Adaptive Multi-Arm Multi-Stage Designs

Cyrus Mehta, Ph.D
Co-Founder and President

Cytel Inc. Cambridge, MA

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

Outline of Talk

1. What are Adaptive MAMS Designs
2. How to Control their FWER
   - Stage-wise MAMS (Bauer and Köhne, 1994)
   - Cumulative MAMS (König et al, 2008; Magirr et al, 2014; Ghosh et al, 2020)
3. Comparison between the two methods
4. When should FWER be controlled?

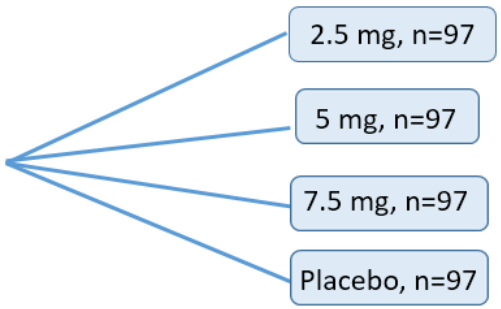
What are Adaptive MAMS?

They are generalizations of 2-arm group sequential designs

- Multiple treatment arms compared to a common control
- Multiple looks at accumulating data
- Early stopping for efficacy or futility
- Treatments may be dropped at each interim look
- Sample size re-estimation permitted at each interim look
- Strong control of Family Wise Error Rate (FWER)

SOCRATES Reduced Trial (Gheorghiade, JAMA 2015)

Chronic HF patients with reduced ejection fraction



	Wbk1:Des1
Mnemonic	MN-MAMS-GS
Test Parameters	
Number of Arms	4
No. of Looks	1
Specified α	0.025
Power	0.8006
Sample Size	
Maximum	388

- 3 doses of Variciguate vs placebo
- Endpoint: wk-12 drop in log NT-proBNP
- N=388 yields 80% power for $\delta = 0.187$ and $\sigma = 0.52$
- But what if (δ, σ) are different?
- Use 4-arm 4-look design GSD with SSR and "Drop the Loser"

How many ways can type-1 error occur?

Null Hypotheses	Type of Incorrect Conclusion
$H^{(1,2,3)}$: $\delta_1 = \delta_2 = \delta_3 = 0$	The selected treatment is declared superior to placebo
$H^{(1,2)}$: $\delta_1 = \delta_2 = 0, \delta_3 > 0$	Treatment 1 or 2 is selected and is declared superior to placebo
$H^{(1,3)}$: $\delta_1 = \delta_3 = 0, \delta_2 > 0$	Treatment 1 or 3 is selected and is declared superior to placebo
$H^{(2,3)}$: $\delta_2 = \delta_3 = 0, \delta_1 > 0$	Treatment 2 or 3 is selected and is declared superior to placebo
$H^{(1)}$: $\delta_1 = 0, \delta_2 > 0, \delta_3 > 0$	Treatment 1 is selected and is declared superior to placebo
$H^{(2)}$: $\delta_2 = 0, \delta_1 > 0, \delta_3 > 0$	Treatment 2 is selected and is declared superior to placebo
$H^{(3)}$: $\delta_3 = 0, \delta_1 > 0, \delta_2 > 0$	Treatment 3 is selected and is declared superior to placebo

FWER =probability of making one or more false claims

Control the FWER by Closed Testing

Goal: test $H^{(i)}$: $\delta_i = 0, i = 1, 2, 3$, with strong FWER control

1. Form the Closed Set of all elementary and intersection hypotheses

$$H^{(1)}, H^{(2)}, H^{(3)}$$

$$H^{(1,2)} = H^{(1)} \cap H^{(2)}, H^{(1,3)} = H^{(1)} \cap H^{(3)}, H^{(2,3)} = H^{(2)} \cap H^{(3)}$$

$$H^{(1,2,3)} = H^{(1)} \cap H^{(2)} \cap H^{(3)}$$

2. To reject any elementary hypothesis at level α, must also reject every intersection hypothesis containing it at its local level-α

Method 1: Stage-Wise MAMS

- Let $\{p_j^{(1)}, p_j^{(2)}, p_j^{(3)}\}$ be unadjusted p-values based only on the incremental data at stages (looks) $j = 1, 2, 3, 4$
- Bonferroni adjusted p-value for $H^{(123)}$ at stage j

$$p_j^{(123)} = 3 \min\{p_j^{(1)}, p_j^{(2)}, p_j^{(2)}\}$$

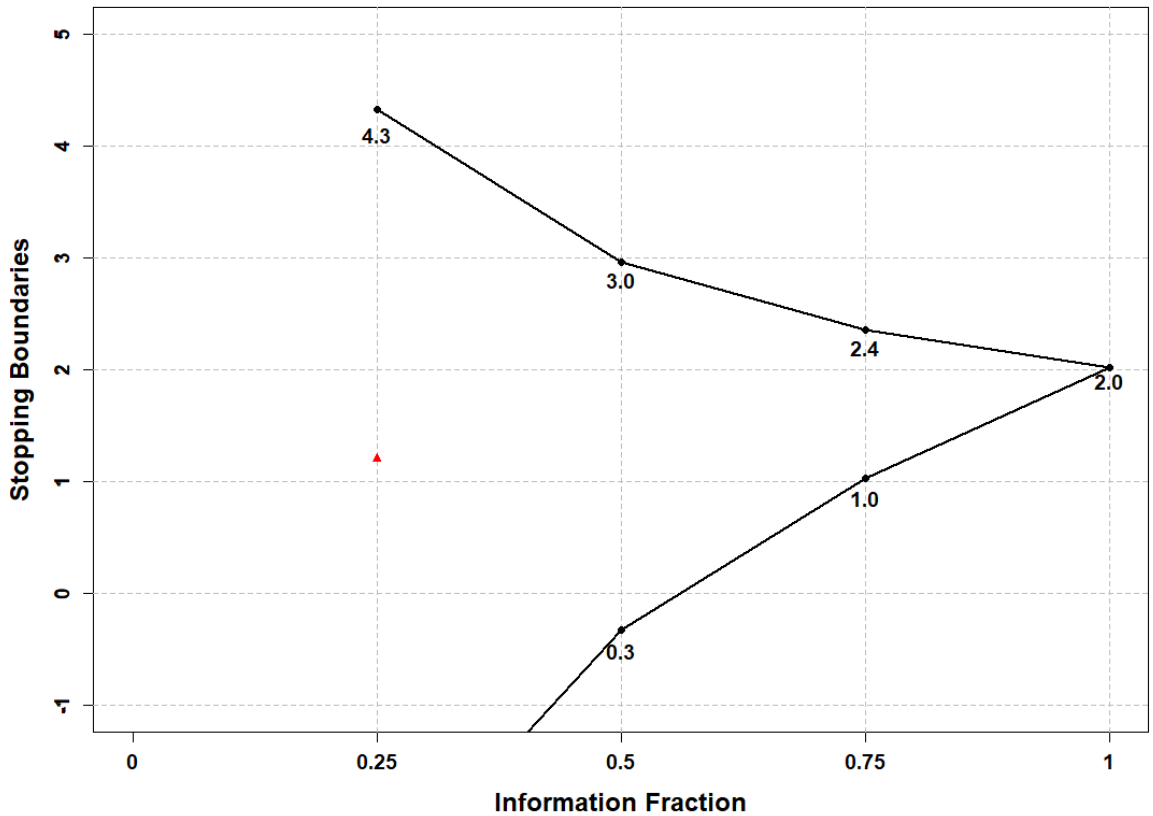
- Simes adjusted p-values for $H^{(123)}$ at stage j

$$p_j^{(1,2,3)} = \min\{3p_j^{(1)}, 1.5p_j^{(2)}, p_j^{(3)}\}$$

- Dunnett adjusted p-value at stage j

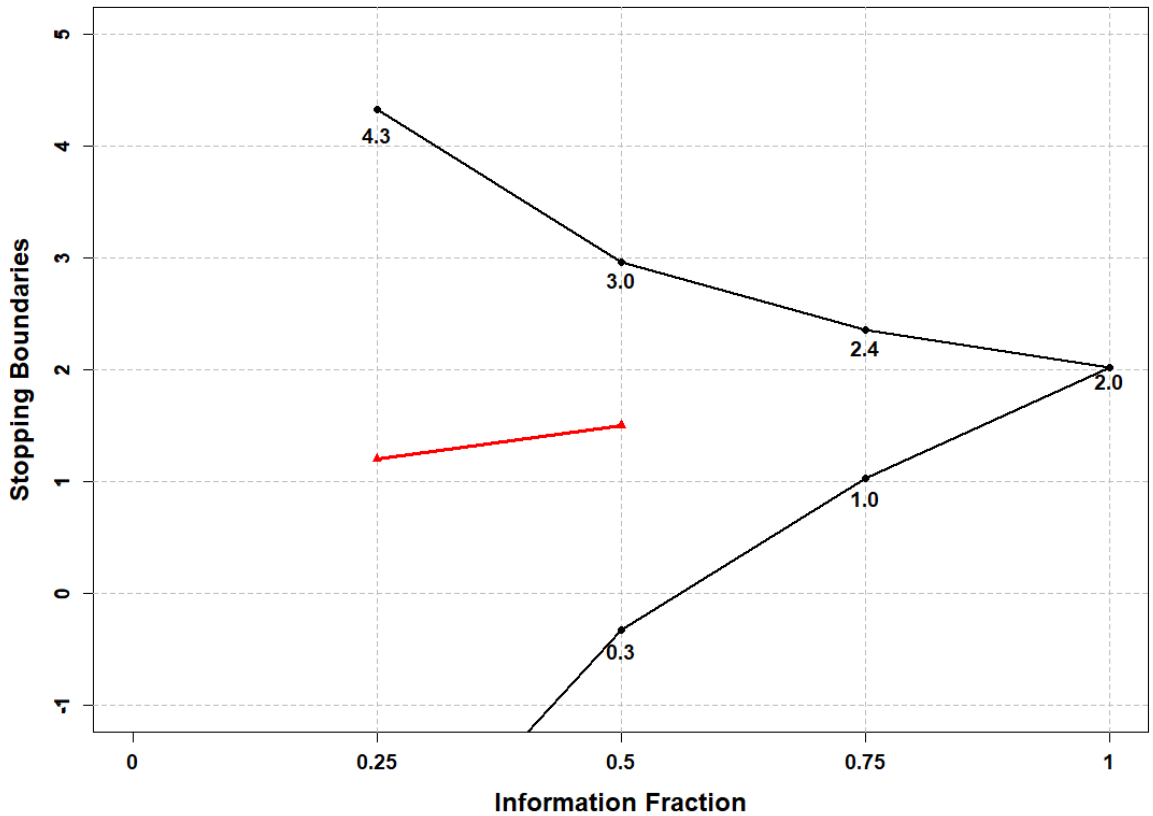
$$p_j^{(1,2,3)} = P\left(\cup_{i=1}^{3} P_i \leq p_i\right)$$

Level-α Test of $H^{(1,2,3)}$: Look 1



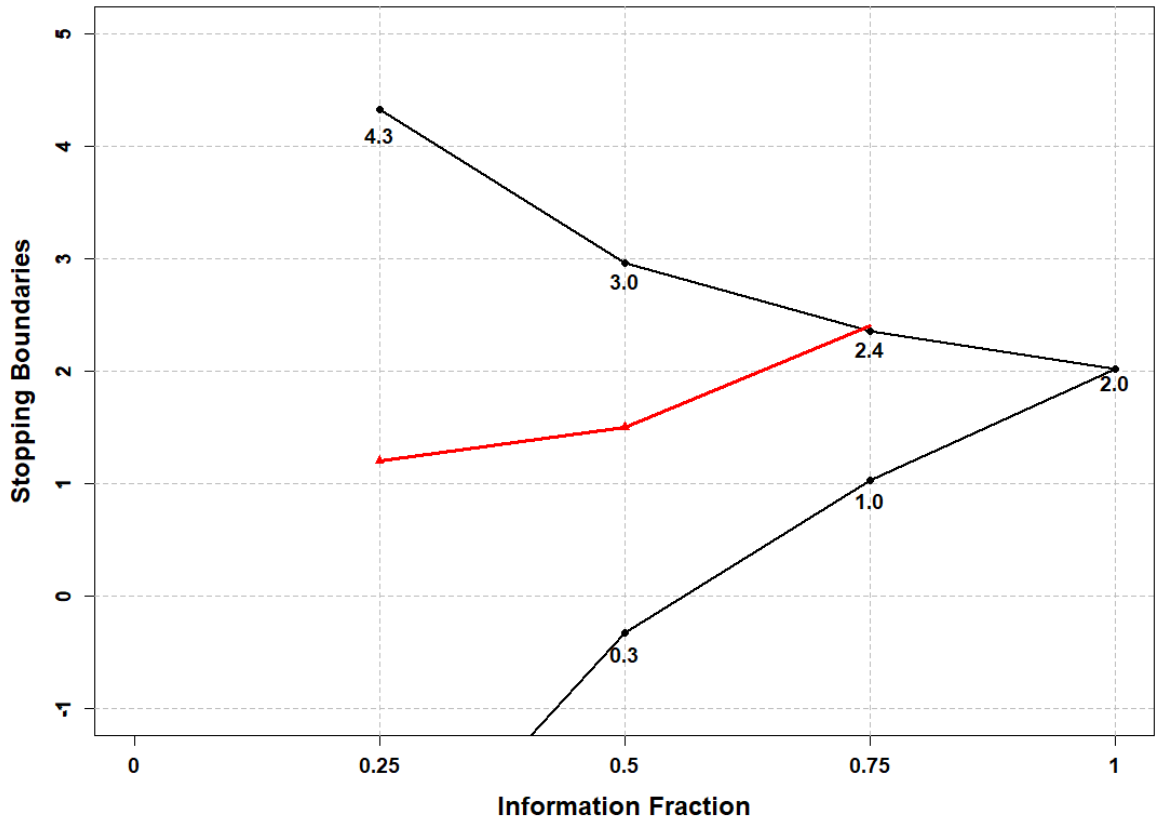
Reject if $Z_1^{(1,2,3)} = \Phi^{-1}(1 - p_1^{(1,2,3)}) \geq 4.3$

Level-α Test of $H^{(1,2,3)}$: Look 2



Reject if $\lambda_{12}\Phi^{-1}(1-p_1^{(1,2,3)}) + \lambda_{22}\Phi^{-1}(1-p_2^{(1,2,3)}) \geq 3.0$

Level-α Test of $H^{(1,2,3)}$: Look 3



Reject if $\lambda_{13}\Phi^{-1}(1-p_1^{(1,2,3)}) + \lambda_{23}\Phi^{-1}(1-p_2^{(1,2,3)}) + \lambda_{33}\Phi^{-1}(1-p_3^{(1,2,3)}) \geq 2.4$

Closed Test Requirements

Repeat this process to test
$H^{(12)}, H^{(1,3)}, H^{(2,3)}, H^{(1)}, H^{(2)}, H^{(3)}$

- Reject $H^{(1)}$ under closed testing if $H^{(123)}$, $H^{(12)}$, $H^{(13)}$ and $H^{(1)}$ are all rejected at level α
- Reject $H^{(2)}$ under closed testing if $H^{(123)}$, $H^{(12)}$, $H^{(23)}$ and $H^{(2)}$ are all rejected at level α
- Reject $H^{(3)}$ under closed testing if $H^{(123)}$, $H^{(13)}$, $H^{(23)}$ and $H^{(3)}$ are all rejected at level α

Flexibility to Adapt

- Can drop treatment arms at each stage
- Can change the number and spacing of future stages
- Can alter the sample size of future stages
- Can change the α-spending function for future stages

In Summary

Stage Wise MAMS is flexible, easy to implement and applicable under a range of distributional assumptions

Method 2: Cumulative MAMS

Exploits asymptotic normality correlations structure

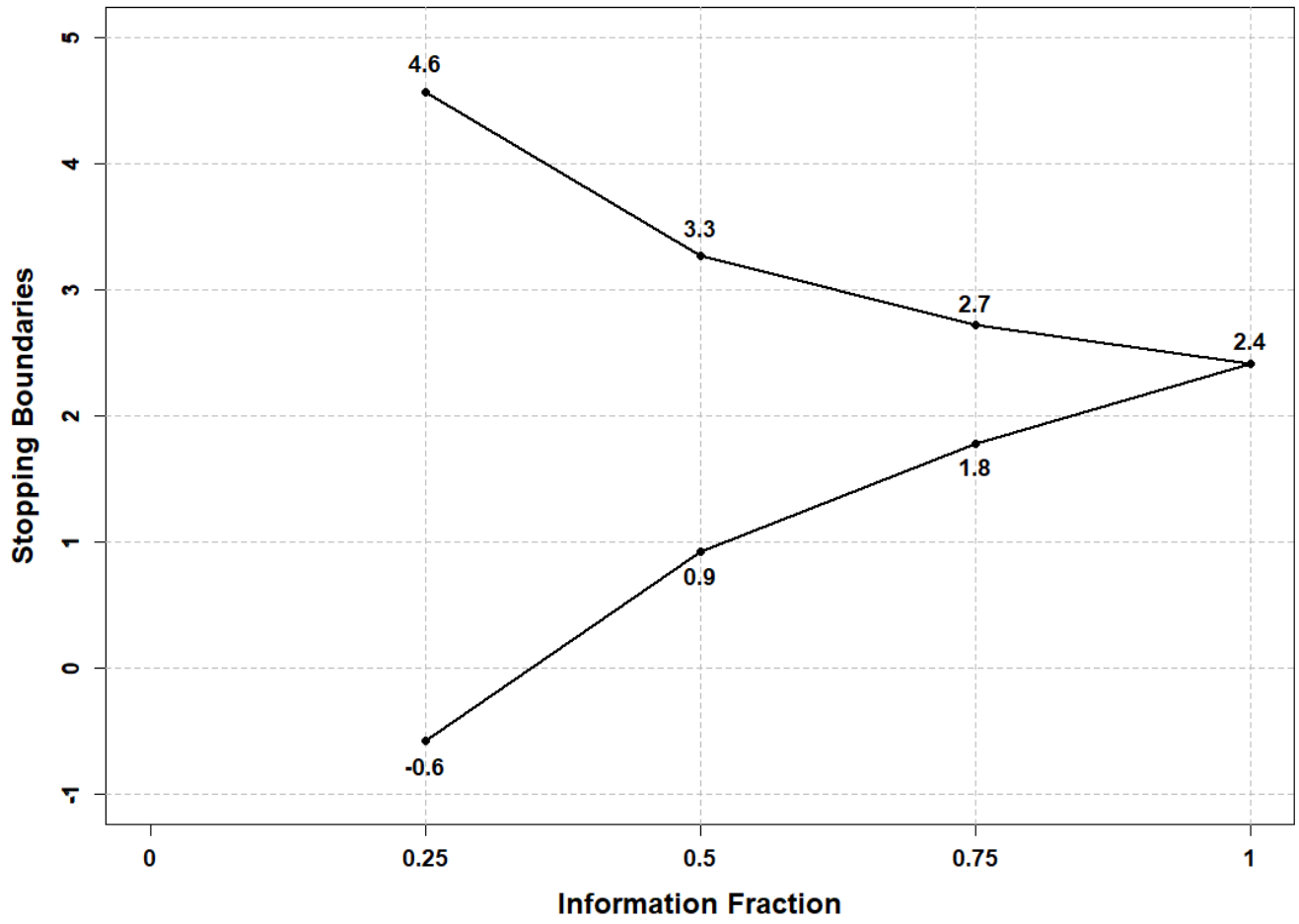
- Cumulative Wald statistic for each dose i vs placebo, at look j

$$Z_{ij} = \frac{\hat{\delta}_{ij}}{\mathsf{se}(\hat{\delta}_{ij})}$$

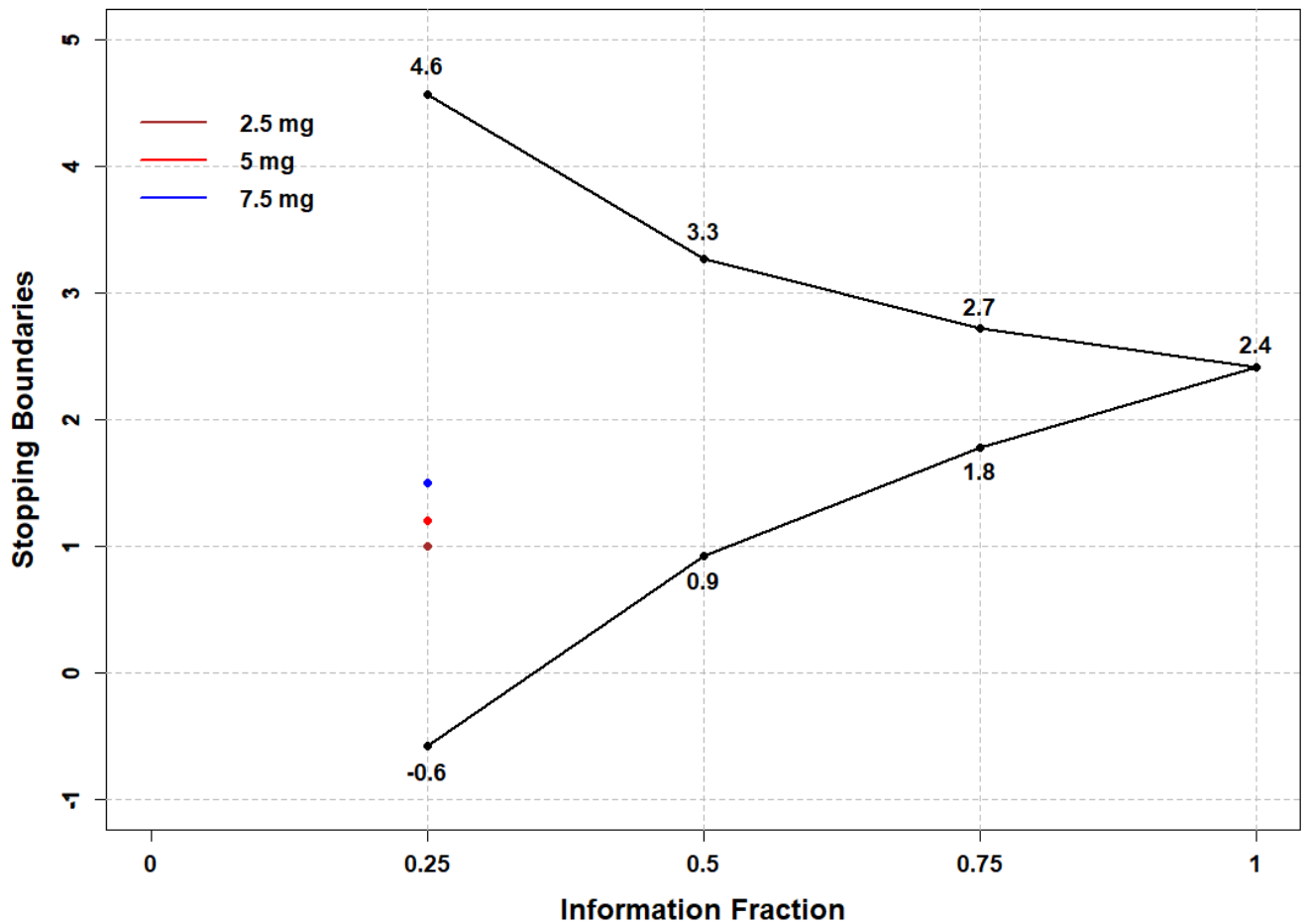
- Construct multiplicity adjusted level-α group sequential efficacy boundaries $\{u_j, j = 1, \ldots, 4\}$

$$P_0\left(\bigcup_{j=1}^{4} \max\{Z_{1j}, Z_{2j}, Z_{3j}\} \geq u_j\right) = \alpha$$

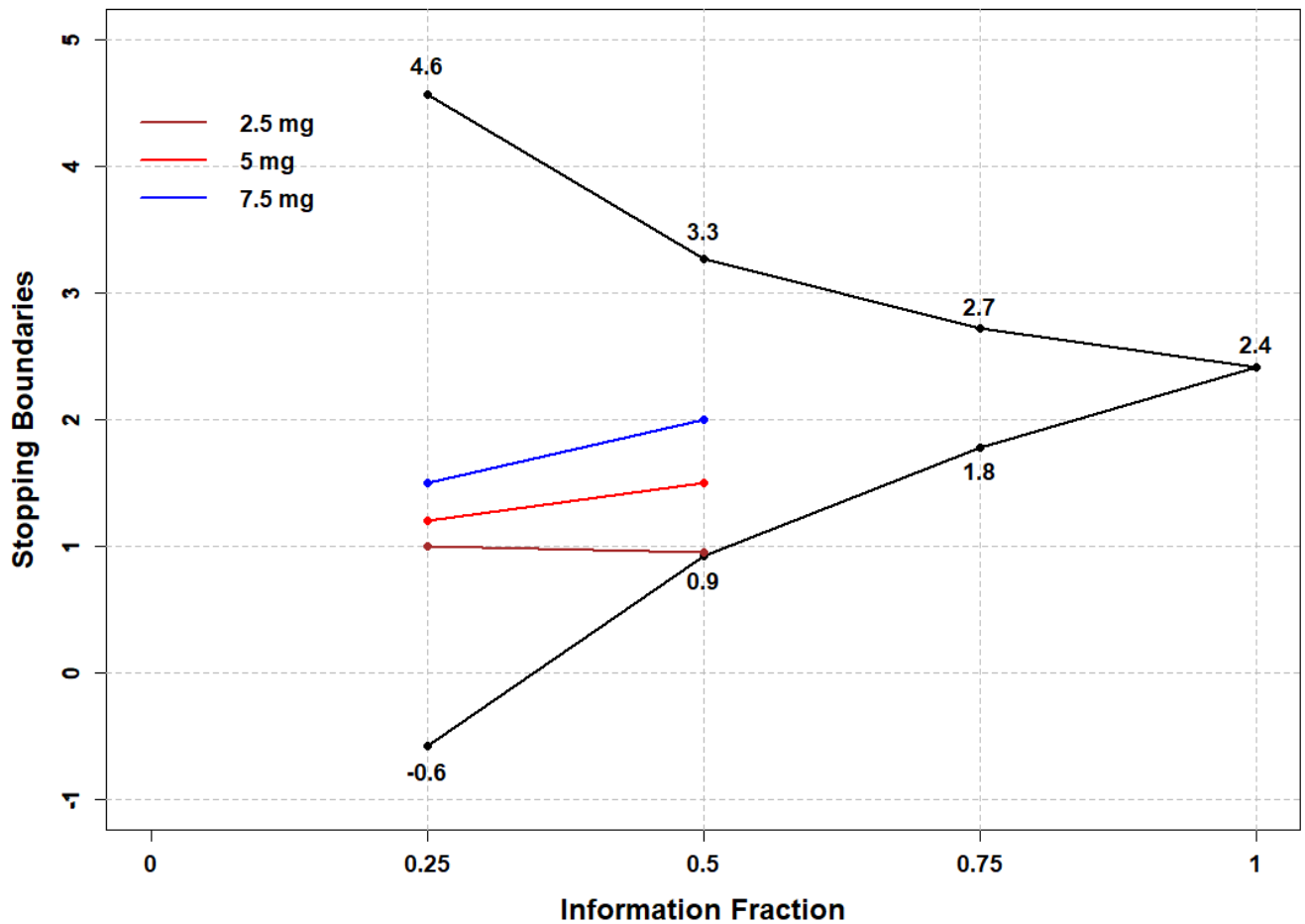
4-Arm 4-Look Boundaries



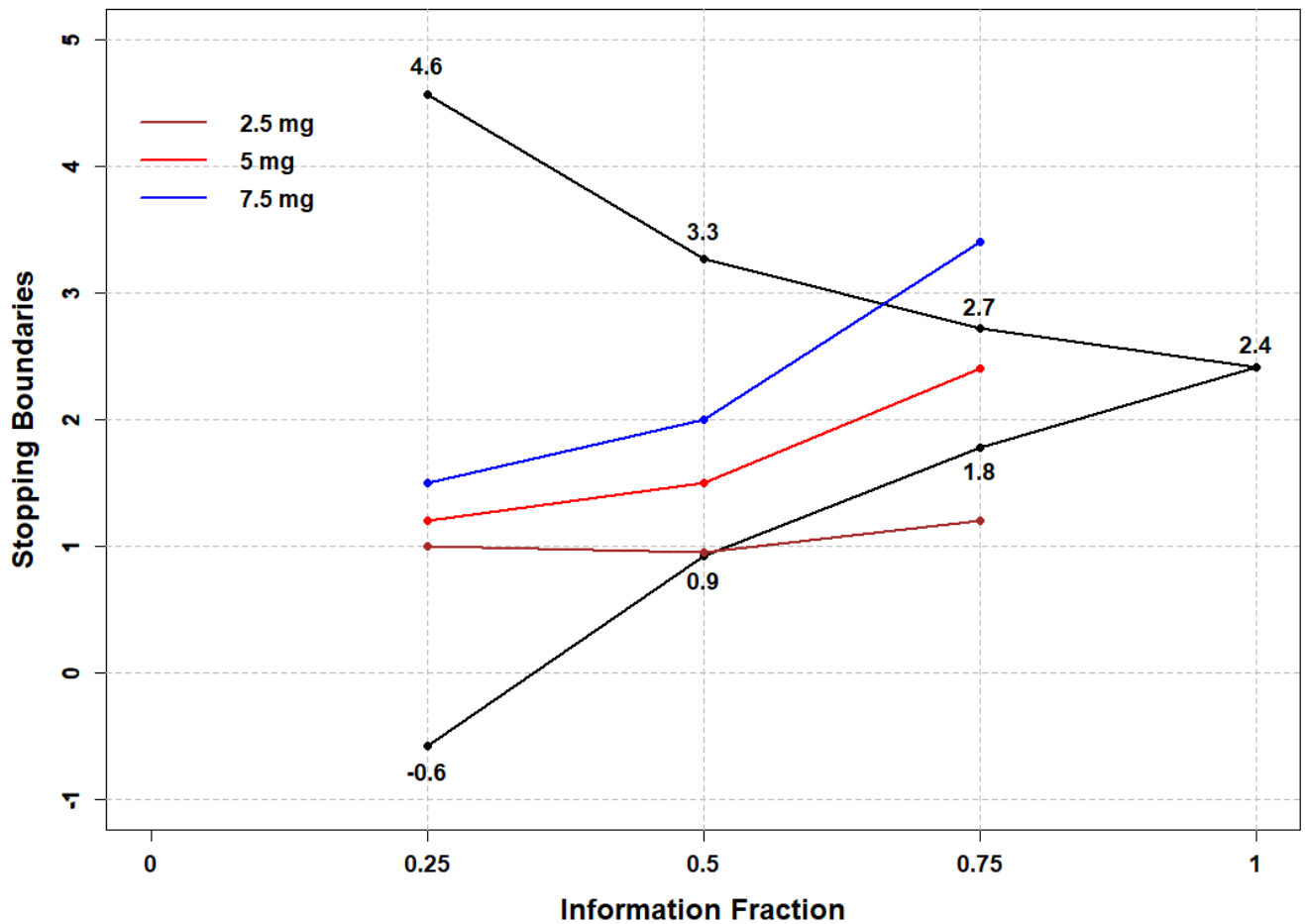
Interim Monitoring at Look 1



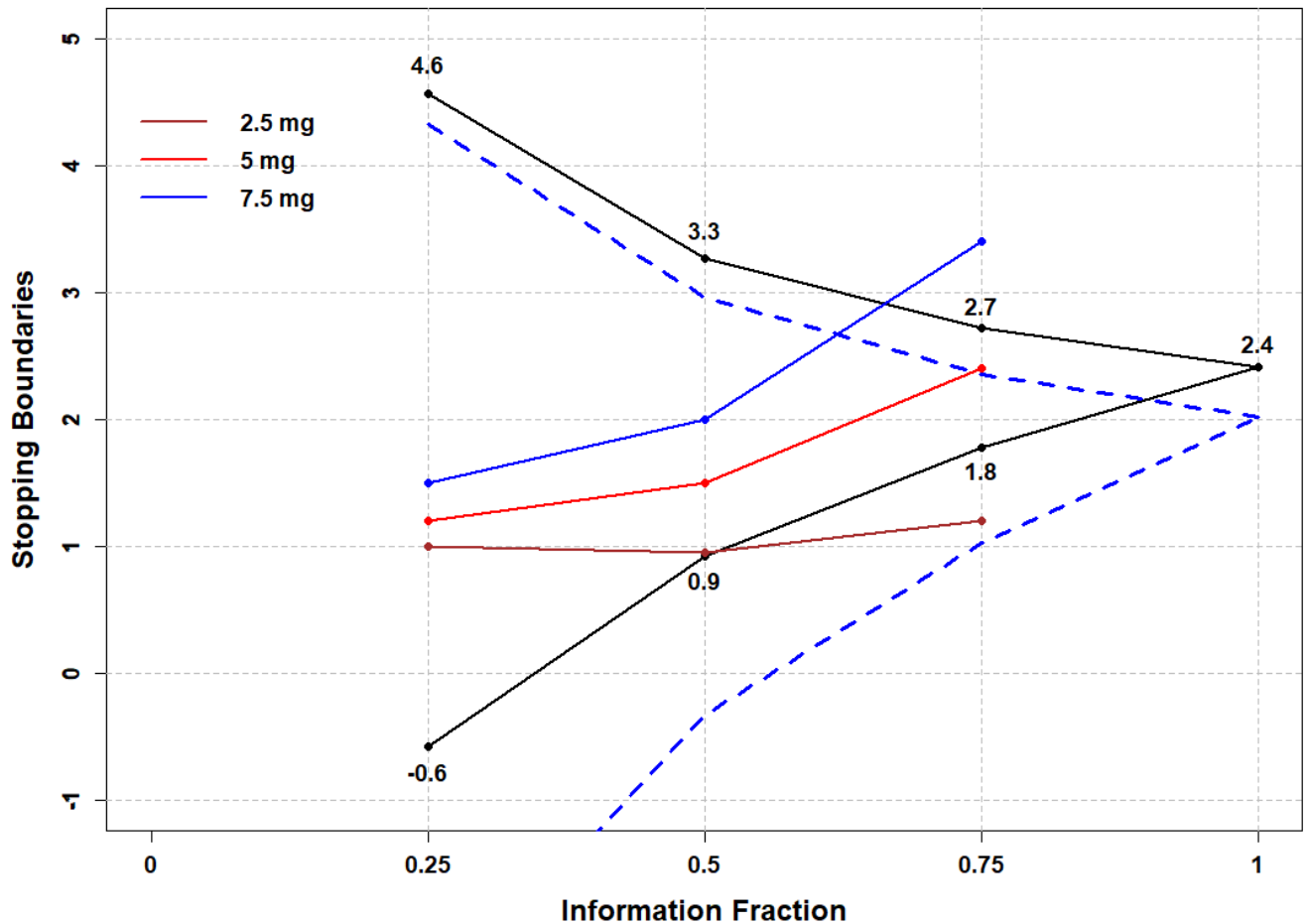
Interm Monitoring at Look 2



Interim Monitoring at Look 3



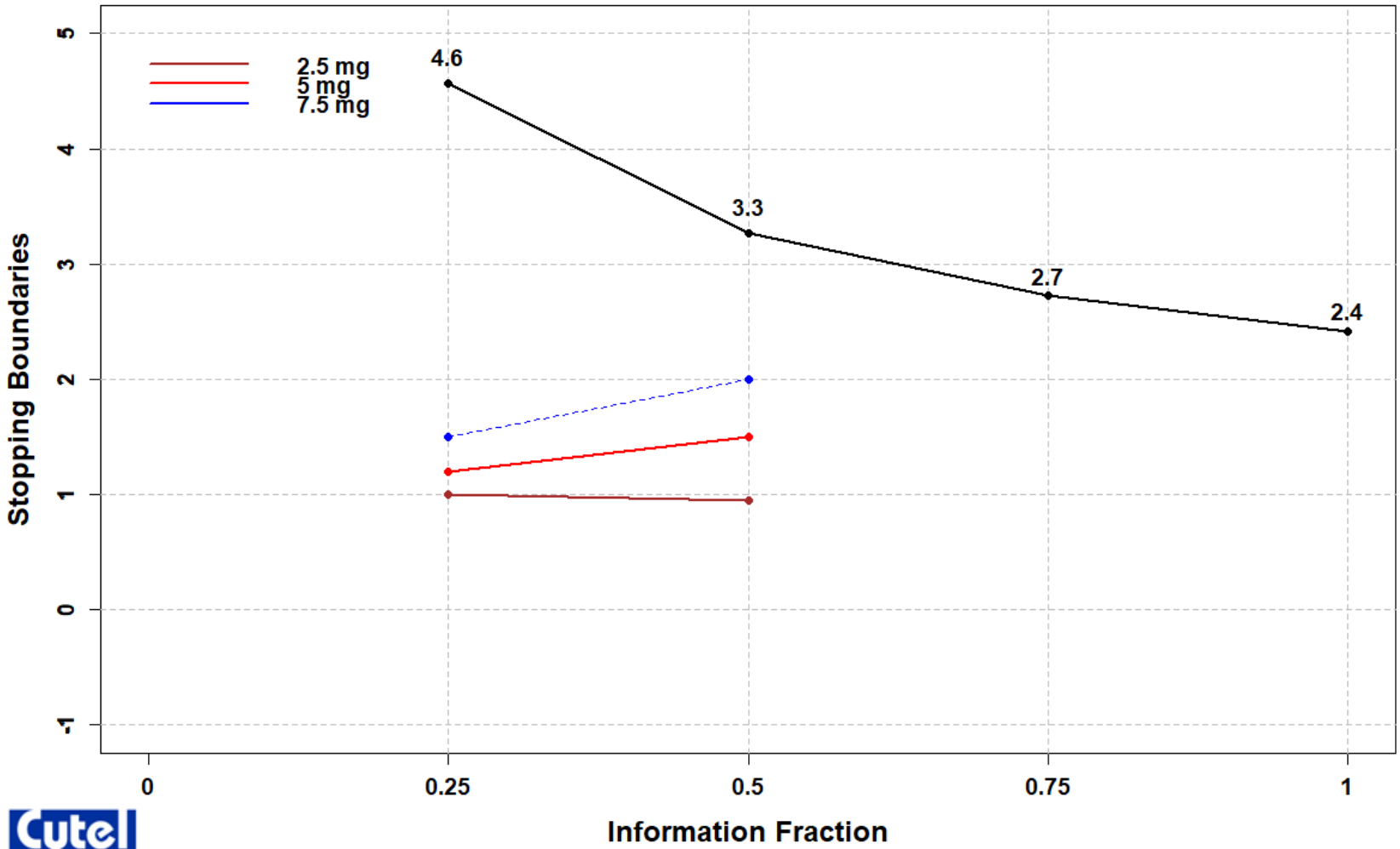
Compare with the 2-Arm Boundaries



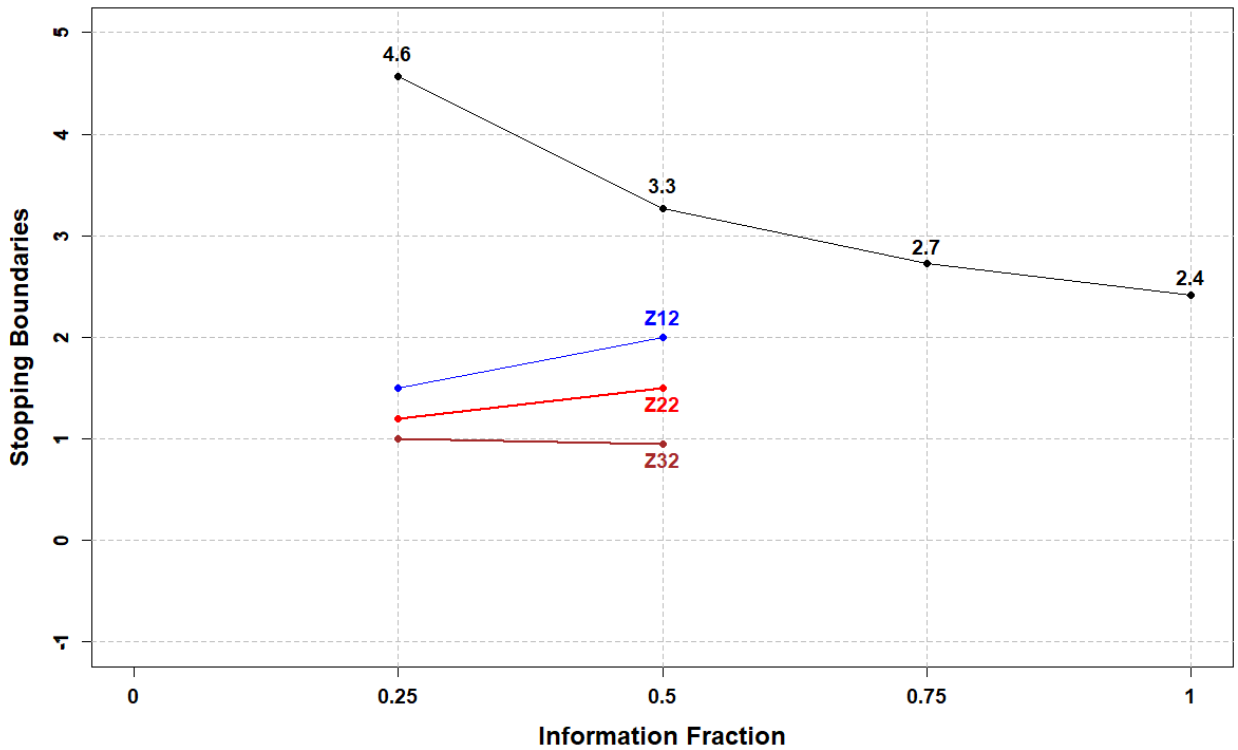
What About FWER Control?

- FWER is automatically controlled if no adaptations
- But what if we have the following adaptations?
  - Drop dose and re-allocate its remaining subjects
  - Re-estimate sample size of future stages
  - Change the number and spacing of future stages
  - Change error spending function for the future stages
- Must recompute boundaries by CER method if adapt

Drop 7.5 mg Dose and Re-allocate Subjects

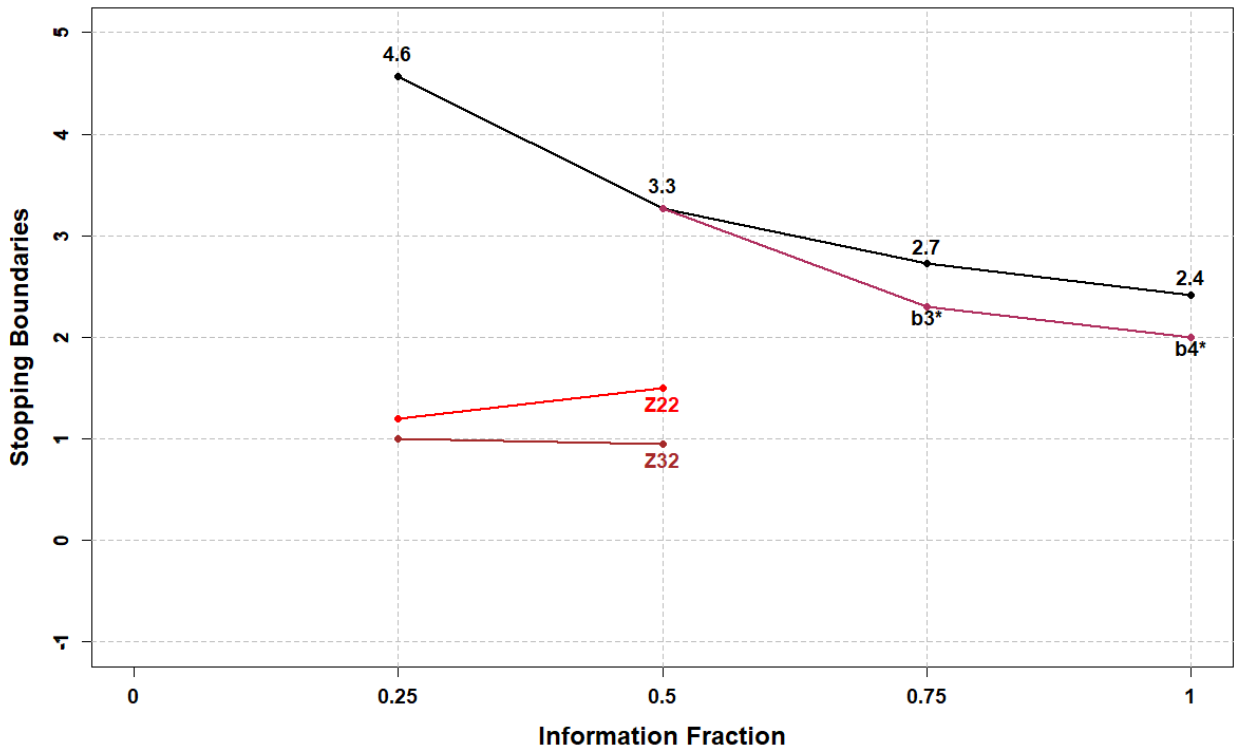


Compute CER as if did not drop dose



$$CER =$$
$$P_{\underline{0}}\left\{\max(Z_{13}, Z_{23}, Z_{33}) \geq 2.7 \text{ or } \max(Z_{14}, Z_{24}, Z_{34}) \geq 2.4 \middle| (z_{12}, z_{22}, z_{32})\right\}$$

Recompute Look 3, Look 4 boundaries that preserve CER



Recompute boundaries $(b_3^*,\ b_4^*)$ so that

$P_{\underline{0}}\left\{\max(Z_{23}^*, Z_{33}^*) \geq b_3^* \text{ or } \max(Z_{24}^*, Z_{34}^*) \geq b_4^* | (z_{22}, z_{32})\right\} = CER$

Key Differences Between the Two Methods

Stage Wise MAMS	Cumulative MAMS
Inverse normal p-value Combination	Cumulative Wald statistic
Track a single statistic	Track one statistic for each dose
Compute two-arm boundaries	Compute multi-arm boundaries
Valid for any general setting	Valid for asymptotically normal setting
If adapt, use pre-specified weights	If adapt, use CER for FWER control

Power Comparisons I: Impact of Heterogeneity

- 4-arm, 2-stage design with $n = 97$/arm
- Drop arms with $\hat{\delta} < 0$ and re-allocate to remaining arms

δ	Stage Wise MAMS			Cumulative MAMS
	Bonferroni	Simes	Dunnett	
(0.187, 0.187, 0.187)	73.6	79.3	80.1	80.3
(0, 0.187, 0.187)	67.8	71.2	74.4	75.8
(0, 0, 0.187)	52.1	54.0	61.5	64.9
(0, 0, 0)	1.5	2.0	2.4	2.5
All table entries are based on 10,000 simulated clinical trials				

Power Comparisons II: Impact of Selection Rules

Power gain of Cumulative MAMS over Stage Wise MAMS

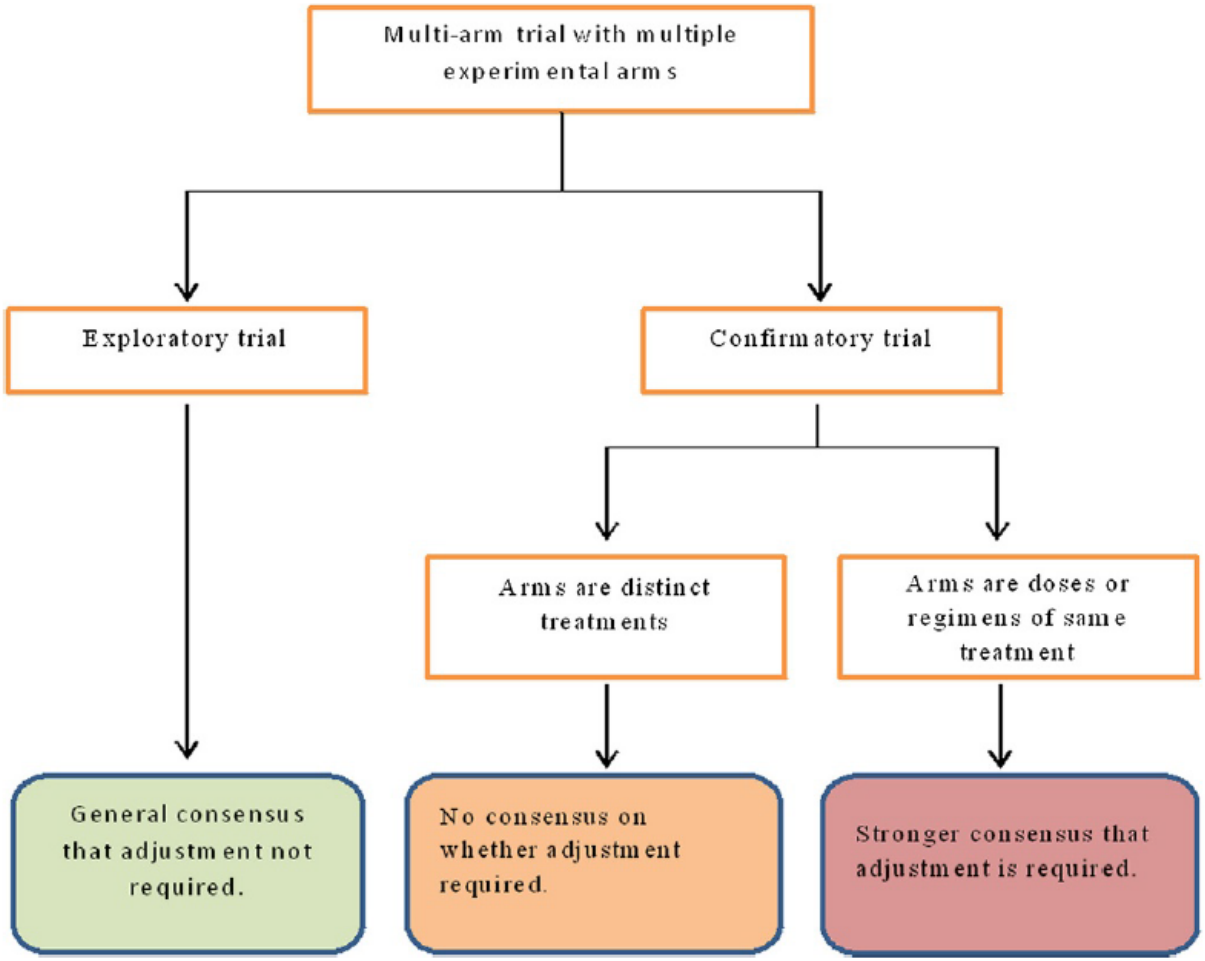
$(\delta_1, \delta_2, \delta_3) = (0, 0, 0.187$ and $\sigma = 0.52$			
Dose Dropping Criterion	Pwr(cumulative) - Pwr(stage wise)		
	Bonferroni	Simes	Dunnett
Any $\hat{\delta}_{i1} < 0$	12.3%	10.4%	3.4%
Any $\hat{\delta}_{i1} < -\sigma$	15.7%	14.0%	8.7%
Any $\hat{\delta}_{i1} < -2\sigma$	18.6%	17.3%	11.6%

$(\delta_1, \delta_2, \delta_3) = (0.187, 0.187, 0.187$ and $\sigma = 0.52$			
Dose Dropping Criterion	Pwr(cumulative) - Pwr(stage wise)		
	Bonferroni	Simes	Dunnett
Any $\hat{\delta}_{i1} < 0$	7.8%	1.2%	-0.1%
Any $\hat{\delta}_{i1} < -\sigma$	6.7%	0.9%	-0.1%
Any $\hat{\delta}_{i1} < -2\sigma$	7.0%	1.3%	0.3%

Summary of Comparisons

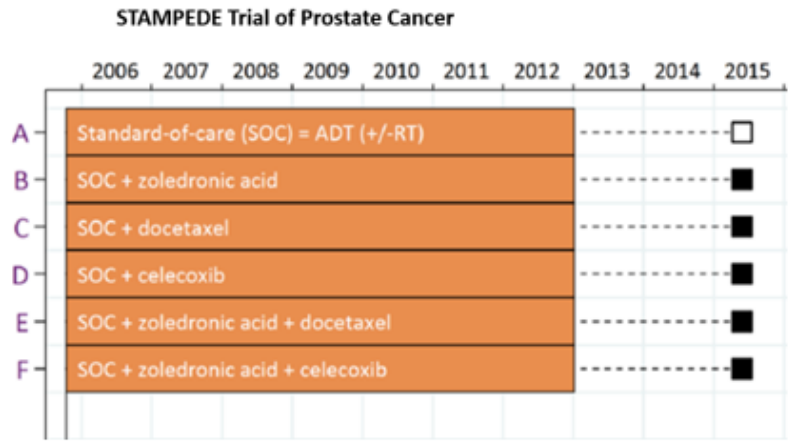
- Greater heterogeneity increases the power gain of Cumulative over Stage Wise MAMS
- More conservative rules for dropping doses increase the power gain of Cumulative over Stage Wise MAMS
- Power gains of Cumulative over Stage Wise MAMS depends on p-value adjustment method
  - 7% to18% gain with Bonferroni adjusted p-values
  - 1% to 17% gain with Simes adjusted p-values
  - 0% to 11% with Dunnett adjusted p-values
- A 5% power gain for SOCRATES trial translates into a sample size saving of 60-80 patients

Is FWER Control Needed for all MAMS Trials?

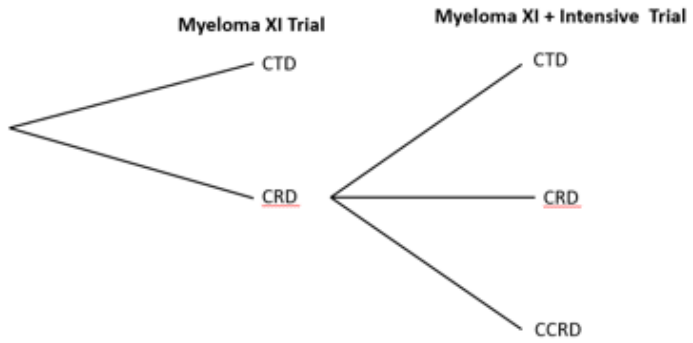


From Wason et al, *Trials*, 2014

Two examples of distinct treatments in one trial



- Single agents from different organizations added on to Standard of Care
- Is FWER required for STAMPEDE



CTD: cyclophosphamine+thalidomide + dexamethasone
CRD: cyclophosphamine+lenalidomide+dexamethasone
CCRD: CRD+carfilzomib

- CTD is the standard of care
- Trial 2 (MAMS trial) was started before results of Trial 1 were known
- Is FWER required for Trial 2?

Final Comments

- Generally speaking Cumulative MAMS would be prefered to Stage Wise MAMS because of greater power
- But Stage Wise MAMS controls FWER even for non-normal data and small sample sizes; hence might be preferable for rare disease trials
- Decision rules for dropping arms play an important role in power comparisons and need further investigation
- MAMS methods can be extended to investigating multiple populations and multiple endpoints with FWER
- FWER control required on a case by case basis