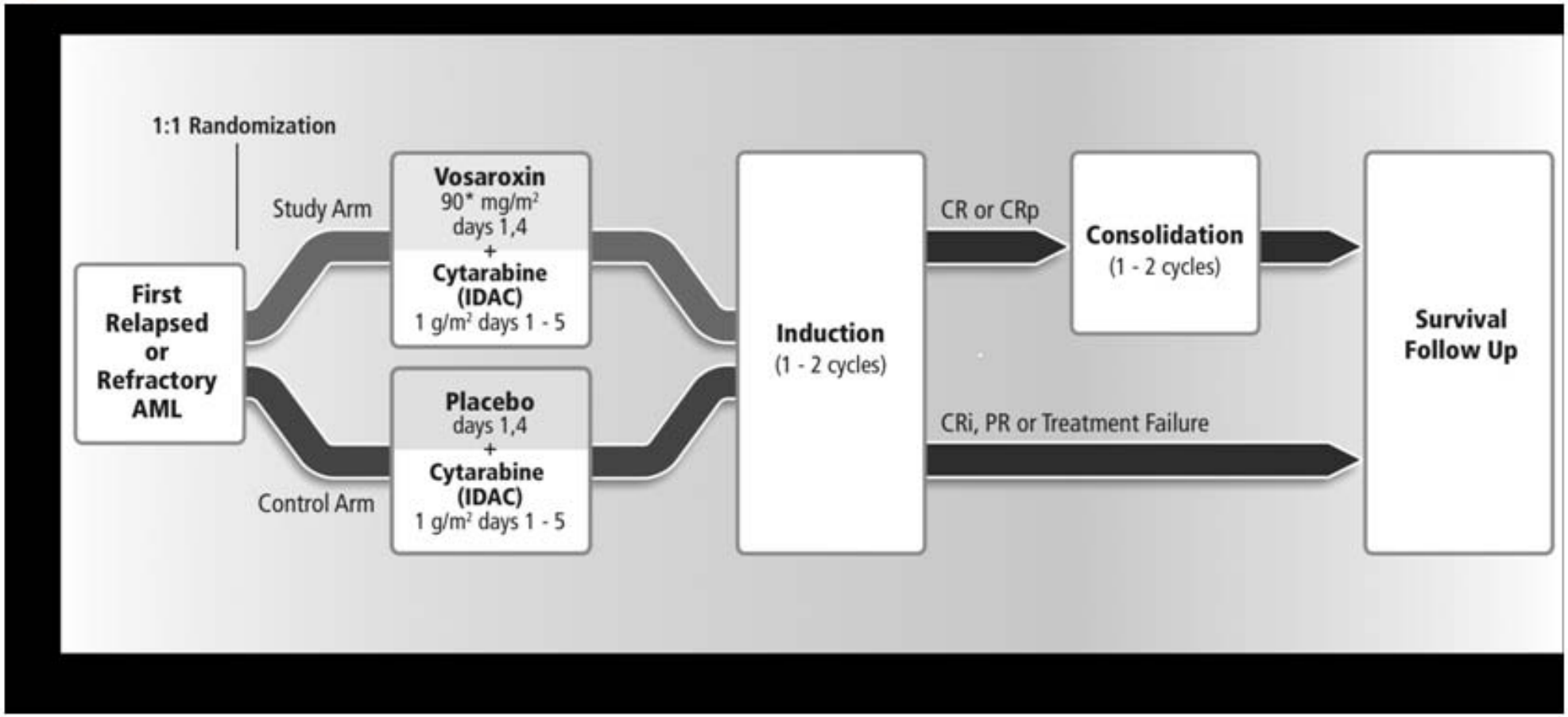
Case 2: Acute Myeloid Leukemia

- **Therapy for relapsed or refractory AML is generally unsatisfactory; no approved drugs; dismal prognosis**
- **Standard of care typically consists of single agent cytarabine or cytarabine-based regimens**
- **NCCN guideline states: new agents need to induce durable remissions with long LFS, improved toxicity profile and improved OS**

The VALOR Trial for AML

- **V**osaroxin and **A**ra-C combination eva**L**uating **O**verall survival in **R**elapsed/refractory AML
- Phase 3, double-blind, placebo-controlled, multinational trial for first-relapsed or refractory Acute Myeloid Leukemia (AML)
- Evaluate efficacy and safety of Vosaroxin plus Cytarabine versus placebo plus Cytarabine
- Vosaroxin is a first-in-class anticancer quinolone derivative under development by Sunesis pharmaceuticals

VALOR Phase 3 Schema



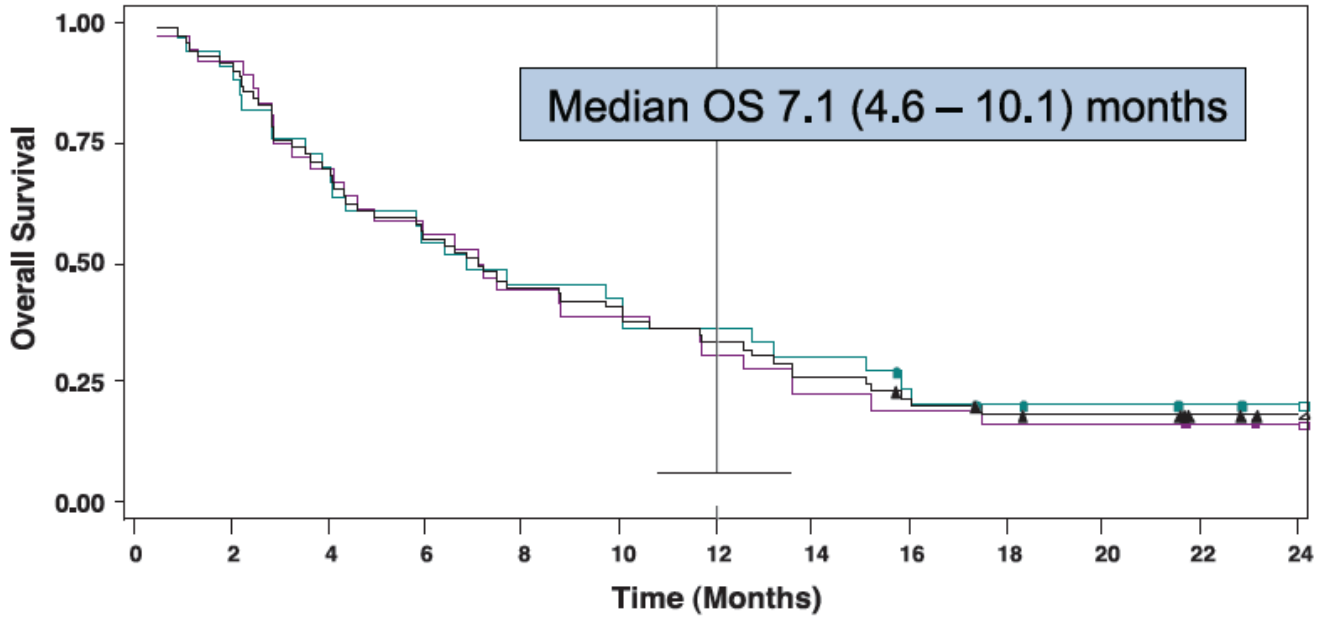
*After cycle 1, all subsequent cycles at 70 mg/m² vosaroxin on days 1 and 4.

Design Objectives

- Primary endpoint is overall survival
- Design for 90% power at 5% significance level
- Complete the trial in 30 months
  - Enroll for 24 months
  - Follow for 6 additional months

Prior Phase 2 Data

- Limited information on Vosaroxin from a single phase 2 trial of 69 patients with no active comparator



- Median OS for Vosaroxin estimated to be 7 months from phase 2 trial
- Median OS for Cytarabine estimated to b 5 months, from meta-analysis of prior studies and consultation with KOLs
- Hazard ratio estimated to be 0.71 amidst considerable uncertainty

Sponsor's Dilemma

- Based on phase 2 data:
  - Assume 5/7 month median on Ctrl/Trtm (HR=0.71)
  - Require 375 events and 450 subjects @ 19/month
- But phase 2 estimates are subject to uncertainty:
  - What if 5/6.5 month median on Ctrl/Trtm (HR=0.77)?
  - HR = 0.77 is still clinically meaningful
  - Require 616 events and 732 subjects @ 31/month
  - Not a feasible option for sponsor

Sponsor is Resource and Time Constrained

True HR	Power if designed with base-case assumption: (HR = 0.71)	Power if designed with alternative assumption: (HR=0.77)
0.71	91%	99%
0.74	83%	97%
0.77	71%	90%
Resources Needed	450 patients @ 19/month	732 patients @ 31/month

Why not design up-front for HR=0.77 (smallest clinically meaningful effect)?

- **Unable to muster resources for large investment with limited phase 2 data**
- **Rule of thumb cost/patient is $50-80K for an oncology trial with OS**
- **Up-front provision for HR=0.77 would overpower study if the more realistic assumption (HR=0.71) holds**

Sponsor Adopts a Strategy of Staged Investment

- Design realistically up-front. Power study to detect HR=0.71 (requires 375 events; 450 subjects @ 19/month)
- One interim analysis after 50% information (187 events)
  - Stop early if overwhelming evidence of efficacy
  - Stop early for futility if low conditional power
  - Increase number of events, sample size and (if possible) rate of recruitment at the interim if results are promising

Key Idea: Invest additional resources and re-power the study to detect HR=0.77 only after seeing interim results

The Promising Zone Design

- Partition the interim outcome into three zones based on the interim estimate of conditional power. For example:

  Unfavorable: HR_hat ≥ 0.86; no change to design

  Promising: $0.74 \leq$ HR_hat < 0.86; increase resources

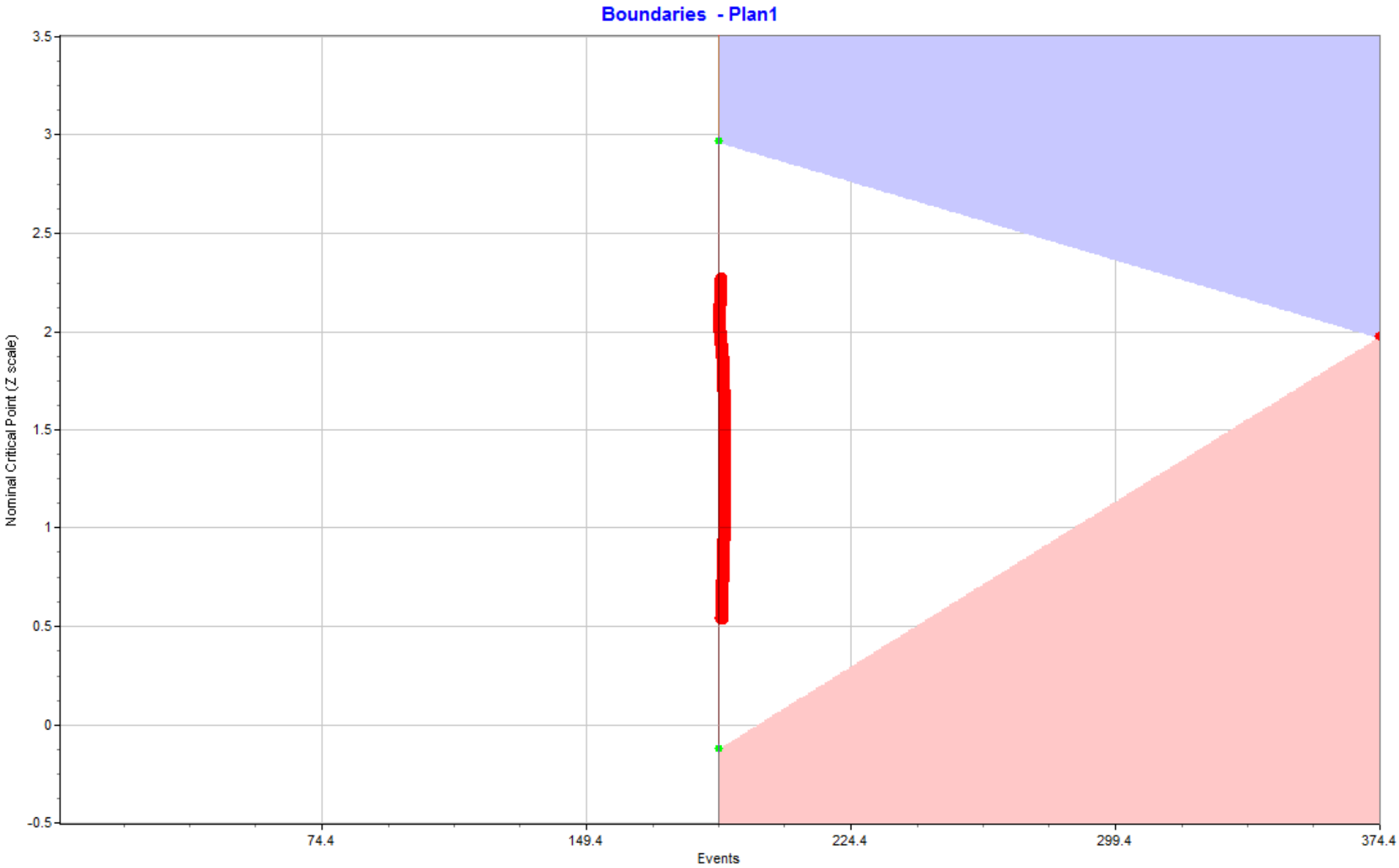
  Favorable: HR_hat ≤ 0.74; no change to design

- Control type-1 error by using Cui, Hung and Wang (1999) weighted statistic modified for survival data

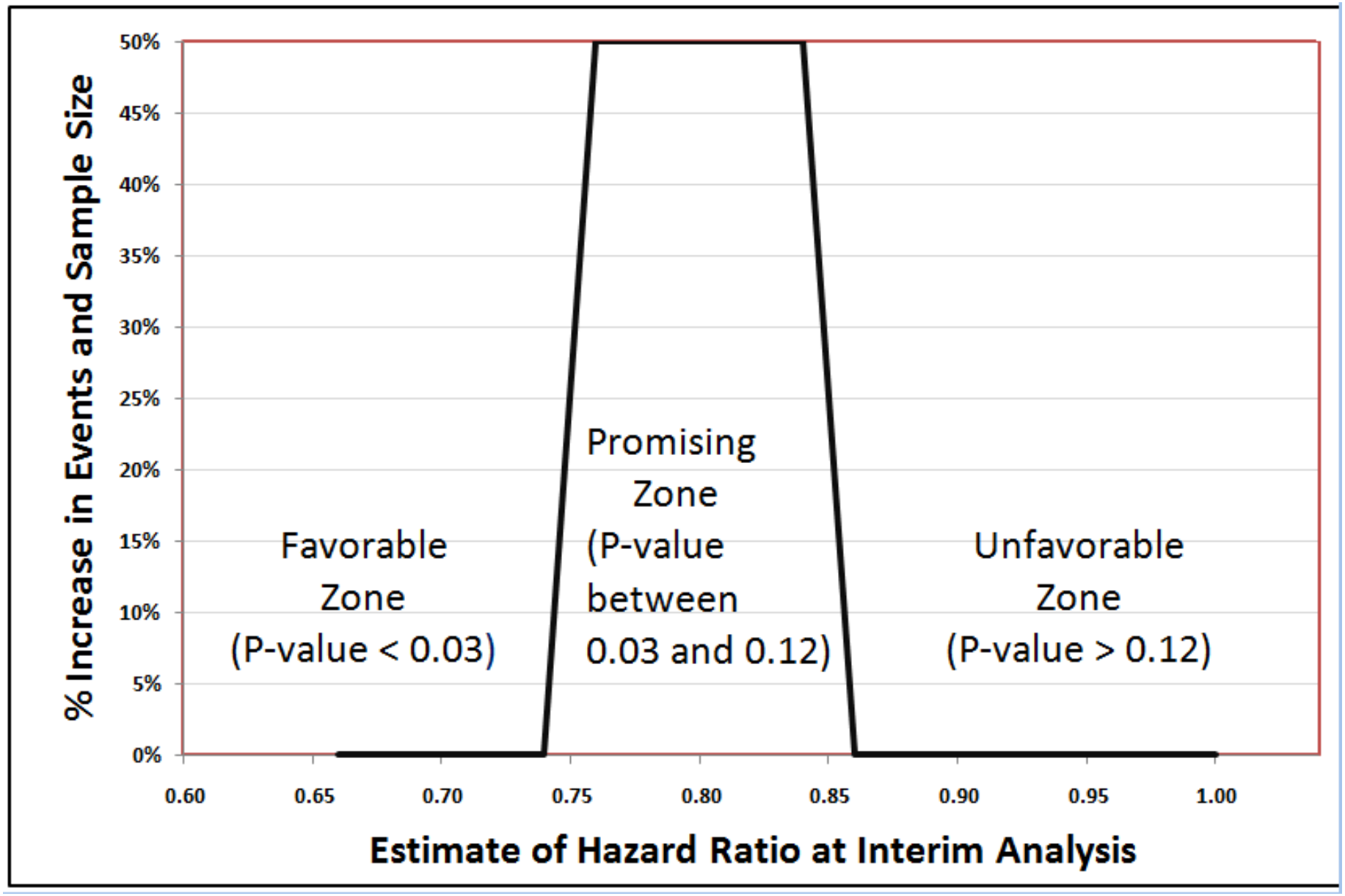
- Evaluate operating characteristics of design by simulation

Note: The cutoffs specified above for the three zones are not the actual cut-offs used in the trial. The actual cut-offs are confidential information

Adaptive Decision Rule: Representation I



Adaptive Decision Rule: Representation II



Preserving the Type-1 Error

- Let D_1 and D_2 be the pre-specified total events at interim and final analysis. (Here $D_1 = 187$ and $D_2 = 375$)
- Let LR_1 and LR_2 be the corresponding logrank statistics
- Suppose D_2 is altered to $D_2^* > D_2$ at the interim
- Let LR_2^* denote the corresponding altered logrank statistic
- Type-1 error is preserved if we use

$$Z_{\text{chw}} = \sqrt{\frac{D_1}{D_2}} \times \text{LR}_1 + \sqrt{\frac{D_2 - D_1}{D_2}} \times \frac{\sqrt{D_2^*}\text{LR}_2^* - \sqrt{D_1}\text{LR}_1}{\sqrt{D_2^* - D_1}}$$

  instead of LR_2^* for the final analysis

Adaptation Principles

- Primary driver of power is number of events
- FDA guidance recommends increase only, not decrease
- Increase events by amount needed to achieve some target conditional power, subject to a cap
- Compute sample size increase necessary to achieve the desired increase in events without undue prolongation of the trial
- Complex relationship exists between increase in events, increase in sample size and study duration. Best evaluated by simulation

Simulate the Design

Adaptation at Look L		1	
Max. Events if Adapt (multiplier; total #)		1.50	561
Max. # of Subjects if Adapt (multiplier, total #)		1.50	677
Upper Limit on Study Duration		90.00	
Shape Parameter for Re-estimating # of Events		0.99	
Promising Zone :	Min CP:	0.30	
	Max CP:	0.90	
HR Used in CP Computations	Estimated HR		
Accrual Rate After Adaptation	No Change		

Show Summary for: Promising

Percentile	Study Duration	Number of Events	Accrual Duration	Number of Subjects
5%	37.0	561	35.1	677
25%	37.8	561	35.6	677
50%	38.3	561	36.0	677
75%	38.9	561	36.3	677
95%	39.7	561	36.7	677
Average	38.3	561	35.9	677

Run | Single Step | Reset | Stop

Simulation Results by Zone

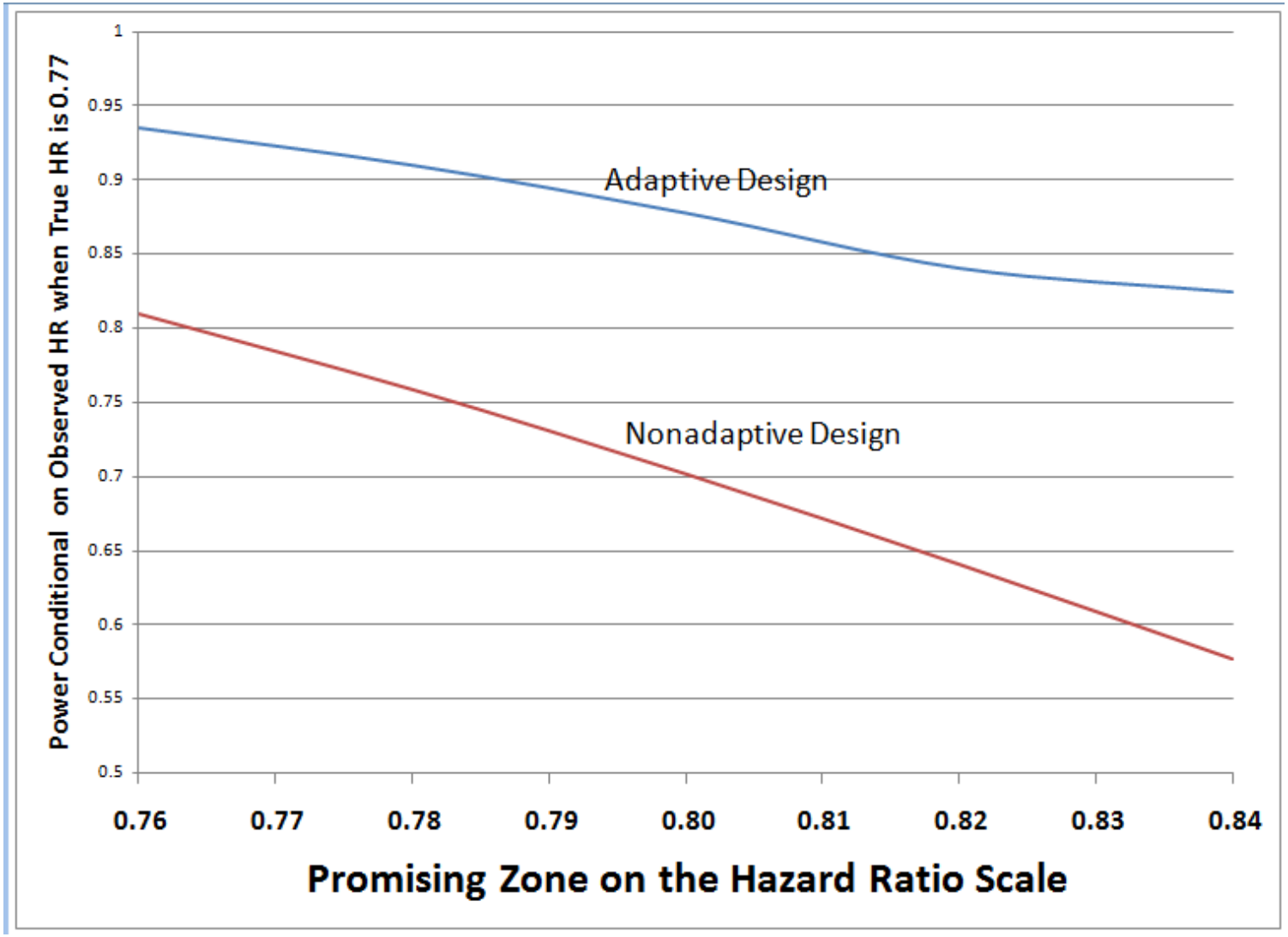
	Zone	Simulations Rejecting H0		Simulations Rejecting H1		Total Simulations		Avg. Study Duration	Avg. Number of Events	Avg. Accrual Duration	Avg. Number of Subjects
		Count	Row %	Count	Row %	Count	Column %				
+	Unfavorable + Futility	858	33.28%	1720	66.72%	2578	25.78%	28.0	355	23.2	437
	Promising: 0.300 ≤ CP < 0.900	3040	89.36%	362	10.64%	3402	34.02%	38.3	561	35.9	677
+	Favorable + Efficacy	3774	93.88%	246	6.12%	4020	40.20%	25.8	319	21.9	413
	All Trials	7672	76.72%	2328	23.28%	10000	100.00%	30.6	410	27.0	509

Operating Characteristics

Under Pessimistic Scenario, HR = 0.77 (10,000 simulations)							
		Power		Duration (months)		SampSize	
Zone	P(Zone)	NonAdpt	Adapt	NonAdpt	Adapt	NonAdpt	Adapt
Unf	25%	33%	33%	28	28	436	439
Prom	34%	71%	90%	29	38	453	680
Fav	41%	95%	95%	26	26	414	413
Total	—	71%	78%	28	31	432	509

- Two-stage investment
- Sponsor unable to invest resources needed for 90% unconditional power at HR=0.77; too risky
- But, if stage-1 results from 172 events (375 subjects) are promising, sponsor can invest needed resources to boost power to 90% at greatly reduced risk

Power Curves of Adaptive and Non-adaptive Designs in Promising Zone



Attractiveness of Approach

- Up-front sample size investment can be modest
- Additional investment is only made if interim results are promising
- If that happens, chances of success are dramatically increased

In Summary: Adaptive design is accompanied by adaptive financing

Milestone-Driven Investment

Sunesis Pharmaceuticals to Implement One-Time Sample Size Increase to Phase 3 VALOR Trial in AML. DSMB Recommends Increase Following Single, Pre-Planned Interim Efficacy and Safety Analysis of VALOR; Enrollment Completion Expected in 2013. DSMB Recommendation Triggers $25.0 Million Investment in Sunesis from Royalty Pharma.

Press Release, September 11, 2012. Sunesis Pharma, South San Francisco

Part 2: Theory of Adaptive Sample Size Reestimation

Ways to Preserve Type-1 Error

1. Use CHW statistic with pre-specified weighting of data from each stage (Cui, Hung & Wang, 1999)
2. Use conventional Wald test if promising interim result are obtained (Chen, DeMets, Lan, 2004; Gao, Ware, Mehta, 2008) (Only valid for two-stage designs)
3. Preserve the conditional type-1 error that would have been obtained had there been no adaptation (Muller & Schafer, 2001)

Method 1. Pre-specified Weights

Method 1. Pre-specified Weights

1. Hypothesis Testing Without Sample Size Increase

STAGE I	STAGE II
sample size n_1	sample size $n^{(2)}$
estimate $\hat{\delta}_1$	estimate $\hat{\delta}^{(2)}$
compute $z_1 = \hat{\delta}_1 / \text{se}(\hat{\delta}_1)$	compute $z^{(2)} = \hat{\delta}^{(2)} / \text{se}(\hat{\delta}^{(2)})$

Reject H_0 if $\sqrt{\frac{n_1}{n_1+n^{(2)}}} z_1 + \sqrt{\frac{n^{(2)}}{n_1+n^{(2)}}} z^{(2)} \geq C_\alpha$

Method 1. Pre-specified Weights

1. Hypothesis Testing Without Sample Size Increase

STAGE I	STAGE II
sample size n_1	sample size $n^{(2)}$
estimate $\hat{\delta}_1$	estimate $\hat{\delta}^{(2)}$
compute $z_1 = \hat{\delta}_1/\text{se}(\hat{\delta}_1)$	compute $z^{(2)} = \hat{\delta}^{(2)}/\text{se}(\hat{\delta}^{(2)})$

Reject H_0 if $\sqrt{\frac{n_1}{n_1+n^{(2)}}}z_1 + \sqrt{\frac{n^{(2)}}{n_1+n^{(2)}}}z^{(2)} \geq C_\alpha$

2. Hypothesis Testing With Sample Size Increase

STAGE I	STAGE II WITH EXTENSION
sample size n_1	sample size $n^{*(2)} > n^{(2)}$
estimate $\hat{\delta}_1$	estimate $\hat{\delta}^{*(2)}$
compute $z_1 = \hat{\delta}_1/\text{se}(\hat{\delta}_1)$	compute $z^{*(2)} = \hat{\delta}^{*(2)}/\text{se}(\hat{\delta}^{*(2)})$

Reject H_0 if $\sqrt{\frac{n_1}{n_1+n^{(2)}}}z_1 + \sqrt{\frac{n^{(2)}}{n_1+n^{(2)}}}z^{*(2)} \geq C_\alpha$

Summary

- Although we have extended the sample size from $n^{(2)}$ to $n^{*(2)}$ at stage-II, the criterion for declaring statistical significance is

$$\sqrt{\frac{n_1}{n_1+n^{(2)}}}z_1+\sqrt{\frac{n^{(2)}}{n_1+n^{(2)}}}z^{*(2)}\geq C_\alpha$$

  instead of

$$\sqrt{\frac{n_1}{n_1+n^{*(2)}}}z_1+\sqrt{\frac{n^{*(2)}}{n_1+n^{*(2)}}}z^{*(2)}\geq C_\alpha$$

- Contribution of the second cohort of patients has been down-weighted
- Also known as the method of "inverse normal weighting of p-values" because significance criterion can be expressed in the form

$$\sqrt{\frac{n_1}{n_1+n^{(2)}}}\Phi^{-1}(1-p_1)+\sqrt{\frac{n^{(2)}}{n_1+n^{(2)}}}\Phi^{-1}(1-p^{*(2)})\leq\Phi^{-1}(1-p_\alpha)$$

Generalize to K-Stage Trial

- Design a K-look group sequential design with boundaries $b_1, b_2, \ldots b_K$ at cumulative sample sizes $n_1, n_2, \ldots n_K$
- Suppose these sample sizes are changed over the course of the trial to $n_1^*, n_2^*, \ldots n_K^*$
- Define the incremental sample sizes $n^{(j)} = (n_j - n_{j-1})$ and $n^{*(j)} = (n_j^* - n_{j-1}^*)$, $j = 1, 2, \ldots K$
- Define the weights $w_j = (n^{(j)}/n_K)$ and $w_j^* = (n^{*(j)}/n_K^*)$
- Let $Z^{*(j)}$ be the incremental Wald statistic based only on the $n^{(*j)}$ new observations between looks $(j-1)$ and j

The CHW Statistic

- The CHW statistic is

$$Z^*_{j,\text{chw}} = \frac{\sqrt{w^{(1)}}Z^{*(1)} + \sqrt{w^{(2)}}Z^{*(2)} + \ldots + \sqrt{w^{(j)}}Z^{*(j)}}{\sqrt{w^{(1)} + w^{(2)} + \ldots + w^{(j)}}}$$

- This statistic is asymptotically normally distributed with mean

$$E(Z^*_{j,\text{chw}}) = \frac{\delta \sum_{l=1}^{j} \sqrt{w^{(l)} I^{*(l)}}}{\sqrt{\sum_{l=1}^{j} w^{(l)}}}$$

  and unit variance, where $I^{*(l)}$ is the incremental information at look l

- Cui, Hung and Wang (1999) and Lehmacher and Wassmer (1999) have shown that

$$P_0(\bigcup_{j=1}^{K} |Z^*_{j,\text{chw}}| \geq b_j) = \alpha \ .$$

- Note: If no sample size change, then $Z_{j,\text{chw}} = Z_{j,\text{wald}}$

RCI's and RPV's

- The repeated confidence intervals (RCI's) at the K looks are given by

$$\frac{(Z^*_{j,\textsf{chw}} \pm b_j)\sqrt{s_j}}{\sum_{l=1}^{j}\sqrt{w^{(l)}I^{*(l)}}},\ j = 1, 2, \ldots K$$

  where $s_i = (n_i/n_K)$

- If δ_0 is the true value of δ then, for all $j = 1, 2, \ldots K$

$$P_{\delta_0}\left\{\bigcap_{i=1}^{j}\left(\frac{(Z^*_{i,\textsf{chw}} - b_i)\sqrt{s_i}}{\sum_{l=1}^{i}\sqrt{w^{(l)}I^{*(l)}}} \le \delta_0 \le \frac{(Z^*_{i,\textsf{chw}} + b_i)\sqrt{s_i}}{\sum_{l=1}^{i}\sqrt{w^{(l)}I^{*(l)}}}\right)\right\} = 1 - \alpha$$

- The repeated p-value at any look j is obtained by iteratively changing α until one of the two extremes of the look j RCI just excludes zero

Conditional Power Calculations

- Adaptive sample size changes are commonly driven by conditional power or the probability, given the current data, of attaining statistical significance by the end of the study
- Suppose we are at some look $L < K$ and the observed value of the test statistic is z_L. Conditional power for a given value of δ and total sample size n_K^* is defined as

$$\mathrm{CP}_\delta(z_L, n_K^*) = P_\delta\{\bigcup_{j=L+1}^{K} (Z_{j,\mathrm{chw}}^* \geq b_j | z_L)\}$$

- One may use either the value of δ specified at the design stage or the value $\hat{\delta}_L$ estimated at look L in the above expression for CP

Handling Time to Event Data

- Test H_0: $\delta = 0$ versus $\delta > 0$, where $\delta = -\ln(\text{HR})$
- Pre-specify that H_0 will be tested by a K-look GSD with monitoring at $D_1, D_{2,\ldots}D_K$ cumulative events
- Data dependent alteration of events to D_j^*, $j = 1, 2, \ldots$
- The incremental statistics,

$$Z^{*(j)} = \frac{\sqrt{D_j^*}\text{LR}_j^* - \sqrt{D_{j-1}^*}\text{LR}_{j-1}^*}{\sqrt{D_j^* - D_{j-1}^*}}, \text{ for } j = 1, 2, \ldots K,$$

are independent and asymptotically normal with mean

$$\text{E}(Z^{*(j)}) = \delta\sqrt{r(1-r)D^{*(j)}}$$

and unit variance, where r is the randomization fraction and LR_j is the logrank statistic

CHW Statistic for Time to Event Data

- Pre-specify the weights for the combination function

$$w^{(j)} = \frac{D^{(j)}}{D_{(K)}}, \text{ for } j = 1, 2, \ldots K,$$

- The CHW statistic is defined in the usual way as

$$Z^*_{j,\text{chw}} = \frac{\sqrt{w^{(1)}}Z^{*(1)} + \sqrt{w^{(2)}}Z^{*(2)} + \ldots + \sqrt{w^{(j)}}Z^{*(j)}}{\sqrt{w^{(1)} + w^{(2)} + \ldots + w^{(j)}}}$$

- All results derived for the normal case hold with $\delta = -\ln(\text{HR})$, $\sigma = 1$, $D^{*(j)}$ replaceing $n^{*(j)}$, and $I^{*(j)} = r(1-r)D^{*(j)}$

Method 2: Use the Conventional Wald Statistic

- Result is due to Chen, Demets and Lan (2004) (CDL method)
- Valid only for two-stage designs in which the sample size may be increased, but not decreased at the interim look
- Use conventional Wald statistic for the final analysis even if the sample size was increased from n_2 to n_{2^*}, provided the interim results were promising
- Interim result is considered promising if $\text{CP}_{\hat{\delta}}(z_1, n_2) \geq 0.5$

Extended CDL Method

- Due to Gao, Ware and Mehta (2008) and Mehta and Pocock (2010)
- Can relax the criterion for a using conventional Wald statistic if $\mathrm{CP}_{\hat{\delta}_1}(z_1, n_2) \geq \mathrm{CP}_{\mathrm{min}}$ as tabulated below:

Sample Size Ratios		$\mathrm{CP}_{\mathrm{min}}$ Values for Targeted Conditional Powers		
Maximum Allowed (N^*_{max}/n_2)	At Interim Look (n_1/n_2)	80%	90%	95%
1.5	0.25	0.42	0.42	0.42
1.5	0.5	0.41	0.41	0.41
1.5	0.75	0.38	0.38	0.38
2	0.25	0.37	0.37	0.37
2	0.5	0.36	0.36	0.36
2	0.75	0.33	0.33	0.33
3	0.25	0.32	0.32	0.32
3	0.5	0.31	0.31	0.30
3	0.75	0.30	0.27	0.27
∞	0.25	0.32	0.28	0.26
∞	0.5	0.31	0.27	0.25
∞	0.75	0.30	0.25	0.23

Advantages of CDL Method

- Uses the familiar Wald statistic instead of the weighted CHW statistic
- Weighting scheme for the two stages of the trial is based on actual enrollment and not enrollment pre-specified at start
- Standard p-value (but not standard confidence interval) can be used at the end of the trial

Method 3. Preserve Conditional Type-1 Error

- Due to Muller and Schafer (2001)
- This method is the most flexible of all
- It gives full freedom to completely re-design a group sequential trial at any interim look. You could:
  - increase the sample size
  - change the spending function
  - alter the number and spacing of future interim looks
- Only Requirement: Preserve the conditional type-1 error computed at the time of the design modification

Preserving the Overall Type-1 Error

In order to preserve the overall type-1 error of this procedure:

1. **Compute what the conditional type-1 error would be if you were to go to the end of the trial without re-designing**
2. **Use this conditional type-1 error as the significance level for the re-designed trial**

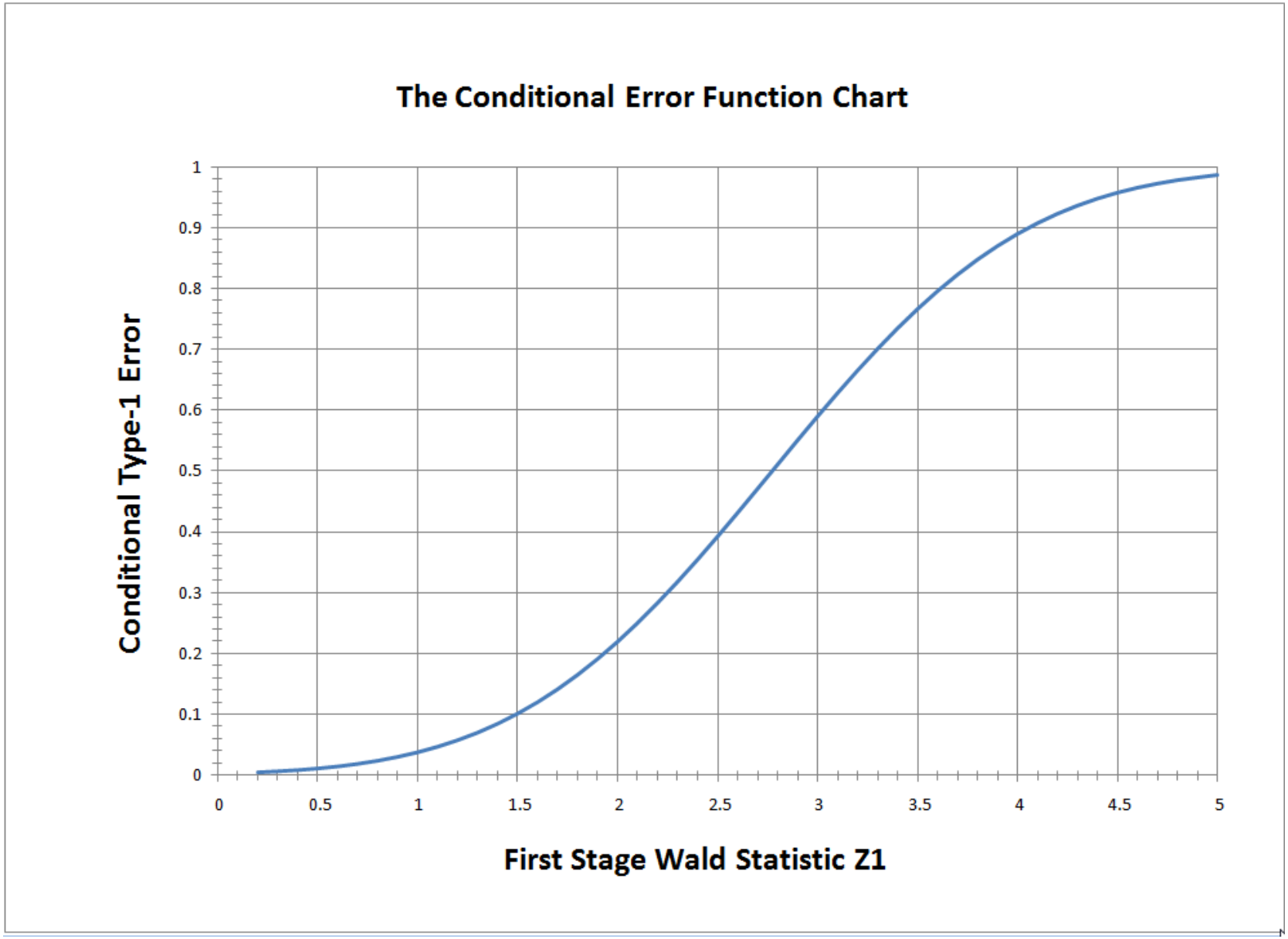
Example: Two-Stage Design

- Normal endpoint, no early stopping
- At 50% information observe $Z_1 = z_1$
- Compute conditional type-1 error

$$\mathrm{CP}_0 = P_0(Z_2 \geq 1.96 | z_1)$$

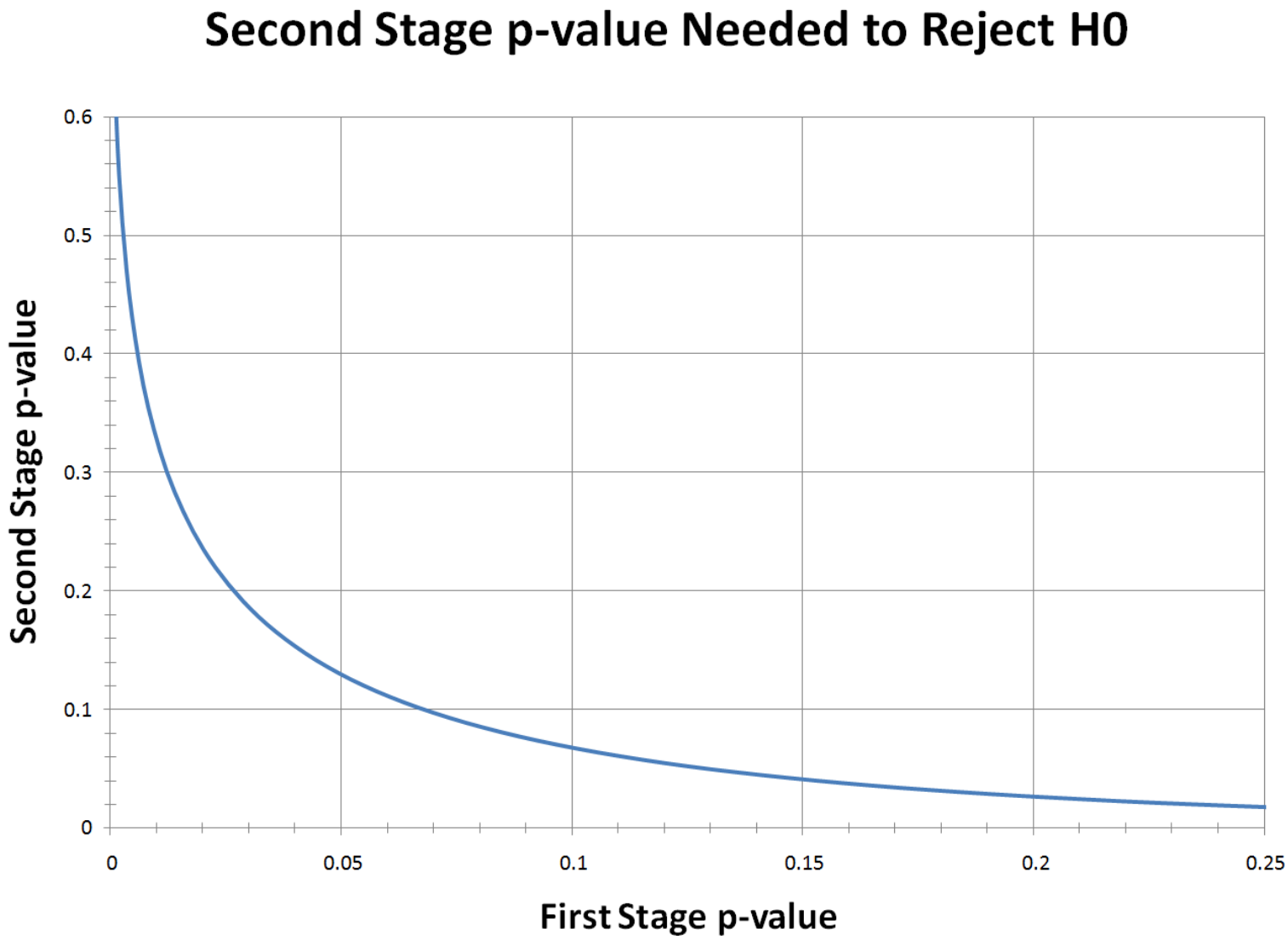
- This will be your available α for the remainder of the trial regardless of the type of adaptation for the second stage

Plot of z_1 versus CP_0 at 50% Information



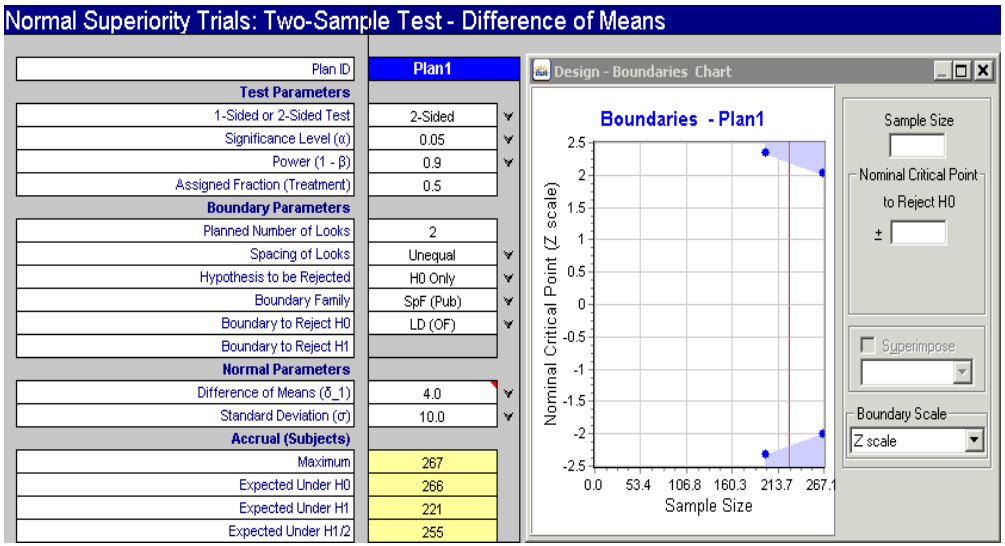
Conditional Error Function

Obtained by transforming the X axis to the p-value scale

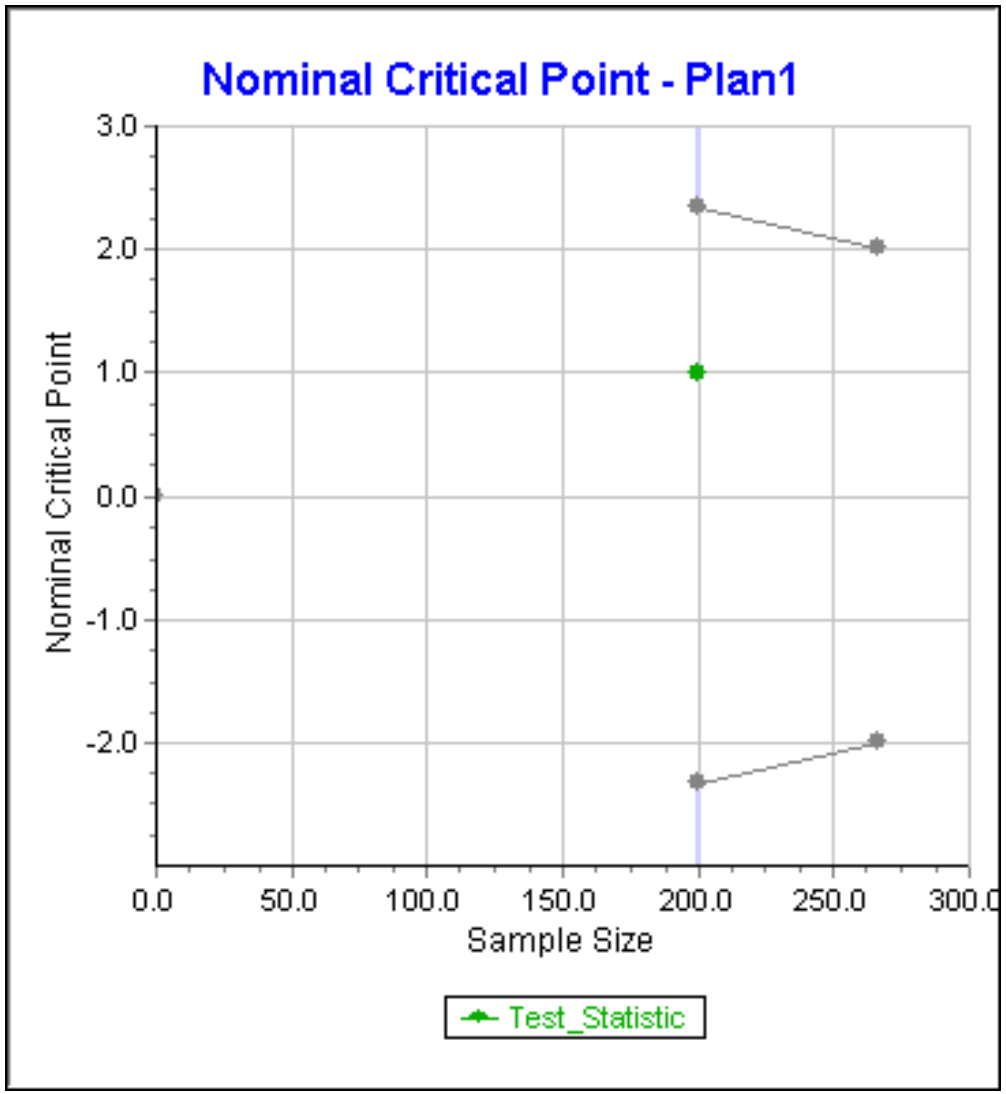


Example of Muller and Schafer Method

- **Normal endpoint trial designed for 90% power to detect $\delta = 4$ with a 2-sided level-.05 test. Initially assume that $\sigma = 10$**
- **Two-look group sequential design with `LD(OF)` spending function and one interim look at 75% of full information**
- **Adapt the sample size and number of future looks at the time of the interim look**

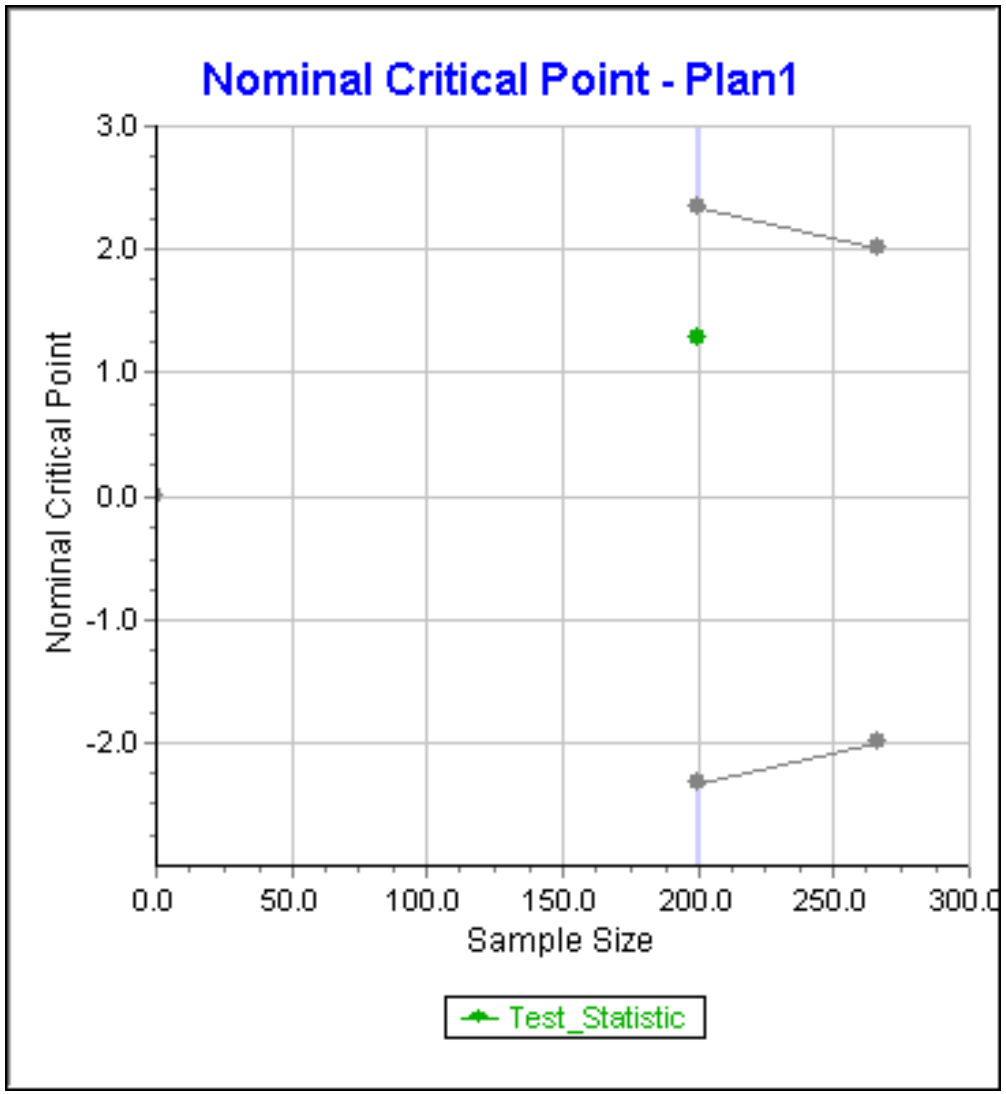


Set the type-1 error of the re-designed trial equal to conditional type-1 error – shown in the table as $\mathrm{CP}_0(z_1, 267)$



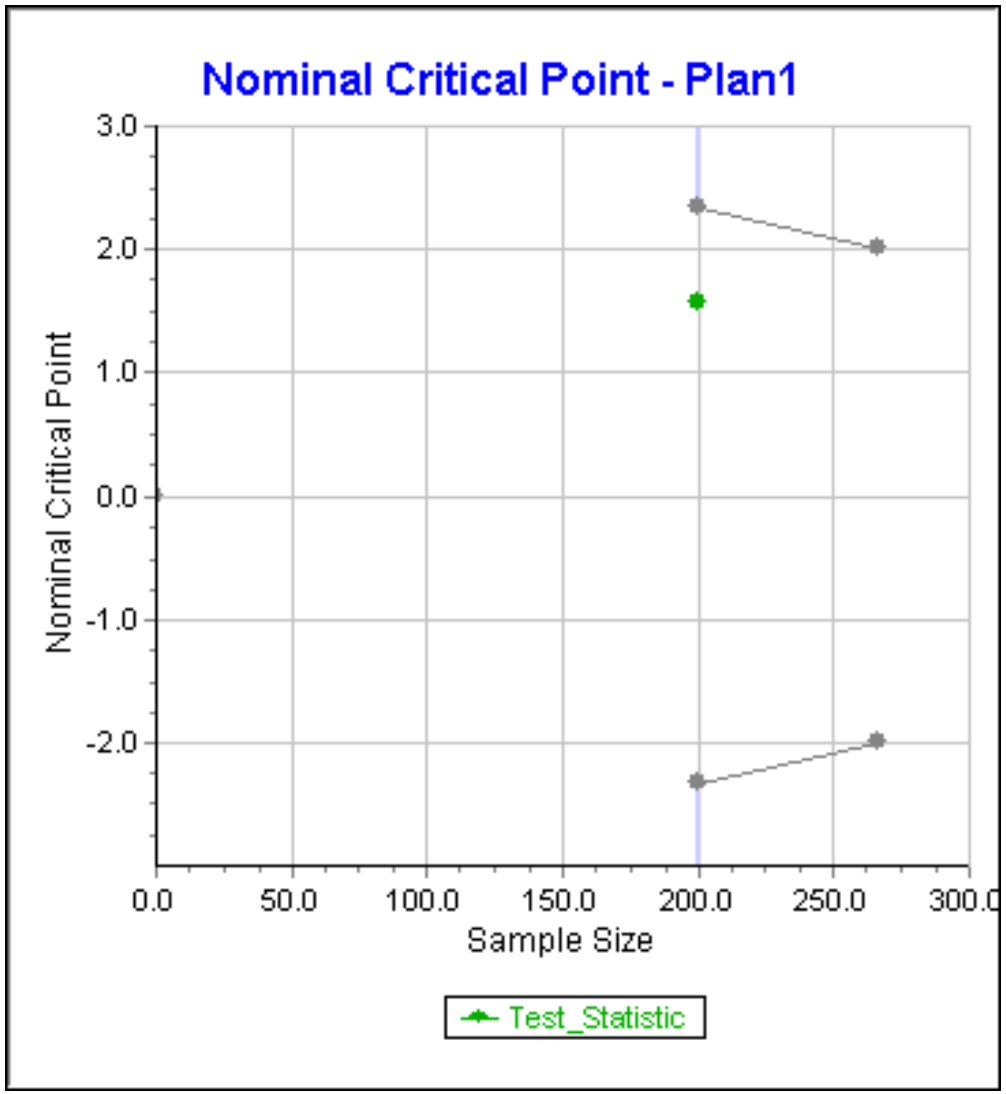
$\hat{\delta}_1$	z_1	$\mathrm{CP}_{\hat{\delta}_1}(z_1, 267)$	$\mathrm{CP}_0(z_1, 267)$
1.4	0.99	0.042	0.011

Set the type-1 error of the re-designed trial equal to conditional type-1 error – shown in the table as $\mathrm{CP}_0(z_1, 267)$



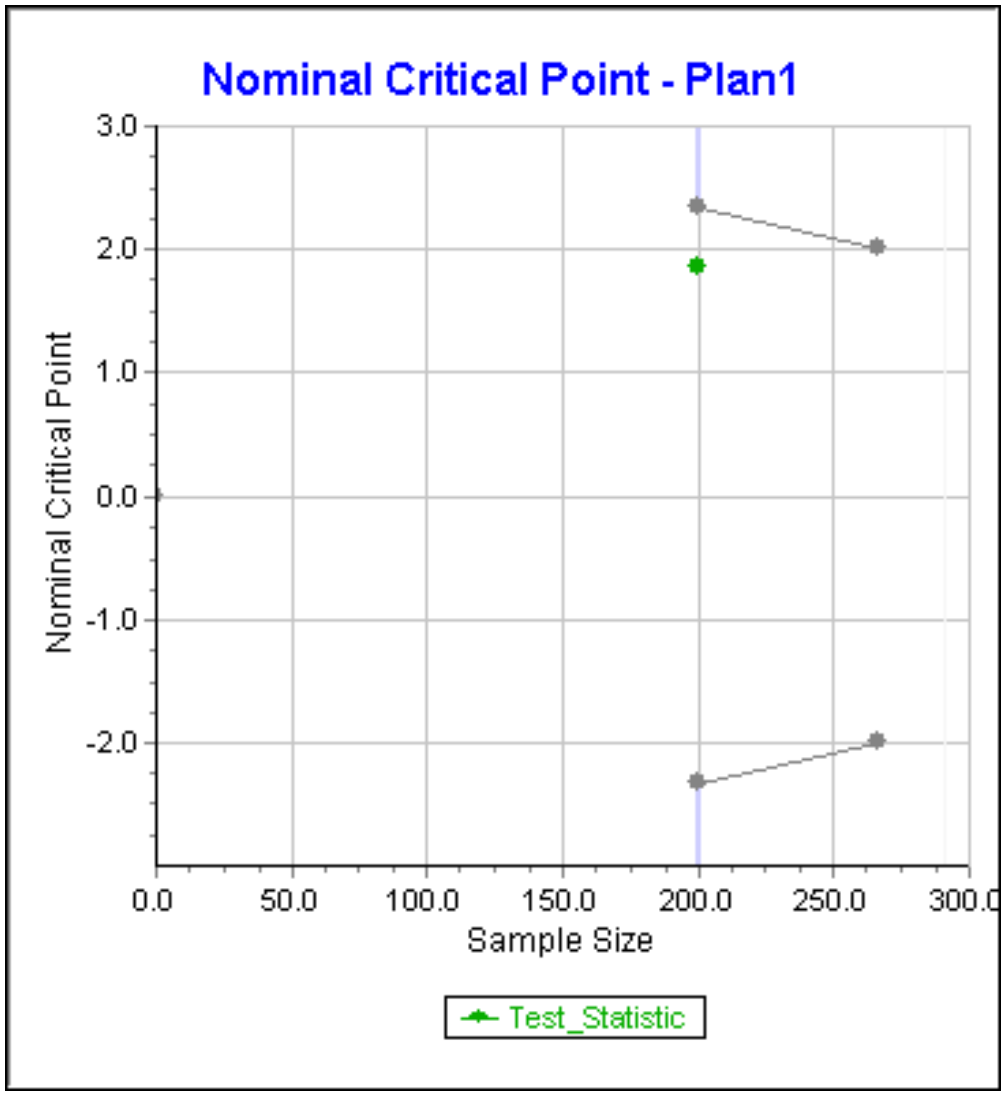
$\hat{\delta}_1$	z_1	$\mathrm{CP}_{\hat{\delta}_1}(z_1, 267)$	$\mathrm{CP}_0(z_1, 267)$
1.4	0.99	0.042	0.011
1.8	1.27	0.14	0.035

Set the type-1 error of the re-designed trial equal to conditional type-1 error – shown in the table as $\mathrm{CP}_0(z_1, 267)$



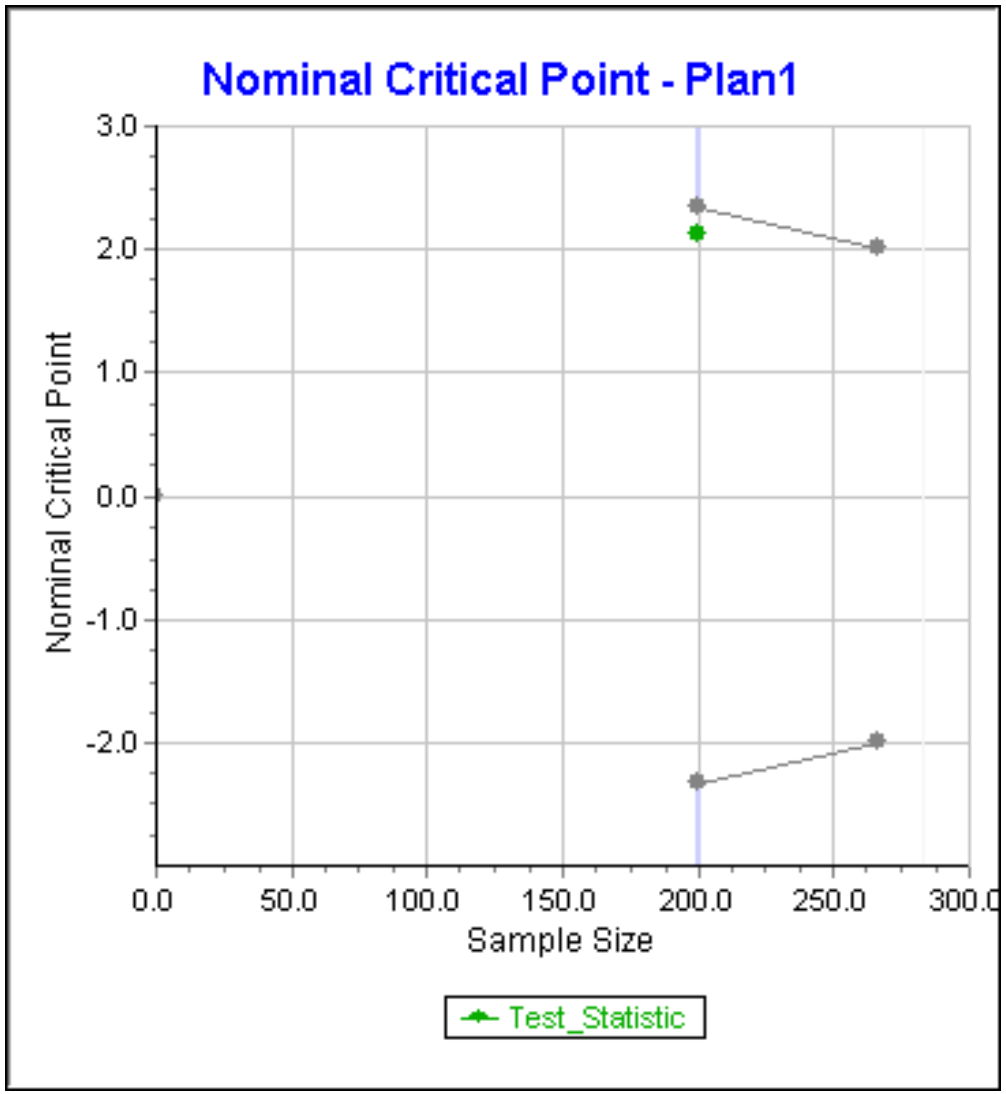
$\hat{\delta}_1$	z_1	$\mathrm{CP}_{\hat{\delta}_1}(z_1, 267)$	$\mathrm{CP}_0(z_1, 267)$
1.4	0.99	0.042	0.011
1.8	1.27	0.14	0.035
2.2	1.56	0.335	0.092

Set the type-1 error of the re-designed trial equal to conditional type-1 error – shown in the table as $\mathrm{CP}_0(z_1, 267)$



$\hat{\delta}_1$	z_1	$\mathrm{CP}_{\hat{\delta}_1}(z_1, 267)$	$\mathrm{CP}_0(z_1, 267)$
1.4	0.99	0.042	0.011
1.8	1.27	0.14	0.035
2.2	1.56	0.335	0.092
2.6	1.84	0.589	0.201

Set the type-1 error of the re-designed trial equal to conditional type-1 error – shown in the table as $\mathrm{CP}_0(z_1, 267)$



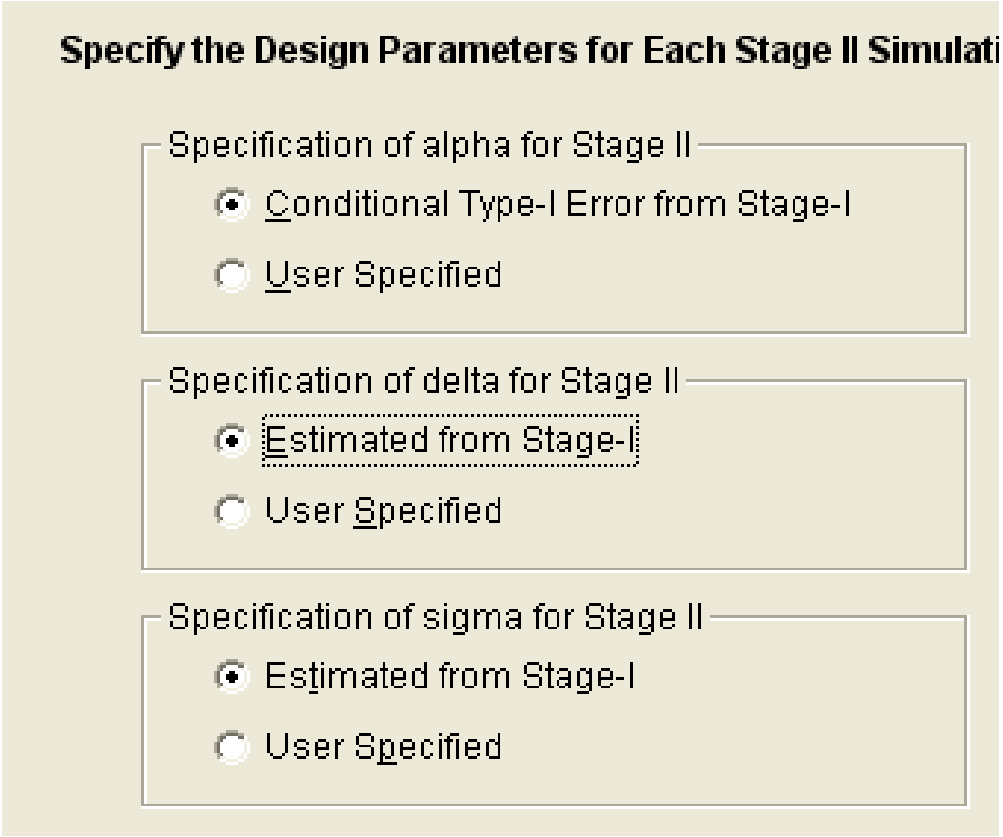
$\hat{\delta}_1$	z_1	$\mathrm{CP}_{\hat{\delta}_1}(z_1, 267)$	$\mathrm{CP}_0(z_1, 267)$
1.4	0.99	0.042	0.011
1.8	1.27	0.14	0.035
2.2	1.56	0.335	0.092
2.6	1.84	0.589	0.201
3.0	2.12	0.810	0.363

Verify Properties by Simulation

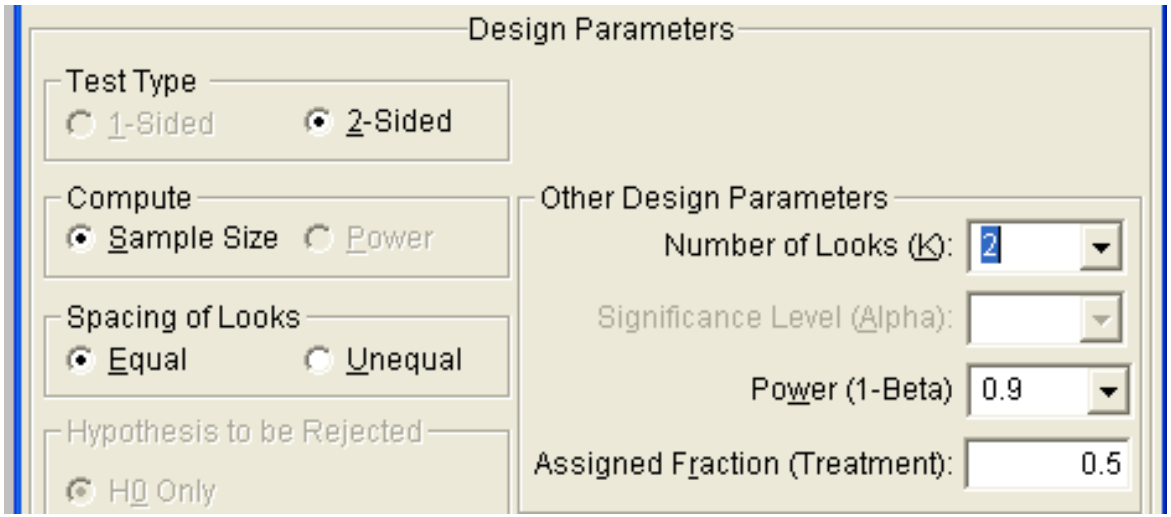
- Take interim look after 200 subjects and compute conditional power $\text{CP}_{\hat{\delta}_1}(z_1, 267)$
- If $\text{CP}_{\min} \leq \text{CP}_{\hat{\delta}_1}(z_1, 267) < \text{CP}_{\max}$, adapt as follows:
  - increase the total number of looks from 2 to 3
  - change spending function from O'Brien-Fleming to Pocock
  - increase total sample size to n_3^* such that $\text{CP}_{\hat{\delta}_1}(z_1, n_3^*) = 0.9$, subject to $267 < n_3^* \leq N_{\max}^*$
- If $\text{CP}_{\hat{\delta}_1}(z_1, n_3^*) < \text{CP}_{\min}$ or $\text{CP}_{\hat{\delta}_1}(z_1, n_3^*) \geq \text{CP}_{\max}$, do not adapt

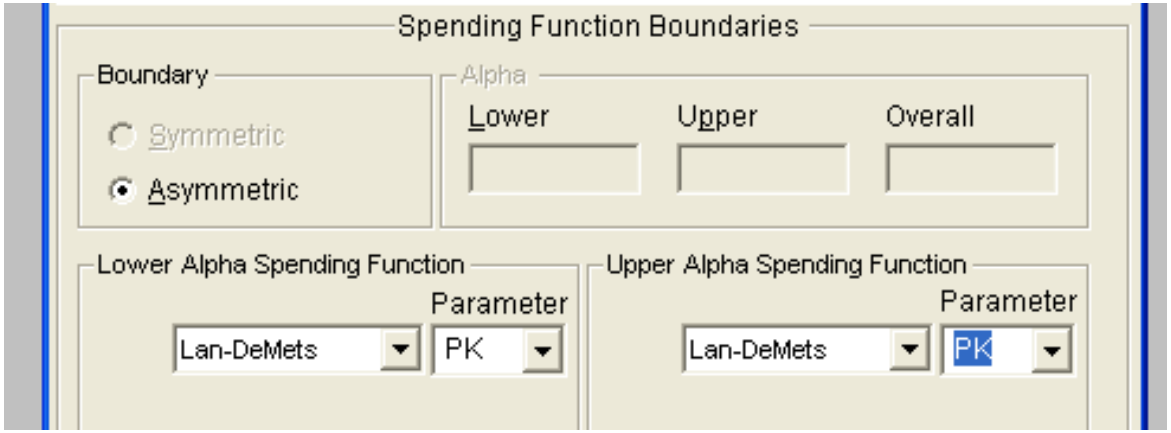
Parameters for Adapted Trial

- α must be current conditional type-1 error
- δ and σ could be currently estimated values or user specified



Number of Future Looks and Spending Function





The Simulation Worksheet

Plan Details for Initial Design	
1-Sided or 2-Sided Test	2-Sided
Significance Level (α)	0.05
Power (1 - β)	0.9
Assigned Fraction (Treatment)	0.5
Planned Number of Looks	2
Spacing of Looks	Unequal
Hypothesis to be Rejected	H0 Only
Boundary Family	SpF (Pub)
Boundary to Reject H0	LD (OF)
Difference of Means (δ_1)	4.0
Standard Deviation (σ)	10.0
Maximum Sample Size	267
Max. Information (Imax)	0.6679

Simulation Boundaries for Combined Trial

Look #	Boundary				
	Sample Size	H0-	H0+	H1-	H1+
1	200.00	-2.3415	2.3415		
2	267.00	-2.0115	2.0115		
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					

Show Stage II Design

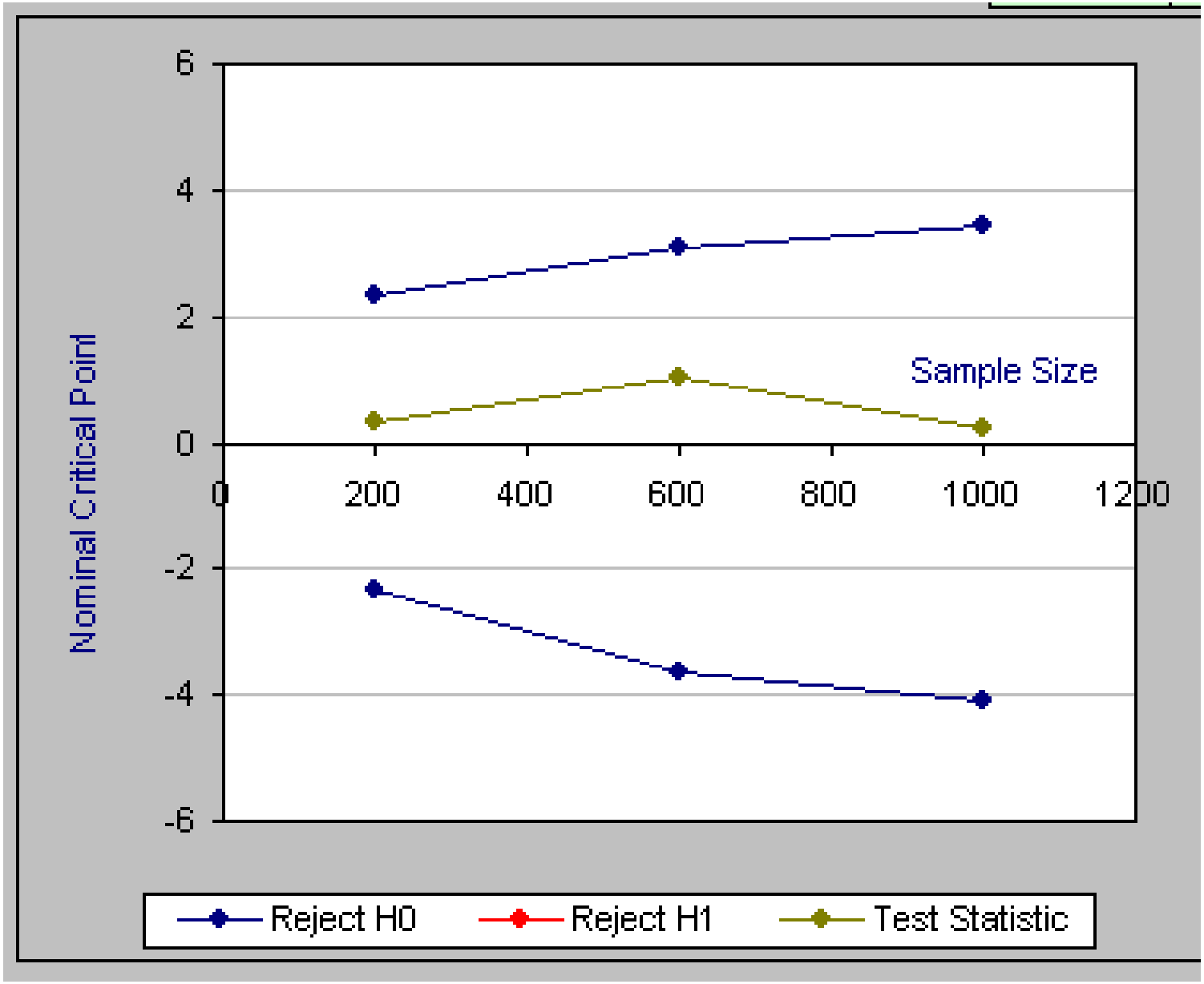
Simulation Boundaries for Stage II Trial

1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Simulation Parameters	
Difference of Means	0.0000
Standard Deviation	10.0000
Use Initial Plan Till 'L' Looks, L =	1
Re-design Criterion: Min. CP	0.0000
Max. CP	0.9000
Min. Sample Size if Re-design	267
Max. Sample Size if Re-design	1000
Number of Trials	10000
Refresh Every 'n' Trials, n =	1000
Simulation Starting Seed	Clock

Show Simulation Results by Zone

Cytel STATISTICAL SOFTWARE & SERVICES


6
4
2
0
-2
-4
-6
0
200
400
600
800
1000
1200
Nominal Critical Point
Sample Size
Reject H0
Reject H1
Test Statistic

Simulation Results for Combined Trial						
Avg. Info.	Avg. Sample Size	# Rejecting H0		# Unable to Reject H0	Total Simulations	
		Up (H0+)	Low (H0-)		Count	%
0.50	200.00	105	93		198	1.98%
0.67	269.46	101	153	4863	5117	51.17%
2.40	958.66	47		4638	4685	46.85%
1.48	590.23	253	246	9501	10000	
		2.53%	2.46%	95.01%		
Simulation Results for Stage II Trial						

Operational and Regulatory Issues

- The protocol should only describe the design in general terms
- Detailed decision rules and statistical methods should be in the DMC charter
- Restrict access to the DMC charter
- Submit the design for regulatory review along with charter, simulation results and software
- Implement internal processes to prevent sponsor organization and investigators from reverse-engineering interim results
- Create an auditable DMC portal for storage of charter, decision rules and interim results

Role of Firewalls

- **"A well-trusted firewall established for trial conduct beyond those established for conventional group sequential trials can help provide assurance that statistical and operational biases have not been introduced."**

  FDA Guidance on Adaptive Design (2010)

- **Such a firewall can reduce risk of operational bias and generate trustworthiness**
  - **Operational Bias: Systematic pattern of patient selection by investigators with knowledge of interim results**
  - **Trustworthiness: Document who saw what and when?**

ACES: Access Control Execution System

- Web based technology to control flow of information and access to confidential documents
- Prevents operational bias by including the actual adaptive algorithm only in restricted appendix to Data Monitoring Committee (DMC) charter and tracking access to this document
- Establishes trustworthiness through secure password protected access to documents, execution of algorithms, and audit trail

Powerpoint Demo of ACES

Importance of Newly Released FDA Adaptive Guidance

- These designs into the category titled, "Adaptive Study Designs whose Properties are Less Well Understood"
- Guidance recommends adapting only for increasing sample size, not for decreasing
- Guidance recommends modest increases in sample size
- Guidance recommends using the method if the primary study objective cannot be achieved by other methods
- Guidance warns of "operational bias". In the present context, one would have to address how the sponsor intends to prevent investigators from "reverse engineering" the treatment effect from knowledge of the adaptive decision

A Final Take-Away Message

Adaptive trials require a considerable amount of planning up-front. One of the most versatile tools for the planning phase is simulation

- **The simulations clarify the risks and benefits of the proposed approach by putting probabilities on each possible outcome**
- **The simulations facilitate better communication with the FDA**
- **The simulations facilitate greater communication between the various members of the study team by showing how different design options and trial outcomes will have different implications for:**
  - **patient recruitment**
  - **drug supply**
  - **economic analyses**
  - **clinical outcomes**
  - **statistical power**
  - **regulatory concerns**

References

Bauer P, Kieser M (1999). Combining different phases in the development of medical treatments within a single trial. Statistics in Medicine, 18:1833-1848.

Bauer P, Kohne (1994). Evaluation of experiments with adaptive interim analysis. Biometrics, 1029-1041.

Brannath W, Mehta CR, Posch M (2009). Exact confidence bounds for adaptive group sequential tests. Biometrics, 65, 539-546.

Chen YHJ, Demets DL, Lan KKG (2004). Increasing the sample size when the unblinded interim result is promising. Statistics in Medicine, 23, 1023-1038.

Cui L, Hung HMJ, Wang S-J (1999). Modification of sample size in group sequential trials. Biometrics, 55, 853-857.

East-5.4 (2010). Software for the Design and Interim Monitoring of Flexible Clinical Trials. Cytel Software Corporation, Cambridge, MA.

Gao P, Liu L, Mehta CR (2013). Exact inference for adaptive group sequential designs. Statistics in Medicine in press.

Gao P, Ware JH, Mehta CR. Sample size re-estimation for adaptive sequential design. J.Biopharmaceutical Statistics. 2008; 18 (6): 1184-96.

Jennison C and Turnbull BW (2000). Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall/CRC, London.

Jennison C, Turnbull BW (2003). Mid-course sample size modification in clinical trials based on the observed treatment effect. Statist. Med., 22, 971-993.

Lehmacher W and Wassmer G (1999). Adaptive sample size calculations in group sequential trials. Biometrics, 55, 1286-1290.

Li G, Shih WJ, Xie T, Lu J (2002). A sample size adjustment procedure for clinical trials based on conditional power. Biostatistics, 3, 277-287.

Mehta CR, Bauer P, Posch M, Brannath W. Repeated confidence intervals for adaptive group sequential trials. Statistics in Medicine. 2008; 26, 5422-5433.

Mehta CR, Patel NR (2006). Adaptive, group sequential and decision theoretic approaches to sample size determination. Statistics in Medicine, 25:3250-3269.

Mehta CR, Tsiatis AA (2001). Flexible sample size considerations using information based interim monitoring. Drug Information Journal, 35, 1095-1112.

Mehta CR, Pocock SJ (2011). Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Statistics in Medicine*, **30 (28):3267-84.**

Muller HH, Schafer H (2001). Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and classical group sequential approaches. Biometrics, 57, 886-891.

O'Brien PC, Fleming TR (1979). A multiple testing procedure for clinical trials. Biometrics, 35, 549-56.

Pampallona S and Tsiatis AA (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. J. Statist. Planning and Inference, 42, 19-35.

Pampallona S, Tsiatis AA, and Kim K (2001). Interim monitoring of group sequential trials using spending functions for the type I and type II error probabilities. Drug Information Journal, 35, 1113-1121.

Pocock SJ (1977). Group sequential methods in the design and analysis of clinical trials. Biometrika, 64, 191-99.

Pocock SJ (2005). When (not) to stop a clinical trial. JAMA, vol 294, No. 17.

Posch M, Koenig F, Brannath W, Dunger-Baldauf C, Bauer P (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. Stat. Medicine, 24:3697-3714.

Proschan MA, Hunsberger SA (1995). Designed extension of studies based on conditional power. Biometrics, 51, 1315-1324.

Scharfstein DO, Tsiatis AA, and Robins JM (1997). Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. JASA, 92, 1342-50.

Shen Y, Fisher LD (1999). Statistical inference for self-designing clinical trials with a one-sided hpothesis. Biometrics, 55, 190-197.

Tsiatis AA, Mehta CR (2003). On the inefficiency of the adaptive design for monitoring clinical trials. Biometrika, 90, 367-378.

Wittes J, Brittain E (1990). The role of internal pilot studies in increasing efficiency of clinical trials. Statist. Med., 9, 65-72.