Design of Multi-Arm Multi-Stage Trial

JSM , Baltimore
August 01, 2017

Pranab Ghosh

Cytel Inc, Cambridge MA
and

Boston University, Boston MA

Acknowledgment

- This is a joint collaboration with
  - Cyrus Mehta, Cytel Inc, Cambridge MA.
  - Lingyun Liu, Cytel Inc, Cambridge MA.
  - Ping Gao, The Medicines Company, Parsippany NJ.
  - Ralph D'Agostino, Boston University, Boston MA.

Outline of Presentation

- Introduction to Multi Arm Multi-Stage(MaMs) design, illustrative example.
- Construction of Basic MAMS design problem.
- Efficient boundary computation in MaMs design, numerical algorithm.
- Comparison of MAMS design against P-value combination method.

What is MAMS Design?

- Generalization of two-arm group sequential designs.
- Pair wise comparison of each arm with a common control.
- Monitor the accruing data as successive looks.
- Possible early stopping or adaptive changes.
  - Stop for efficacy if any arm crosses the efficacy boundary.
  - Stop for futility if all arms cross the futility boundary.
  - Permit dropping of losers that cross futility.
  - Modifying sample size re-estimation or patient randomization.
- Alternative to method of combining p-values (Posch et. al., 2005).
- Saves sample size, by not running separate trials to do pairwise comparison.

Properties of MAMS design

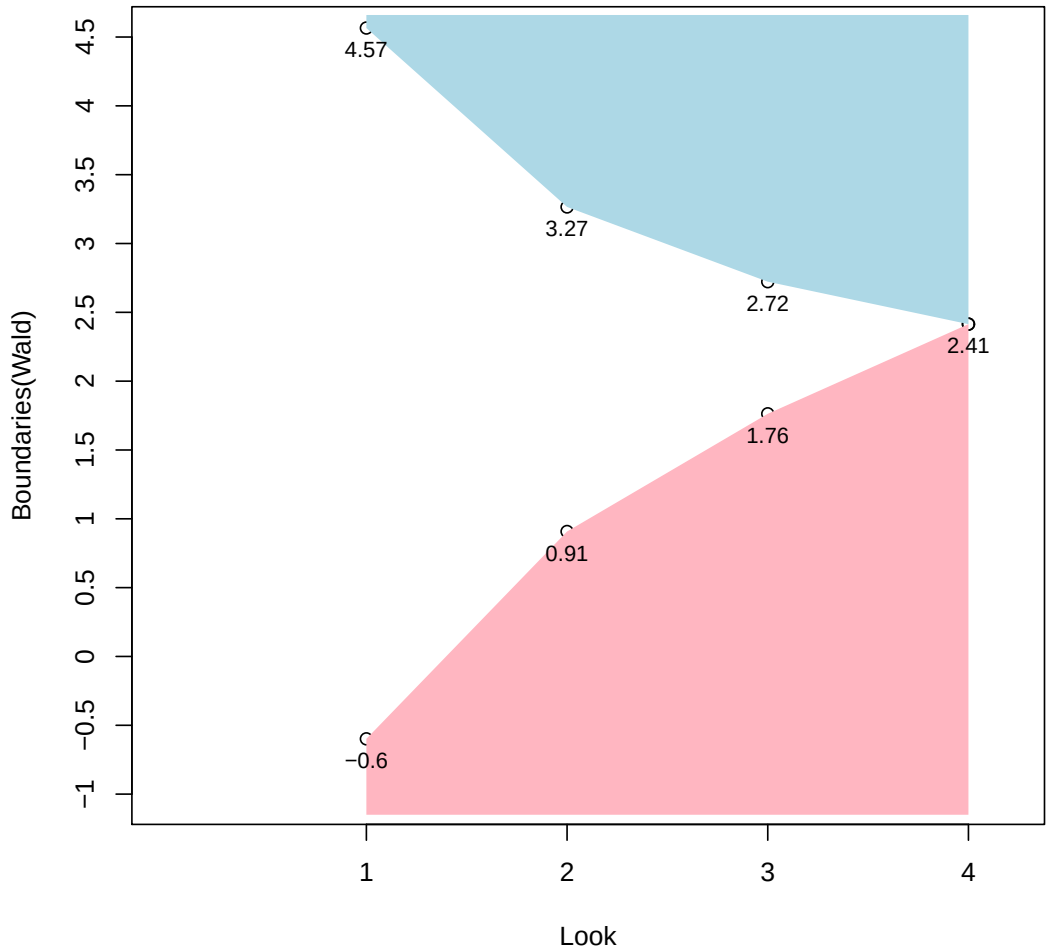
- Extended version of Dunnett's test from single look to multiple look.
- Extend two arm group sequential design to compare multiple (> 2) arms.
- Closed testing is not required, test is based on maximum statistics.
- Design will control family wise error rate (FWER).
- Dropping of arms at interim are allowed, FWER will be controlled.
- Even one arm crosses the efficacy boundary, trial can be continued with remaining arms.

Example: INHANCE Trial

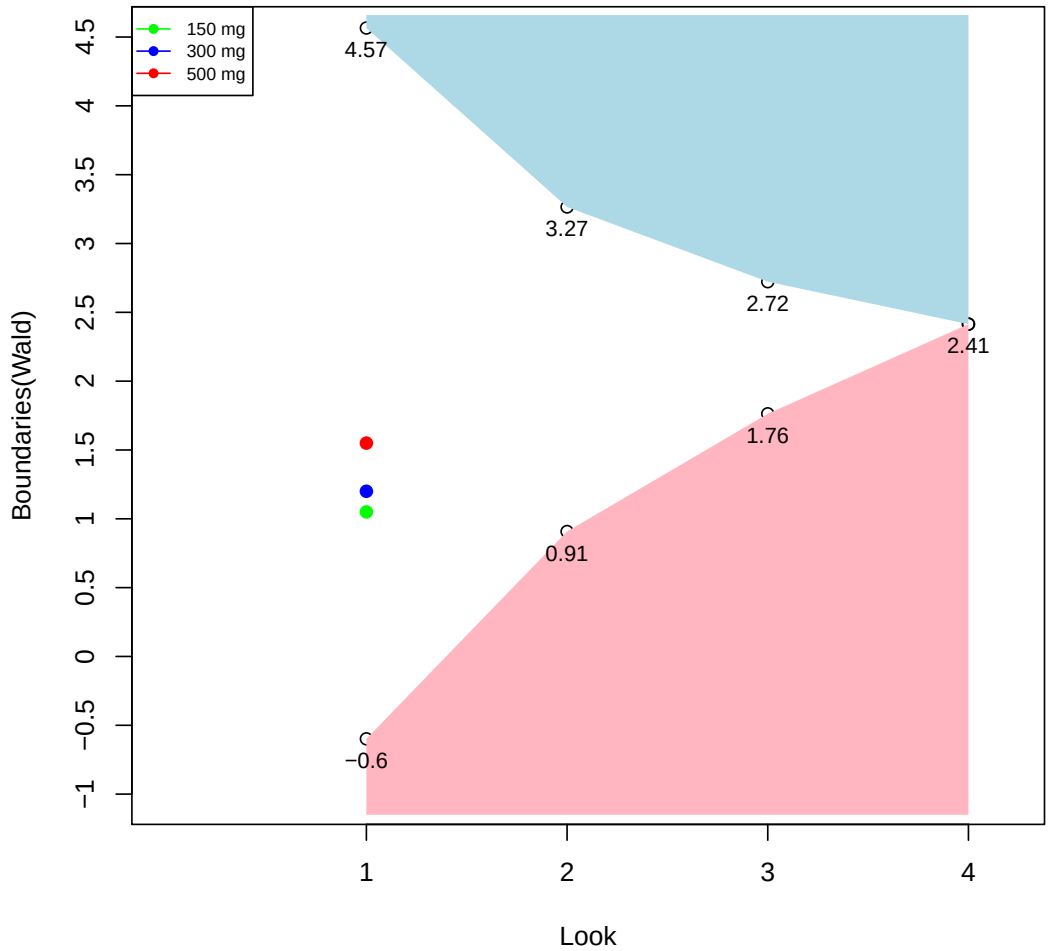
- Treatment for chronic obstructive pulmonary disease (COPD).
- Three doses (150 mg, 300 mg, 500 mg) of Indacaterol vs Placebo.
- Endpoint: Week 12 change from baseline in 24 hour trough FEV1.
- Differences from placebo are between 0.14 and 0.18 liters with $\sigma = 0.5$.
- Design a 4-arm-4-look trial for 90% power at one-sided $\alpha = 0.025$.
- 1:1 allocation between each treatment arm with placebo.
- Use O'Brien-Fleming efficacy and futility boundaries.
- Require 171 patients on each arm.

INHANCE Trial: adapted from Donohue et al, Am J Respi Crit Care, Vol 182, pp 155-162, 2010

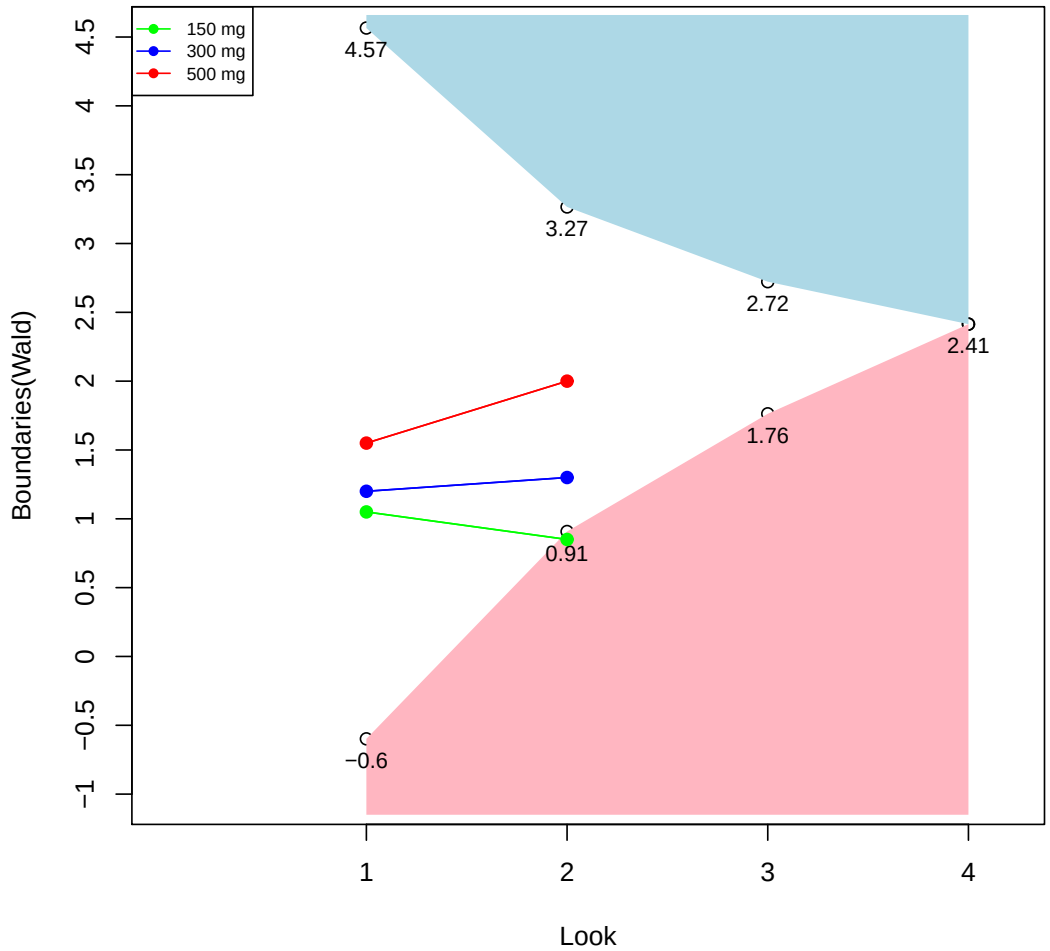
MAMS Design Boundaries



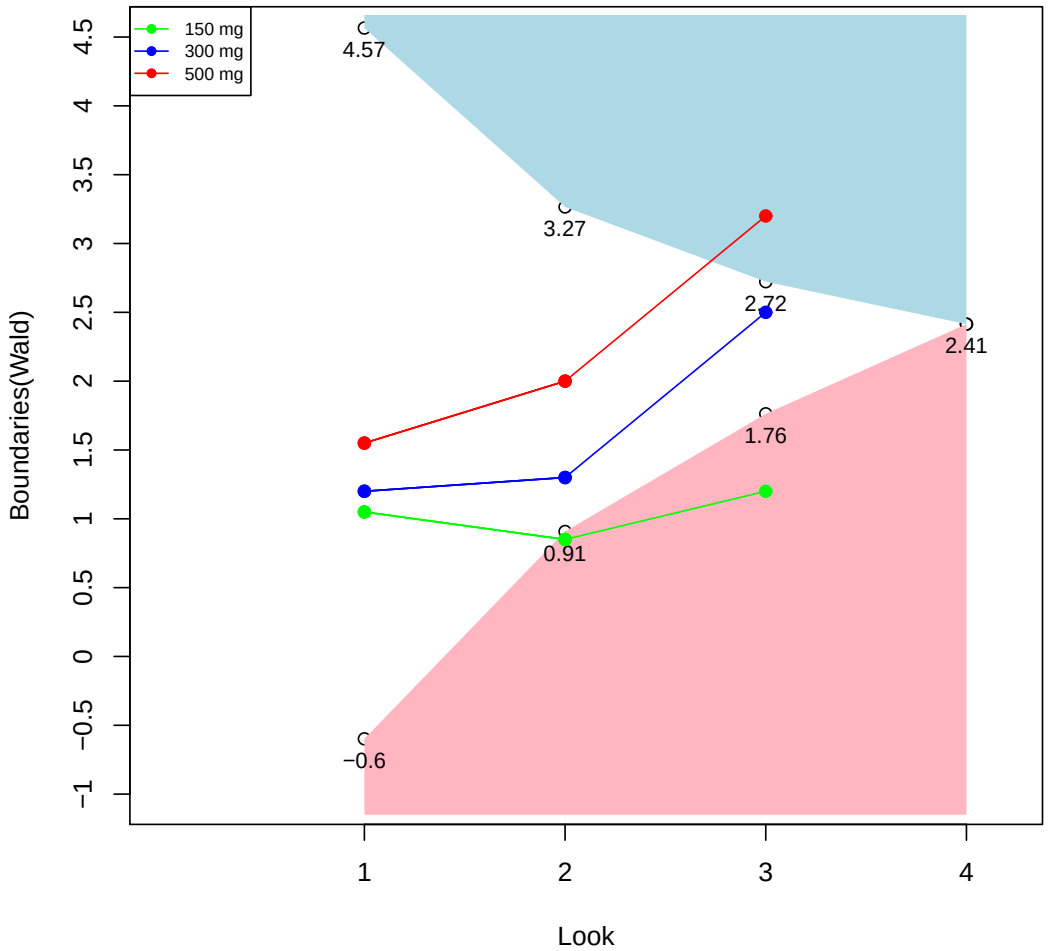
Monitor the Trial : Look 1



Monitor the Trial : Look 2



Monitor the Trial : Look 3



Problem Formulation

- Multiple-Arm :
  - Pairwise comparison of D active treatments against a common placebo.
  - δ_i be the treatment effect of i^{th} arm against placebo, $i = 1, \ldots D$.

$$\begin{aligned} H_0 &: \delta_i \leq 0 \text{ for all } i \\ H_A &: \delta_i > 0 \text{ for at least one } i \end{aligned}$$

- Multiple-Stage
  - K looks at accumulating data indexed by $j = 1, 2, \ldots K$
  - Score statistics for the i$^{\text{th}}$ treatment at look j is $W_{ij} = \hat{\delta}_{ij} I_{ij}$.
- Construct efficacy boundaries under H_0 that provide strong control of FWER at level-α.

Type I Error, Type II Error

- Let $e_1, \ldots e_K$ be the efficacy boundaries and $f_1, \ldots f_K$ $(e_K = f_K)$ are the non-binding futility boundaries. Stop at look j due to -
  - early efficacy if $W_{ij} \geq e_j$, for at least one $i = 1, 2, \ldots, D$.
  - early futility if $W_{ij} \leq f_j$, $\forall i = 1, 2, \ldots, D$.
- Efficacy boundaries must satisfy the following criteria

$$\sum_{j=1}^{K} P_{H_0}\left(\bigcap_{l=1}^{j-1} \max_i\{W_{il}\} < e_l \text{ and } \max_i\{W_{ij}\} \geq e_j\right) = \alpha$$

- Type II error

$$\sum_{j=1}^{K} P_{H_A}\left(\bigcap_{l=1}^{j-1} f_l < \max_i\{W_{il}\} < e_l \text{ and } \max_i\{W_{ij}\} \leq f_j\right) = \beta$$

Distribution of the Score Statistics

- $\underline{W}_j = (W_{1j}, \ldots, W_{Dj})$ is a multivariate discrete Brownian motion, indexed by look number j.
- W_{ij} follows multivariate normal distribution with
  - $E(W_{ij}) = \delta_i I_{ij}$
  - $\text{Cov}(W_{i_1 j}, W_{i_2 j}) = \begin{cases} I_{i_1 j} & \text{if } i_1 = i_2 \\ n_{0j}\sigma_0^2 \Lambda_{i_1}\Lambda_{i_2} & \text{if } i_1 \neq i_2 \end{cases}$
- $\Lambda_i = \left(\sigma_0^2 + \frac{\sigma_i^2}{\lambda_i}\right)^{-1}$ and $I_{ij} = n_{0j}\Lambda_i$
- For $j_1 < j_2$, $Cov(\underline{W}_{j_1}, \underline{W}_{j_2}) = Var(\underline{W}_{j_1})$. This implies $\underline{W}_{(j+1)} = \underline{W}_{j+1} - \underline{W}_j$ and $\underline{W}_j$ are independent.

Complexity in Computation

- For Computing boundary crossing probability at look j, we need integrate the joint density of $\underline{W}_1, \underline{W}_2, \ldots, \underline{W}_j$, which will be of the form

$$P_j(b_1, \ldots, b_j) = \oint_{\underline{w}_1 < b_1} \cdots \oint_{\underline{w}_j < b_j} f\left(\underline{w}_1, \ldots, \underline{w}_j\right) d\underline{w}_j \ldots d\underline{w}_1$$

- Computing this probability requires integration of multivariate density of $(\underline{W}_1, \ldots \underline{W}_j)$ with dimension $j \times D$.
- Using any numerical quadrature method with G points on each dimension, will require $G^{j \times D}$ times evaluation of the joint density function.

Computing Step I - Scaling Score Statistics

- Scale score statistics by $\frac{1}{\sqrt{\mathcal{I}_{max}}}$, where $\mathcal{I}_{max} = n_{0K} * \Lambda_{max}$; $(\Lambda_{max} = \max_i \Lambda_i)$
- $\underline{U}_j = \frac{1}{\sqrt{\mathcal{I}_{max}}} \underline{W}_j \sim N(t_j \vec{\eta}, t_j \rho)$.
  - $t_j = \frac{n_{0j}}{n_{0K}}$, information fraction at look j.
  - $\eta_i = \delta_i \sqrt{\mathcal{I}_{max}} \frac{\Lambda_i}{\Lambda_{max}}$, drift parameter for the i^{th} treatment arm.
  - $$\rho_{i_1 i_2} = \begin{cases} \frac{\Lambda_{i_1} \Lambda_{i_2}}{\Lambda_{max}} \sigma_0^2 & i_1 \neq i_2 \\ \frac{\Lambda_{i_1}}{\Lambda_{\max}} & i_1 = i_2 \end{cases}$$
- Also Cov $\left(\underline{U}_{j_1}, \underline{U}_{j_2}\right) = t_{j_1} \rho$, $j_1 < j_2$.
- Preserve Brownian process (independent increment) properties of the score statistics.
- Efficacy boundaries (under H_0) will not depend on sample size.

Computation Steps II: Independent Increment

- $\underline{U}_{(j)} = \underline{U}_j - \underline{U}_{j-1} \sim N(t_{(j)}\eta, t_{(j)}\rho)$ and is independent of $\underline{U}_{j-1}$.
- Using the independent property of the Brownian process for $\vec{U}_j$, we can write this as integration of dimension D only, with recursive in nature.

$$P_j = \oint_{\underline{u}_1 < \frac{b_1}{\sqrt{\mathcal{I}_{\max}}}} f_{\underline{U}_1}(\underline{u}_1) \oint_{\underline{u}_{(2)} < \frac{b_2}{\sqrt{\mathcal{I}_{\max}}} - \underline{u}_1} f_{\underline{U}_{(2)}}(\underline{u}_{(2)}) \cdots \oint_{\underline{u}_{(j)} < \frac{b_j}{\sqrt{\mathcal{I}_{\max}}} - \underline{u}_{j-1}} f_{\underline{U}_{(j)}}(\underline{u}_{(j)}) d\underline{u}_{(j)} \ldots d\underline{u}_{(2)} d\underline{u}_1$$

Computation Steps III : Transformation

- Series of integral transformation, due to a suggestion by Alen Genz(1992),
  - $\frac{\underline{u}_{(j)} - t_j \underline{\eta}}{\sqrt{t_{(j)}}} = C \underline{y}_j$; $\rho = CC^T$; reduces the computation to recursive univariate normal integration.
  - Gaussian transformation $\Phi(y_{ij}) = x_{ij}$ to get finite integration range.

$$P_j = \int_0^1 e_{11} \cdots \int_0^1 e_{D1} \cdots \int_0^1 e_{1j} \cdots \int_0^1 e_{Dj} d\vec{x}_j \cdots d\vec{x}_1$$

$$e_{ik} = \Phi\left\{ \frac{1}{C_{ii}} \left[\frac{1}{t_{(k)}} \left(\frac{b_k}{\sqrt{\mathcal{I}_{\max}}} - t_k \eta_i - \sum_{m=1}^{i} C_{im} p_{mk-1} \right) - \sum_{m=1}^{i-1} C_{im} \Phi^{-1}(e_{mk} x_{mk}) \right] \right)$$

$$p_{mk} = \sum_{l=1}^{k} \sqrt{t_{(l)}} \Phi^{-1}(e_{ml} x_{ml})$$

Computation Steps IV : Quasi-Monte Carlo

- Quasi Monte Carlo method was used, which provide a higher convergence rate than regular Monte Carlo ($O(N^{-1})$ against $O(N^{-0.5})$).
- Also provides the **accuracy** in estimation which depends on number of sample points (N).

Speed and Accuracy of Computing Algorithm: I

Our Algorithm				R Package(†)
K	D	$3 \times \sigma$ Accuracy of Probability Estimates	Computing Time (secs)	Computing Time (secs)
2	3	0.000075	1	2
	4	0.000156	1	2
	5	0.000302	2	2
	6	0.000421	2	2
3	3	0.000359	1	138
	4	0.000495	1	148
	5	0.001042	2	156
	6	0.000637	2	158
(†) https://cran.r-project.org/web/packages/MAMS/index.html				

Speed and Accuracy of Computing Algorithm: II

Our Algorithm				R Package(†)
K	D	$3 \times \sigma$ Accuracy of Probability Estimates	Computing Time (secs)	Computing Time (secs)
4	3	0.000585	1	$>$ 8 hrs
	4	0.000581	2	$>$ 8 hrs
	5	0.001848	2	$>$ 8 hrs
	6	0.00097	3	$>$ 8 hrs
5	3	0.000739	1	$>$ 8 hrs
	4	0.001324	2	$>$ 8 hrs
	5	0.001823	2	$>$ 8 hrs
	6	0.000995	4	$>$ 8 hrs
(†) https://cran.r-project.org/web/packages/MAMS/index.html				

Comparison of MAMS and P-value Combination

- P-value Combination Method
  - Uses **closed testing** to guarantee strong control of FWER.
  - Combines the multiplicity adjusted p-values(Bonferroni, Simes, Dunnet) from the two stages with pre-specified weights and combination function.
  - Does not utilize correlation between p-values (except Dunnett test).
- MAMS Method
  - Boundaries are constructed under global null hypothesis.
  - Strong control of type-1 error is nevertheless guaranteed.
  - Boundaries constructed from distribution of the maximum statistic.
  - Exploits the correlation between arms for added efficiency.

Power Comparison: MAMS vs P-value Combination

- Global power of 2-stage design with 50 patients/arm
  - $\alpha = 0.025$
  - $\alpha_1 = 0.01$ at 50% interim analysis
  - $\delta/\sigma = 0.5$ for all comparisons

Number of Arms	Disjunctive Power			
	Bonferroni	Simes	Dunnett	MAMS
2	0.70	0.72	0.73	0.75
3	0.70	0.75	0.75	0.78
4	0.69	0.76	0.76	0.80

Concluding Remarks

- MAMS designs natural extension of 2-arm group sequential design.
- Availability of MAMS software has been the major hurdle to their acceptance in the past. Powerful new algorithms have been developed that overcome this hurdle.
- MAMS designs appear to be competitive in terms of power with P-value Combination designs.