
Flexible Adaptive Design: Five Case Studies

American Statistical Association
Princeton–Trenton Chapter

November 8, 2006

Cyrus R. Mehta

President, Cytel Inc.

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

Five Case Studies

Rheumatoid Arthritis Phase II adaptive dose ranging trial with a binary endpoint

CNS Seamless Phase II/III trial with continuous endpoint

Cardiology Phase III group sequential trial with a mortality endpoint

Psoriasis Phase III trial with sample size adjustment due to uncertainty about the placebo response

Schizophrenia Phase III trial with sample size adjustment due to uncertainty about the primary effect size

Part 1: Phase II Dose-Ranging Trials

Role of Adaptive Methods in Dose-Ranging Trials

Phase II dose-ranging trials attempt to answer the following questions:

Safety: are the higher doses safe?

Proof of Concept: is there any evidence of trend?

Shape of Dose-Response Curve: where is the interesting range of doses?

Dose Selection: which dose or doses should we pick for confirmatory Phase III testing?

Phase II Trial for Rheumatoid Arthritis

Current Non-Adaptive Design for this Study

- Double-blind, randomized, parallel group
- Placebo (0) + four doses (2.5, 5, 10, 20 mg)
- Fixed sample size of 400 subjects
- Expected enrollment period is 24 weeks
- Allocated equally: 80 subjects/treatment arm
- DAS28 measured bi-weekly through week 12
- Primary Endpoint: dichotomised value of DAS28 at week 12

Goals of Study

- **Proof of concept.** Demonstrate that there is indeed a dose-response relationship through a statistically significant test of trend
- Model the dose-response curve
- Identify doses to carry forward for Phase III testing

Limitations of Current Design

No opportunity to learn about dose-response relationship while trial is still on-going. Hence:

- Assign the same number of subjects (80) to each dose regardless of efficacy
- Cannot eliminate ineffective doses prior to completion of trial

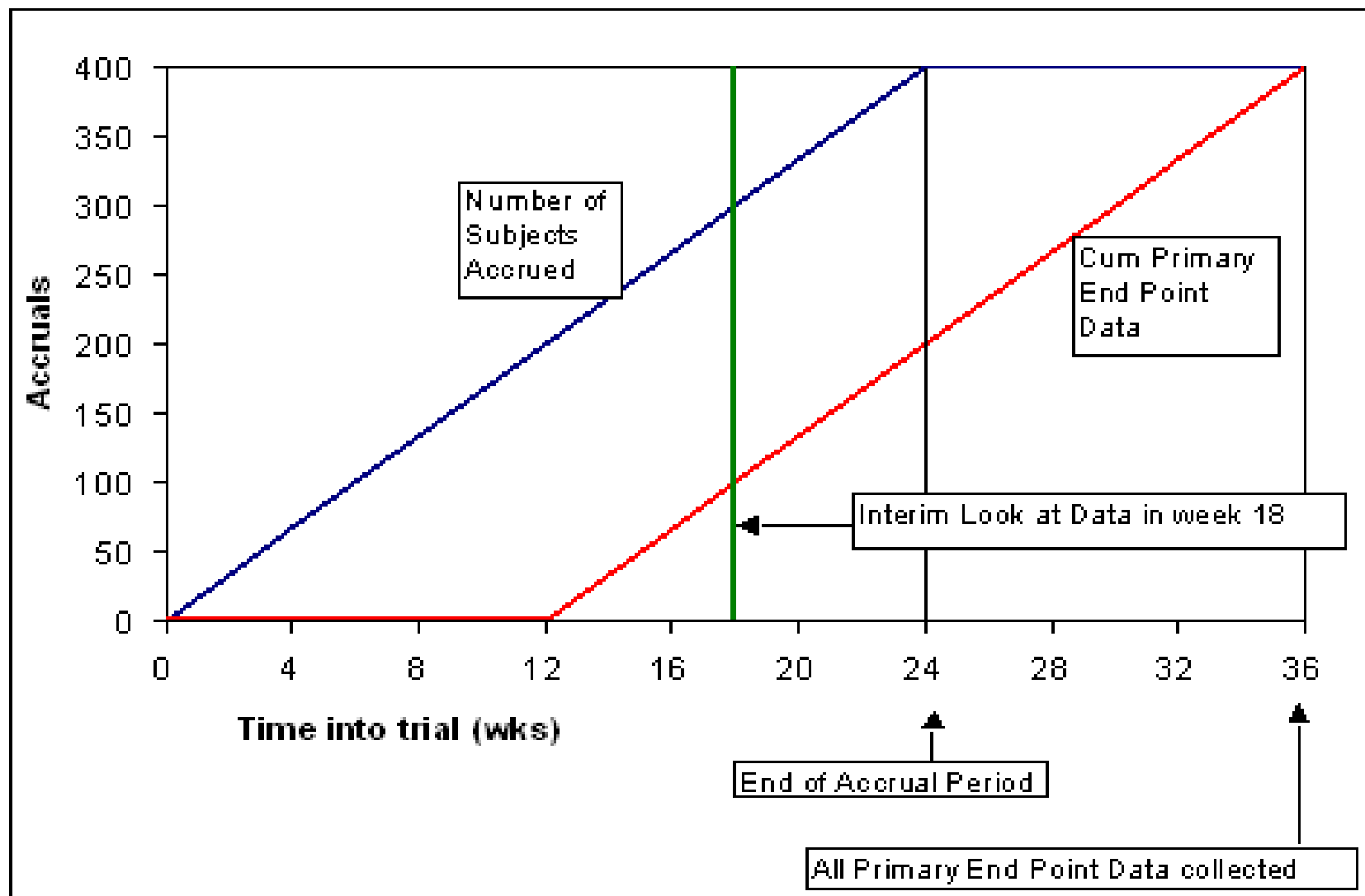
Alternative Adaptive Design

- **Stage 1**
 - Randomize equally to each dose through the end of stage1
 - compute the predictive probability of success with each dose for the future patients
- **Stage 2** Use the stage-1 findings to:
 - Change the randomization fractions for future patients
 - eliminate doses for future patients

Statistical Validity

- This approach uses the intuitive Bayesian concept of predictive probability to allocate patients to doses
- But it uses standard frequentist hypothesis testing methods to determine which doses differ from placebo. Specifically:
 - The p-values from the two stages can be combined (Bauer and Keiser, 1999)
 - Adjustments for multiple comparisons are available (Hochberg and Tamhane, 1987)
- Therefore the statistical methods do not raise any regulatory difficulties

Depiction of Two-Stage Strategy



At Week 18: 300 enrolled; 100 with 12-week data; 100 yet to enroll

Improving the Allocations for Remaining 100 Subjects

- Input the current data into a Bayesian model and compute the predictive probability of a response at week 12
- Based on these predictive probabilities:
 - Drop unpromising arms from further patient allocation
 - Re-allocate the remaining 100 subjects to the remaining arms
- The rules for dropping arms and re-allocating subjects should be pre-specified and their properties should be verified by simulations

Dose Dropping and Patient Allocation Rules

Rule 1: If the predictive probability for a dose is less than 0.4 drop the dose from future randomizations

Rule 2: If the predictive probability for two or more doses exceeds 0.8 retain the smallest dose and drop up to two higher doses according to the following criteria:

- If rule 1 led to dropping one dose then drop the highest dose
- If rule 1 led to dropping no doses drop the two highest doses.
- If Rule 1 led to dropping two doses do not drop any doses using Rule 2

The doses that remain after applying the above two rules will be allocated to the 100 patients yet to be randomized equally.

Typical Scenario at Week 18

Underlying Dose-Response Curve



Observed Response Proportions at Week 18

Dose	Placebo	2.5 mg	5 mg	10 mg	10 mg
Response	2/20	6/20	12/20	19/20	20/20
Percentage	10%	30%	60%	95%	100%

Resulting Bayesian Predictive Probabilities

Dose	Pr(Response)
0 mg	0.1154
2.5 mg	0.3106
5 mg	0.6026
10 mg	0.9222
20 mg	0.9678

Invoking the pre-specified rules we would

- Drop the 2.5 mg arm
- Drop the 20 mg arm
- Re-allocate the remaining 100 subjects equally to the 5 mg and 10 mg doses

Evaluate Adaptive Patient Allocation by Simulation

Patient Allocations in 1000 Simulated Trials					
Placebo	2,5 mg	5 mg	10 mg	20 mg	# Simulations
80	60	100	100	60	930
80	60	60	100	100	30
80	100	100	60	60	30
80	87	87	86	60	10

Average Patient Allocation

80	61	99	99	61
----	----	----	----	----

Generate similar patient allocation tables with other plausible dose-response curves prior to finalizing the protocol

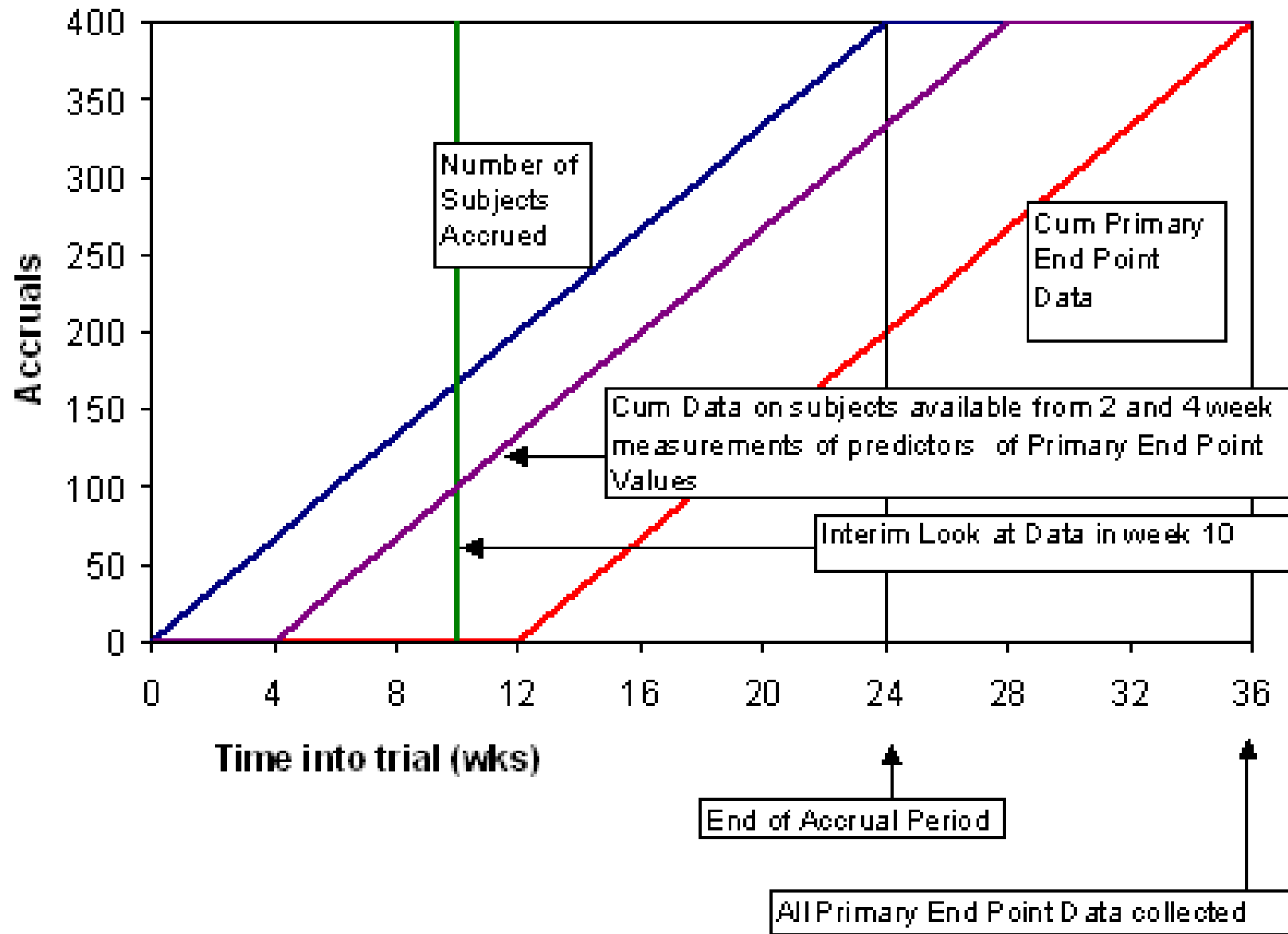
Proof of Concept Objective was Not Compromised

The table below shows that the type-1 error and power of the **overall test of trend** is not compromised by allocating patients adaptively instead of in a fixed proportion

Scenario	Dose Response Curves					Power	
	0 mg	2.5 mg	5 mg	10 mg	20 mg	Fixed	Adaptive
No Trend	0.3	0.3	0.3	0.3	0.3	5%	5%
Trend 1	0.378	0.395	0.413	0.450	0.525	47%	49%
Trend 2	0.467	0.488	0.508	0.550	0.630	60%	60%
Trend 3	0.466	0.494	0.523	0.579	0.684	89%	89%

Scope for Further Improvement: Use of Surrogates

- The results at week-4 might be correlated with the results at week-12
- We will know this in May 2006 when results of prior study are unblinded
- We could use the correlation to make adaptive improvements earlier and thereby allocate more subjects to the better performing doses



Benefits of Adaptive Design

- The adaptive design performs just as well as the fixed design in terms of proof of concept based on the overall trend test
- However, the adaptive design allocates more patients to the more interesting parts of the dose response curve
- The adaptive design is better able to identify the dose or doses that should be tested further in a Phase III confirmatory trial
- The adaptive design requires very careful planning and up-front simulations of different scenarios

Part 2: Seamless Phase II/III Trials

Phase II/III Trial in CNS

- Two-stage adaptive design with a 4-week endpoint, 5 doses, and **pairwise comparisons to placebo**
- Plan to enroll 170 patient per arm
- Enroll 50/arm at Stage 1. Based on pre-specified **conditional power** criteria:
 - Drop up to three doses for futility
 - Re-estimate sample size for doses continuing to Stage 2
- At end of Stage 2 identify the statistically significant doses using appropriate adjustments for multiple comparisons and adaptive changes in sample size

Pre-Specify All Rules for Adaptive Changes and for Final Analysis

For dose with most favorable response let $CP(170)$ and $CP(250)$ be the conditional power if the sample sizes were 170/arm and 250/arm, respectively (based on $\alpha = 0.01$)

- **Rules for increasing sample size**

If $CP(170) \geq 90\%$, then set $N = 170$

If $CP(170) < 90\%$ and $CP(250) > 90\%$, set N so that $CP(N) = 90\%$

If $50\% \leq CP(250) \leq 90\%$ set $N = 250$

If $CP(250) < 50\%$, terminate trial for futility

- **Rules for dropping arms**

Once N has been determined as above, drop any arm if $CP(N)$ for that arm is less than 50% (based on $\alpha = 0.05$)

- **Over-ruling the decision to drop arms**
 - No more than 3 doses may be dropped
 - Do not drop a higher dose if a lower dose is continuing
- **Final analysis at Stage 2**
 - Compute p-values for Stage 1 and Stage 2 for each dose
 - Combine the two p-values of for each dose by **inverse normal weighting**
 - For dropped doses the combined p-value is 1
 - Sort the p-values in descending order and use Hochberg's closed testing procedure to identify the statistically significant ones

Verify Operating Characteristics by Simulation

In any adaptive design it is a regulatory requirement that simulations be submitted along with the protocol. Preferably submit the simulation software as well

Simulation of Seamless Phase II/III Design

# of Arms (Excluding Placebo)	5
2 * 1-sided Overall Alpha	0.05

Simulation Parameters						
	Placebo	Trtm1	Trtm2	Trtm3	Trtm4	Trtm5
Mean	0	0	0	0	0	0
Std. Dev.	4	4	4	4	4	4

	Hypothesis Rejected					At Least One	All Accepted
	H1	H2	H3	H4	H5		
Count							
%							

	Treatments Dropped						
	Trtm1	Trtm2	Trtm3	Trtm4	Trtm5	None	All
Count							
%							

Average Sample Size Per Arm						Overall Average
Placebo	Trtm1	Trtm2	Trtm3	Trtm4	Trtm5	

Total Simulations	10000
Current Simulations	

Run Reset Adaptive Rules

Part 3: Phase III Group Sequential Trials

Group Sequential Methods in Phase III Trials

- Well established methods. Fully endorsed in ICH-E9 Regulatory Guidance
- Maximum impact in long term mortality trials; especially in the **oncology, cardiology and HIV** therapeutic areas
- Such trials typically run for 2 to 5 years
- **A group sequential design can easily save one or more years of study duration**

Example: The RALES Trial

- The randomized aldactone evaluation study (RALES) was a double-blind multicenter clinical trial of aldeosterone-receptor blocker vs placebo (NEJM, vol 341, 10, pages 709-717, 1999)
- Open to patients with severe heart failure due to systolic left ventricular dysfunction
- Primary endpoint is all-causes mortality
- The anticipated accrual rate is 960 patients/year
- The mortality rate for the placebo group is 38%
- Investigators want 90% power to detect a 17% reduction in the mortality rate for the Aldactone group (from 38% to 31.54%) with $\alpha = 0.05$, 2-sided
- Six DSMB meetings were planned

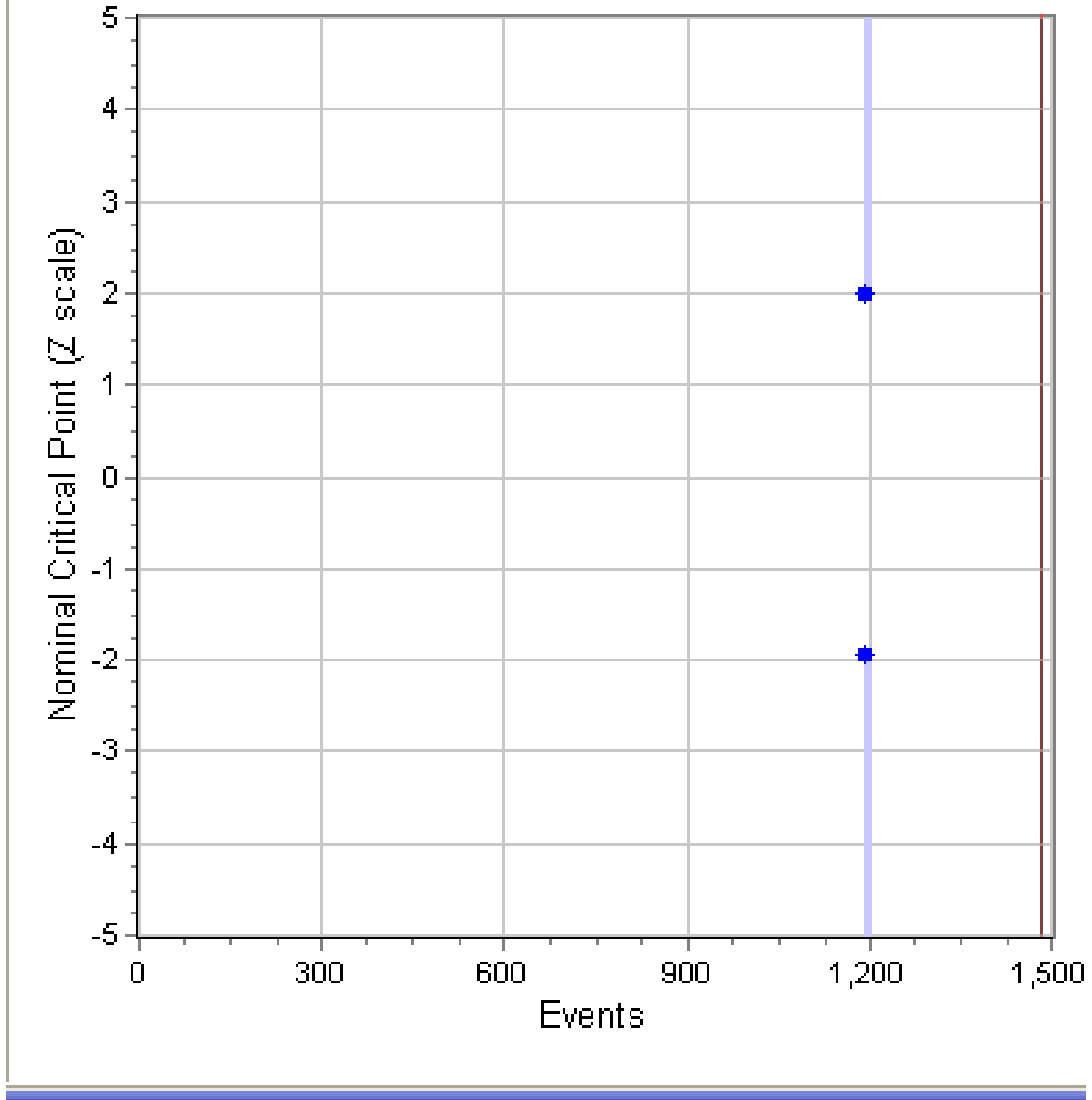
Conventional Single-Look Design

The goal was to design a trial that would be completed in 4.5 years. This could be achieved as follows:

- Enroll 1660 patients over 1.7 years
- Follow these patients until 1200 events (deaths) are observed
- Then perform the final analysis

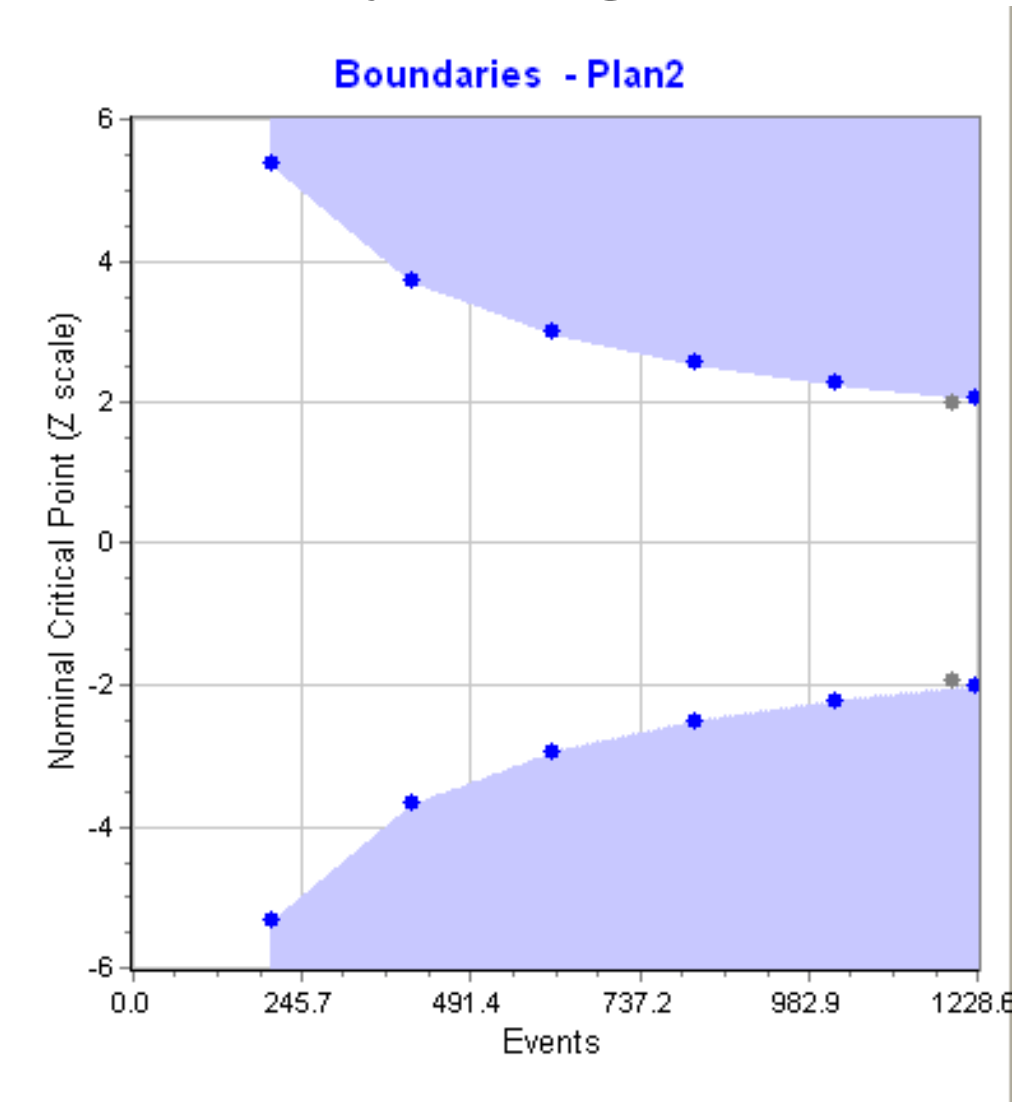
The study was activated in January 1996. Based on the specified design parameters, we would expect to see the required 1200 events in June 2000 and would then perform the final analysis

Boundaries - Plan1

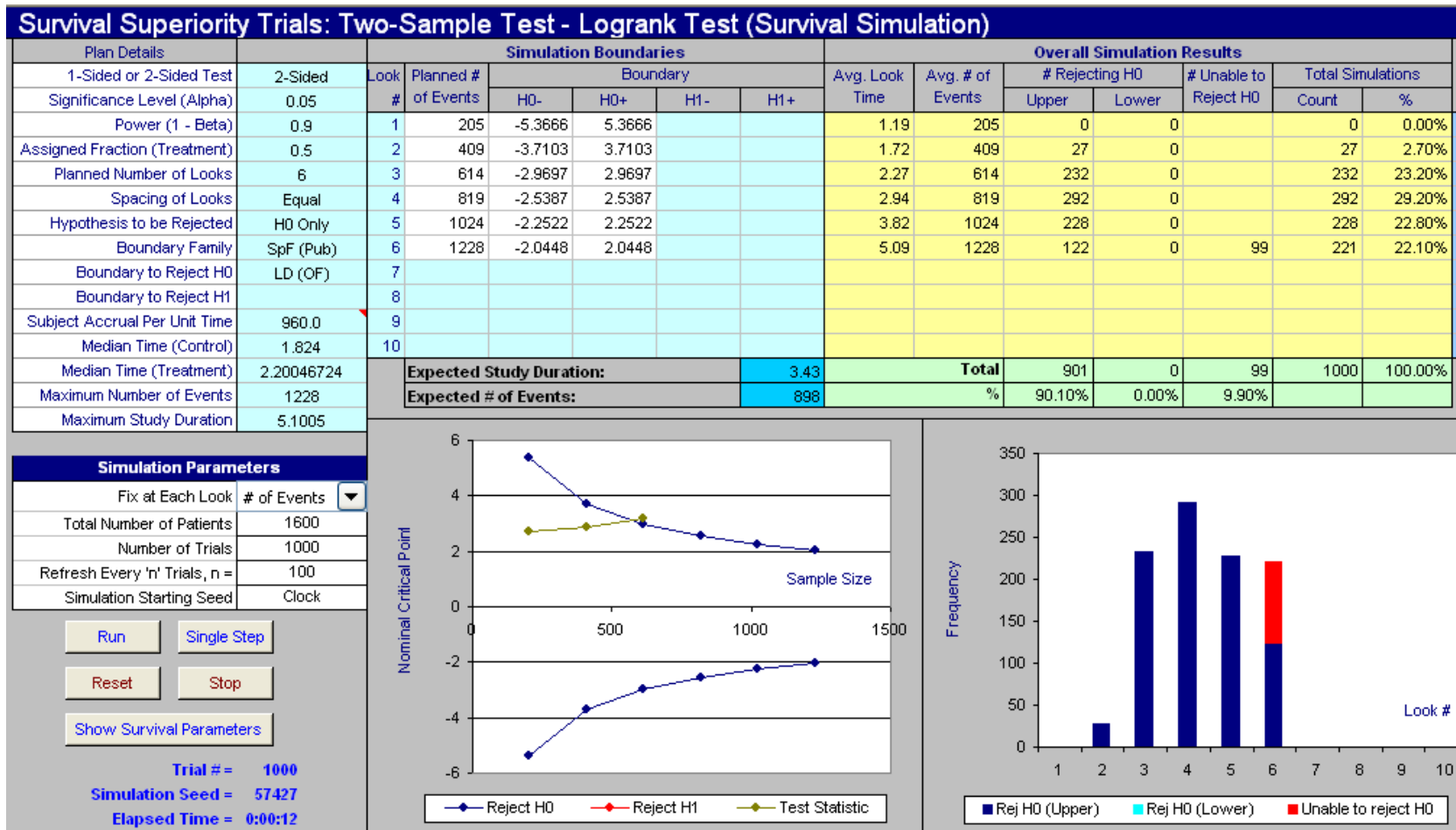


Six-Look Group Sequential Design

Instead of waiting till the end, DSMB will monitor the accruing efficacy data up to six times to see if early stopping is possible



Verify Operating Characteristics by Simulation



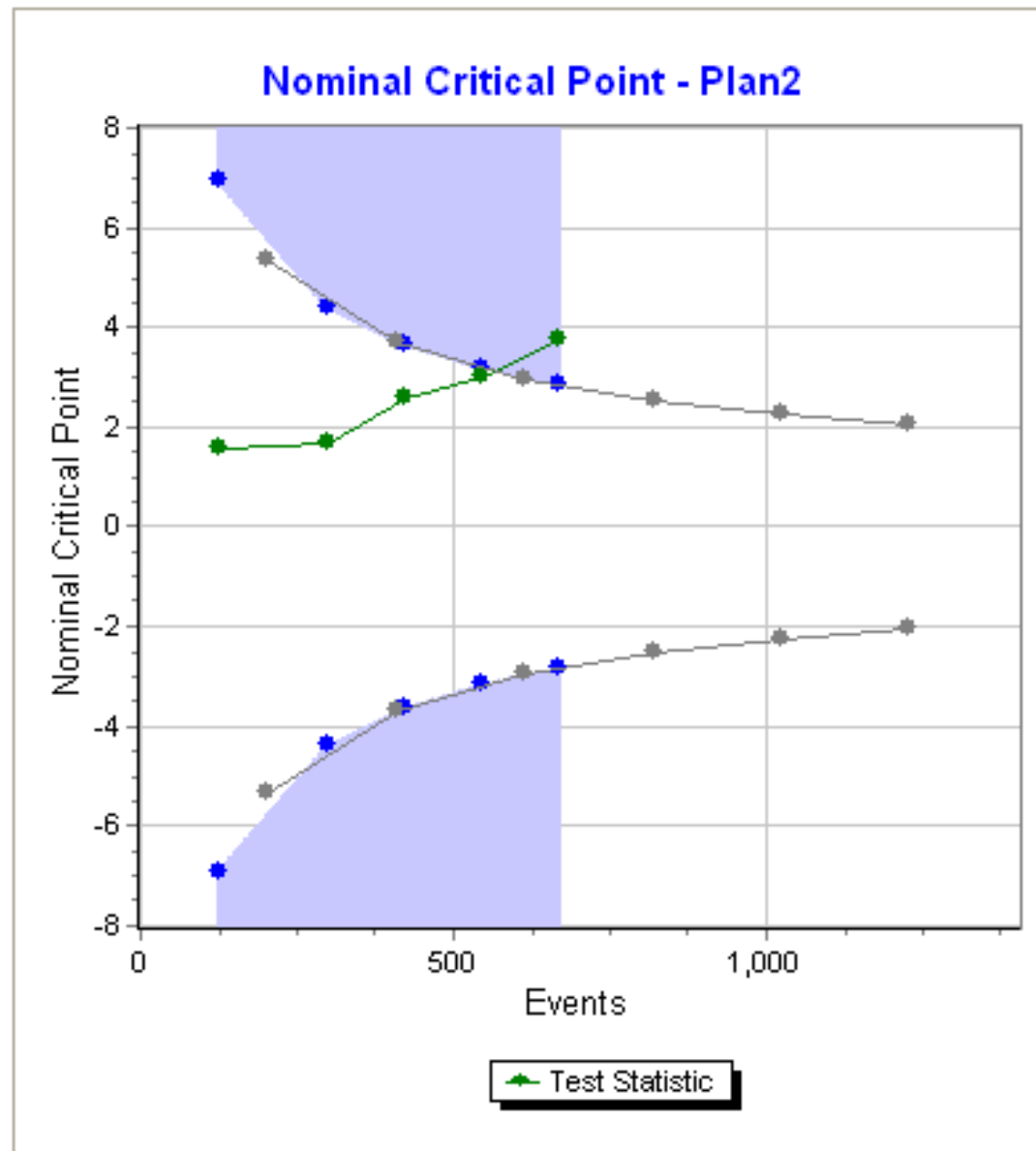
Interim Monitoring of RALES

The data monitoring committee stopped the trial at look-5 and thereby saved nearly two years of study duration

Date	Total Deaths	$\hat{\delta}$ (†)	$se(\hat{\delta})$ (†)	Z-Statistic
Aug 96	125	0.283	0.179	1.581
Mar 97	299	0.195	0.116	1.681
Aug 97	423	0.248	0.097	2.557
Mar 98	545	0.259	0.086	3.012
Aug 98	670	0.290	0.077	3.766

(†) Estimated from Z-Stat and Total Deaths (D): $se(\hat{\delta}) \approx \sqrt{4/D}$; $\hat{\delta} = Z \times se(\hat{\delta})$

Trial Stopped for Efficacy in August 1998 instead of June 2000



Cost versus Benefit of Group Sequential Design

Design Element	Single-Look Design	Six-Look Design
Sample Size	1660	1660
Required Final P-value	0.05	0.041
Maximum Study Duration	4.6 years	5.1 years
Expected Study Duration ^(†)	4.6 years	3.4 years
(†) If the new treatment works		

In fact the trial was terminated for efficacy in August 1998, nearly two years ahead of schedule

Part 4: Sample Size Re-Estimation in Phase III Trials

Motivation for Sample Size Re-Estimation in Confirmatory Trials

Despite best efforts, some of the crucial information used to design a confirmatory trial is unavailable at the design stage. For example:

- How variable are the data?
- What is the response rate of the placebo arm?
- How severe a problem is patient compliance and patient drop-out?
- For what treatment effect should the trial be powered?

Two Types of Uncertainties

1. Uncertainty about Nuisance Parameters

- Placebo response rate is uncertain
- Variability of population is uncertain
- Drop-out rate due to inability to tolerate drug is unknown

2. Uncertainty about Treatment Effect

- **New product not studied enough in the target population; (e.g., new product for treatment of HIV is a drug cocktail containing newly approved anti-retroviral medications)**
- **Changing medical practice might significantly affect an event rate**
- **External results might influence current thinking**
- **Conscious decision by sponsor to start out with a small initial commitment but add additional resources after seeing some interim data**

Psoriasis Example: Uncertainty about Nuisance Parameters

- Treatment versus placebo for high-need **psoriasis** patients. Primary endpoint is attainment of PASI-75 by week 16
- Design for 95% power to detect a 10% improvement with new treatment relative to placebo; **i.e. $\delta = 0.1$**
- 2-sided test at 5% significance level
- Current best guess is that placebo rate is 7.5%. But it could be anywhere in the range 5% to 15%

Sensitivity of Sample Size to Placebo Response Rate

- Investigators wish to power the trial to detect an improvement of 5% over placebo
- But the placebo rates could range anywhere between 5% and 15%
- All these different placebo rates will lead to different sample sizes

	Plan1	Plan2	Plan3
Plan ID			
Test Parameters			
1-Sided or 2-Sided Test	2-Sided	2-Sided	2-Sided
Significance Level (α)	0.05	0.05	0.05
Power (1 - β)	0.95	0.95	0.95
Assigned Fraction (Treatment)	0.5	0.5	0.5
Boundary Parameters			
Planned Number of Looks	3	3	3
Spacing of Looks	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)	LD (OF)	LD (OF)
Boundary to Reject H1			
Binomial Parameters			
Proportion Response (Control: π_c)	0.075	0.15	0.25
Difference in Proportions (δ_1)	0.1	0.1	0.1
Variance of Standardized Test Statistic	UnPooled	UnPooled	UnPooled
Accrual (Subjects)			
Maximum	562	828	1091
Expected Under H0	559	824	1086
Expected Under H1	426	627	827
Expected Under H1/2	533	785	1034

A flexible approach to sample size

- Let $\hat{\delta}$ be the estimate of δ from the trial and $\text{se}(\hat{\delta})$ be the standard error of this estimate
- The larger the sample size, N , the smaller will be $\text{se}(\hat{\delta})$
- The smaller the standard error, $\text{se}(\hat{\delta})$, the greater the statistical power
- We can show that to achieve 90% power to detect $\delta = 0.1$

$$\text{se}(\hat{\delta}) = 0.0275$$

Allow the sample size to float until desired $se(\hat{\delta})$ is attained

- For this example, which is powered at $\delta = 0.1$

$$se(\hat{\delta}) = \sqrt{\frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{N/2} + \frac{(\hat{\pi}_e)(1 - \hat{\pi}_e)}{N/2}}$$

- So keep the study open with a floating sample size until N is sufficiently large so that

$$se(\hat{\delta}) = 0.0275$$

Tabulate a Range of Sample Sizes in the Protocol

Since N is allowed to float until $\text{se}(\hat{\delta}) = 0.0275$, tabulate range of plausible values of N in the protocol.

π_c	0.05	0.075	0.10	0.15
N	460	562	657	828
$\text{se}(\hat{\delta})$	0.0275	0.0275	0.0275	0.0275

Note: Information is defined as $I = [\text{se}(\hat{\delta})]^{-2}$. This is known as a **maximum information design** because we are fixing the maximum information at $I = 0.0275^{-2} = 1313$.

Final Comments on Maximum Information Designs

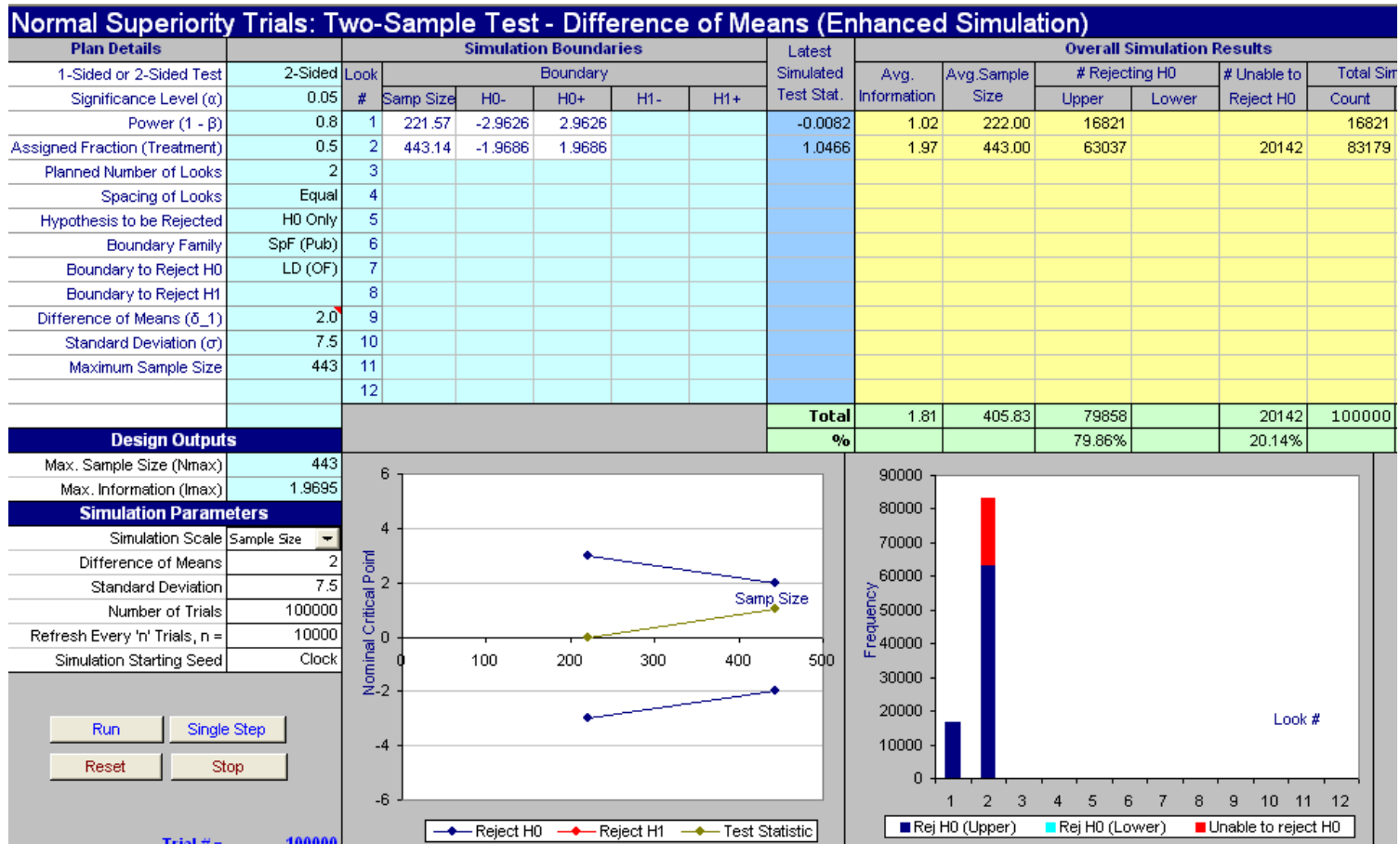
- Utilizes interim monitoring not just for early stopping but also to adjust the sample size
- By allowing sample size to float until the correct amount of information is attained, the study is always adequately powered
- Handles uncertainty about variance, placebo response rate, drop-outs and other nuisance parameters in a natural way using data from the trial itself
- Accepted by regulatory authorities

Schizophrenia Example: Uncertainty about Treatment Effect

- New drug versus placebo for treatment of a mental disorder
- Based on the limited information available sponsor powers the trial to detect a 2-point superiority ($\delta = 2$ with $\sigma = 7.5$) for new drug

Preliminary Design and Simulations

Normal Superiority Trials: Two-Sample Test - Difference	
Plan ID	Plan1
Test Parameters	
1-Sided or 2-Sided Test	2-Sided
Significance Level (α)	0.05
Power ($1 - \beta$)	0.8
Assigned Fraction (Treatment)	0.5
Boundary Parameters	
Planned Number of Looks	2
Spacing of Looks	Equal
Hypothesis to be Rejected	H0 Only
Boundary Family	SpF (Pub)
Boundary to Reject H0	LD (OF)
Boundary to Reject H1	
Normal Parameters	
Difference of Means (δ_1)	2.0
Standard Deviation (σ)	7.5
Accrual (Subjects)	
Maximum	443
Expected Under H0	442
Expected Under H1	407
Expected Under H1/2	438

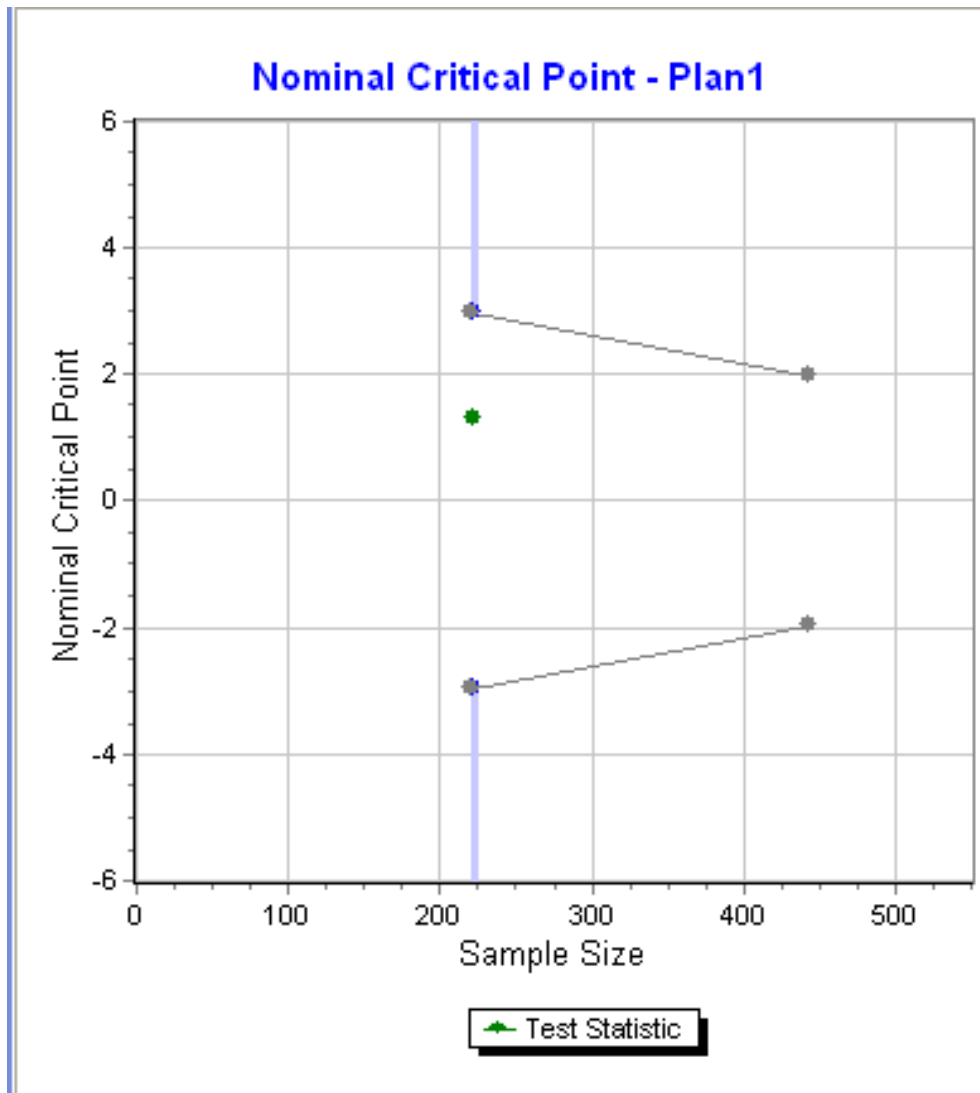


Run simulations under different scenarios to study impact of misjudging δ and σ

Conditional Power Calculations

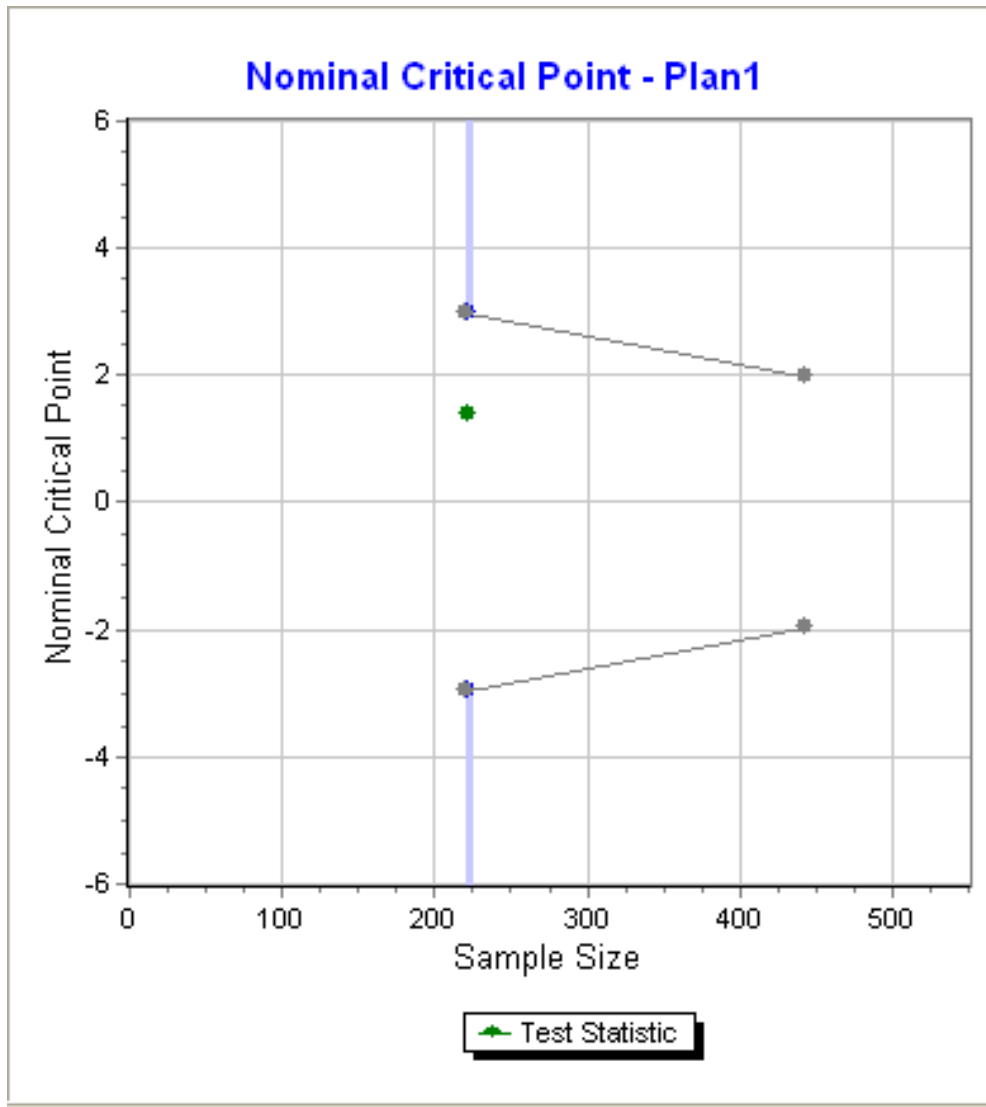
- At the interim analysis you will have data from your own study
- This will give you the opportunity to re-consider two crucially important design parameters:
 - Is it still reasonable to assume that $\sigma = 7.5$?
 - Should the study still be powered to detect $\delta = 2$?
- Should you change your mind about either or both these design parameters you would want to examine **conditional power** at different sample sizes

Sample size required to attain 80% conditional power: 986



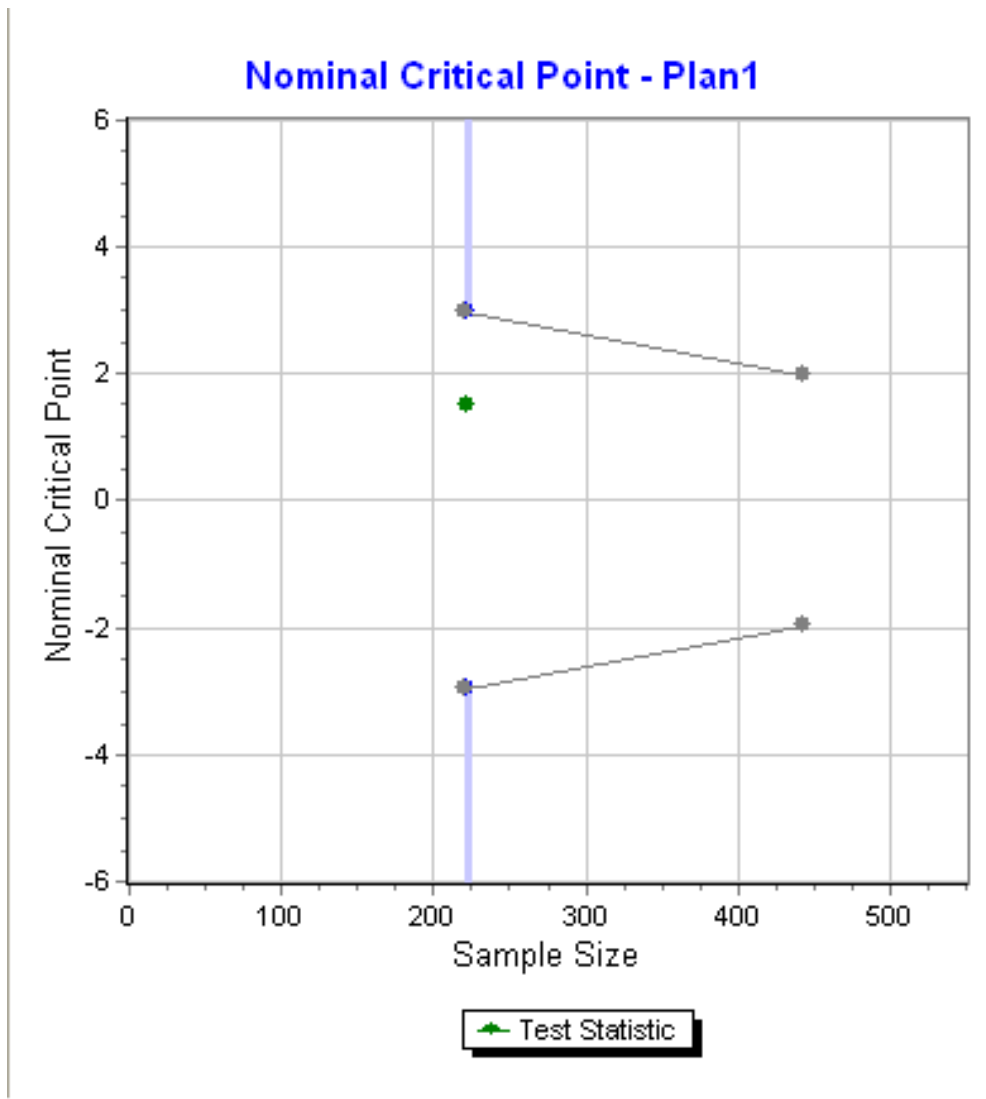
$\hat{\delta}_1$	z_1	$CP(\hat{\delta}_1)$
1.3	1.291	0.419

Sample size required to attain 80% conditional power: 828



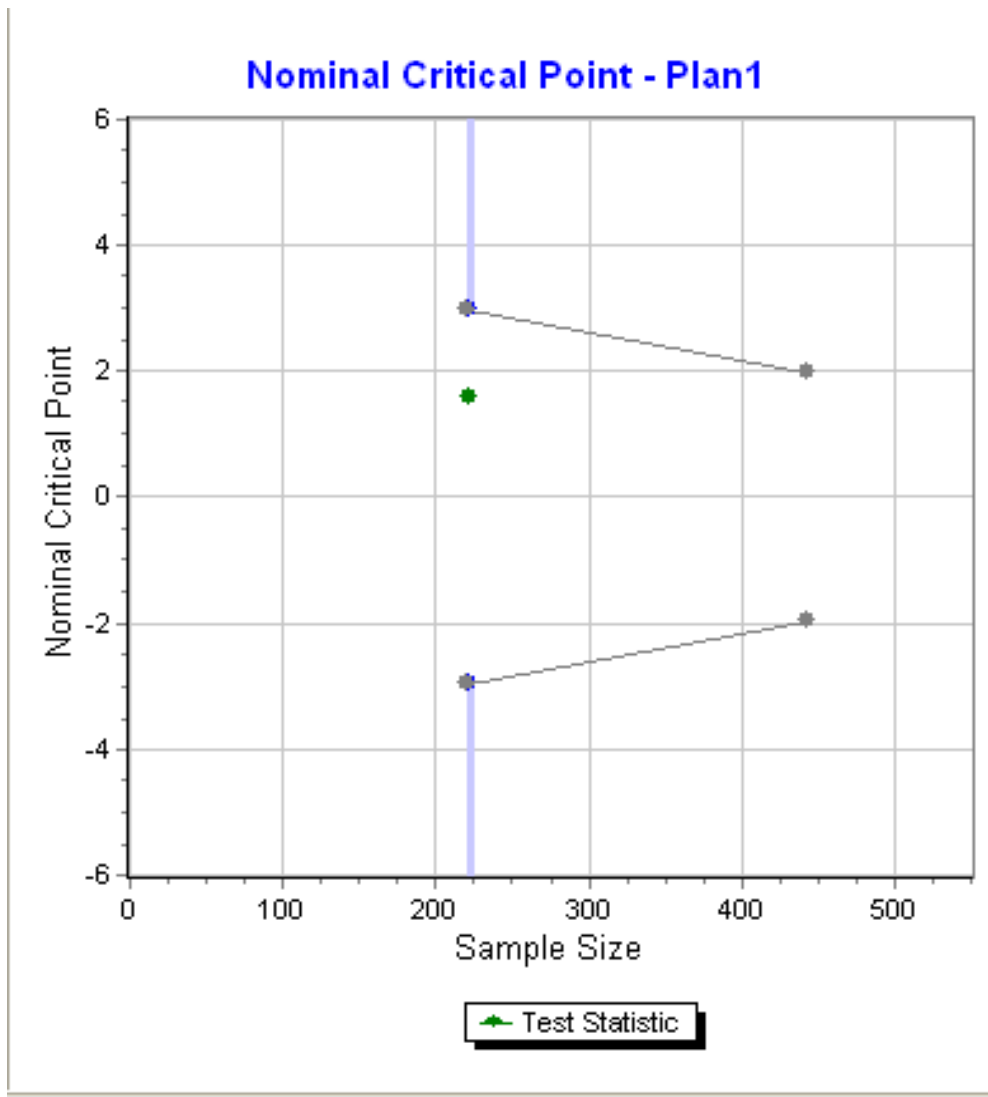
$\hat{\delta}_1$	z_1	$CP(\hat{\delta}_1)$
1.3	1.291	42%
1.4	1.391	50%

Sample size required to attain 80% conditional power: 704



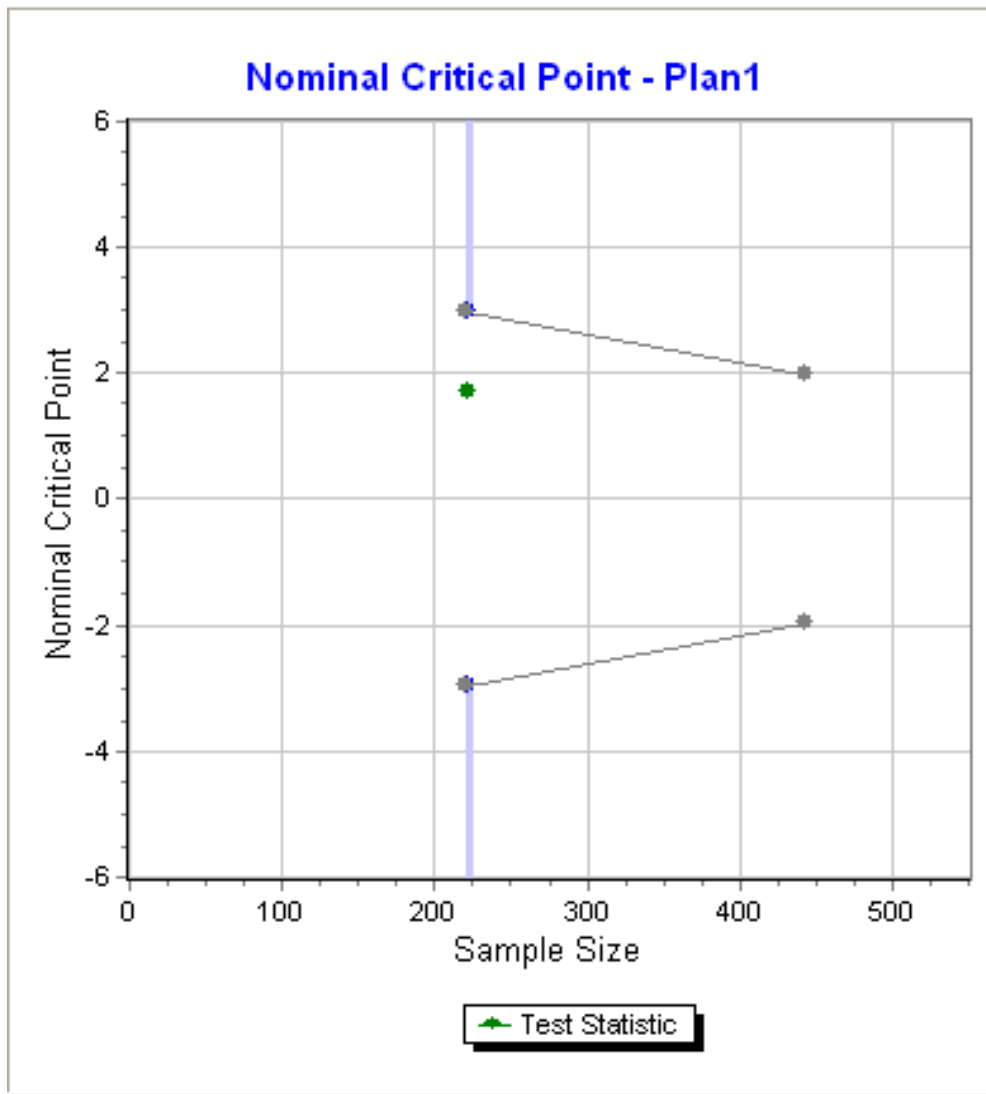
$\hat{\delta}_1$	z_1	$CP(\hat{\delta}_1)$
1.3	1.291	42%
1.4	1.391	50%
1.5	1.489	58%

Sample size required to attain 80% conditional power: 610



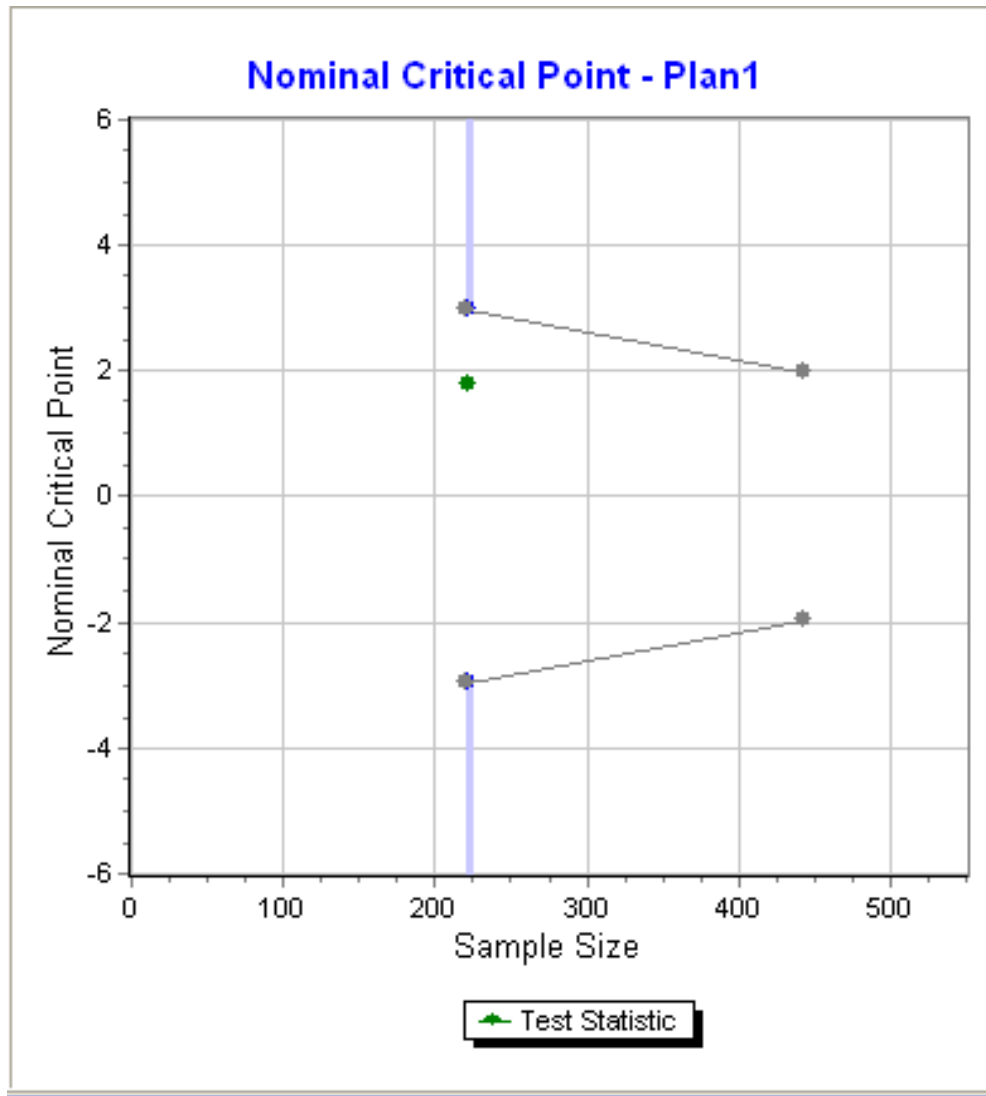
$\hat{\delta}_1$	z_1	$CP(\hat{\delta}_1)$
1.3	1.291	42%
1.4	1.391	50%
1.5	1.489	58%
1.6	1.589	65%

Sample size required to attain 80% conditional power: 534



$\hat{\delta}_1$	z_1	$CP(\hat{\delta}_1)$
1.3	1.291	42%
1.4	1.391	50%
1.5	1.489	58%
1.6	1.589	65%
1.7	1.689	72%

Sample size increase is not required



$\hat{\delta}_1$	z_1	$CP(\hat{\delta}_1)$
1.3	1.291	42%
1.4	1.391	50%
1.5	1.489	58%
1.6	1.589	65%
1.7	1.689	72%
1.8	1.788	79%

Developing Sample Size Re-Estimation Rules

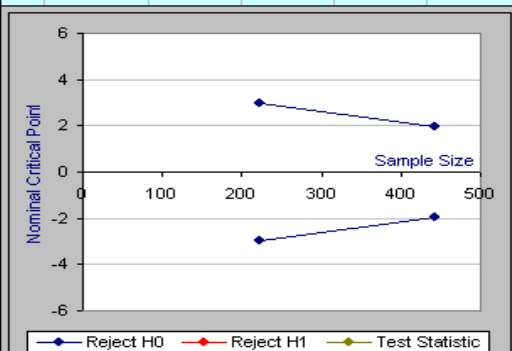
- Sponsor reviews resources he wishes to commit to this trial and develops a set of rules for how much, and under what conditions, sample size will be increased
- These rules are developed in consultation with the independent trial management committee
- Care is taken to ensure that the action recommended by the trial management committee cannot be “reverse engineered” so as to unblind the treatment effect

Typical Sample Size Re-Estimation Rules

- Do not exceed a maximum sample size of 900
- If $\hat{\delta} \geq 1.8$, leave the current maximum sample size unchanged
- If $1.5 < \hat{\delta} < 1.8$, alter the sample size to get as close as possible to 80% conditional power, subject to the maximum sample size limit of 900
- If $0 \leq \hat{\delta} \leq 1.5$, leave the current sample size (444) unchanged
- If $\hat{\delta} < 0$, terminate the trial for futility

Simulate the Trial

Normal Superiority Trials: Two-Sample Test - Difference of Means (CHW Simulation)													
Plan Details		Simulation Boundaries						Latest Simulated Test Stat.	Overall Simulation Results				
1-Sided or 2-Sided Test	2-Sided	Look #	Boundary				Avg. Info.		Avg. Sample Size	# Rejecting H0		# Unable to Reject H0	Total Simulation
Significance Level (α)	0.05	#	Sample Size	H0-	H0+	H1-	H1+			Upper	Lower	Count	%
Power (1 - β)	0.8	1	221.57	-2.9626	2.9626								
Assigned Fraction (Treatment)	0.5	2	443.14	-1.9686	1.9686								
Planned Number of Looks	2	3											
Spacing of Looks	Equal	4											
Hypothesis to be Rejected	H0 Only	5											
Boundary Family	SpF (Pub)	6											
Boundary to Reject H0	LD (OF)	7											
Boundary to Reject H1		8											
Difference of Means (δ_1)	2.0	9											
Standard Deviation (σ)	7.5	10											
Maximum Sample Size	443	11											
		12											
Design Outputs													
Max. Sample Size (Nmax)	443												
Max. Information (Imax)	1.9695												
Simulation Parameters													
Difference of Means	2.0000												
Standard Deviation	7.5000												
Use Nmax Till 'L' Looks, L =	1												
Criterion for Rescaling: Min. δ_{hat}	2.0000												
Max. δ_{hat}	2.0000												
Min. Usable Sample Size	444												
Max. Usable Sample Size	444												
Desired Conditional Power (Cp)	0.9000												
Number of Trials	1000												
Refresh Every 'n' Trials, n =	100												
Simulation Starting Seed	Clock												
								Total					
								%					
Simulation Results for Adapted Trial Only													
								Total					
								%					



Membership of the Independent Trial Management Committee

- Membership consist of an experienced biostatistician and at least of clinician who is familiar with the disease and (preferably) the drug
- This committee does not deal with safety issues. These are dealt with by the regular data monitoring committee.
- If safety is a concern, the recommendations of the trial management committee can be overruled by the regular data monitoring committee
- The sponsor can attend the deliberations of the trial management committee as long as only blinded results are being discussed. The sponsor may not attend the executive session in which unblinded results are discussed

Interactions with the FDA

- Many adaptive designs are complex. Engage the the FDA in discussions early in the design stage. They may take a long time to respond
- In addition to theoretical arguments, adaptive simulations that demonstrate that the overall type-1 error is preserved despite the data dependent sample size change should be made available to the FDA

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