
Adaptive Sample Size Re-estimation for Confirmatory Time to Event Trials

Joint Statistical Meetings
Vancouver, Canada
August 4, 2010

Cyrus R. Mehta, Ph.D
Cytel Inc., Cambridge, MA

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

Motivation: arises from oncology trials (but is broadly applicable)

- Primary endpoint is usually overall survival (OS)
- Small gains in OS (e.g. hazard ratios between 0.75 and 0.8) are nevertheless clinically meaningful
- Sample size requirements for such small gains are large, and pose a major design challenge

Promising Zone designs resolve this difficulty by requiring a smaller up-front sample size commitment, to be followed up by a larger commitment only if interim results are promising

Lung Cancer Example

- Two arm double-blind multicenter trial with second line therapy for metastatic non-small cell lung cancer
- Primary endpoint is overall survival (OS)
- Median for control arm is 8 months
- Require 90% power to detect $HR = 0.7$ (median = 11.4 months on experimental arm)
- One-sided level 0.025 test with one interim look for early efficacy or futility stopping
- Design 24 month enrollment and 12 months additional follow-up

Group Sequential Design

Survival Dependency Trials: Two Sample Test - Logrank Test: Given Accrual Duration and Sta

Plan ID	Plan1	Plan2
Test Parameters		
1-Sided or 2-Sided Test	1-Sided	1-Sided
Significance Level (Alpha)	0.025	0.025
Power (1 - Beta)	0.9	0.9
Assigned Fraction (Treatment)	0.5	0.5
Boundary Parameters		
Planned Number of Looks	2	2
Spacing of Looks	Equal	Equal
Hypothesis to be Rejected	H0 or H1 (NB)	H0 or H1 (NB)
Boundary Family	SpF (Pub)	SpF (Pub)
Boundary to Reject H0	LD (OF)	LD (OF)
Boundary to Reject H1	Gm (-5)	Gm (-5)
Survival Parameters		
-Log-hazard Ratio	0.3567	0.2614
Number of Hazard Pieces	1	1
Number of Accrual Periods	1	1
Variance of -Log-hazard Ratio	Null	Null
Committed Accrual		
Committed Accrual (Duration)	24.0	24.0
Committed Accrual (Subjects)	417	763
Max. Duration and Events		
Maximum Study Duration	36.0	36.0
Maximum Number of Events	333	620
Expected Values under...		
Expected Accrual (Subjects)	H0: 377 H1: 400 H 1/2: 407	H0: 692 H1: 730 H 1/2: 743
Expected Study Duration	26.348 31.852 31.981	26.949 31.821 32.5
Expected Number of Events	258 290 311	480 539 578

Uncertainty about Treatment Effect

- If $HR=0.7$ we need 333 events for 90% power; enroll 417 subjects over 24 months; follow for 12 additional months
- Could easily have underestimated hazard ratio due to improved standard of care, or weaker treatment effect
- If $HR = 0.77$ we need 620 events for 90% power
 - enroll 763 subjects over 24 months and follow for 12 additional months
 - or enroll 672 subjects over 24 months and follow for 24 additional months
- **Sponsor cannot make such large up-front commitments**

Sponsor is Resource and Time Constrained

- Unable to invest up-front to protect power in case of pessimistic scenario
- But willing to invest additional resources if interim results are promising

True HR	Power of Optimistic Design (design for HR=0.7)	Power of Pessimistic Design (design for HR=0.77)
0.7	90%	99%
0.75	74%	95%
0.77	66%	90%

Sponsor Adopts an Adaptive Strategy

- Design optimistically (HR=0.7; 333 events; 417 subjects)
- One interim analysis after 50% information
 - Stop early if overwhelming evidence of efficacy
 - Stop early for futility if low conditional power
 - Increase number of events, sample size at the interim **if interim results fall in a promising zone**
- Can define promising zone equivalently in terms of **conditional power, or HR, or Z-statistic**

Conditional Power Calculator

The screenshot shows a software window titled "Conditional Power Calculator". It contains several input fields and radio buttons. The "Input" section has "Current Look" set to 1 and "Current # of Event" set to 167. The "Input/Output" section has "HR to be Used in Conditional Power Computation" with radio buttons for "User Defined (HR)" and "Estimated (HR, z)", where "Estimated (HR, z)" is selected. Below this, there are two more radio buttons: "Computed Value of HR" (selected) and "Computed value of z". The "Computed Value of HR" field contains 0.830311, and the "Computed value of z" field contains 1.201534. The "Conditional Power" field contains 0.35, and the "# of Events (Overall)" field contains 333. At the bottom, there are "Recalc" and "Close" buttons. A note at the bottom states: "* Use the radio button to select the quantity to be computed."

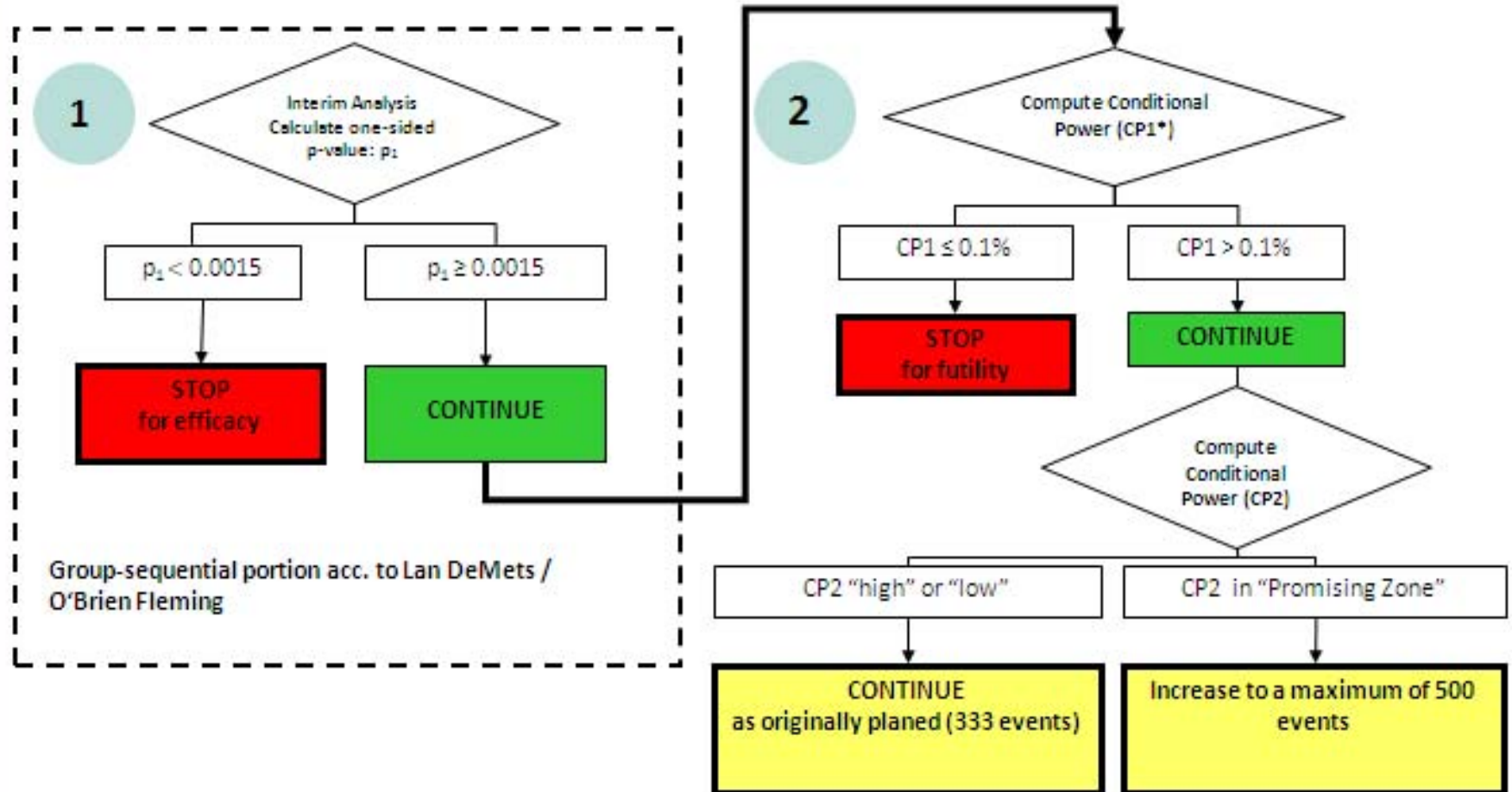
Field	Value
Current Look	1
Current # of Event	167
HR to be Used in Conditional Power Computation	Estimated (HR, z)
Computed Value of HR	0.830311
Computed value of z	1.201534
Conditional Power	0.35
# of Events (Overall)	333

(In this example 35% CP at interim corresponds to $HR=0.83$)

The Promising Zone Design

- Estimate the conditional power at the interim look
 - Unfavorable:** $CP < 35\%$; no change in design
 - Promising:** $35\% \leq CP < 90\%$; increase resources
 - Favorable:** $CP \geq 90\%$; no change in design
- Use simulation to experiment with promising zones
- Use simulation to experiment with sample size re-estimation rules
- Use Cui, Hung, Wang (CHW) method or Chen, DeMets Lan (CDL) method to control type-1 error

Schema of Trial Design



Adaptive Distribution Theory

- Let δ be the mean difference of two normal distributions with common variance σ^2
- Test $H_0: \delta = 0$ versus $H_1: \delta > 0$ with a K-look group sequential design
- For $j = 1, \dots, K$ define:
 - $b_j =$ level- α stopping boundaries for initial design
 - $n_j =$ cumulative sample sizes of initial design
 - $n_j^* =$ cumulative sample sizes after adaptation
 - $n^{(j)} = n_j - n_{j-1}$, $n^{*(j)} = n_j^* - n_{j-1}^*$; incremental data
 - $(\hat{\delta}^{(j)}, \hat{\delta}^{*(j)}) =$ estimates based on incremental data

Cui, Hung and Wang (CHW) Test

The CHW statistic is formed by combining the incremental Wald statistics

$$Z^{*(l)} = \frac{\hat{\delta}^{*(l)}}{\text{se}(\hat{\delta}^{*(l)})} = \hat{\delta}^{*(l)} \sqrt{I^{*(l)}}, \quad l = 1, 2, \dots, j,$$

with the **prespecified** weights

$$w^{(l)} = \frac{n^{(l)}}{n_K} \quad l = 1, 2, \dots, j$$

so as to form the weighted statistic

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

CHW Test; continued

- This statistic is asymptotically normal 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. Thus, under H_0 , $Z_{j,\text{chw}}^* \sim N(0, 1)$

- As long as the weights $w^{(1)}, w^{(2)}, \dots, w^{(K)}$ are pre-specified,

$$\text{corr}(Z_{j_1,\text{chw}}^*, Z_{j_2,\text{chw}}^*) = \sqrt{\frac{n^{(j_1)}}{n^{(j_2)}}}, \quad 1 \leq j_1 < j_2 \leq K$$

- It follows that, regardless of adaptive sample size change,

$$P_0\left(\bigcup_{j=1}^K Z_{j,\text{chw}}^* \geq b_j\right) = \alpha$$

Repeated Confidence Intervals

We can show (Lehmacher and Wassmer, 1999) that the K repeated confidence intervals for δ are given by

$$\frac{(Z_{j,\text{chw}}^* \pm b_j) \sqrt{s_j}}{\sum_{l=1}^j \sqrt{w^{(l)} I^{*(l)}}}, \quad j = 1, 2, \dots, K$$

where $s_j = n_j/n_K$ is the information fraction at look j based on the pre-specified sample sizes. Thus, if δ_0 is the true value of δ then, for all $j = 1, 2, \dots, K$,

$$P_{\delta_0} \left\{ \bigcap_{i=1}^j \left(\frac{(Z_{i,\text{chw}}^* - b_i) \sqrt{s_i}}{\sum_{l=1}^i \sqrt{w^{(l)} I^{*(l)}}} \leq \delta_0 \leq \frac{(Z_{i,\text{chw}}^* + b_i) \sqrt{s_i}}{\sum_{l=1}^i \sqrt{w^{(l)} I^{*(l)}}} \right) \right\} \geq 1 - \alpha$$

Repeated p-values at each look are obtained by manipulating α such that the RCI just excludes $\delta = 0$

Conditional Power Calculation

- Sample size increase is based on CP (promising zone design)
- Suppose an interim look is taken at look $L < K$ and the observed value of the test statistic is $Z_{L,\text{chw}}^* = z_L$. Then

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

- East provides a CP calculator to perform this computation
- For the simulations, however, East ignores all intermediate looks between L and K . The approximate CP is given by

$$\text{CP}_\delta(z_L) \approx 1 - \Phi \left\{ b_K \sqrt{1 + \frac{n_L}{n_K - n_L}} - z_L \sqrt{\frac{n_L}{n_K - n_L}} - \frac{\delta \sqrt{n_K^* - n_L}}{\sqrt{2}\sigma} \right\}$$

- Either the design δ or the estimated δ may be used for the CP calculations required by the simulations

Special Case of Survival Studies

- Treatment effect $\delta = -\ln(\text{HR})$; $\sigma = 1$
- Let D_j (D_j^*) denote the number of events in the initial (adapted) designs at looks $j = 1, 2, \dots, K$.
- Event driven trial. D_j plays the role of n_j and D_j^* plays the role of n_j^* for trial design
- Let LR_j be the logrank statistic based on all data upto and including look j and define

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

where r is the randomization fraction

- With the above substitutions all the previous results for normal and binomial endpoints carry over to the survival setting

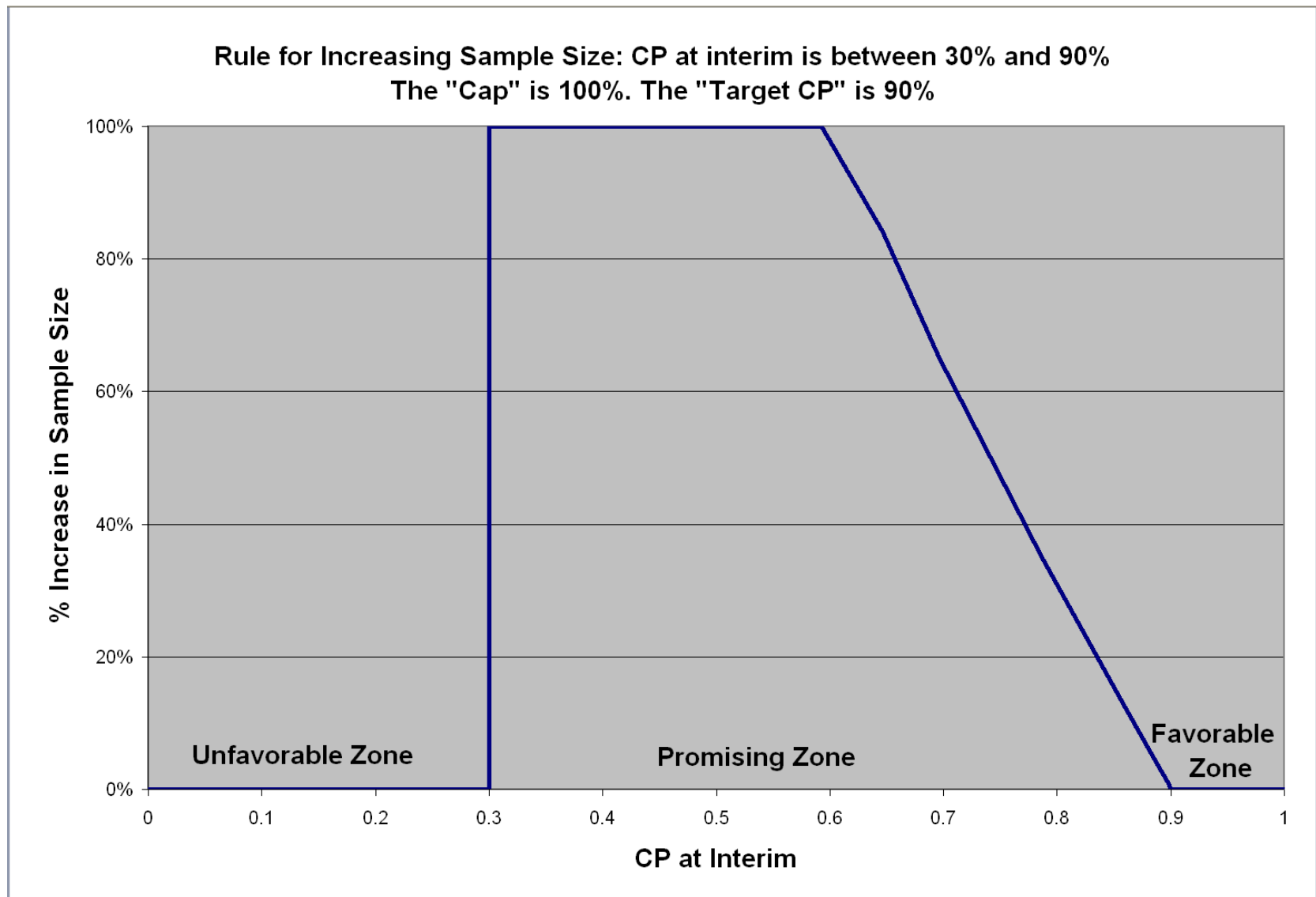
Return to NSCLC Example

- Sponsor cannot make the up-front commitment needed to design for $HR=0.77$
- Instead, sponsor designs for $HR=0.7$ with 333 events and 418 subjects
- But, sponsor might be willing to invest more resources at the interim look, if results fall in promising zone ($0.35 \leq CP < 0.9$) because then the chances of success would increase dramatically (as we now show)

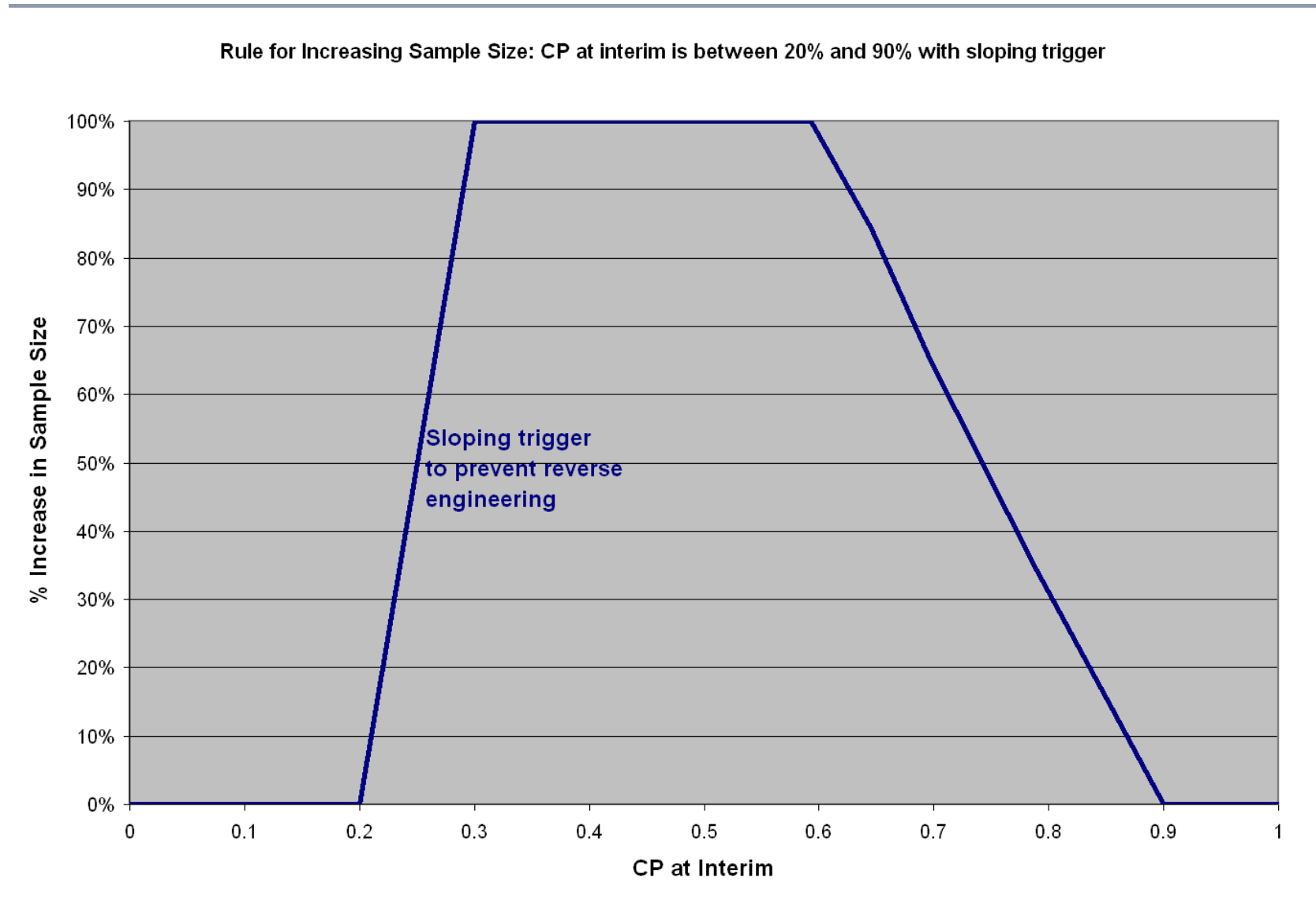
Adaptation Principles

- **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**

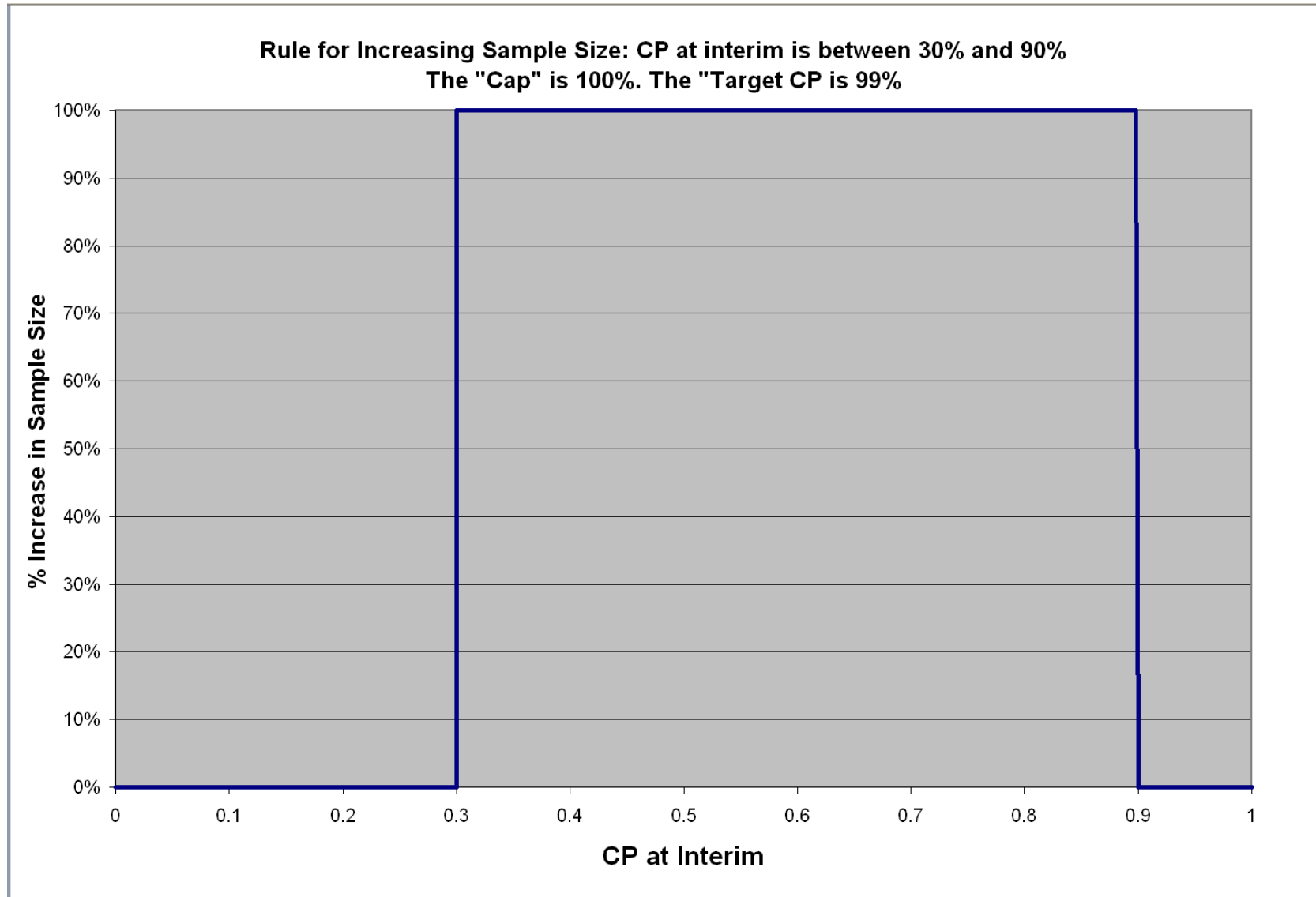
Increasing Number of Events: 1



Increasing Number of Events: 2



Increasing Number of Events: 3



Simulation Based Operating Characteristics

1. Generate first-stage data with accrual and hazard rates $(a, \lambda_c, \lambda_t)$ as specified in the simulation inputs worksheet
2. At interim analysis, determine the desired maximum number of events D^* based on the conditional power rules
3. Generate the second-stage data and stop the trial when required number of events are obtained
4. Examine power, study duration and sample size, zone by zone and also unconditionally

Explore changes in promising zone and adaptation rules

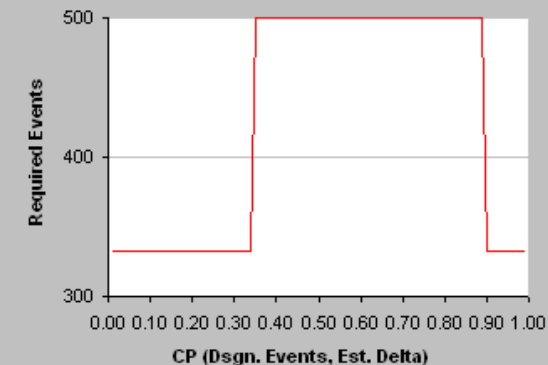
Adaptive Simulation Worksheet

Survival Superiority Trials: Two Sample Test - Logrank Test: Given Accrual Duration and Study Duration (Survival CHW Simulation)

Perform Adaptation if Necessary (During Simulations)

Input Parameters	
Adaptation at Look L	1
Max. Events if Adapt (multiplier; total #)	1.50 500
Sample Size if Adapt (multiplier; total #)	1.50 627
Expected Study Duration if Adapt	
Upper Limit on Study Duration	108.00
Shape Parameter for Re-estimating # of Events	0.99
Promising Zone :	
Min CP:	0.35
Max CP:	0.90
Type of Adaptation	Increase Sample Size
Accrual Rate After Adaptation	No Change

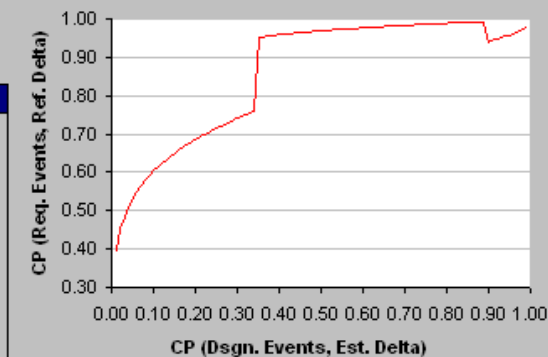
Output for all Trials				
Show Summary for: All Trials				
Percentile	Study Duration	Number of Events	Accrual Duration	Number of Subjects
5%	19.6	167	19.5	341
25%	33.7	333	23.9	418
50%	35.3	333	24.0	418
75%	42.1	500	35.6	627
95%	44.0	500	36.4	627
Average	35.4	364	27.2	475



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 %				
Futility			330	100.0%	330	3.3%	19.7	167	19.7	344
Unfavorable: CP < 0.350	988	34.9%	1844	65.1%	2832	28.3%	34.7	333	23.9	418
Promising: 0.350 ≤ CP < 0.900	2843	89.0%	351	11.0%	3194	31.9%	42.9	500	35.9	627
Favorable: CP ≥ 0.900	2362	89.8%	269	10.2%	2631	26.3%	34.8	333	23.9	418
Efficacy	1013	100.0%			1013	10.1%	19.8	167	19.8	346
All Trials	7206	72.1%	2794	27.9%	10000	100.0%	35.4	364	27.2	475



Reference HR 0.7000 Refresh Charts

Simulation Inputs: Overall Simulation Outputs: Overall Simulation Inputs and Outputs: Adaptive Exploration

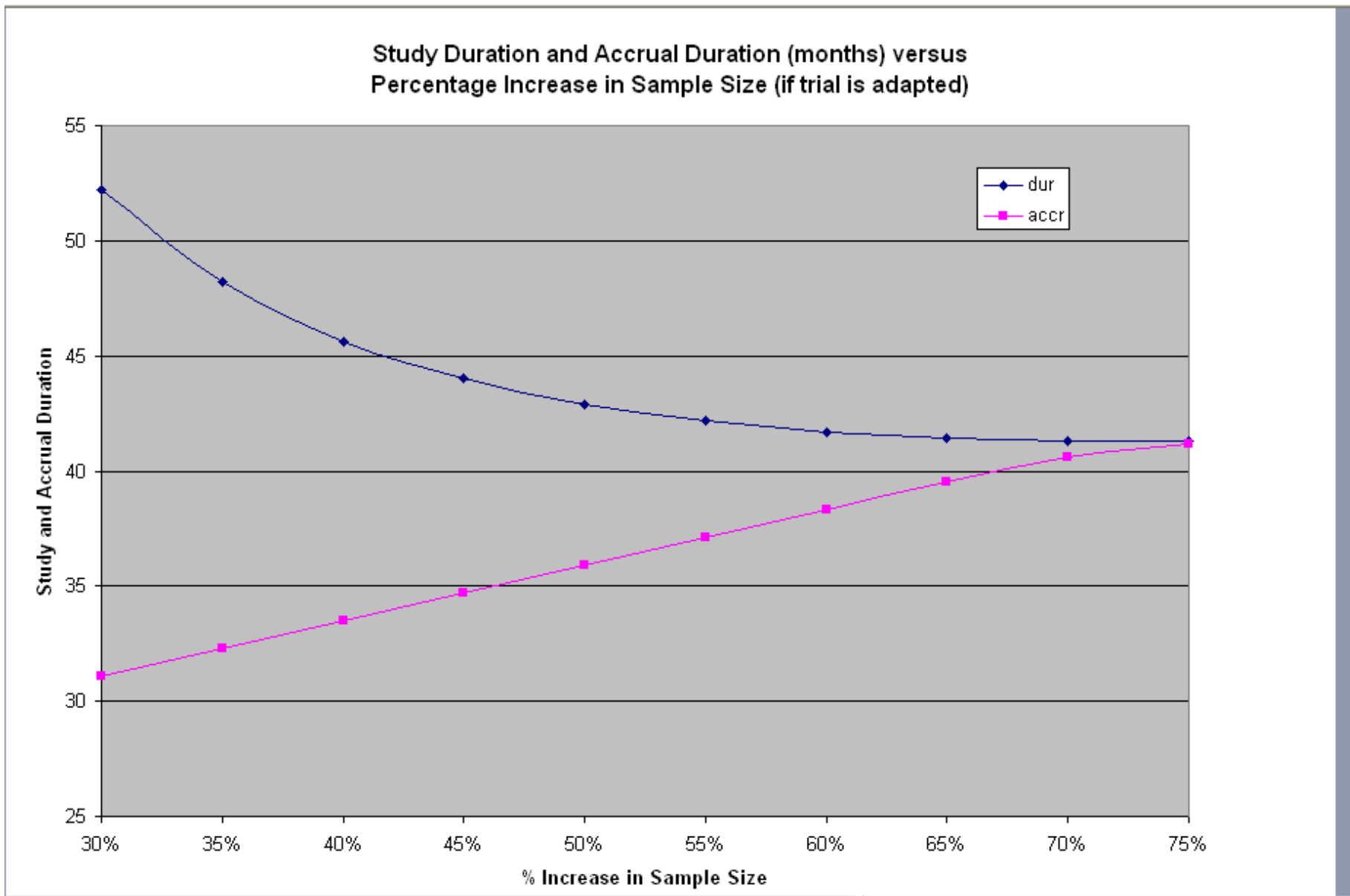
Design Plan1_SurvivalCHW_Sim/

Ready

Sample Size versus Study Duration Trade-Off

- Suppose you have entered the promising zone and the new number of events is D_{\max}^*
- How will you pick the new sample size, N_{\max}^* to go along with the new number of events?
 - If N_{\max}^* is too small, the trial will be excessively prolonged
 - If N_{\max}^* is too large, the trial costs will be excessive while time savings might be marginal
- Obtain an accrual-duration chart by simulation and choose N_{\max}^* by inspection

Show Summary for		Promising						
Type of Adaptation		Simulation Count	Power	Avg. Study Duration	Avg. Accrual Duration	Avg. Number of Events	Avg. Number of Subjects	
Multiplier	Total							
1.30	543	3188	88.0%	52.2	31.1	500	543	
1.35	564	3261	87.9%	48.2	32.3	500	564	
1.40	585	3288	88.0%	45.6	33.5	500	585	
1.45	606	3297	88.2%	44.0	34.7	500	606	
1.50	627	3103	87.3%	42.9	35.9	500	627	
1.55	648	3250	89.1%	42.2	37.1	500	648	
1.60	669	3206	88.3%	41.7	38.3	500	669	
1.65	690	3239	88.7%	41.4	39.5	500	690	
1.70	711	3158	88.2%	41.3	40.6	500	709	
1.75	732	3232	88.3%	41.3	41.2	500	719	



Unnecessary to increase sample size by more than 50%

Operating Characteristics of Optimistic Design (Powered to Detect HR=0.7)

1. Simulations Under Pessimistic Scenario, HR = 0.77 (10,000 simulations)

Zone	P(Zone)	Power		Duration (months)		SampSize	
		NonAdpt	Adapt	NonAdpt	Adapt	NonAdpt	Adapt
Unf	32%	31%	31%	33	33	409	409
Prom	32%	69%	88%	35	43	418	627
Fav	36%	93%	93%	31	31	398	398
Total	—	66%	72%	33	35	408	476

2. Simulations Under Optimistic Scenario, HR = 0.7 (10,000 simulations)

Zone	P(Zone)	Power		Duration		SampSize	
		NonAdpt	Adapt	NonAdpt	Adapt	NonAdpt	Adapt
Unf	14%	57%	57%	35	35	414	414
Prom	26%	88%	98%	36	44	418	627
Fav	60%	98%	98%	29	29	390	390
Total	—	90%	93%	32	34	401	454

Simulations at Interim Analysis Stage

- At interim analysis, DMC has access to unblinded data
- They will know the exact value of D_{\max}^*
- They will have updated estimates of $(a, \lambda_c, \lambda_t)$
- With this information DMC can update the chart of percentage increase in sample size versus study duration
- This could result in a recommendation to decrease the sample size if pre-specified in DMC charter

Interim Analysis Worksheet

Survival Superiority Trials: Two Sample Test - Logrank Test: Given Accrual Duration and Study Duration (Plan1) (CH

Plan Details		Look #	Incr. Events	Incr. Statistic	Cumul. Events	Prespec. Weights	Weighted Statistic	Prespecified Nominal Critical Points				Repeated 87.50% CI for HR		Repeated p-value
								Reject H0		Reject H1		Lower	Upper	
1-Sided or 2-Sided Test	1-Sided	1	175	1.475956	175	0.5000	1.475956		2.962588	-0.126569	0.549386	1.252019	0.200064	
Significance Level (Alpha)	0.025	2				0.5000			1.968596	1.968596				
Power (1 - Beta)	0.9	3												
Assigned Fraction (Treatment)	0.5	4												
Planned Number of Looks	2	5												
Spacing of Looks	Equal	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	Gm (-5)	10												
-Log-hazard Ratio	0.357													
Variance of -Log-hazard Ratio	Null													
Maximum Number of Events	333													
Maximum Study Duration	36													

Conditional Power Calculator

Input

Current Look:

Current # of Event:

Current Weighted Test Statistic:

Input/Output

Value of HR:

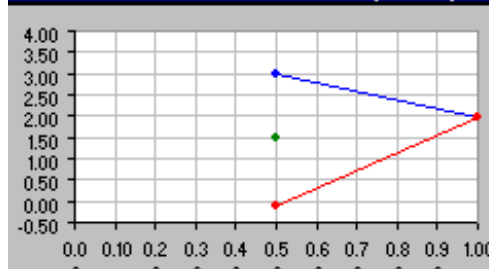
Computed Conditional Power:

of Events (Overall):

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

Recalc Plot Close

Nominal Critical Point Chart (Select) ↑



The CDL Method

- Some have objected to using the weighted statistic instead of the conventional statistic for performing the hypothesis test
- Chen, DeMets and Lan (2004) have shown that if promising zone starts at $CP \geq 0.5$ it is ok to use the conventional statistic.
- Mehta and Pocock (2010) have extended this result
- Depending on event multiplier, target conditional power, and time of interim look, the promising zone can be widened as shown on the following table

CP_{min} for Various Design Options

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

Simulating the CDL Method

Input Parameters		
Adaptation at Look L		1
Max. Events if Adapt (multiplier; total #)	10.00	3300
Sample Size if Adapt (multiplier; total #)	10.00	4140
Expected Study Duration if Adapt		
Upper Limit on Study Duration		500.00
Shape Parameter for Re-estimating # of Events		0.90
Promising Zone :	Min CP:	0.00
	Max CP:	0.90
Use Wald Stat. if CP(330) >=		0.00
Type of Adaptation	Increase Sample Size	
Accrual Rate After Adaptation	No Change	

Output for all Trials				
Show Summary for				All Trials
Percentile	Study Duration	Number of Events	Accrual Duration	Number of Subjects
50%	34.2	330	24.0	414
75%	200.5	3300	200.5	3479
90%	203.2	3300	203.2	3503
95%	204.1	3300	204.0	3511
99%	205.5	3300	205.4	3524
Average	91.7	1368	87.3	1507

Simulation Results by Zone										
Zone	Simulations Rejecting H0		Simulations not Rejecting H0		Total Simulations		Avg. Study Duration	Avg. Number of Events	Avg. Accrual Duration	Avg. Number of Subjects
	Count	Row %	Count	Row %	Count	Column %				
Futility										
Unfavorable: CP < 0.000										
Promising: 0.000 ≤ CP < 0.900	188	1.92%	9596	98.08%	9784	97.84%	93.0	1391	88.7	1531
Favorable: CP ≥ 0.900	92	42.59%	124	57.41%	216	2.16%	32.3	330	23.9	414
Efficacy										
All Trials	280	2.80%	9720	97.20%	10000	100.00%	91.7	1368	87.3	1507

References

1. Chen YHJ, DeMets DL, Lan KKG. Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine* 2004; 23, 1023-1038.
2. Cui L, Hung HMJ, Wang S-J. Modification of sample size in group sequential trials. *Biometrics* 1999; 55, 853-857.
3. Gao P, Ware JH, Mehta CR. Sample size re-estimation for adaptive sequential design. *J. Biopharmaceutical Statistics* 2008; 18, 1184-1196.
4. Lehmacher W, Wassmer G. Adaptive sample-size calculations in group sequential trials. *Biometrics* 1999; 55, 1286-1290.