

Inference for Multi-Arm Multi-Stage Adaptive Clinical Trials

Cyrus Mehta, Ph.D.

Cytel Inc.

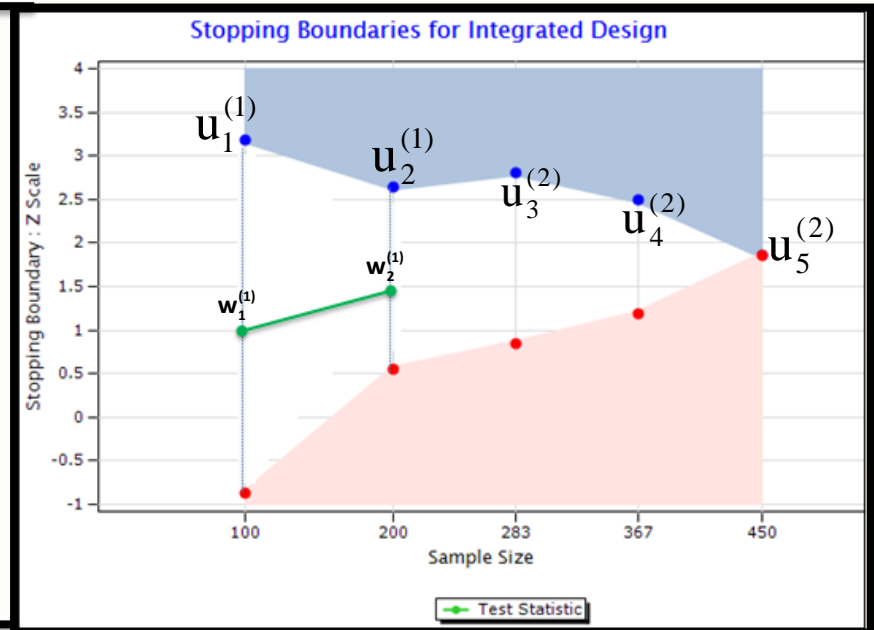
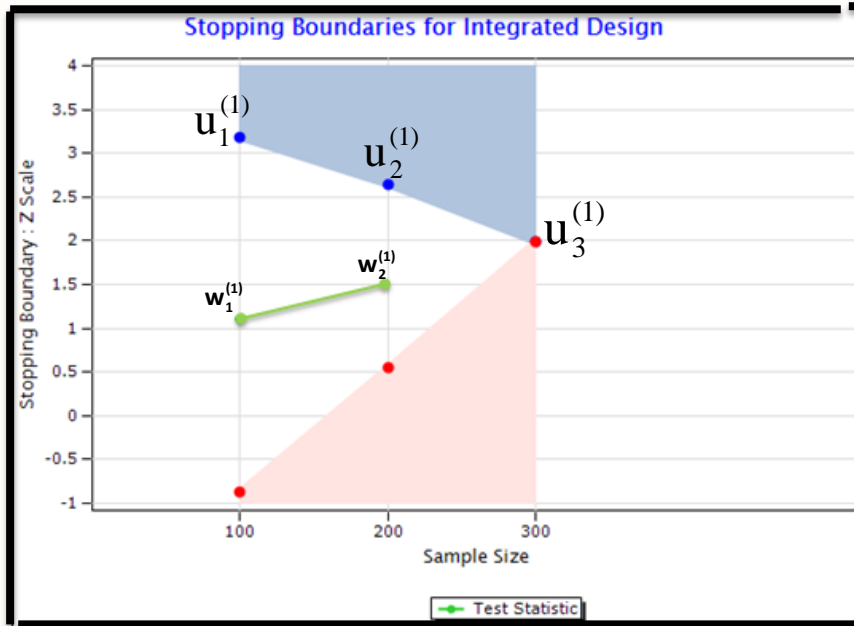
- This is joint work with
 - Ping Gao, The Medicines Company
 - Lingyun Liu, Cytel Inc
 - Pranab Ghosh, Cytel Inc

1. Quick introduction to Adaptive Group Sequential Designs
2. Preservation of α and parameter estimation in 2-arm **classical** GSD
3. Preservation of α and parameter estimation in 2-arm **adaptive** GSD
4. Extension to multi-arm adaptive GSD
5. Future directions

- An independent data monitoring committee monitors the accruing data:
 - Monitoring is unblinded
 - Monitoring times are pre-specified but flexible
- Possible decisions at each monitoring time
 - Stop early due to overwhelming efficacy
 - Stop early for futility
 - Make changes to future course of trial: (increase sample size, change number of looks, drop arms)

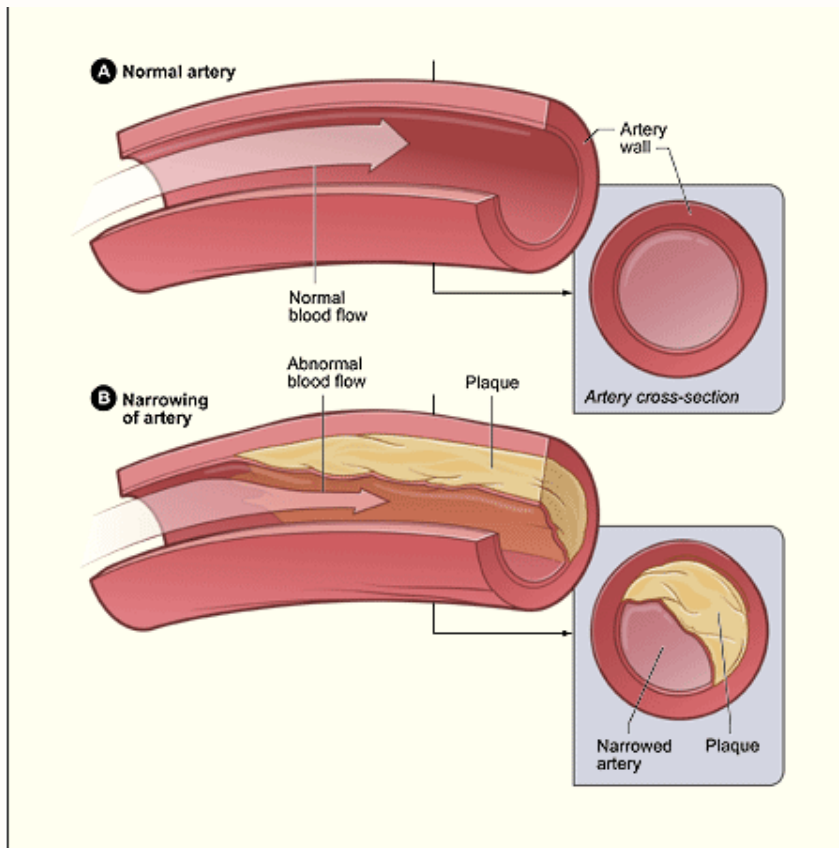
Design (1): Non-Adaptive

Design (2): Adaptive



Sample size, number and spacing of future looks were adapted at look 2 of GSD

Plaque build-up in arteries



- MDCO-216 is a macromolecular complex of a recombinant protein and phospholipid
- MDCO-216 is delivered by IV infusion
- Reduces plaque burden (as measured by ultrasound)
- This clinical trial will compare four doses (4mg/kg, 8 mg/kg, 16 mg/kg, 32 mg/kg) of MDCO-216 to a standard of care control arm
- Three analyses with possible dose selection and sample size re-estimation

- Methods for α error control in adaptive trials are well developed
- Methods for p-values, confidence intervals and parameter estimates not well developed
- Available Methods for two-arm multi stage:
 - Extension of RCI (Mehta et. al. , 2008): do not exhaust the α , hence conservative
 - Extension of SWCI (Brannath et al 2009, Gao et al 2013): exhaust the α , hence provide exact coverage
- Nothing available so far for multi-arm multi-stage adaptive group sequential designs

- For a K-look group sequential design define:

δ = $\mu_1 - \mu_0$ is the treatment effect

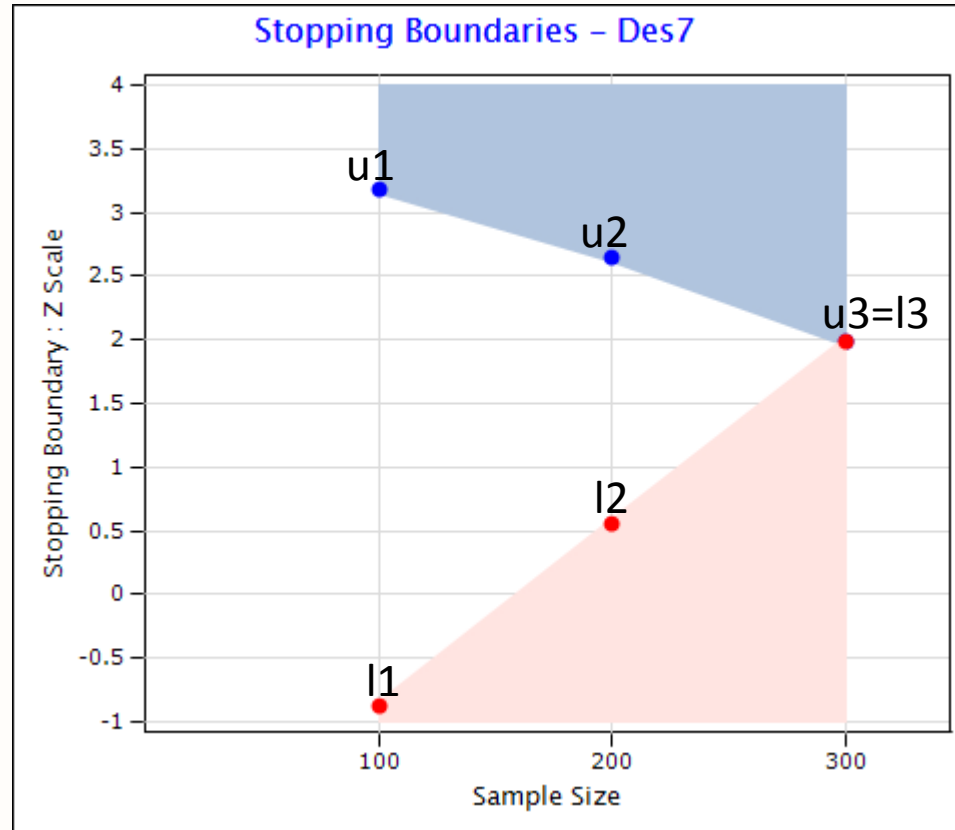
$\hat{\delta}_j$ = the mle of δ at look j , for $j = 1, 2, \dots, K$

I_j = the Fisher information for δ at look j

W_j = $\hat{\delta}_j \sqrt{I_j}$ is the score statistic at look j

$W_j \sim N(\hat{\delta}_j I_j, I_j)$ with independent increments

- Independent increments means that W_j and $(W_k - W_j)$ are independent for any $k > j$
- This property enables us to create the design efficiently for any value of K

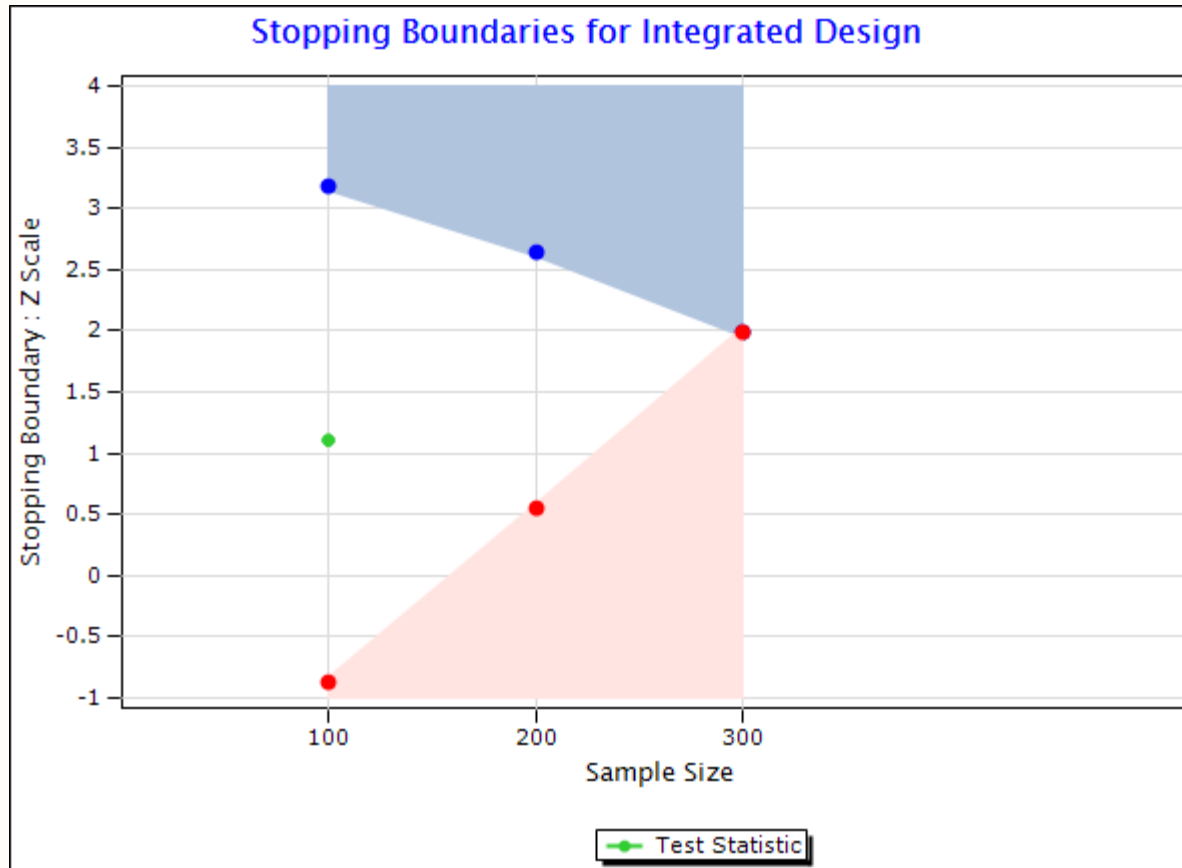


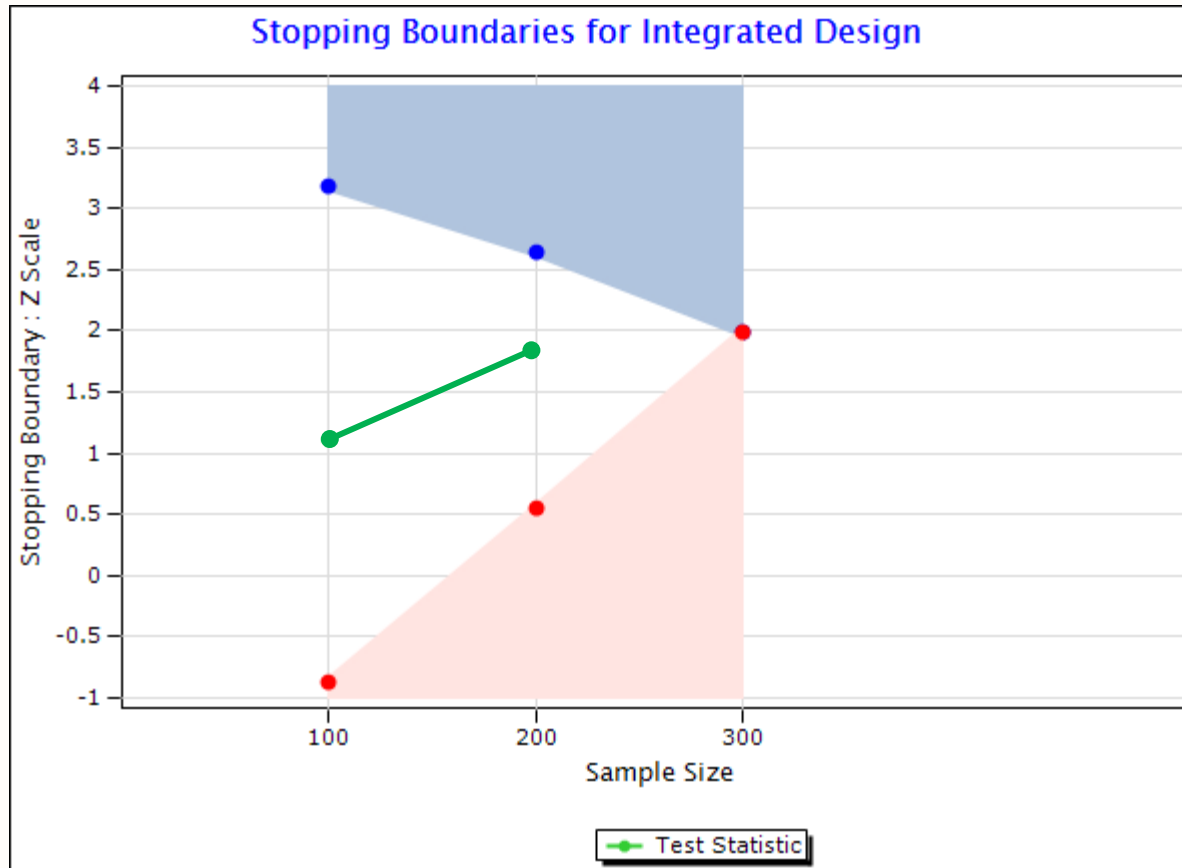
Boundaries (u_1, l_1) , (u_2, l_2) , (u_3, l_3) and I_{\max} obtained from:

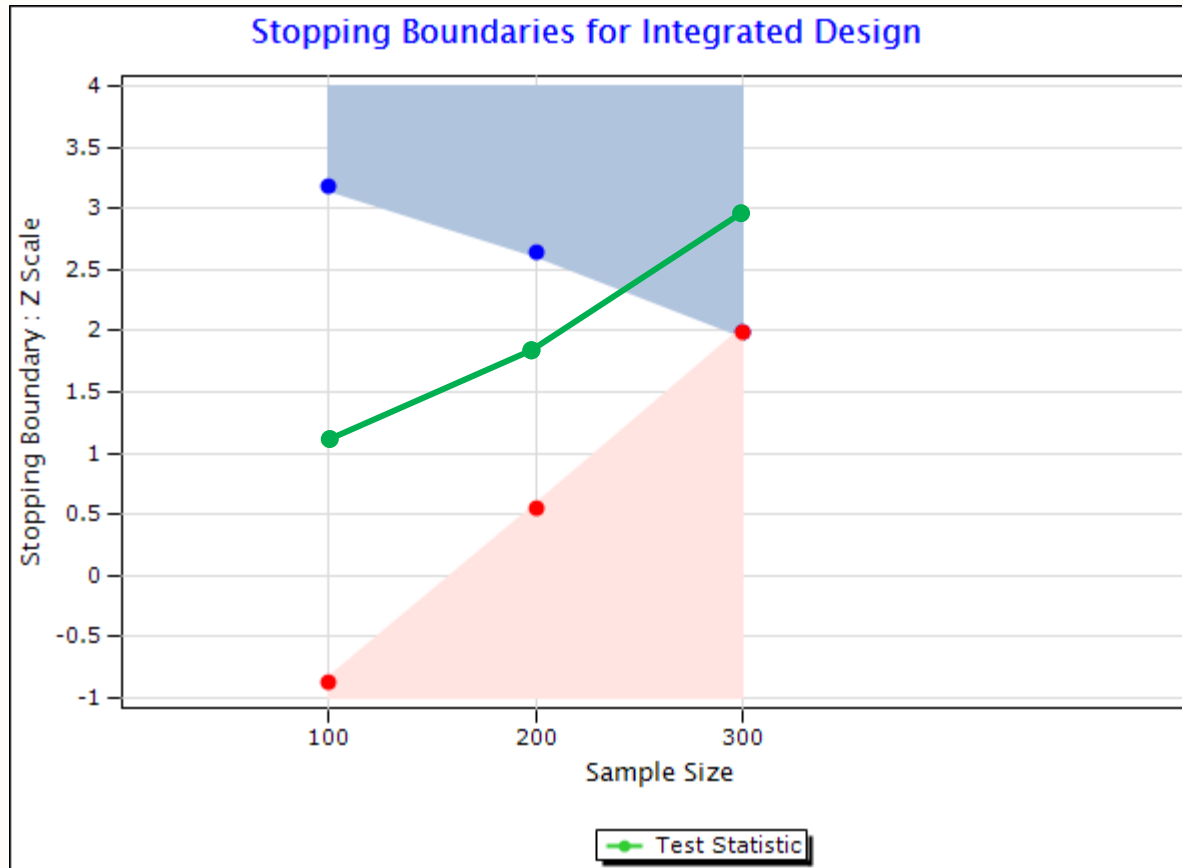
$$P_0(W_1 \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq u_3) = \alpha$$

$$P_{\delta\sqrt{I_{\max}}}(W_1 \leq l_1 \text{ or } W_2 \leq l_2 \text{ or } W_3 \leq u_3) = \beta$$

Look 1 of Classical Three-Look Design



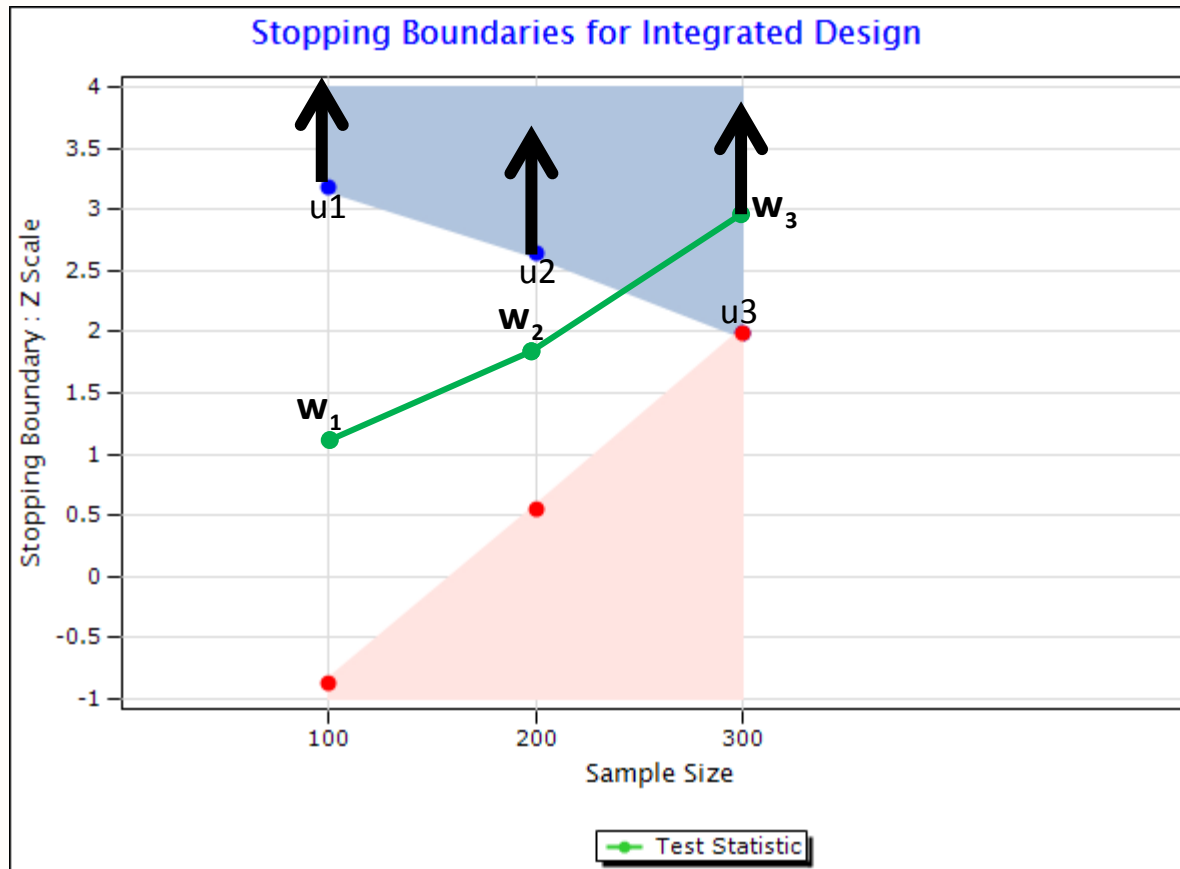




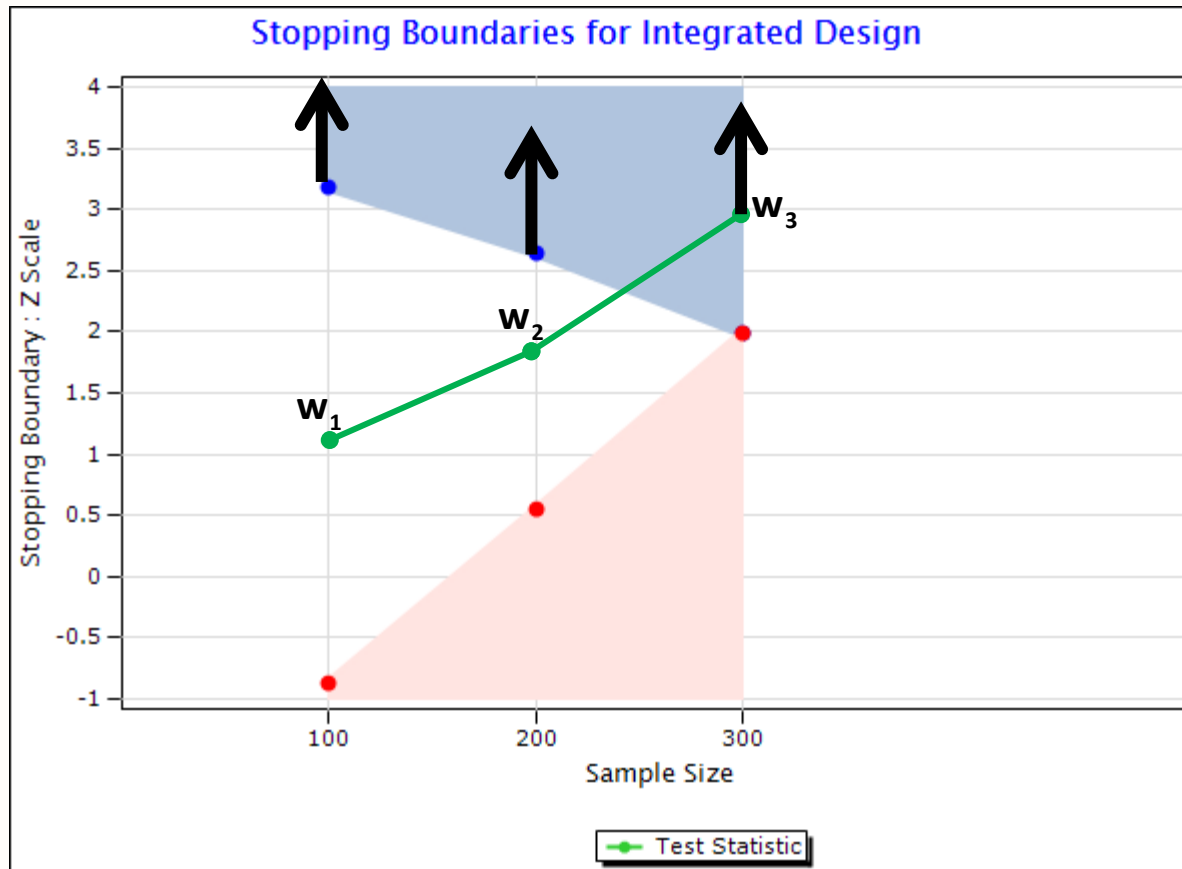
Null hypothesis that $\delta=0$ is rejected. But what about P-value and CI?

- Stage wise ordering of sample space
 - Stopping at the same look with larger value of test statistic is more extreme
 - Stopping at an earlier look is more extreme than stopping at a later look
- Compute the probability of the more-extreme region based on this ordering

Classical SWCI Method: Tsiatis, Rosner, Mehta (*Biometrics*, 1984)



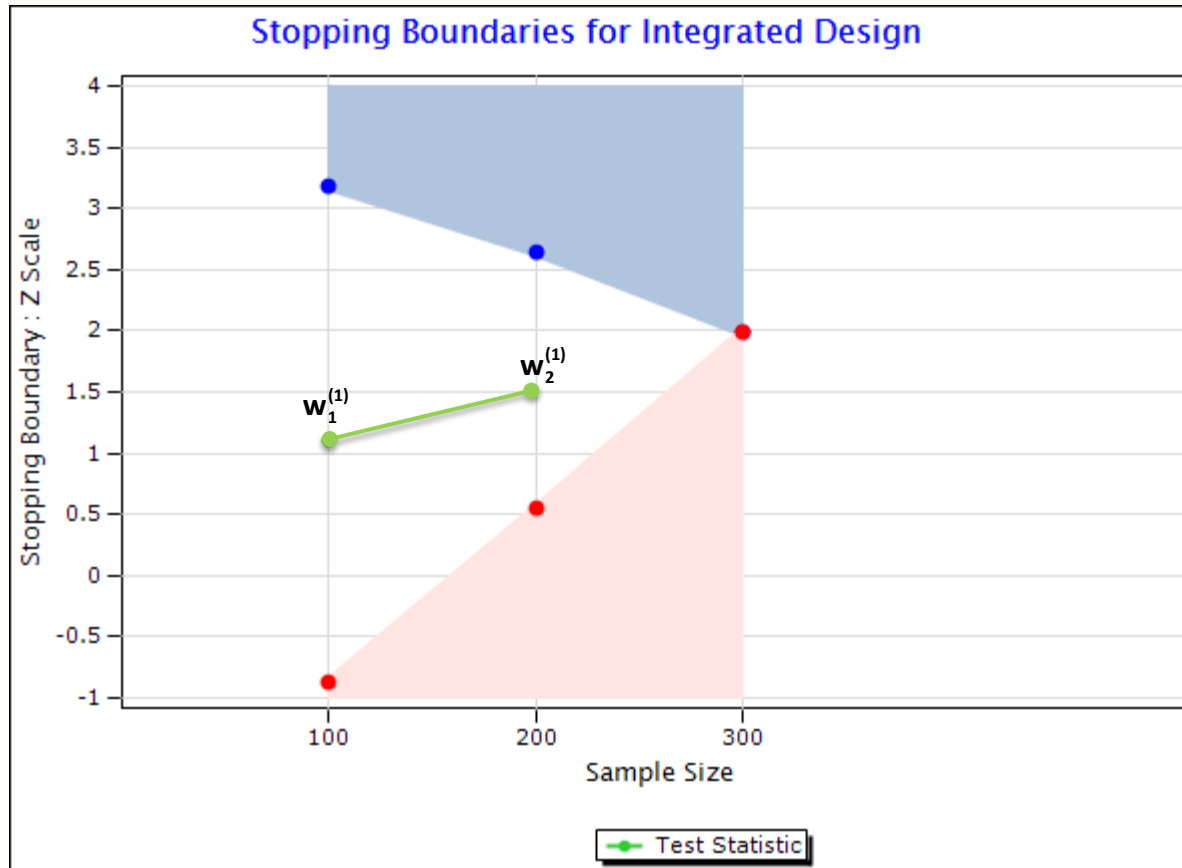
$$P_{\text{value}} = P_0 (W_1 \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq w_3)$$



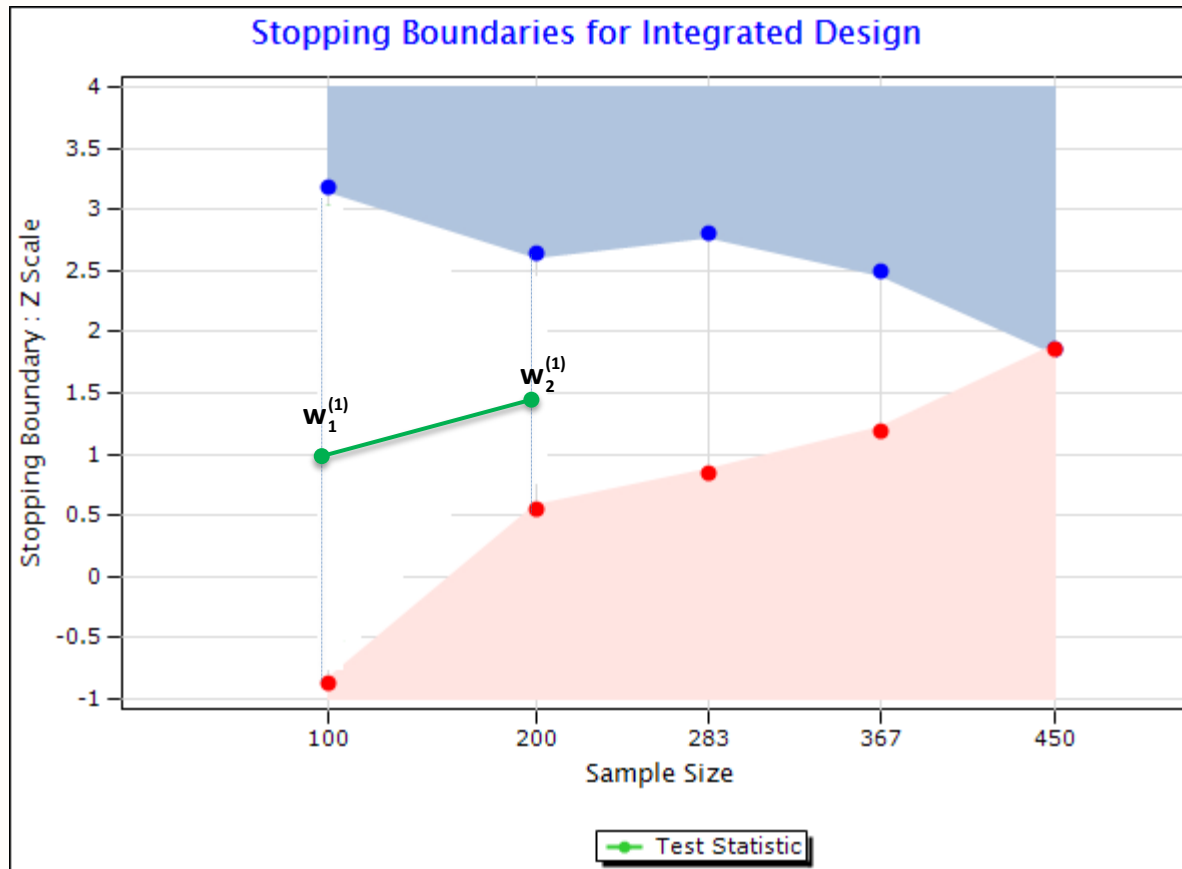
Find δ_{low} such that: $P_{\delta_{low}}(W_1 \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq w_3) = 0.975$

Find δ_{up} such that: $P_{\delta_{up}}(W_1 \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq w_3) = 0.025$

What if we adapt at look 2?



Adaptive Design at Look 2



Created two additional looks, changed the spending function and increase sample size

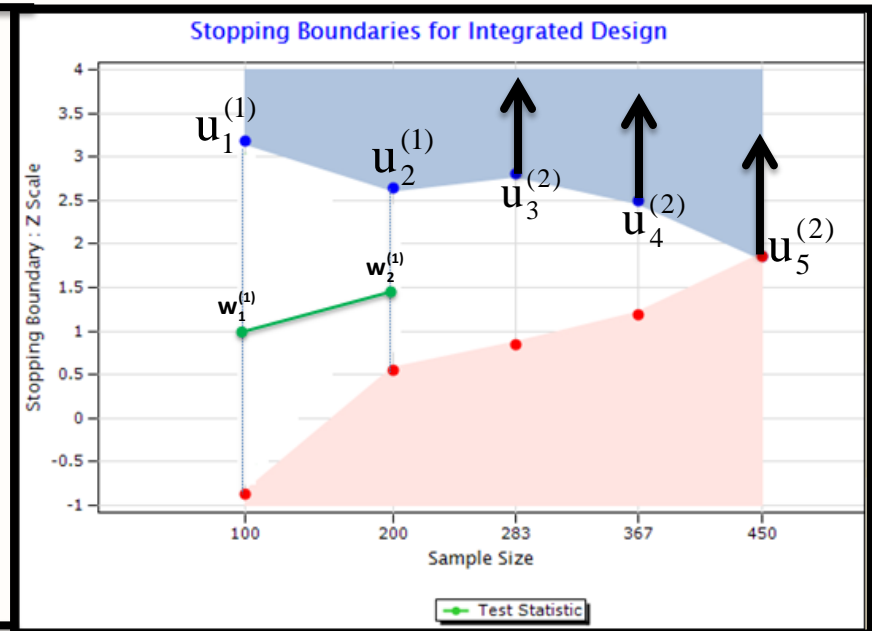
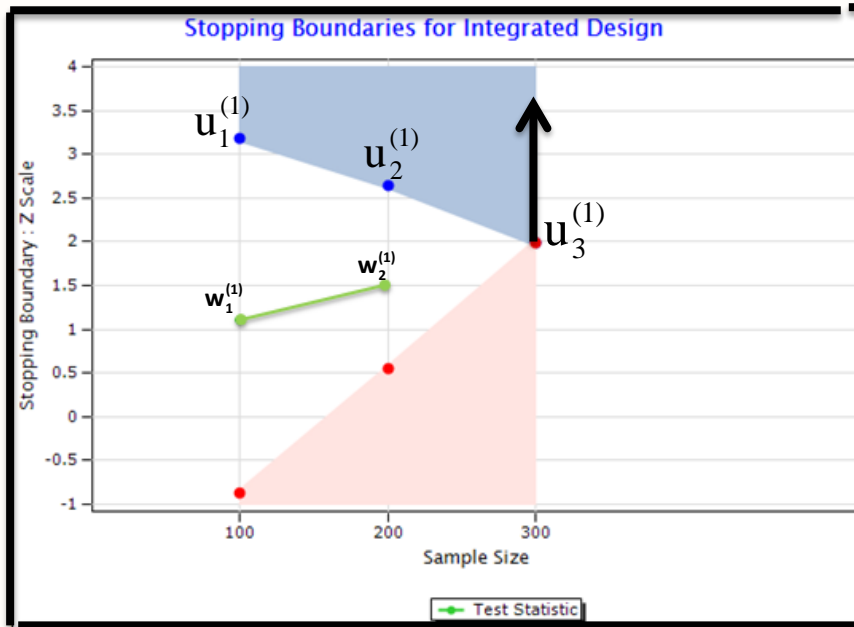
How Type-1 Error is Preserved

Design (1): Non-Adaptive

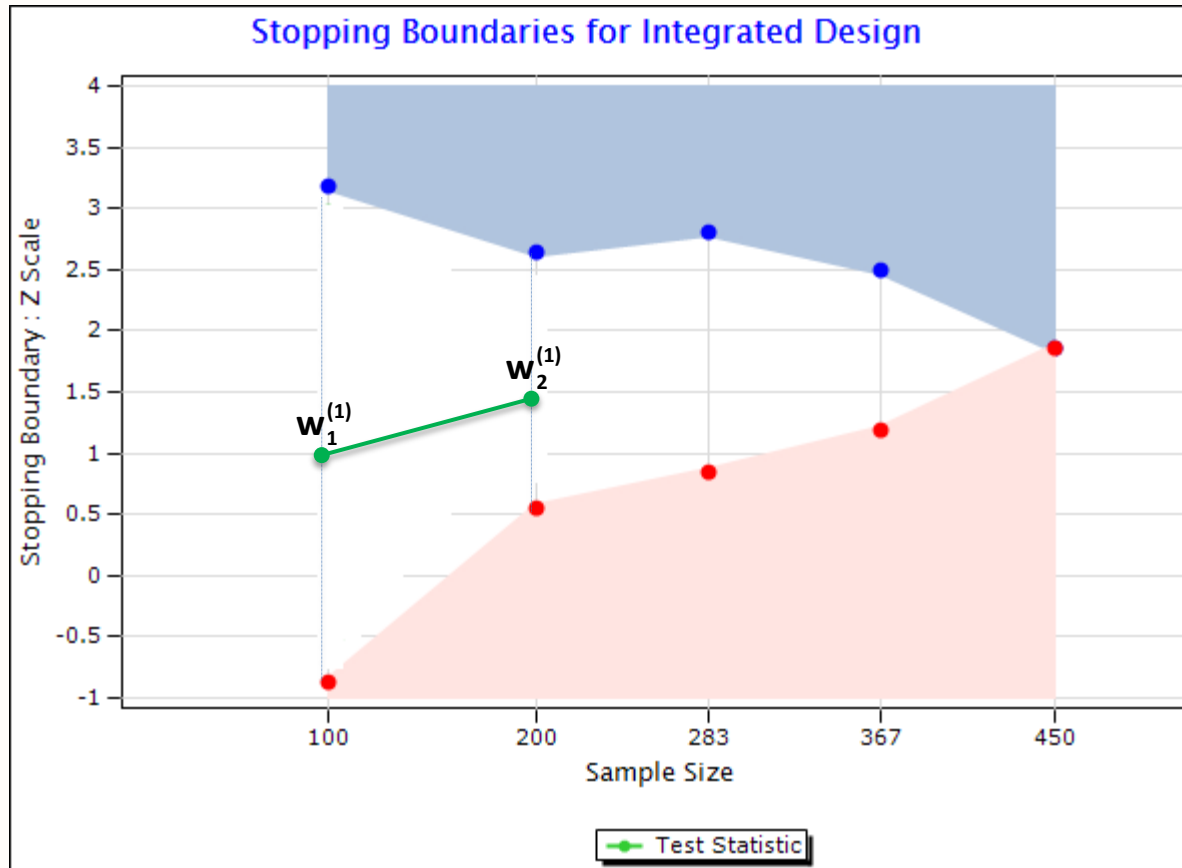
$$P_0(W_3^{(1)} \geq u_3^{(1)} | w_2^{(1)}) = \varepsilon_0^{(1)}$$

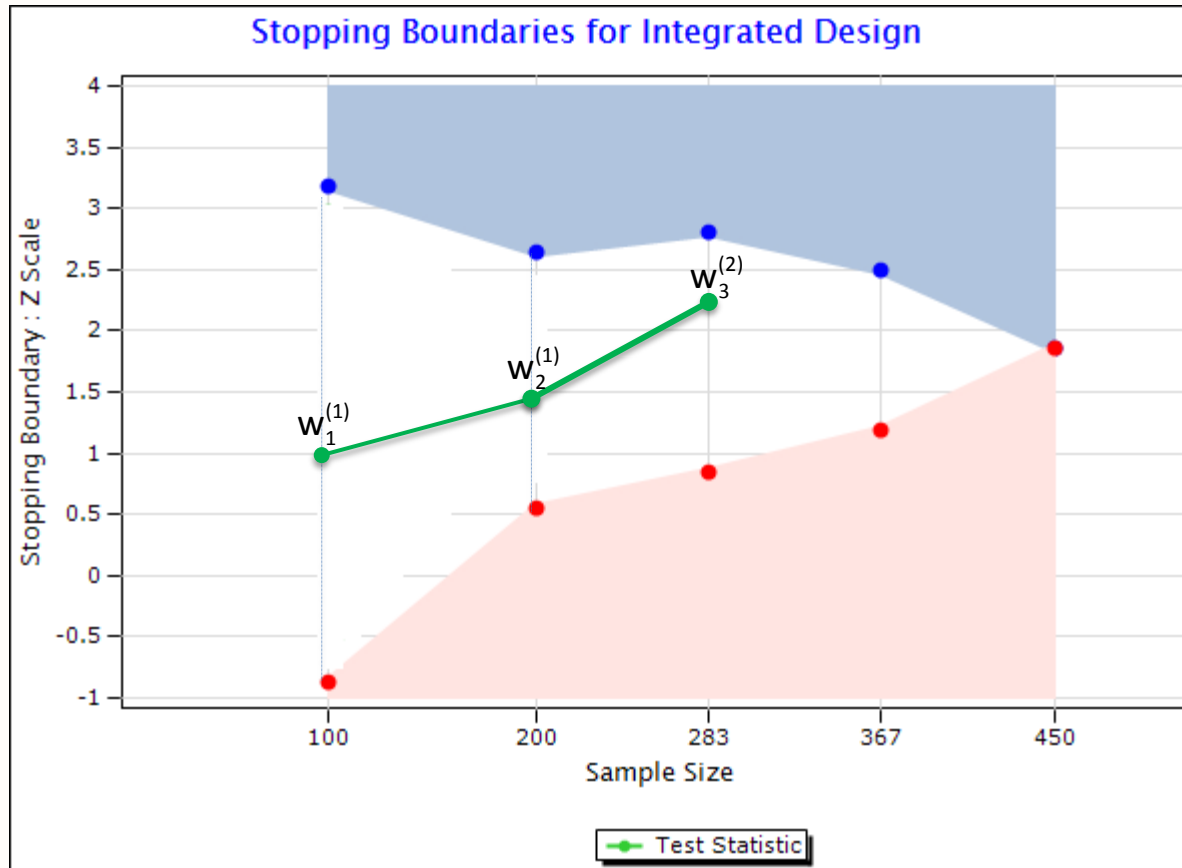
Design (2): Adaptive

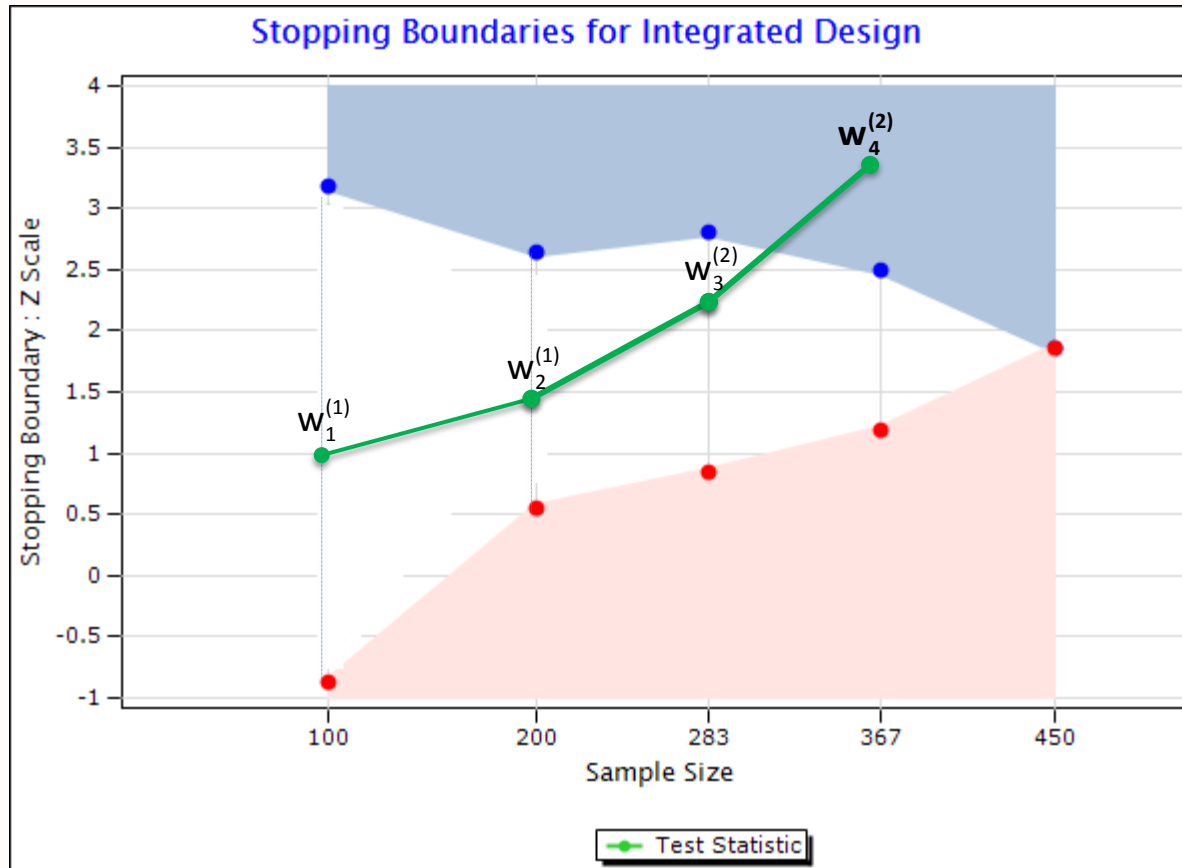
$$P_0 \cup (W_3^{(2)} \geq u_3^{(2)}, W_4^{(2)} \geq u_4^{(2)}, W_5^{(2)} \geq u_5^{(2)} | w_2^{(1)}) = \varepsilon_0^{(2)}$$



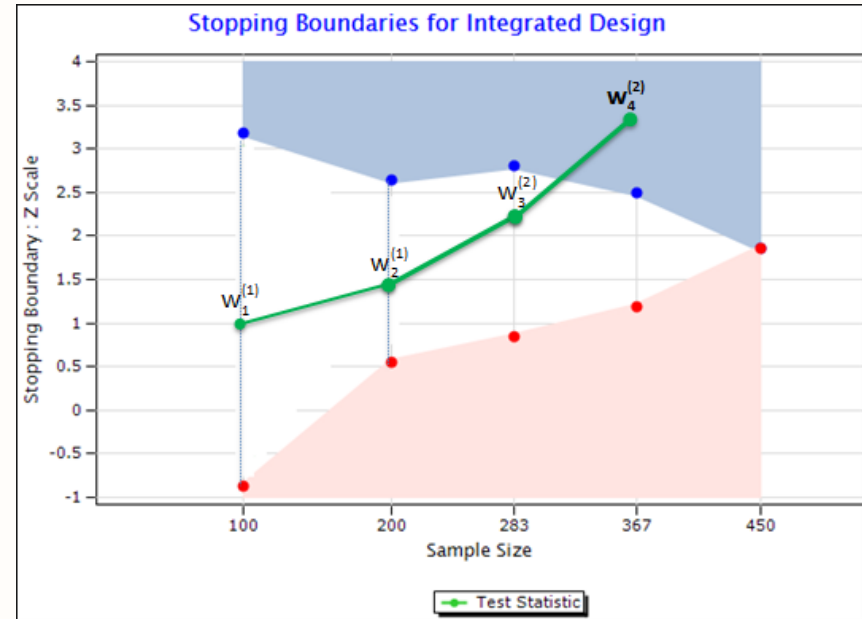
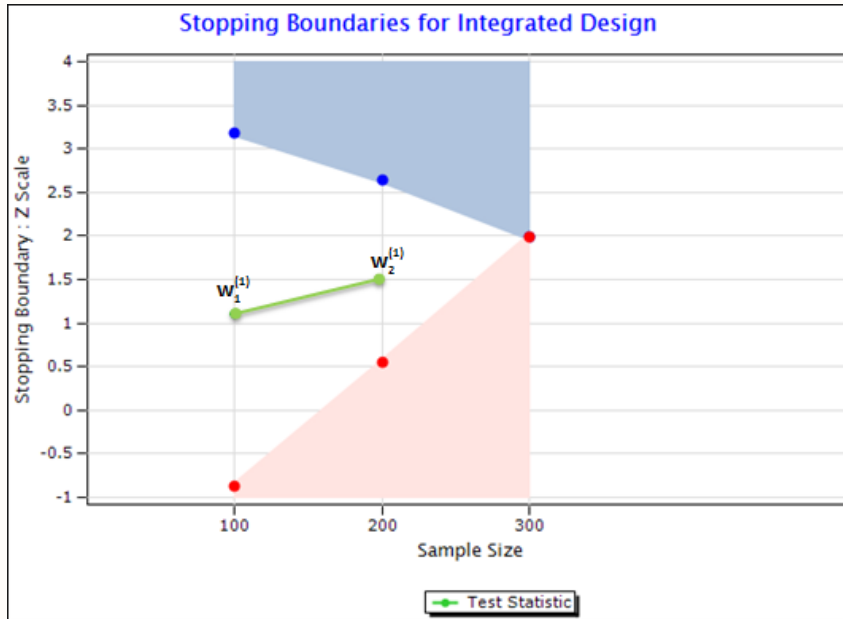
Identical Conditional Errors: $\varepsilon_0^{(1)} = \varepsilon_0^{(2)}$



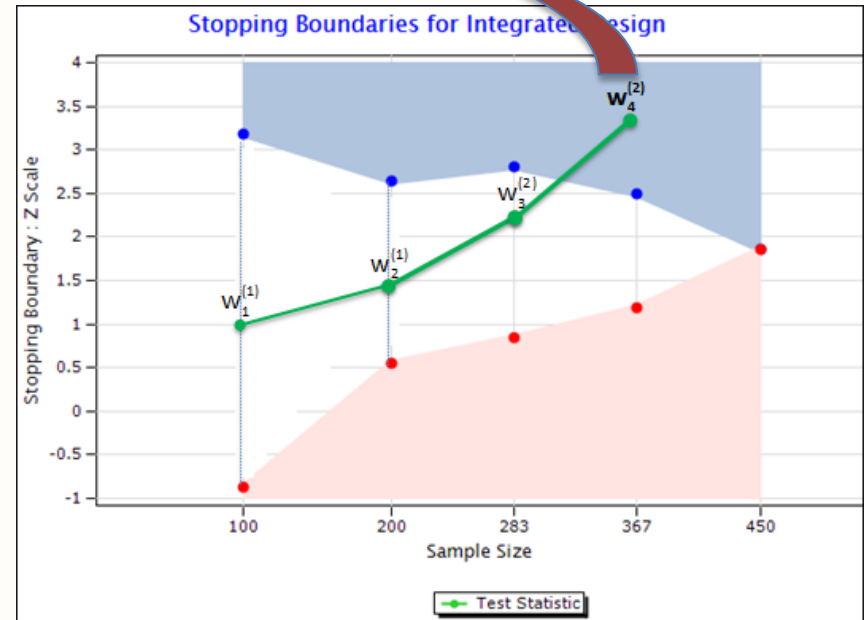
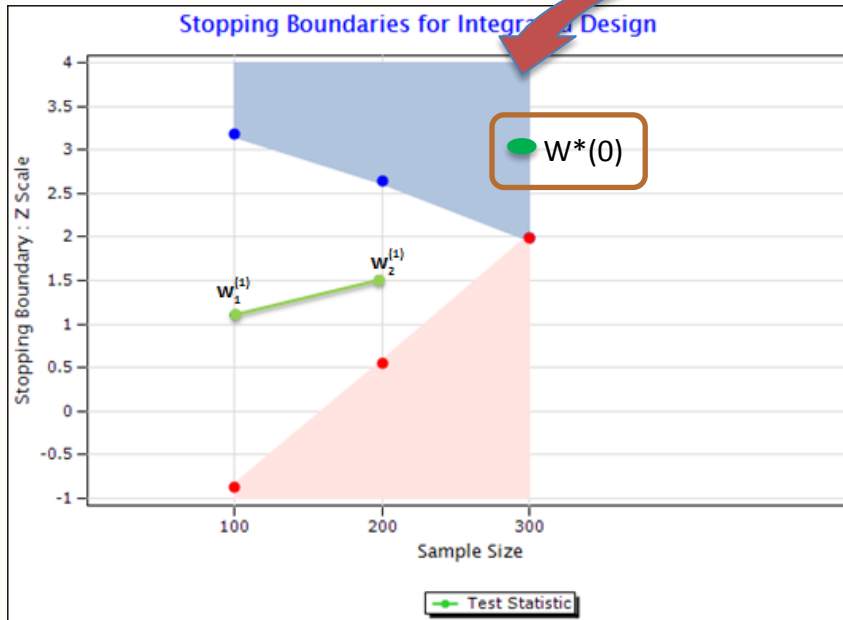




- Null hypothesis that $\delta=0$ is rejected
- What about P-value and confidence interval?



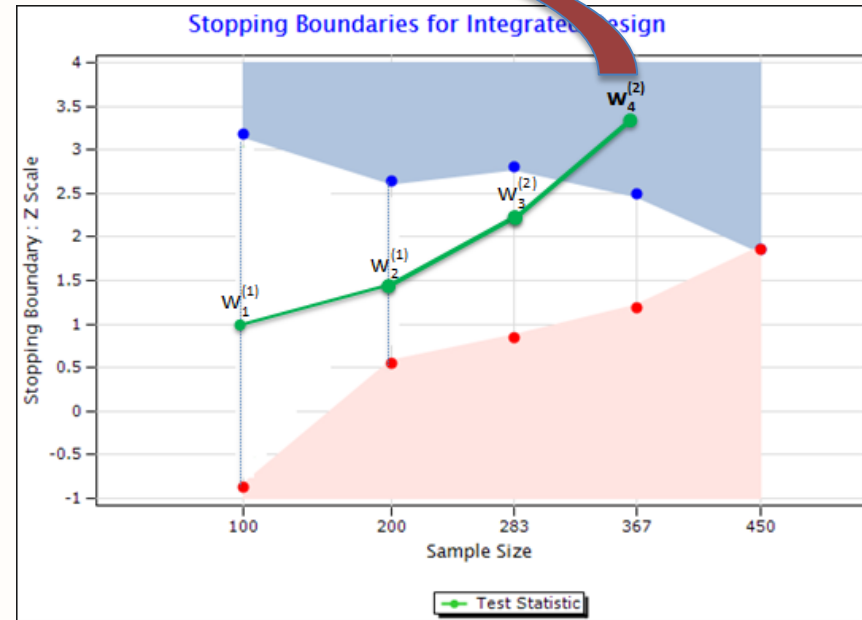
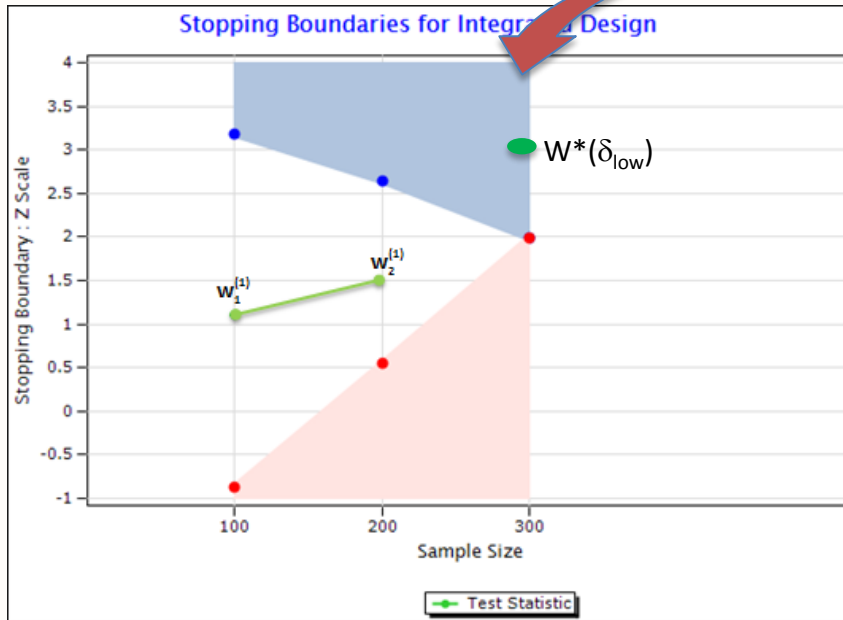
1. Find the backward image, in the classical sample space, of the statistic $w_4^{(2)}$ obtained in the adaptive sample space
2. Use the classical SWCI method to compute P-value and CI for the backward image



- Find $w^*(0)$ such that:

$$P_0(W_3^{(1)} \geq w^*(0) | w_2^{(1)}) = P_0(W_3^{(2)} \geq u_3^{(2)} \text{ or } W_4^{(2)} \geq w_4^{(2)} | w_2^{(1)})$$

- Then find the p-value corresponding to $w^*(0)$ by classical SWCI method



$$1. P_{\delta} (W_3^{(1)} \geq w^*(\delta) | w_2^{(1)}) = P_{\delta} (W_3^{(2)} \geq u_3^{(2)}, W_4^{(3)} \geq w_4^{(3)} | w_2^{(1)})$$

$$2. P_{\delta_{low}} (W_1^{(1)} \geq u_1^{(1)} \text{ or } W_2^{(1)} \geq u_2^{(1)} \text{ or } W_3^{(1)} \geq w^*(\delta_{low})) = 0.975$$

Results for Backward Image (BWCI) Method

Simulation of a 3-look LD(OF) GSD with adaption at look 2 to a 3-look LD(PK) GSD (100,000 simulations)

True Value of δ	Median of 100,000 estimates of δ	Prop. of 90% CI's containing δ	Proportion of 90% CI's excluding δ	
			From below	From above
-0.15	-0.14972	0.90007	0.05022	0.04971
0.00	0.00027	0.90073	0.04920	0.05007
0.15	0.14986	0.89866	0.04955	0.05179
0.30	0.29999	0.90087	0.04940	0.04973
0.45	0.44963	0.89929	0.05083	0.04988

Comparison with other adaptive methods

- Simulation of a 3-look LD(OF) GSD with adaption at look 1 to a 2-look LD(OF) GSD (100,000 simulations)

True value of δ	Probability of lower 95% CL $> \delta$			Probability of upper 95% CL $< \delta$		
	BWCI	SWCI	RCI	BWCI	SWCI	RCI
-0.15	0.02505	0.0256	0.01905	0.02529	NA	0.02324
0.0	0.02462	0.0251	0.02448	0.02524	NA	0.02339
0.15	0.02473	0.0256	0.02585	0.02511	NA	0.02238
0.3	0.02411	0.0253	0.00654	0.02527	NA	0.01749
0.45	0.02470	0.0259	0.00075	0.02594	NA	0.01050

- SWCI Method: Brannath, Mehta, Posch (*Biometrics*, 2009)
- RCI Method: Mehta, Bauer, Posch, Brannath (*Statist. Med.*, 2007)

Research Article

Statistics
in Medicine

Received 26 February 2013,

Accepted 16 April 2013

Published online 19 May 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5847

Exact inference for adaptive group sequential designs

Ping Gao,^a Lingyun Liu^b and Cyrus Mehta^{b,c,*†}

**Next we consider extending the approach to
Multi-Arm Multi-Stage (MAMS) Designs**

- Suppose there are D dose comparisons with respect to a common control arm
- Select the best dose at end of look 1
- Future looks are then based on a two-arm trial

$$\begin{aligned}\vec{\delta} &= (\delta_1, \delta_2, \dots, \delta_D) \text{ where effect of } i\text{th dose is} \\ \delta_i &= \mu_1 - \mu_0 \text{ for } i = 1, 2, \dots, D \\ \vec{W}_1 &= (\hat{\delta}_{11}\sqrt{I_{11}}, \hat{\delta}_{21}\sqrt{I_{21}}, \dots, \hat{\delta}_{D1}\sqrt{I_{D1}}) \\ &= \text{score statistics for the } D \text{ dose groups at look 1} \\ W_{\max,1} &= \max(\hat{\delta}_{11}\sqrt{I_{11}}, \hat{\delta}_{21}\sqrt{I_{21}}, \dots, \hat{\delta}_{D1}\sqrt{I_{D1}})\end{aligned}$$

Boundaries $(u_1, l_1), (u_2, l_2), (u_3, l_3)$ and I_{\max} obtained from:

$$\begin{aligned}P_0(W_{\max,1} \geq u_1 \text{ or } W_2 \geq u_2 \text{ or } W_3 \geq u_3) &= \alpha \\ P_{\delta\sqrt{I_{\max}}}(W_{\max,1} \leq l_1 \text{ or } W_2 \leq l_2 \text{ or } W_3 \leq u_3) &= \beta\end{aligned}$$

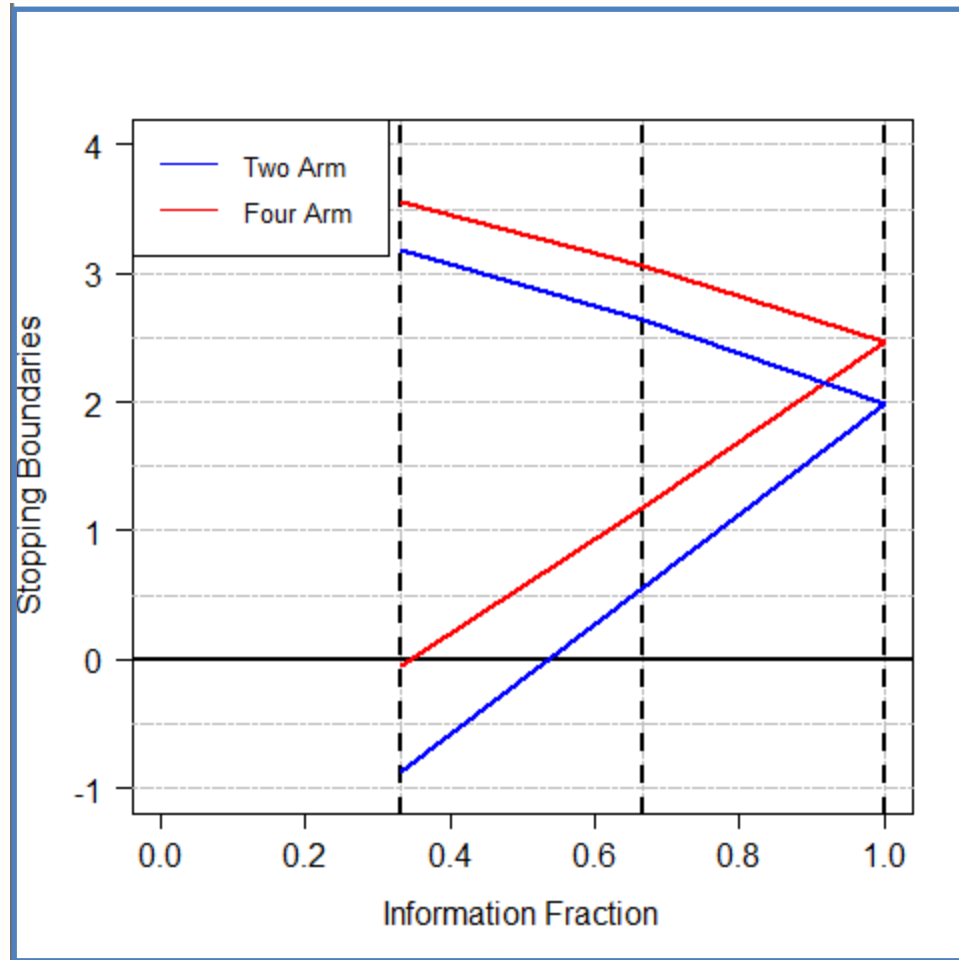
Two arms, 3 looks, $\gamma(-5)$ error spending functions, $\alpha=0.025$

Cum. α Spent	Efficacy Boundary	Cum. β Spent	Futility Boundary
0.0007	3.18319	0.0028	-0.86843
0.0046	2.64034	0.0177	0.55167
0.0250	1.98335	0.0967	1.98335

Five arms, 3 looks, $\gamma(-5)$ error spending functions, $\alpha=0.025$

Cum. α Spent	Efficacy Boundary	Cum. β Spent	Futility Boundary
0.0007	3.55470	0.0037	-0.05339
0.0046	3.05130	0.0233	1.17190
0.0250	2.46592	0.1270	2.46592

Superimpose the two boundaries



Simulation Results (100,000 simulations)

Three-look design; two active doses vs control; select best dose at end of look 1; increase sample size at end of look 2 if conditional power is between 30% and 90%

True Parameter (δ_1, δ_2)	97.5% lower CI for δ_s	Probability that (50% CI is $\leq \delta_s$)
(1.5, 0)	0.97581	0.50043
(2, 0)	0.97535	0.50092
(2.5, 0)	0.97535	0.50092
(1.5, 2.5)	0.97540	0.49750
(2, 2.5)	0.97538	0.49765
(2, 2)	0.97520	0.49878

These and other results have been submitted to a technical journal

- For Two Arm Trials

- Backward image method is the only published method with exact two-sided coverage
- Applicable to SSR, alteration of number and spacing of interim looks and spending function
- SWCI method (Brannath et. al. 2009) provides comparable results but only for one-sided case

- For Multi-Arm Trials

- Special Case (select best dose at look 1): results shown here have been submitted for publication
- Results hold conservatively if best dose not selected
- General Case (select any number of doses and drop losers along the way): still being investigated. Hypothesis testing problem published (Gao et al 2015) and available in East.

- Stage wise adjusted confidence intervals for group sequential designs provide exact coverage and median-unbiased point estimates
- These methods extend in a natural way to adaptive group sequential design
- Also extend to MAMS trials in which ONE dose is selected for further testing versus control
- Fully general case in which any number of doses can be selected at each stage, still under investigation