Platform Trials in Confirmatory Settings

Pranab Ghosh, PhD.

Cytel Inc, Cambridge

Biopharmaceutical Applied Statistics Symposium, 2019

Outline of Presentation

- **What is Platform trial?**
- **Example with STAMPEDE trial.**
- **Type I error control, FWER and PWER.**
- **Currently available statistical methodology, Comparison.**
- **Conclusions.**

Platform Trials : Features

- Multiple treatment arms get compared to a common placebo.
- Multiple interim looks along the way.
- Early identification of efficacious arms.
- Early stopping of arms that show lack of benefit
- Possibility of adding new treatments in between the trial.
- Re-estimating sample size at interim looks, if needed.
- Control of type I error (FWER or PWER)

Motivation of PLATFORM Trial

Latest version of the Prescription Drug User Fee Act(PDUFA VI)

encourages the use of complex and innovative trial design that helps efficient and informative decision making in drug development and this requires both:

1. **Speed up development of new Agents and the combinations.**
2. **More efficient use of available resource (patients and money).**

1. Speed up Evaluation of new Agents

- **Controlled testing of many new therapies simultaneously.**
- **Eliminate non-performing agents quickly.**
- **Identify multiple winners earlier, in a single trial,**

2. Efficient Use of Resources

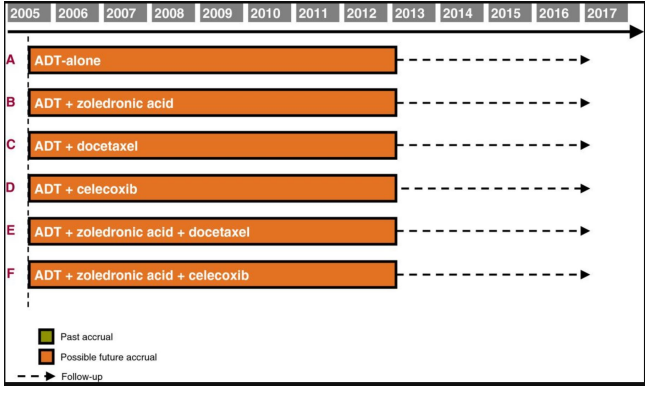
- **Common control arm saves sample size.**
- **Reduction of administrative burden.**
- **Common platform instead of trial-specific infrastructures**
- **Continue using same machinery even if adding new arms**
- **Attractive option for patients, hence speeds up enrollment**

Error Control in Platform Trial

- **PairWise Error Rate (PWER): The null hypothesis for each treatment arm is controlled separately at level-α. Straightforward application of two-sample methodology!**
- **FamilyWise Error Rate (FWER): Under any combination of null and alternative hypothesis for the treatment arms, the probability of making one or more false claims controlled at level-α. Statistically challenging!**

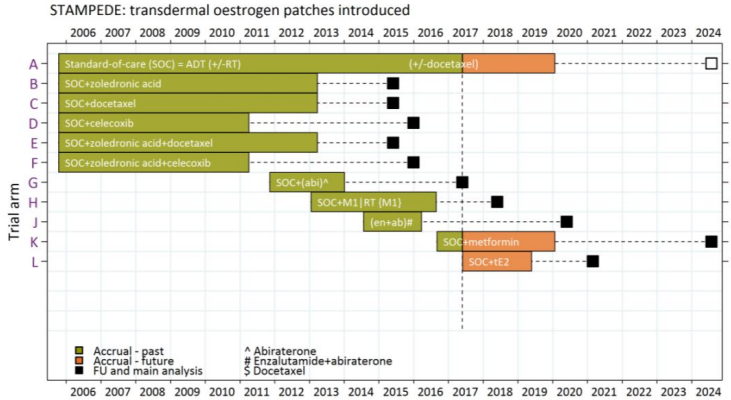
STAMPEDE Trial (Activate in 2006)

(Systemic Therapy for Advanced Metastatic Prostate cancer: Evaluation of Drug Efficacy)



Primary Efficacy Endpoint: Overall Survival (OS)
Futility Criterion: Disease-free Survival (DFS)

STAMPEDE Trial (2006 to 2017+)



Note: dotted line represents activation of this protocol version

STAMPEDE : Important Features

- Treatment vs. Control comparisons are concurrent in time
- No early efficacy stopping of treatments at interim looks
- Interim looks are only used to terminate futile arms
  - Terminate futile arms based on DFS criterion, not OS
  - Drop arm if $H_0 : HR_{dfs} = 1$ cannot be rejected
- Final analysis after pre-planned number of events
- Type I error control
  - Pair-wise error rate (PWER) controlled at $\alpha = 0.025$.
  - Family-wise error rate (FWER) estimated only

When FWER Control Necessary

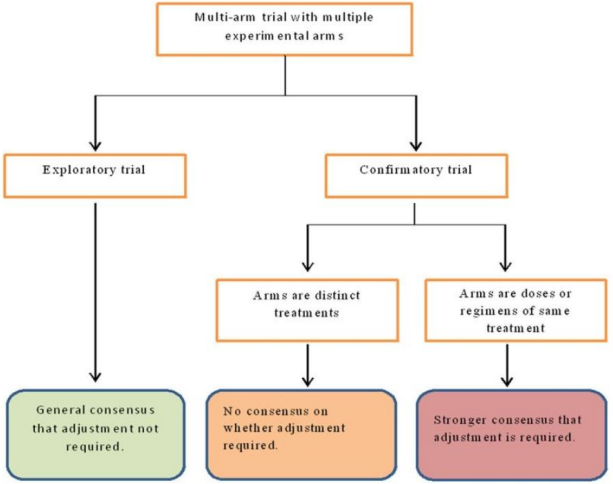


Figure 2
Flowchart representing consensus from literature on under what circumstances multiple-testing adjustment is necessary.

Statistical Methodology for FWER Control

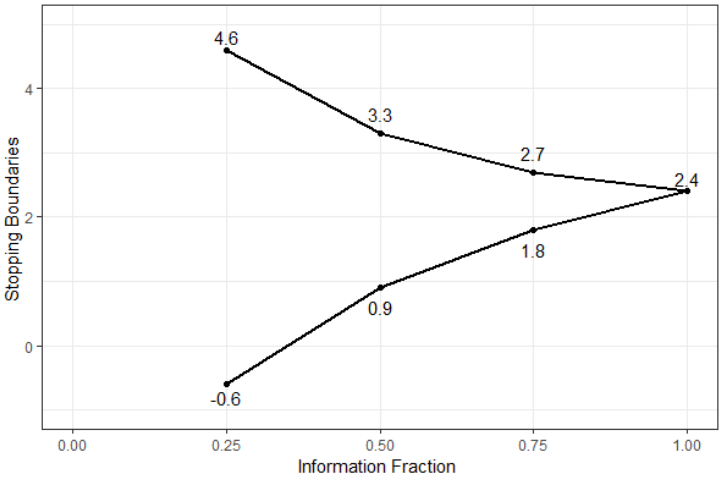
Two Competing methods are available -

1. **Cumulative Multi-Arm Multi-Stage (MAMS) Design :**
   - **Use cumulative data to construct each statistic stage by stage.**
   - **Construct separate test statistic for each treatment.**
   - **Group Sequential boundaries using distribution of maximum statistic. No closed is required in absence of adaptation.**
   - **For adaptive changes; type I error control requires-**
     - **Closed testing.**
     - **Preserve type I error using conditional error rate principle.**
2. **Stagewise Multi-Arm Multi-Stage (MAMS) Design:**
   - **Construct a single test-statistic by combining multiplicity adjusted p-values, computed from stagewise incremental data.**
   - **2-arm group sequential boundary should be used**
   - **Closed testing is must to achieve strong FWER control.**

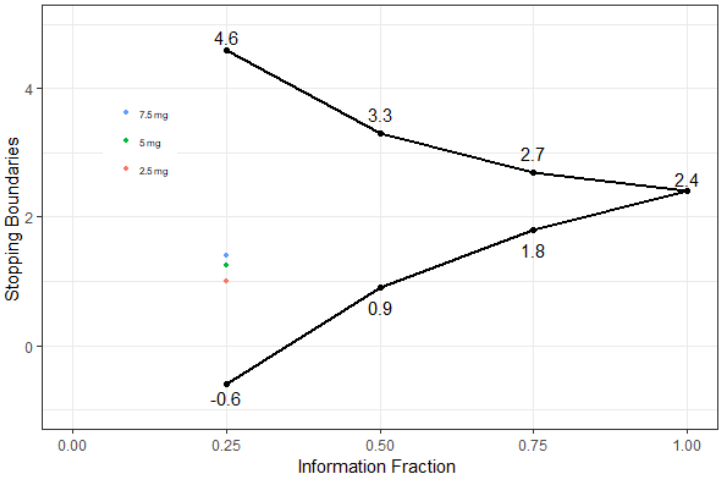
Example : Socrates REDUCED

- Multi-center, randomized trial targeting patients with worsening chronic heart failure after clinical stabilization
- Randomized to either placebo or three doses (2.5, 5 and 7.5mg) of vericiguat.
- Primary Endpoint : Change from baseline to week 12 in log-transformed level of N-terminal pro-B-type natriuretic peptide (NT-proBNP).
- Up to four looks at the accumulating data
- Drop non-performing arms along the way.
- identify arms with early efficacy.

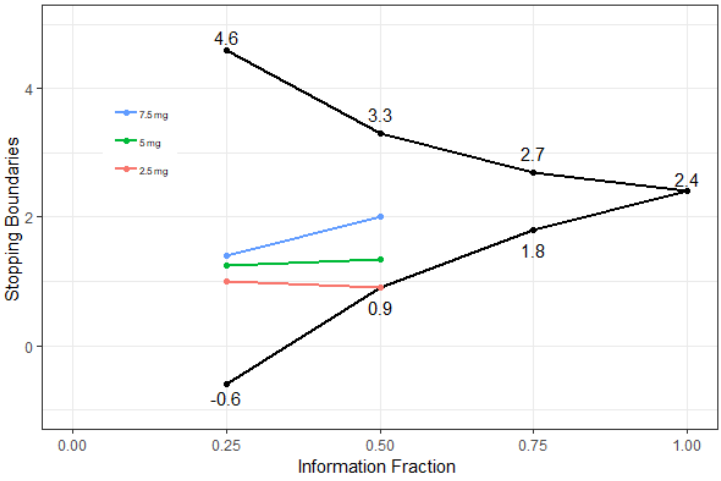
1. Cumulative MAMS Approach



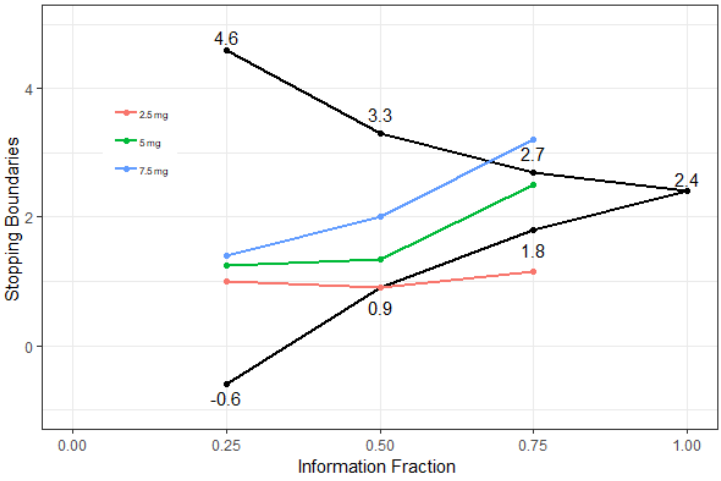
Interim Monitoring at Look 1



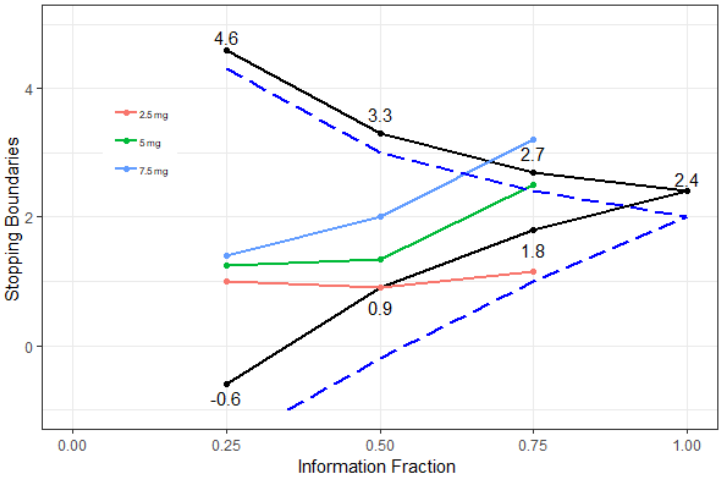
Interim Monitoring at Look 2



Interim Monitoring at Look 3



PWER : 2-Arm Boundaries



Main Statistical Issue

- How to generate the boundaries with a strong control of FWER.
- How to modify the boundaries if there is adaptation?
  - Dropping arms at an interim look
  - Altering sample size at an interim look, if needed.

Reference for Boundary Computation

BIOMETRIC METHODOLOGY | Open Access

Design and monitoring of multi-arm multi-stage clinical t

Pranab Ghosh, Lingyun Liu, P. Senchaudhuri, Ping Gao, Cyrus Mehta

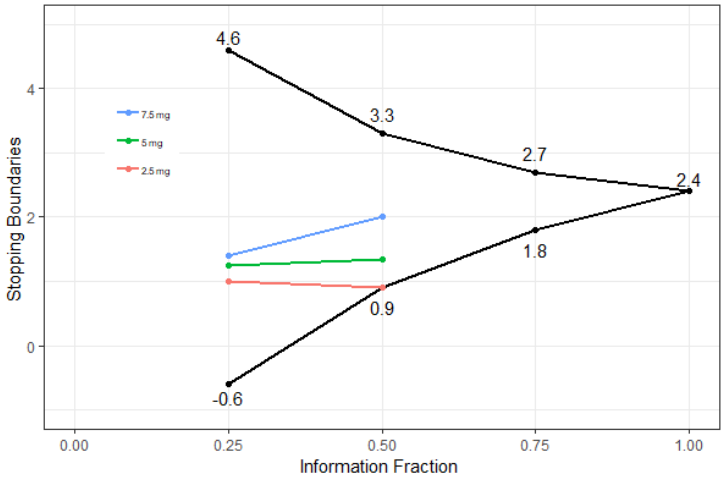
First published: 27 March 2017 | https://doi.org/10.1111/biom.12687 | Cited by: 9

Adaption to the Original Design

Drop one or more arms at an interim look and randomize all the remaining patients among the remaining arms. Thereby sample size on selected arms gets increased and strong control of FWER achieved by:

- **Re-computing the efficacy stopping boundaries**
- **Preservation of conditional error rates**
- **Closed testing**

FWER Control for Adaptive Changes



Drop 2.5mg dose, randomize all the remaining patients to available arms.

Closed Testing Requirement

- Test $H_0^{(H)} : \delta_H \leq 0$ and $H_0^{(M)} : \delta_M \leq 0$ only with strong control over FWER.
- To reject $H_0^{(H)}$ with strong control over FWER we must reject

$$H_0^{(H)}, H_0^{(H,M)}, H_0^{(H,L)}, H_0^{(H,M,L)}$$

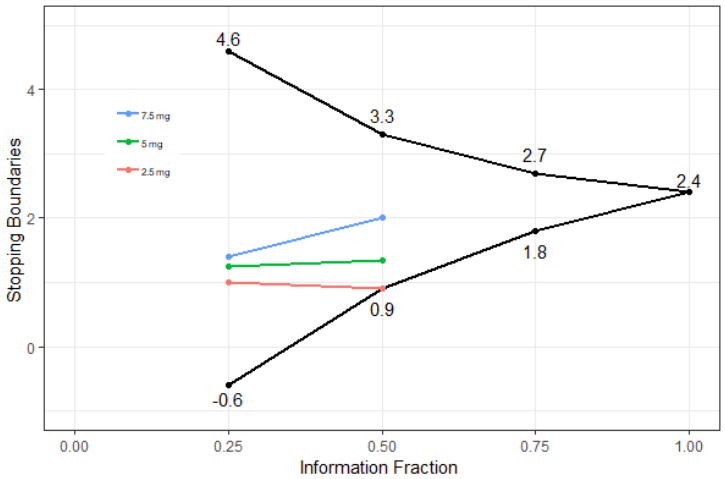
  all with valid level α test.
- To reject $H_0^{(M)}$ with strong control over FWER we must reject

$$H_0^{(M)}, H_0^{(H,M)}, H_0^{(M,L)}, H_0^{(H,M,L)}$$

  all with valid level α test.

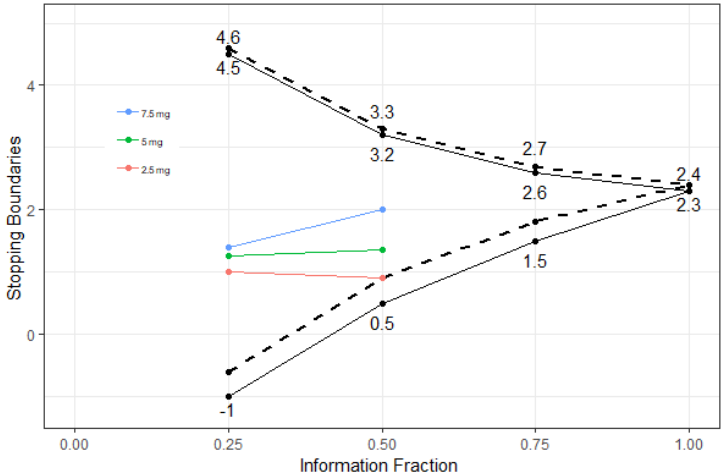
Example : A Valid Level-α Test of $H_0^{(H,M)}$

Original boundaries for 4-arm design.



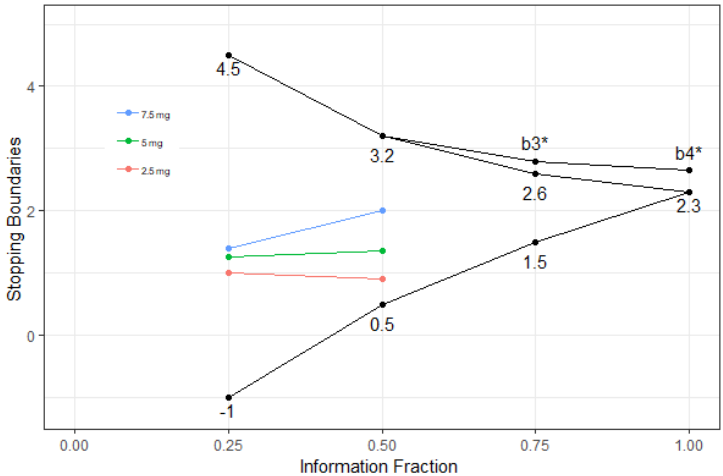
Example : A Valid Level-α Test of $H_0^{(H,M)}$ contd.

Step 1 Create a 3-arm design for (5 mg(M), 7.5 mg(H) vs control



Example : A Valid Level-α Test of $H_0^{(H,M)}$ contd.

Step 2 Recompute (b_3^*, b_4^*) to preserve conditional error rate



2. StageWise MAMS Approach

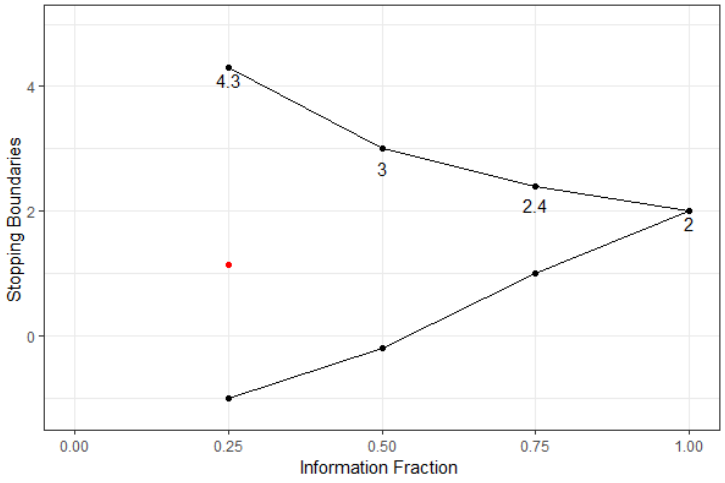
Consider again testing of $H_0^{(H,M)}$

- Let $p_j^{(H,M)}$ be the multiplicity adjusted p-value for testing $H_0^{(H,M)}$ based on the incremental data between stage $j-1$ and j.
- For example, adjustment could be by Bonferroni, Simes or Dunnett
- Combine the adjusted p-values stage by stage with inverse normal combination function

$$Z_j^{(H,M)} = \lambda_1 \Phi^{-1}(1-p_1^{H,M}) + \lambda_2 \Phi^{-1}(1-p_2^{H,M}) + \cdots + \lambda_j \Phi^{-1}(1-p_j^{H,M})$$

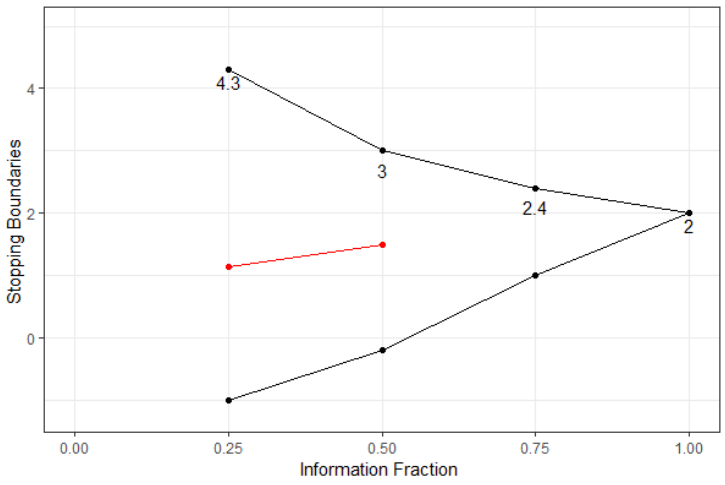
- Reject $H_0^{(H,M)}$ if $Z_j^{(H,M)}$ crosses the usual two-sample efficacy boundary.

Level-α Test of $H_0^{(H,M)}$: Look 1



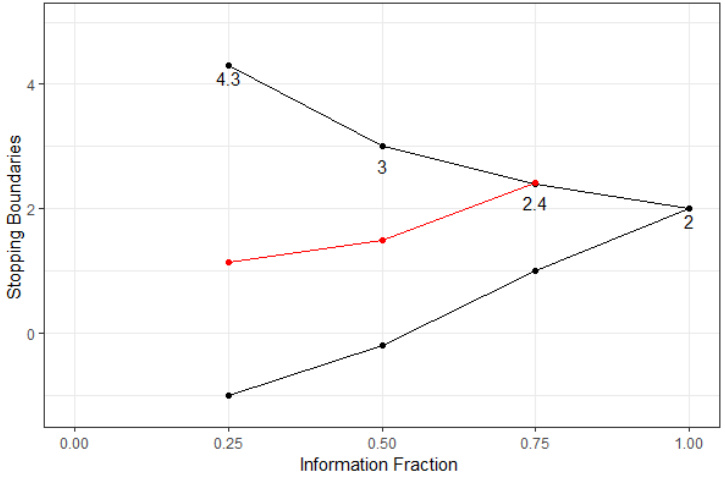
Reject if $Z_1^{(H,M)} = \Phi^{-1}(1 - p_1^{(H,M)}) > 4.3$

Level-α Test of $H_0^{(H,M)}$: Look 2



Reject if $Z_2^{(H,M)} = \lambda_1 \Phi^{-1}(1 - p_1^{(H,M)}) + \lambda_2 \Phi^{-1}(1 - p_2^{(H,M)})) > 3.0$

Level-α Test of $H_0^{(H,M)}$: Look 3



Reject $Z_2^{(H,M)} = \lambda_1 \Phi^{-1}(1 - p_1^{H,M}) + \lambda_2 \Phi^{-1}(1 - p_2^{H,M}) + \lambda_3 \Phi^{-1}(1 - p_3^{H,M}) > 2.4$

Power Comparison I : Analytical

- Two-stage design
- Two active doses versus control
  - $\delta_1 = \{0, 0.2, 0.4\}, \delta_2 \in [0, 0.4]$
  - Between-patient $\sigma^2 = 1$.
  - Sample size 200 on each arm, (Total N = 600)
- No early stopping and no sample size adaptation
- Exact analytical comparison of Cumulative vs Stage Wise

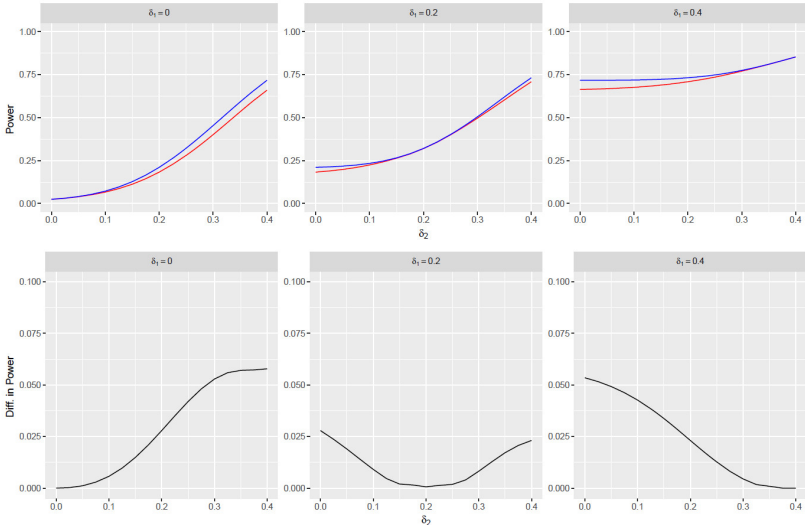


Figure : Power comparisons between CUMULATIVE and STAGE-WISE

Power Comparison II : Simulation

- Two-stage design
- Three active doses versus control
  - $\delta_1 \in [0, 0.3]$
  - $\delta_2 = \in [0, 0.3]$
  - $\delta_3 = 0.3$
  - Between-patient $\sigma^2 = 1$.
  - Sample size 200 on each arm, (Total N = 800)
- No early stopping, but drop dose i if $\hat{\delta}_i \leq -0.1$.
- Increase sample size in selected arms to maintain total sample size fixed.
- 10,000 simulations at every (δ_1 X δ_2 X δ_3) combinations.

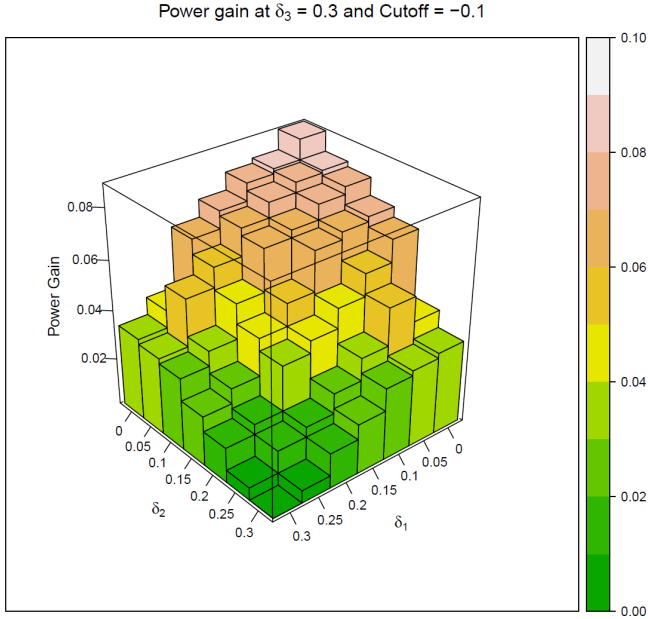


Figure : Power comparisons between CUMULATIVE and STAGE-WISE

Summary of Comparisons

- **Cumulative MAMS dominates over Stage Wise MAMS**
- **Greater heterogeneity increases dominance of Cumulative MAMS**
- **Very large power gains for CUMUL over STAGE are possible**
  - **CUMUL gains between 7% and 18% power over STAGE with Bonferroni adjusted p-values**
  - **CUMUL gains between 1% and 15% power relative to STAGE with Simes adjusted p-values**
  - **CUMUL gains upto 10% relative to STAGEWISE with Dunnett adjusted p-values**

Conclusion

- Platform trials can evaluate a larger number of new agents in a shorter duration and with fewer patients.
- managing a multi-center clinical trial is complex and the administrative machinery needed to set up .
- Platform trials are more effective for phase 3 testing than for phase 2
  - Many times Phase 2 endpoints are short-term (PFS, DFS) and not accepted for registration
  - Thus successful arms in phase 2 have to be tested all over again in phase 3 where they may fail.
  - In contrast a platform trial can quickly identify and drop inactive arms based on short-term endpoints (PFS, DFS), while following active arms for long-term endpoints (OS).
  - Robust statistical methodology exists for strong control of FWER.