
EastAdapt: Software for Adaptive Sample Size Re-estimation

Cytel Webinar
Cambridge, MA

Cyrus R. Mehta

President, Cytel Corporation

May 16, 2009

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

Goal of this Presentation

Show how to use EastAdapt to design, simulate and monitor clinical trials that permit unblinded mid-course sample size corrections, without undermining their statistical validity

- Why have unblinded sample size re-estimation?
- Methods built into EastAdapt
 - Cui, Hung and Wang (1999)
 - Lemacher and Wassmer (1999)
 - Chen, DeMets and Lan (2004)
 - Müller and Schäfer (2001)
 - Mehta et. al. (2008, 2009)
- Negative symptoms schizophrenia example

Motivation for Mid-Course Sample Size Correction in Pivotal Trials

We don't know what value of δ 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
- Better safety profile at interim might justify smaller δ
- Opportunity to combine internal and external data

Negative Symptoms Schizophrenia

- New drug versus placebo for negative symptoms schizophrenia
- Primary endpoint is the change in negative symptoms assessment (NSA) **at week 26** relative to the baseline
- Smallest clinically meaningful effect size is 0.2
- Sponsor prefers to power for effect size of 0.27

Why Sponsor Chose $\delta = 0.27$

Normal Superiority Trials: Two-Sample Test - Difference of Means		
	Plan1	Plan2
Plan ID		
Test Parameters		
1-Sided or 2-Sided Test	2-Sided	2-Sided
Significance Level (α)	0.05	0.05
Power ($1 - \beta$)	0.9	0.9
Assigned Fraction (Treatment)	0.5	0.5
Boundary Parameters		
Planned Number of Looks	1	1
Spacing of Looks		
Hypothesis to be Rejected		
Boundary Family		
Boundary to Reject H0		
Boundary to Reject H1		
Normal Parameters		
Difference of Means (δ_{11})	0.27	0.2
Standard Deviation (σ)	1.0	1.0
Accrual (Subjects)		
Maximum	577	1051
Expected Under H0		
Expected Under H1		
Expected Under H1/2		

As is typical, the sample size decision is heavily influenced by sponsor's resource constraints. Sponsor can free up resources for at most 600 subjects **up-front**. Working backwards, study is powered at $\delta = 0.27$

What's the Right Sample Size?

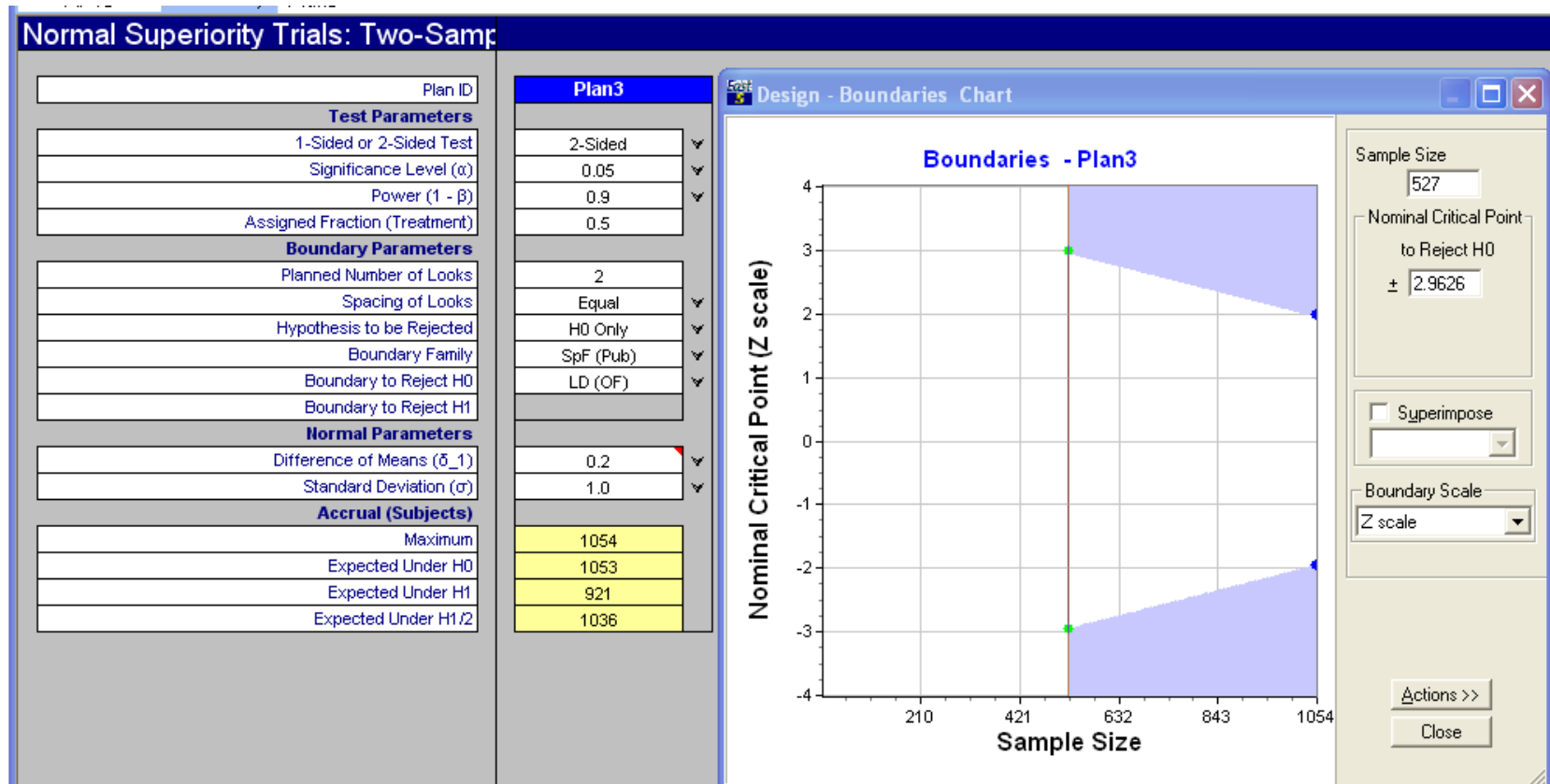
Table 1: Operating Characteristics of Plan1 and Plan2

δ	Plan1		Plan2	
	Sample Size	Power	Sample Size	Power
0.27	577	90%	1051	99%
0.25	577	85%	1051	98%
0.23	577	79%	1051	96%
0.21	577	71%	1051	93%
0.20	577	67%	1051	90%

- Plan1 is adequately powered if $\delta = 0.27$ but underpowered if $\delta = 0.2$
- Plan2 is adequately powered if $\delta = 0.2$ but overpowered if $\delta = 0.27$

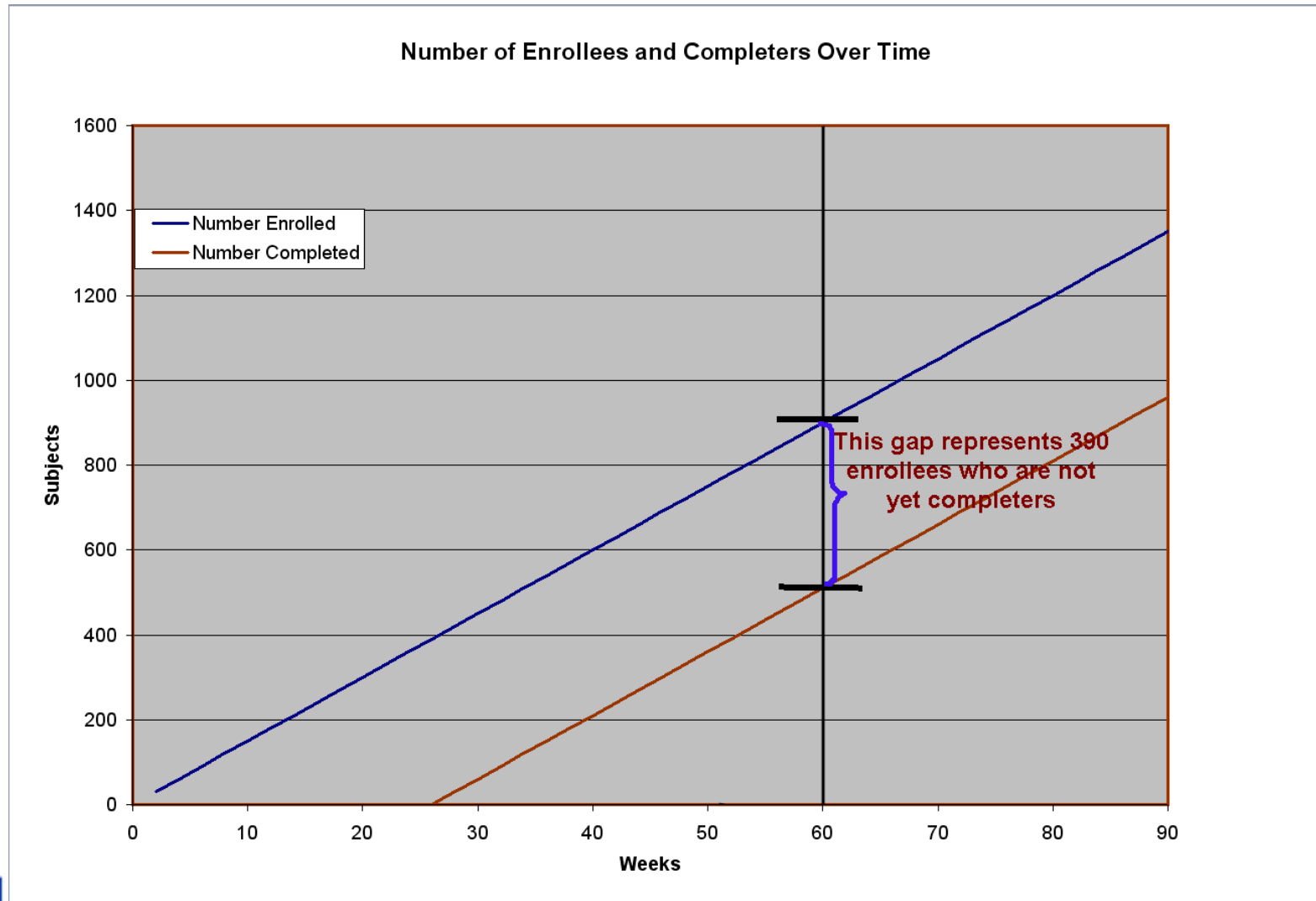
Try a Group Sequential Design

Design for $\delta = 0.2$ possible early stopping if interim results are compelling, as they will be if in truth $\delta = 0.27$



The Problem of Overruns

26-week endpoint; 15/week enrollment; overrun is $15 \times 26 = 390$



Comparison of Plans

Table 2: Comparison of Fixed Sample Plan1 (577 Subjects), Fixed Sample Plan2 (1051 Subjects), and Group Sequential Plan3

δ	Plan1		Plan2		Plan3 (GrpSequential)	
	SampSiz	Power	SampSiz	Power	E(SampSiz) [†]	Power
0.27	577	90%	1051	99%	977	99%
0.25	577	85%	1051	98%	991	98%
0.23	577	79%	1051	96%	1002	96%
0.21	577	71%	1051	93%	1014	93%
0.20	577	67%	1051	90%	1019	90%

[†] Group sequential sample sizes incorporate an overrun of 390 subjects

Conclusion from Plan Comparisons

- After accounting for overruns Plan 3 (group sequential) has only marginal advantage over Plan2 (fixed)
- Both plans protect power if $\delta = 0.2$, but sponsor cannot meet their large up-front commitment (> 1050 subjects)
- Sponsor still prefers to design for $\delta = 0.27$. **But can he reduce the risk of a failed trial if δ is closer to 0.2?**

Adaptive Design

1. Commit 577 subjects up-front (90% power at $\delta = 0.27$)
2. Perform an interim analysis at week 36 with
 $15 \times 36 = 540$ enrolled
 - 150 completers and 390 still in follow-up
3. Perform interim analysis only on the 150 completers
4. Increase the sample size, and hence the power, if the interim results fall in a “promising” zone

Sample Size Re-estimation Rule

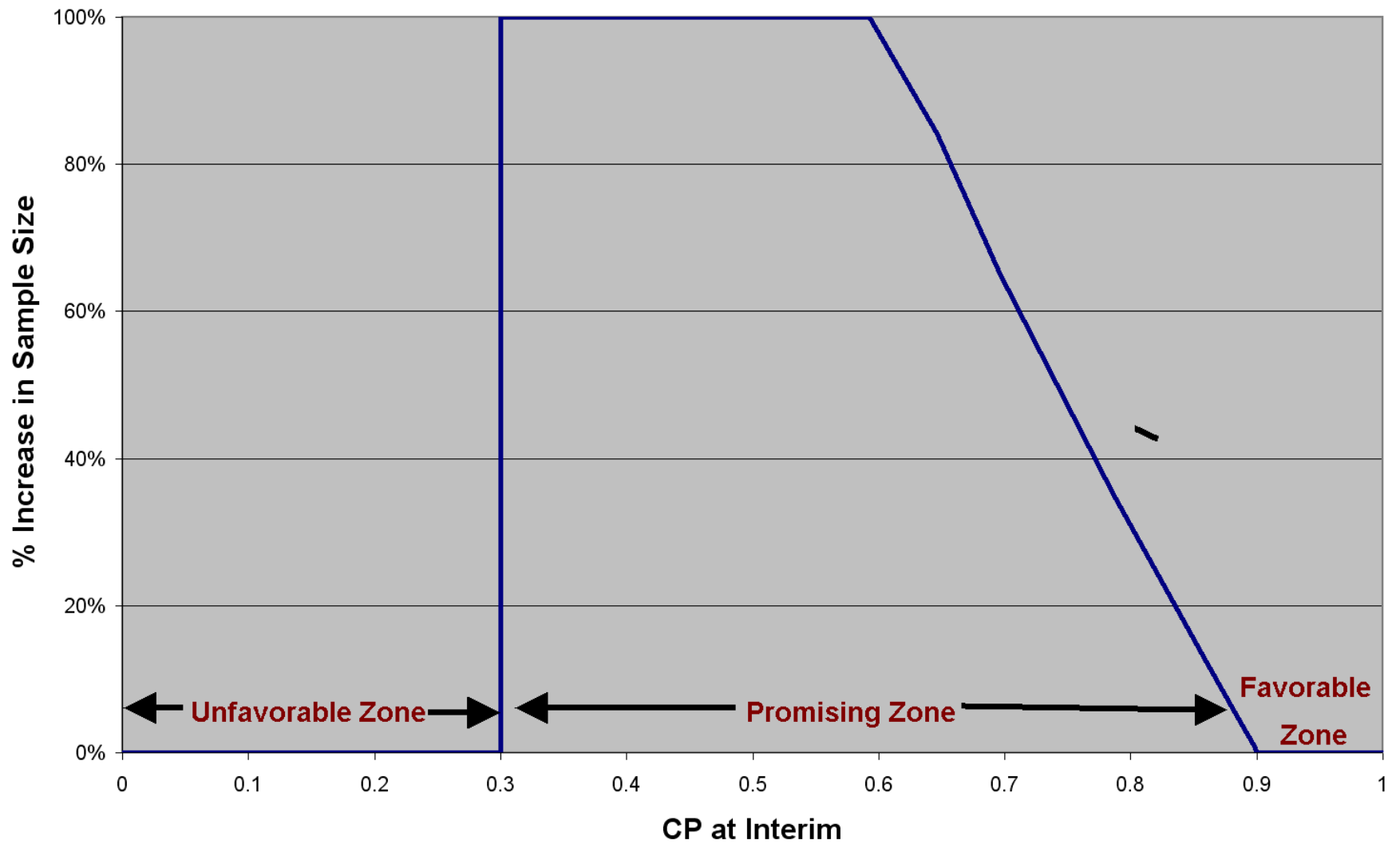
- Partition interim outcome into unfavorable, promising and favorable zones, based on **conditional power**
- Only increase sample size if interim outcome is promising
- Specify a **target conditional power** and calculate sample size needed to achieve the target
- Set minimum and maximum limits on sample size change

Inputs to EastAdapt

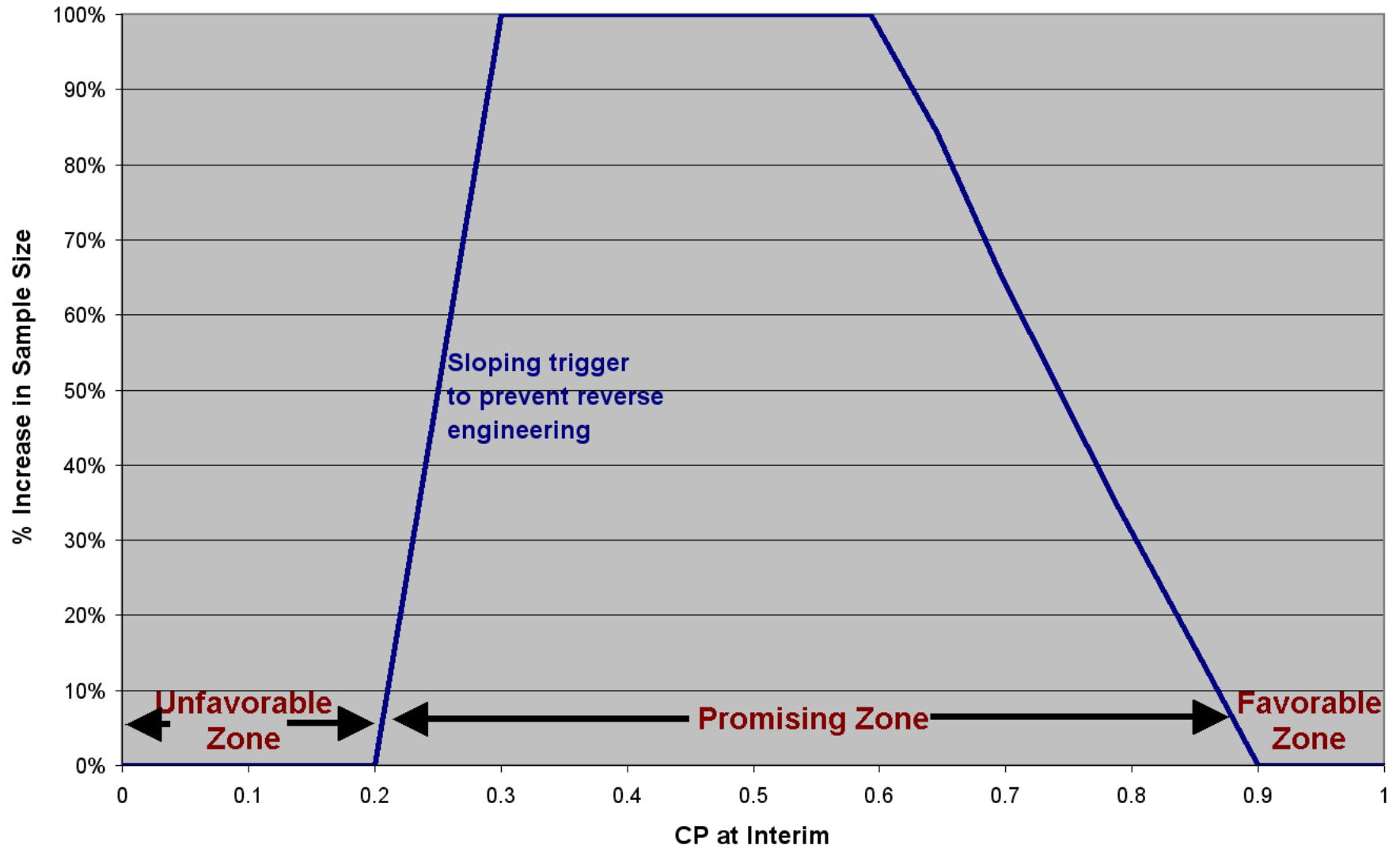
Criterion for Adapting: Min. CP	0.3000
Max. CP	0.9000
Min. Usable Sample Size	577
Max. Usable Sample Size	1154
Desired Conditional Power (CP)	0.9000

- If CP at the interim is between 0.3 and 0.9, **the outcome is considered promising** and sample size is increased
- New sample size is set so as to boost the CP to 0.9
- Range of new sample size is 577 (original) to 1154 (double)

Rule for Increasing Sample Size: CP at interim is between 30% and 90%



Rule for Increasing Sample Size: CP at interim is between 20% and 90% with sloping trigger



Results of 100,000 Simulated Trials: Unconditional and by Zone

Overall Simulation Results					
Avg. Info.	Avg. Sample Size	# Rejecting H0	# Unable to Reject H0	Total Simulations	
				Count	%
173.94	695.76	73284	26716	100000	100.00%
173.94	695.76	73284	26716	100000	
		73.28%	26.72%		
Simulation Results for Adapted Trial Only					
240.85	963.40	27208	3526	30734	100.00%
240.85	963.40	27208	3526	30734	
		88.53%	11.47%		

Simulation Results by Zone							
Zone	Avg. Sample Size	Simulations Rejecting H0		Simulations not Rejecting H0		Total Simulations	
		Count	%	Count	%	Count	%
Unfavorable: CP < 0.300	577.00	14503	44.87%	17819	55.13%	32322	32.32%
Promising: 0.300 <= CP < 0.900	963.40	27208	88.53%	3526	11.47%	30734	30.73%
Favorable: CP >= 0.900	577.00	31573	85.46%	5371	14.54%	36944	36.94%

Table 3: Operating Characteristics of Fixed Sample and Adaptive Designs

Value of δ	Fixed Sample		Adaptive	
	Power	N	Power	E(N)
0.27	90%	577	93%	677
0.23	79%	577	83%	689
0.20	67%	577	73%	695

- Power gain is offset by corresponding sample size increase
- But adaptive design reduces the sponsor's risk:
 - it only increases sample size if interim result is promising
 - if that happens, the payoff to sponsor is huge
 - if not, sponsor is no worse off than before

Table 4: Operating Characteristics Conditional on Interim Outcome

δ	Interim Outcome	Probability of Interim Outcome	Power Conditional on Interim Outcome		Expected Sample Size	
			Fixed	Adaptive	Fixed	Adaptive
0.27	Unfavorable	15%	74%	74%	577	577
	Promising	31%	88%	98%	577	944
	Favorable	53%	97%	97%	577	577
0.23	Unfavorable	22%	57%	57%	577	577
	Promising	34%	77%	94%	577	956
	Favorable	44%	92%	92%	577	577
0.2	Unfavorable	28%	45%	45%	577	577
	Promising	36%	67%	88%	577	963
	Favorable	37%	86%	86%	577	577

The Value Proposition

- With fixed 577 subjects, trial only has 67% power
- With fixed 1051 subjects, trial has 90% power; but up front commitment too large
- With adaptive trial sponsor's risk is reduced:
 - Only commits resources for 577 subjects initially
 - If interim result is promising, then commits up to 577 additional subjects
 - Happy to make the additional commitment since it raises the power substantially

Interim Monitoring: CHW Method

Seminal paper by Cui, Hung and Wang (Biometrics, 1999)

- Initial design with total sample size = 577:
150 observations at stage 1; 427 observations at stage 2
- Pre-specified weights:
 $w^{(1)} = (150/577) = 0.26$; $w^{(2)} = (427/577) = 0.74$
- Actual weights if total sample size increases to 750:
 $w^{*(1)} = (150/750) = 0.2$; $w^{*(2)} = (600/750) = 0.8$
- Let $(Z^{(1)}, Z^{(2)})$ be Wald statistics for (Stage 1, Stage 2)
- $Z_{\text{chw}} = \sqrt{0.26}Z^{(1)} + \sqrt{0.74}Z^{(2)}$
- **Reject H_0 if $Z_{\text{chw}} \geq 1.96$**

Two-Look CHW IM Worksheet

Normal Superiority Trials: Two-Sample Test - Difference of Means (Plan4) (CHW IM)

Plan Details	Look #	Incr. Accr.	Incr. Statistic	Cumul. Accr.	Prespec. Weights	Weighted Statistic	Prespecified Nominal Critical Point				Repeated 97.50% CI for δ		Repeated p-value
							Reject H0		Reject H1		Lower	Upper	
							Lower	Upper	Lower	Upper			
1-Sided or 2-Sided Test	1	150	1.2247	150	0.2600	1.2247		6.1150			-0.799	Infy	1.00000
Significance Level (α)	2				0.7400			1.9600					
Power ($1 - \beta$)	3												
Assigned Fraction (Treatment)	4												
Planned Number of Looks	5												
Spacing of Looks	6												
Hypothesis to be Rejected	7												
Boundary Family	8												
Boundary to Reject H0	9												
Difference of Means (δ_1)	10												
Standard Deviation (σ)													
Maximum Sample Size													

Nominal Critical Point Chart (Select) ↑	
	Cum. Wts 0.260 1.000

Error Spending Chart (Select) ↑	
	Cum. Wts 0.260 1.000

Conditional Power Calculator

Input

Current Look:

Current Cumulative Sample Size:

Current Weighted Test Statistic:

Input/Output

Value of δ :

Value of σ :

Value of δ/σ :

Desired Conditional Power:

Computed Sample Size (Overall):

* Use the radio button to select the quantity to be computed.

Recalc Plot Details Close

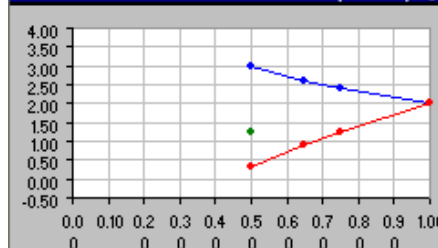
(Select) ↑			
Cum. Wts.	Lower	Upper	
0.2600	-0.799	Infy	

4-Look CHW IM Worksheet

Normal Superiority Trials: Two-Sample Test - Difference of Means (Plan5) (CHW IM)

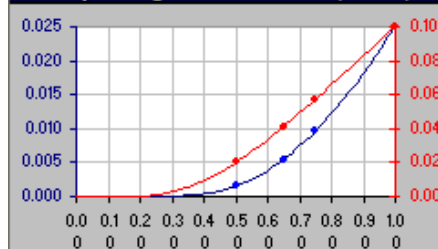
Plan Details		Look #	Incr. Accr.	Incr. Statistic	Cumul. Accr.	Prespec. Weights	Weighted Statistic	Prespecified Nominal Critical Point				Repeated 87.50% CI for δ		Repeated p-value		
								Reject H0		Reject H1		Lower	Upper		Lower	Upper
								Lower	Upper	Lower	Upper					
1-Sided or 2-Sided Test	1-Sided	1	150	1.2247	150	0.5000	1.2247		2.9626	0.3389	-0.2838	0.4147	0.25886			
Significance Level (α)	0.025	2				0.1500			2.5786	0.9226						
Power ($1 - \beta$)	0.9	3				0.1000			2.4116	1.2289						
Assigned Fraction (Treatment)	0.5	4				0.2500			2.0178	2.0178						
Planned Number of Looks	4	5														
Spacing of Looks	Unequal	6														
Hypothesis to be Rejected	H0 or H1 (NB)	7														
Boundary Family	SpF (Pub)	8														
Boundary to Reject H0	LD (OF)	9														
Boundary to Reject H1	LD (OF)	10														
Difference of Means (δ_1)	0.27	Effect Size under H1: $\delta_1 = (\mu_1 - \mu_0) / \sigma$														
Standard Deviation (σ)	1.0															
Maximum Sample Size	628															

Nominal Critical Point Chart (Select) ↑



Cum. Wts.	H0-	H0+	H1-	H1+
0.5000		2.9626		0.3389
0.6500		2.5786		0.9226
0.7500		2.4116		1.2289
1.0000		2.0178		2.0178

Error Spending Chart (Select) ↑



Cum. Wts.	Alpha	Beta
0.5000	0.0015	0.0200
0.6500	0.0054	0.0413
0.7500	0.0096	0.0575
1.0000	0.0250	0.1000

Conditional Power Calculator

Input

Current Look:

Current Cumulative Sample Size:

Current Weighted Test Statistic:

Input/Output

Value of δ :

Value of σ :

Value of δ/σ :

Computed Conditional Power:

Sample Size (Overall):

* Use the radio button to select the quantity to be computed.

Recalc Plot Details Close

CDL Method for 2-Stage Designs

Schizophrenia trial: initial sample size 577, increased to 750, based on interim results from 150 subjects

- Some statisticians do not like to use CHW statistic

$$Z_{\text{chw}} = \sqrt{0.26}Z^{(1)} + \sqrt{0.74}Z^{(2)}$$

- Prefer to use the Wald statistic after sample size increase

$$Z_{\text{wald}} = \sqrt{0.2}Z^{(1)} + \sqrt{0.8}Z^{(2)}$$

because it is the sufficient statistic for δ

- This is ok if $CP \geq 0.5$; (Chen, DeMets, Lan, 2004)
- CDL conditions relaxed by Gao, Ware, Mehta (2008)

CP Calculator Verifies CDL Conditions

Use East's regular IM worksheet if CDL conditions are met

Look #	Cumul. Accr.	Test Stat.	Info Fract.	Nominal Critical Point				Repeated 95.00% CI for δ	
				Reject H0		Reject H1		Lower	Upper
				Lower	Upper	Lower	Upper	Lower	Upper
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									

Relaxing the CDL Conditions

Sample Size Ratios		CP _{min} Values for		
Maximum Allowed	At Interim Look	Targeted Conditional Powers		
(N_{\max}^*/n_2)	(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

Results obtained by using method of Gao, Mehta, Ware (2008)

CDL Simulation Worksheet

Normal Superiority Trials: Two-Sample Test - Difference of Means (CDL Simulation)

Plan Details		Simulation Boundaries						Latest Simulated Test Stat.	Overall Simulation Results						
1-Sided or 2-Sided Test	1-Sided	#	Boundary				Avg. Info.		Avg. Sample Size	# Rejecting H0	# Unable to Reject H0	Total Simulations			
Significance Level (α)	0.025		Sample Size	H0-	H0+	H1-		H1+				Count	%		
Power (1 - β)	0.9	1	150.00		6.1150			-0.2742							
Assigned Fraction (Treatment)	0.5	2	577.00		1.9600			1.3320	174.39	697.57	709	291	1000	100.00%	
Planned Number of Looks	2	3													
Spacing of Looks	Unequal	4													
Hypothesis to be Rejected	H0 Only	5													
Boundary Family	SpF (Pub)	6													
Boundary to Reject H0	Gm (-24)	7													
Difference of Means (δ_1)	0.27	8													
Standard Deviation (σ)	1.0	9													
Maximum Sample Size	577	10													
		11													
		12													
Recompute															
Design Outputs								Total	174.39	697.57	709	291	1000		
Max. Sample Size (Nmax)	577							%			70.90%	29.10%			
Max. Information (Imax)	144.1348							Simulation Results for Adapted Trial Only							
Simulation Parameters															
Difference of Means	0.2000								241.80	967.18	278	31	309	100.00%	
Standard Deviation	1.0000														
Use Nmax Till 'L' Looks, L =	1														
Criterion for Adapting: Min. CP	0.3000														
Max. CP	0.9000														
Min. Usable Sample Size	577														
Max. Usable Sample Size	1153														
Use Wald Stat. if CP(577) >=	0.5000														
Desired Conditional Power (CP)	0.9000														
Number of Trials	1000														
Refresh Every 'n' Trials, n =	100														
Simulation Starting Seed	Clock														
<input type="button" value="Run"/> <input type="button" value="Single Step"/> <input type="button" value="Stop"/> <input type="button" value="Reset"/>															

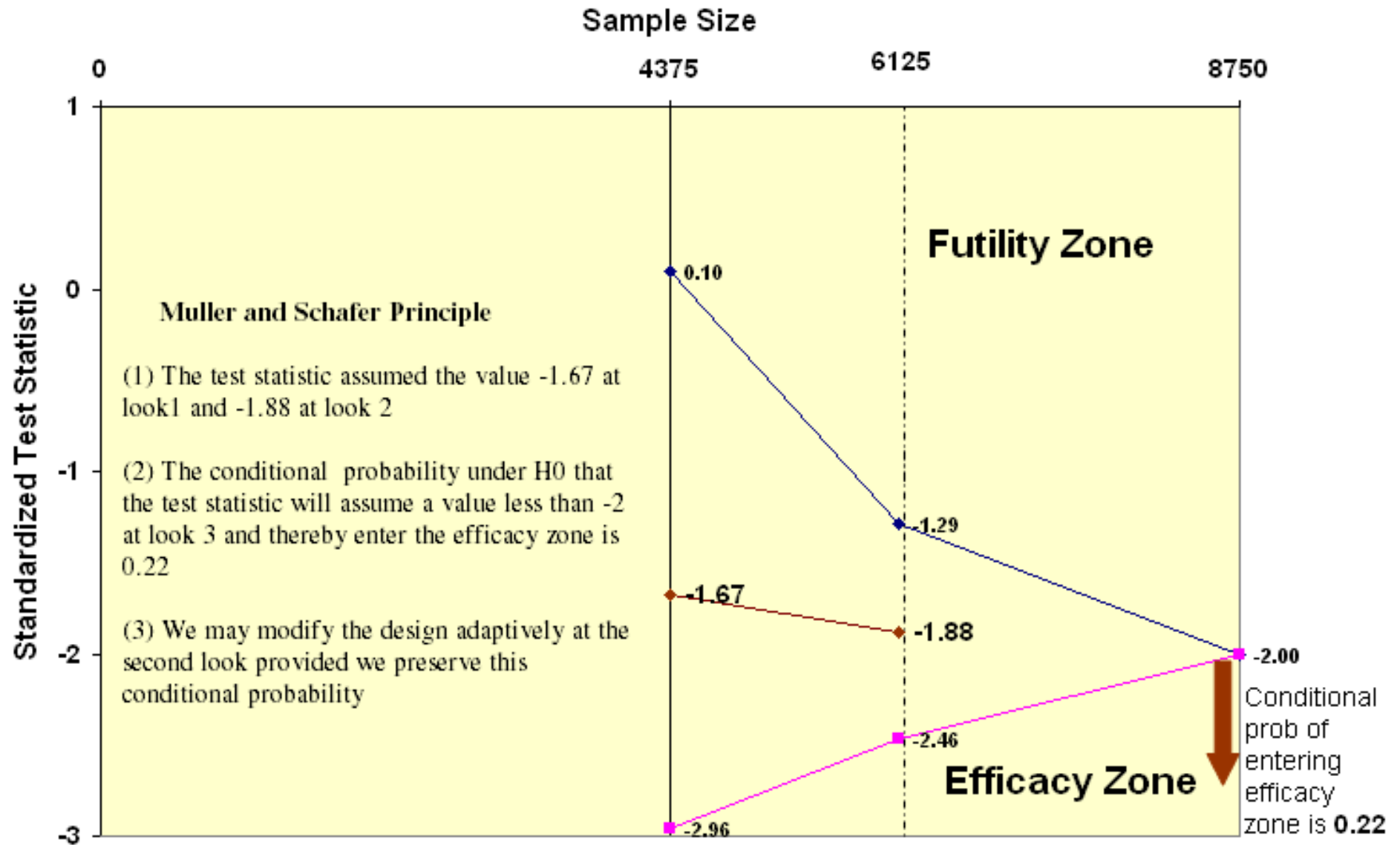
[Show Simulation Results by Zone](#)

- Simulate to obtain operating characteristics
- Simulate to determine power loss (if any) compared to CHW method

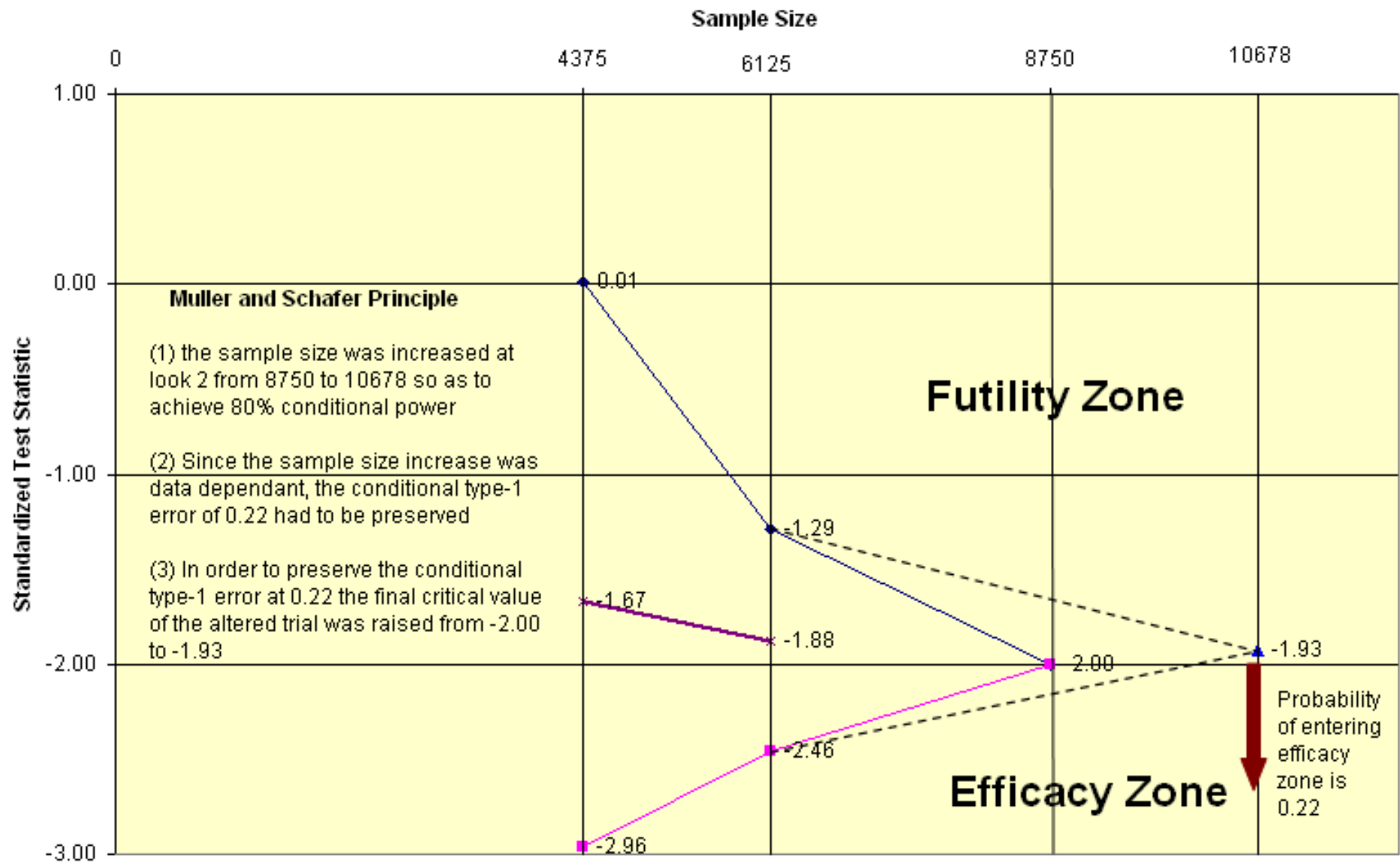
Müller and Schäfer Method

- More general than CHW or CDL methods
- Permits changes to sample size, number of looks, spacing of looks, population enrichment, and any other data dependent trial modification
- **Basic Principle: preserve the conditional type-1 error of the original design in the modified design**
Müller and Schäfer (Biometrics, 1999)

Group Sequential Design; 8750 Patients; Conditional Type-1 Error is 0.22



Adaptive Sample Size Increase Keeping Conditional Type 1 Error Unchanged



Interim Monitoring of Schizophrenia by Müller and Schäfer Method

- First interim look after 150 subject
- Observe $\hat{\delta} = 0.18$
- Compute conditional type-1 error = 0.05208
- Shoot for 90% conditional power by increasing sample size but preserving the conditional type-1 error
- Achieved by creating a second trial in East with $\alpha = 0.05298$

Normal Superiority Trials: Two-Sample Test - Difference of Means (Plan5)

Plan Details		Look #	Cumul. Accr.	Test Stat.	Info Fract.	Nominal Critical Point				Repeated 97.50% CI for δ		Interim Outputs			
1-Sided or 2-Sided Test	1-Sided					Reject H0		Reject H1		Lower	Upper	Lower	Upper	CP at INLP	0.59241
Significance Level (α)	0.025 <th>Lower</th> <th>Upper</th> <th>Lower</th> <th>Upper</th> <th>Lower</th> <th>Upper</th> <th>Lower</th> <th>Upper</th> <th>Ideal Next Look Position</th> <td>576</td>					Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Ideal Next Look Position	576
Power (1 - β)	0.9	1	150	1.1023	0.2602		6.1142			-0.8185	Infy				
Assigned Fraction (Treatment)	0.5	2													
Planned Number of Looks	2	3													
Spacing of Looks	Unequal	4													
Hypothesis to be Rejected	H0 Only	5													
Boundary Family	SpF (Pub)	6													
Boundary to Reject H0	Gm (-24)	7													
Difference of Means (δ_1)	0.27	8													
Standard Deviation (σ)	1.0	9													
Maximum Sample Size	577	10													

Eff Size	CP INLP
0.0000	0.05208
0.0331	0.09964
0.0661	0.17304
0.0992	0.27399
0.1322	0.39766
0.1598	0.51001
0.1873	0.62157
0.2149	0.72378
0.2424	0.81019
0.2700	0.87761

Info. fr	Lower	Upper
0.2602	-0.8185	Infy

Conditional Power Calculator

Input

Current Look:

Current Cumulative Sample Size:

Current Test Statistic:

Input/Output

Value of δ :

Value of σ :

Value of δ/σ :

Computed Conditional Power:

Sample Size (Overall):

Recalc Plot Details Close

Normal Superiority Trials: Two-Samp

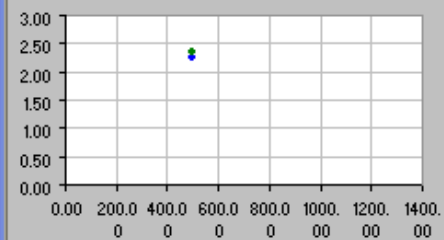
	Plan5	Plan6	Plan7
Plan ID			
Test Parameters			
1-Sided or 2-Sided Test	1-Sided	1-Sided	1-Sided
Significance Level (α)	0.025	0.0521	0.0521
Power (1 - β)	0.9	0.9	0.9
Assigned Fraction (Treatment)	0.5	0.5	0.5
Boundary Parameters			
Planned Number of Looks	2	1	2
Spacing of Looks	Unequal		Equal
Hypothesis to be Rejected	H0 Only		H0 Only
Boundary Family	SpF (Pub)		SpF (Pub)
Boundary to Reject H0	Gm (-24)		Gm (-2)
Boundary to Reject H1			
Normal Parameters			
Difference of Means (δ_1)	0.27	0.18	0.18
Standard Deviation (σ)	1.0	1.0	1.0
Accrual (Subjects)			
Maximum	577	1043	1071
Expected Under H0	577		1064
Expected Under H1	577		828
Expected Under H1/2	577		1005

Normal Superiority Trials: Two-Sample Test - Difference of Means (Plan7)

Plan Details		Look #	Cumul. Accr.	Test Stat.	Info Fract.	Nominal Critical Point				Repeated 94.79% CI for δ		Final Outputs at Look #1	
1-Sided or 2-Sided Test	1-Sided					Reject H0		Reject H1		for δ		Adjusted p-value	0.009
Significance Level (α)	0.052					Lower	Upper	Lower	Upper	Lower	Upper	Adj. Pt. Est. for δ	0.2100
Power ($1 - \beta$)	0.9	1	500	2.3479	0.4668		2.2388		0.0098	Infty	Adj. 94.79% CI for δ		
Assigned Fraction (Treatment)	0.5	2									Lower Confidence Bound	0.0647	
Planned Number of Looks	2	3									Upper Confidence Bound	Infty	
Spacing of Looks	Equal	4									Post-Hoc Power		
Hypothesis to be Rejected	H0 Only	5											
Boundary Family	SpF (Pub)	6											
Boundary to Reject H0	Gm (-2)	7											
Difference of Means (δ_1)	0.18	8											
Standard Deviation (σ)	1.0	9											
Maximum Sample Size	1071	10											

Effect Size under H1: $\delta_1 = (\mu_t - \mu_c) = 0.1800$

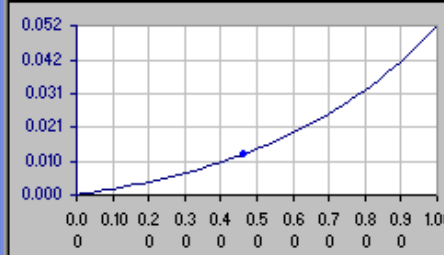
Nominal Critical Point Chart (Select) ↑



Conditional Power Chart (Select) ↑

SamSiz	H0-	H0+	H1-	H1+
500		2.2388		

Error Spending Chart (Select) ↑



Info. fr	Alpha	Beta
0.4668	0.0126	

Muller and Schafer Parameter Estimation Calculations

Primary/Secondary Trial

Choose Primary Trial: Plan5

Choose Secondary Trial: Plan7

Method of Estimation

RCI Method

SWACI Method

Output

Point Estimate (MUE) of Δ : 0.2027

Overall P-Value: 0.00492

97.5 % Lower Confidence Limit on Δ : 0.0488

Recalc Close

Parameter Estimation Methods

RCI Method Extends the repeated confidence intervals of Jennison and Turnbull (1989) to the adaptive Müller and Schäfer setting

- Developed by Mehta, Bauer, Posch and Brannath (2007)
- CI provides conservative coverage of true δ
- Point estimate is negatively biased

SWACI Method Extends the stage wise adjusted confidence intervals of Tsiatis, Rosner, Mehta (1984) to adaptive Müller and Schäfer setting

- Developed by Brannath, Mehta and Posch (2009)
- CI provides exact coverage of true δ
- Point estimate is median unbiased
- Methodology only developed for 1-sided tests

Endpoints Handled by EastAdapt

Normal Endpoints: All features fully implemented

Binomial Endpoints: All features fully implemented

Survival Endpoints: Only partially implemented

- Interim monitoring capability available indirectly through worksheet for normal endpoints
- Simulation of mid-course changes to number of events available indirectly through simulation worksheet for normal endpoints
- Still need to develop tools to evaluate trade-off between study duration, sample size and mid-course change in number of events

For Future Release

- **Additional capabilities for survival endpoints, especially sample size increase to shorten trial duration**
- **Multi-arm trials with dose selection and possible sample size increase at interim**
- **Population enrichment trials with possible restriction of enrollment to pre-specified subgroups at interim**

Logistical and Operational Issues

- **Form an independent interim analysis review committee (IARC)**
- **Create a detailed Charter for the IARC with guidance for the adaptive change and some flexibility to overrule if circumstances warrant**
- **Submit a Special Protocol Assessment (SPA) for regulatory approval of trial. Include simulation results (or software) and IARC charter as part of the SPA**
- **Successful implementation requires centralized randomization, EDC, efficient data clean-up, and management of drug-supply to numerous remote sites**

References

Brannath W, Mehta CR, Posch M (2009). Exact confidence bounds for adaptive group sequential tests. *Biometrics* (In Press).

Chen YH, DeMets DL, Lan KKG (2004). Increasing the sample size when the unblinded interim results are promising. *Statistics in Medicine* 23, 1023-1038.

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

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

Jennison C, Turnbull BW (1989). Interim analysis: the repeated confidence interval approach. *J. Roy. Stat. Soc., Series B* 51, 305-361.

Lehmacher W, Wassmer G (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* 45, 1286-2508.

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

Müller and Schäfer (2001). Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and classical group sequential approaches. *Biometrics* 23, 2497-2508.

Tsiatis AA, Rosner GL, Mehta CR (1984). Exact confidence intervals following a group sequential test. *Biometrics* 40, 797-803.