Comparing MAMS and P-value Combination Tests

Cyrus Mehta, PhD.

Cytel Inc, Cambridge MA

and

Harvard TH Chan School of Public Health

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

Acknowledgements

This is a joint collaboration with

Pranab Ghosh, Cytel Inc and Boston University

Lingyun Liu, Cytel Inc

Ping Gao, The Medicines Company

Ralph D'Agostino Sr., Boston University

Outline of Talk

1. MAMS procedure
   - Generalization of 2-arm group sequential boundaries
   - FWER control for adaptions
     - recompute group sequential boundaries
     - use closed testing and conditional error rate methods
2. P-Value Combination procedure
   - FWER control by closed testing
   - Boundary recomputation not necessary
3. Analytical comparison of MAMS and P-Value Combo
4. Design of SOCRATES-REDUCED clinical trial

The Problem

- D treatments are compared to a common control
- $\underline{\delta} = (\delta_1, \delta_2, \ldots \delta_D)$ = mean treatment effect
- Test $H_0 : \delta_i = 0$ for all $i = 1, 2, \ldots D$, versus the 1-sided alternative that $\delta_i > 0$ for at least one i
- Two-stage design with treatment selection, possible early stopping and possible SSR at Stage 1

MAMS Utilizes Score Statistic

- $\hat{\delta}_{ij}$ = mle of δ_i at stage $j = 1, 2$
- $\mathcal{I}_{ij}$ = Fisher information for δ_i at stage $j = 1, 2$
- $W_{ij} = \hat{\delta}_i \mathcal{I}_{ij}$ = score statistic for treatment i at stage j
- $\underline{W}_j = (W_{1j}, W_{2j}, \ldots W_{Dj})$ is Brownian process
- $\underline{W}_j$ has independent increments across stages but dependence between treatments within each stage

Boundaries for MAMS Procedure

- **Split available α into α_1 and $\alpha - \alpha_1$**
- **Find the stopping boundaries (b_1, b_2) such that**

$$P_0(\max\{\underline{W}_1\} \geq b_1) = \alpha_1$$

$$P_0(\max\{\underline{W}_1\} < b_1 \cup \max\{\underline{W}_2\} \geq b_2) = \alpha - \alpha_1$$

- **Monitor and claim efficacy if $\max\{\underline{W}_j\} \geq b_j$, $j = 1$ or 2**
- **These boundaries provide strong control of FWER**

(Ghosh et al, Biometrics, 2017: Generalization to K-stage MAMS)

What about Adaptive MAMS?

- Possible adaptive changes at end of stage 1:
  - Select a subset of the treatments for stage 2
  - Change the sample size for stage 2
- Control FWER by recomputing the boundary b_2 with:
  - closed testing
  - preservation of conditional error rates
- Note: Original cut-off b_2 also protects the FWER provided there is no SSR. But closed testing is more efficient

P-Value Combination Procedure

- Combine independent p-values from each stage
- Flexible. Use any valid p-values
- Control multiplicity by closed testing
- For example, to reject H_2: $\delta_1 = 0$, we must have

$$\Phi^{-1}\{1 - C[p_{2(1)}, p_{2(2)}]\} \le b$$

$$\Phi^{-1}\{1 - C[p_{2j(1)}, p_{2j(2)}]\} \le b \text{ for all } j \ne 2$$

$$\Phi^{-1}\{1 - C[p_{2jk(1)}, p_{2jk(2)}]\} \le b \text{ for all } j, k \ne 2$$

$$\vdots$$

Must reject all intersection hypotheses that include H_2

Illustrate FWER Control for $D = 3$

- Let $\mathcal{D} = \{1, 2, 3\}$ denote the three treatment indices. Suppose $\mathcal{S} = \{2, 3\}$ are selected for stage 2 with SSR
- Test $H^{(2)}$: $\delta_2 = 0$ and $H^{(3)}$: $\delta_3 = 0$ with strong FWER
- To reject $H^{(2)}$ with strong FWER we must reject

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

  all with valid level-α tests
- To reject $H^{(3)}$ with strong FWER we must reject

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

  all with valid level-α tests

Note: $\boldsymbol{H^{(a,b)} = H^{(a)} \cap H^{(b)}}$

Testing $H^{(I)}$: MAMS Approach

- For any $I \in \{(2),(3),(1,2),(2,3),(1,2,3)\}$ let $I_{\mathcal{S}} = I \cap \mathcal{S}$
- Let $\underline{W}_{Ij} = \{W_{qj}; q \in I\}$ = scores for treatments in I only
- A valid level-α test of $H^{(I)}$ must preserve the conditional error rate

1. Recompute the boundaries (b_{I1}, b_{I2}) that are appropriate for $H^{(I)}$

$$P_0(\max\{\underline{W}_{I1}\} \geq b_{I1}) = \alpha_1$$

$$P_0(\max\{\underline{W}_{I1}\} < b_{I1} \cap \max\{\underline{W}_{I2}\} \geq b_{I2}) = \alpha - \alpha_1$$

2. Compute critical cut-off b^*_{I2} that preserves conditional error rate

$$P_0(\max\{\underline{W}^*_{I_{\mathcal{S}}2}\} \geq b^*_{I2} | \underline{w}_{I_{\mathcal{S}}1}) \leq P_0(\max\{W_{I2}\} \geq b_{I2} | \underline{w}_{I1})$$

3. Reject $H^{(I)}$ if observed $\max\{\underline{W}^*_{I_{\mathcal{S}}2}\}$ exceeds b^*_{I2}

Testing $H^{(I)}$: P-value Combo

- As before let $\mathcal{D} = \{1, 2, 3\}$, $\mathcal{S} = \{2, 3\}$ and $I_{\mathcal{S}} = I \cap \mathcal{S}$
- Recall that to reject $H^{(2)}$ we must reject $H^{(I)}$ for all $I \in \{(2), (1, 2), (2, 3), (1, 2, 3)\}$
- P-value combo differs from MAMS in how $H^{(I)}$ is tested
  - Combines independent p-values
  - Considerable flexibility exists in choice of p-values
  - Bonferroni or Simes p-values control FWER conservatively; less sensitive to normality assumption
  - $\max\{\underline{W}_{Ij}\}$ (Dunnett) p-values control FWER exactly but sensative to normality assumption

Test $H^{(I)}$: Pvalue Combo

1. Compute independent p-values for the two stages
   $p_{I1} = P_0(\max\{\underline{W}_{I1}\} \geq \max\{w_{I1}\})$
   $p_{I(2)} = P_0(\max\{\underline{W}_{I_{\mathcal{S}}(2)}\} \geq \max\{w_{I_{\mathcal{S}}(2)}\})$
   where $I(2), I_{\mathcal{S}}(2)$ denote incremental stage 2 data

2. Combine with pre-specified weights h_1 and h_2
   $C(p_{I1}, p_{I(2)}) = 1 - \Phi\{h_1\Phi^{-1}(1 - p_{I1}) + h_2\Phi^{-1}(1 - p_{I(2)})\}$

3. Reject $H^{(I)}$ if $C(p_{I1}, p_{I(2)}) < c$ where c is such that
   $\int_{\alpha_1}^{1} \int_{0}^{1} 1_{[C(x,y) \leq c]} dy dx = \alpha - \alpha_1$

Note: p_{I1} and $p_{I(2)}$ are Dunnett-type p-values

Exact Analytical Power Comparisons: MAMS vs P-Value Combo

- **Three-arm trial with normally distributed data**
- **No early stopping and no sample size adaptation**
- **Power functions:**

P(MAMS) =

$$1-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{w_{1(2)}=-\infty}^{b_2-w_{11}}\int_{w_{2(2)}=-\infty}^{b_2-w_{21}} f_{(2)}\left(w_{1(2)},w_{2(2)}\right)dw_{2(2)}dw_{1(2)}\right)f_1(w_{11},w_{21})dw_{21}dw_{11}$$

P(COMB) =

$$1-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{w_{1(2)}=-\infty}^{F_{(2)}^{-1}(g)}\int_{w_{2(2)}=-\infty}^{F_{(2)}^{-1}(g)} f_{(2)}\left(w_{1(2)},w_{2(2)}\right)dw_{2(2)}dw_{1(2)}\right)f_1(w_{11},w_{21})dw_{21}dw_{11}$$

where $g = \Phi\left(\frac{Z_\alpha - h_1 Z_{p_1}}{h_2}\right)$ is a function of w_{11}, w_{21}

- **The only difference is in the limits of integration**

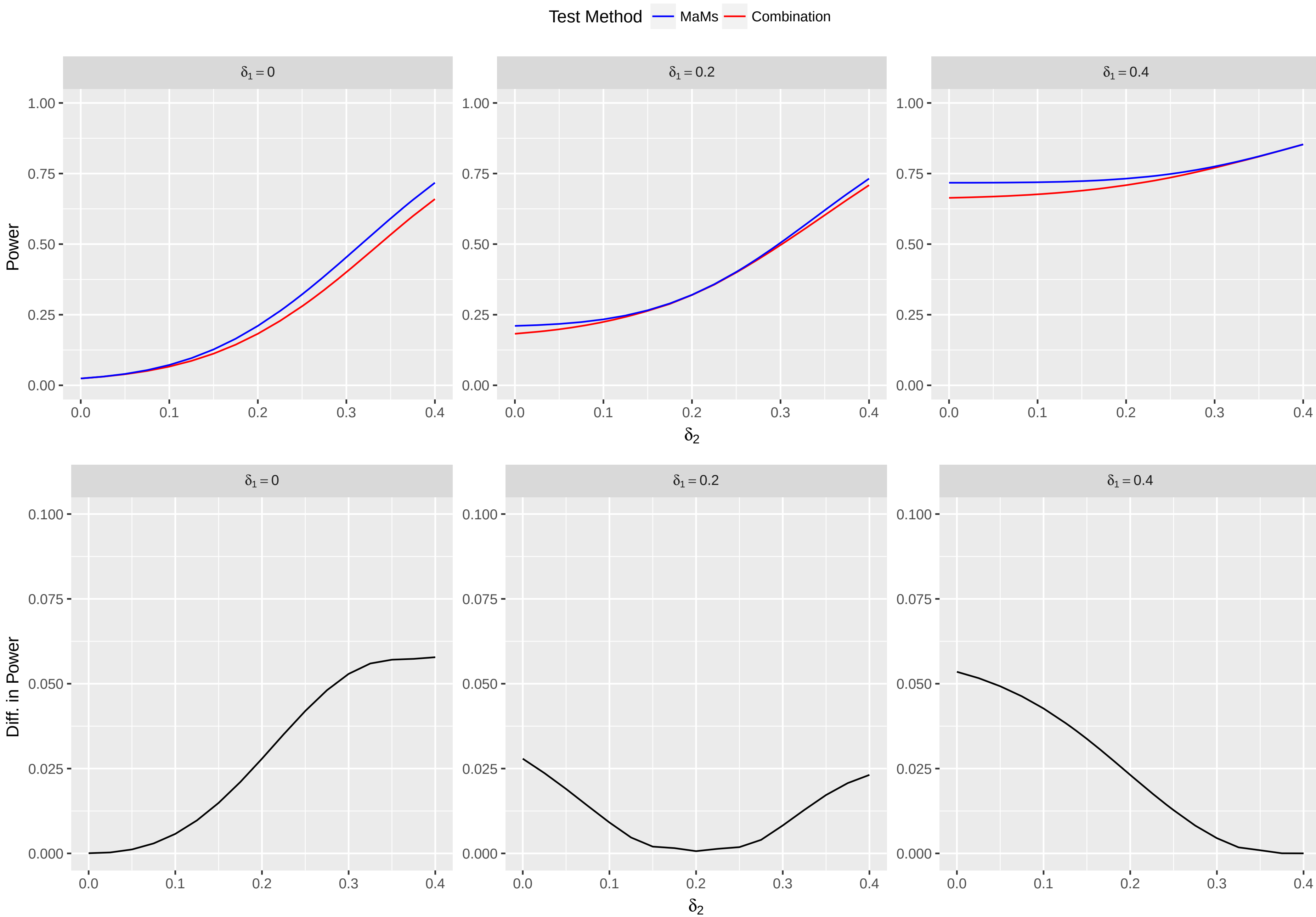


Figure 1: Power comparisons between MAMS and COMBO

Summary of Comparisons

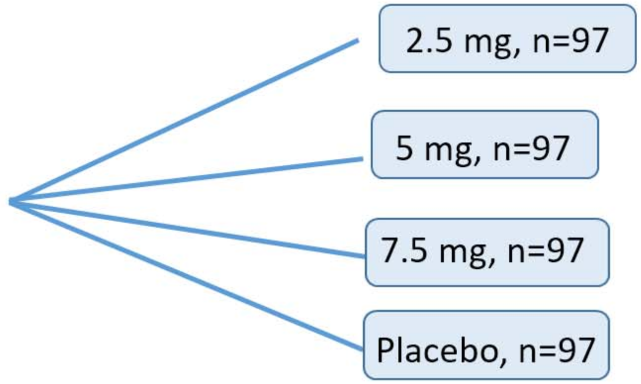
- **Two treatments were compared to a common control**
- **Ranges:** $\delta_1 = (0, 0.2, 0.4)$; $(0 \leq \delta_2 \leq 1)$; $\sigma^2 = 1$
- **When** $\delta_1 = \delta_2$ **the two methods have the same power**
- **The more the** δ**'s differ, the greater the power gain for MAMS**
- **When** $\delta_i = 0$ **and** $\delta_j = 0.4$ **MAMS has 5% more global power than P-val Combo**

SOCRATES-REDUCED Randomized Trial

- **Vericuguat (2.5 mg, 5 mg, 10 mg) compared to placebo**
- **Endpoint: week-12 change from baseline in log NT-proBNP**
- **Trial enrolled 65 patients/arm and planned to pool the treatment arms for the final analysis**
- **Observed** $\hat{\delta}_1 = 0.039, \hat{\delta}_2 = 0.073, \hat{\delta}_3 = 0.248$
- **Pooling diluted the treatment effect and trial failed**
- **Re-design as a 4-arm adaptive trial**

Ref: Gheorghiade et al, JAMA 2015

Re-design as 4-arm Adaptive Trial



	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

- **Base design (97/arm) has 80% power at 1-side $\alpha = 0.025$ if $\delta = 0.187$ for all three arms versus placebo**
- **But power will deteriorate if all the δs are not 0.187**
- **Use 2-stage P-value Combo and MAMS adaptive designs**
  - **Drop arms with $\hat{\delta} < 0$ at stage 1**
  - **Re-allocate sample size to remaining arms**
  - **No early stopping**

Table 1: Power Comparison for SOCRATES-REDUCED, using Multiple Arm designs

δ	Power (%)				
	Single Look	Adaptive P-value Combination			Adaptive Group Sequential
		Bonferroni	Simes	Dunnett	
(0.04, 0.073, 0.25)	84.1	80.7	82.5	86.1	88.9
(0.187, 0.187, 0.187)	80.4	73.6	79.3	80.1	80.97
(0, 0.187, 0.187)	73.1	67.8	71.2	76.8	78.85
(0, 0.094, 0.187)	57.1	50.9	55.2	61.3	64.86
(0, 0, 0.187)	59.1	52.1	54.0	62.7	64.66
(0, 0, 0)	2.502	1.52	2.01	2.53	2.418

All table entries are based on 10,000 simulated clinical trials

Conclusions

- **MAMS dominates over all other designs**
- **Under homogeneity of δs Single Look and MAMS are equivalent (because of no early stopping)**
- **Bonferroni and Simes are not competitive with Dunnett or MAMS despite adjusting the nominal α of the latter two**