Adpative MAMS Design

Lingyun Liu

27 April 2019

Outline

Outline

Outline

Outline

Outline

Outline

Challenges with Drug Development

- Drug development is a lengthy, complex, and costly process
- Entrenched with a high degree of uncertainty that a drug will actually succeed
- DiMasi JA, Grabowski HG, Hansen RA (2016)
  - Innovation in the pharmaceutical industry: new estimates of R&D costs
  - Developing a new prescription medicine that gains marketing approval is estimated to cost $2.6 billion
  - Rate of success from phase I to approval is only 12%
- More efficient approaches to drug development process

Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics
Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (CDER) Lee Pai-Scherf at 301-796-3400 or (CBER) the Office of Communication, Outreach, and Development at 800-835-4709 or 240-402-8010.

Adaptive Designs for Clinical Trials of Drugs and Biologics
Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Scott N. Goldie at 301-796-2055, or (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010.
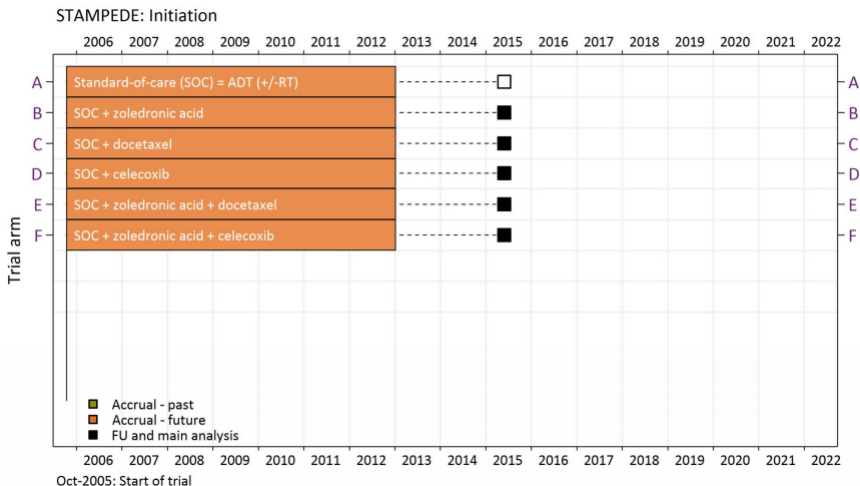
Multi-arm Multi-stage (MAMS) Trial Design

- Multi-arm—several treatments/doses are simultaneously assessed against a common control group within a single randomised trial

- Multi-stage—patient recruitment is discontinued to research arms that are not showing sufficient activity based on a series of pre-planned interim analyses

STAMPEDE – Multi-arm Multi-stage (MAMS) Design

- Ongoing multi-arm multi-stage design trial for men with locally advanced or metastatic prostate cancer
- Early stopping of ineffective arms
- Adding various experimental arms as knowledge increases facilitates rapid study of new therapeutic strategies
- Target 25% relative improvement in overall survival HR=0.75
- Interim analysis 3 lack-of-benefit analyses
- Requires ˜400 control arm deaths
- Power: 90%
- One-sided α: 0.025

STAMPEDE – Multi-arm Multi-stage (MAMS) Design

STAMPEDE: Initiation

Trial arm

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |

A — Standard-of-care (SOC) = ADT (+/-RT)
B — SOC + zoledronic acid
C — SOC + docetaxel
D — SOC + celecoxib
E — SOC + zoledronic acid + docetaxel
F — SOC + zoledronic acid + celecoxib

- Accrual - past
- Accrual - future
- FU and main analysis

Oct-2005: Start of trial

Two Approaches for Preserving Type I Error

Design and Monitoring of Multi-Arm Multi-Stage Clinical Trials

Pranab Ghosh,[1,2] Lingyun Liu,[1] P. Senchaudhuri,[1] Ping Gao,[3] and Cyrus Mehta[1,4,*]

[1]Cytel Inc., Cambridge, Massachusetts, U.S.A.
[2]Boston University, Boston, Massachusetts, U.S.A.
[3]The Medicines Company, Parsippany, New Jersey, U.S.A.
[4]Harvard T.H. Chan School of Public Health, Boston, Massachusetts, U.S.A.
*email: mehta@cytel.com.

SUMMARY. Two-arm group sequential designs have been widely used for over 40 years, especially for studies with mortality endpoints. The natural generalization of such designs to trials with multiple treatment arms and a common control (MAMS designs) has, however, been implemented rarely. While the statistical methodology for this extension is clear, the main

Group Sequential Approach

Testing and estimation in flexible group sequential designs with adaptive treatment selection

Martin Posch[1,*,†], Franz Koenig[1], Michael Branson[2], Werner Brannath[1], Cornelia Dunger-Baldauf[2] and Peter Bauer[1]

[1]Section of Medical Statistics, Core Unit for Medical Statistics and Informatics,
Medical University of Vienna, Spitalgasse 23, A-1090 Wien, Austria
[2]Novartis Pharma AG, Biostatistics and Statistical Reporting, Lichtstrasse 35, CH-4002 Basel, Switzerland

P-value Combo Approach

Outline

Mathematical Framework

- K-look group sequential design to compare one active treatment arm to control

$$\sum_{i=1}^{K} P_0 \left(\cap_{j=1}^{i-1} [W_j < e_j] \text{ and } [W_j \geq e_i] \right) = \alpha$$

- Dunnett's test:

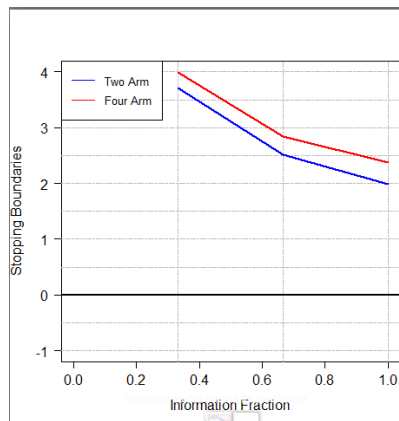$$P_0 \left(\max \{ W_1 \ldots W_D \} \geq e \right) = \alpha$$

- K-look MAMS design
  - Generalization of two-arm group sequential design to multiple arms (D comparisons to common control made K times)
  - Generalization of Dunnett's test to multiple looks

$$\sum_{i=1}^{K} P_0 \left(\sum_{j=1}^{i-1} [\max \{ W_{j1} \ldots W_{jD} \} < e_j] \text{ and } [\max \{ W_{i1} \ldots W_{iD} \}] \geq e_i \right) = \alpha$$

Higher Hurdle with 4-arm Trial

Look	Info Fraction	Two Arm	Four Arm
1	0.333	3.704	3.976
2	0.667	2.514	2.856
3	1.0	1.992	2.391

Possible Adaptations

- Trial can be stopped for efficacy if any arm cross the efficacy boundary
- Permit dropping the ineffective treatment arms
- Permit sample size re-estimation

Outline

Methodology Framework

- Consider two-stage design with one interim analysis to select the best arm
- Suppose s is the selected arm at Stage 1
- Then the Wald statistic for the final analysis can be written as

$$Z_s = \sqrt{\frac{n^{(1)}}{n^{(1)} + n^{(2)}}} Z_s^{(1)} + \sqrt{\frac{n^{(2)}}{n^{(1)} + n^{(2)}}} Z_s^{(2)}$$

- $Z_s^{(1)}$ is the maximum of multiple Wald statistics
- Thus Z_s is not $N(0,1)$ under H_0 and α is not preserved

$$P_0 \left(Z_s > 1.96 \right) > 0.025$$

Null Hypotheses	Type of Incorrect Conclusion
$H^{(1,2,3)}$: $\theta_1 = \theta_2 = \theta_3 = 0$	The selected treatment is declared superior to placebo
$H^{(1,2)}$: $\theta_1 = \theta_2 = 0, \theta_3 > 0$	Treatment 1 or 2 is selected and is declared superior to placebo
$H^{(1,3)}$: $\theta_1 = \theta_3 = 0, \theta_2 > 0$	Treatment 1 or 3 is selected and is declared superior to placebo
$H^{(2,3)}$: $\theta_2 = \theta_3 = 0, \theta_1 > 0$	Treatment 2 or 3 is selected and is declared superior to placebo
$H^{(1)}$: $\theta_1 = 0, \theta_2 > 0, \theta_3 > 0$	Treatment 1 is selected and is declared superior to placebo
$H^{(2)}$: $\theta_2 = 0, \theta_1 > 0, \theta_3 > 0$	Treatment 2 is selected and is declared superior to placebo
$H^{(3)}$: $\theta_3 = 0, \theta_1 > 0, \theta_2 > 0$	Treatment 3 is selected and is declared superior to placebo

Strong control means that probability of making a false claim is less than α no matter which of the above null hypotheses is applicable

Closed Testing and Combination Test

- Suppose Arm 3 is selected
- Claim sinificance on Arm 3 if reject $H^{(123)}, H^{(13)}, H^{(23)}$ and $H^{(3)}$ at their respective local $\alpha = 0.025$ levels

Reject $H^{(3)}$ if $C(p_3, q_3) = 1 - \Phi\left[w_1\Phi^{-1}(1 - p_3) + w_2\Phi^{-1}(1 - q_3)\right] < 0.025$

Reject $H^{(13)}$ if $C(p_{13}, q_{13}) = 1 - \Phi\left[w_1\Phi^{-1}\left(1 - p^{(13)}\right) + w_2\Phi^{-1}\left(1 - q^{(13)}\right)\right] < 0.025$

Reject $H^{(23)}$ if $C(p_{23}, q_{23}) = 1 - \Phi\left[w_1\Phi^{-1}\left(1 - p^{(23)}\right) + w_2\Phi^{-1}\left(1 - q^{(23)}\right)\right] < 0.025$

Reject $H^{(123)}$ if $C(p_{123}, q_{123}) = 1 - \Phi\left[w_1\Phi^{-1}\left(1 - p^{(123)}\right) + w_2\Phi^{-1}\left(1 - q^{(123)}\right)\right] < 0.025$
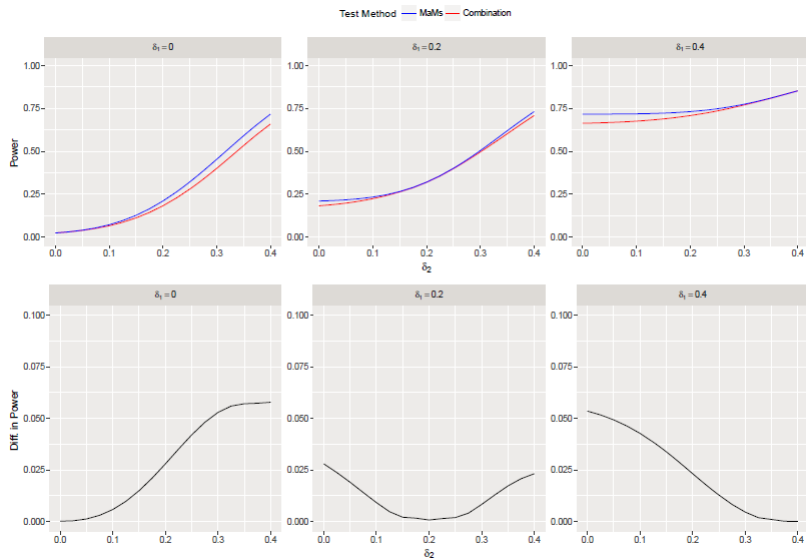
- Could use Simes test to compute the adjusted p-values $p^{(13)}$, $p^{(23)}$ and $p^{(123)}$

Outline

Power Comparison with Three Arms

- Three-arm trial with normally distributed data:
  - $\delta_1 \in [0, 0.4], \delta_2 \in [0, 0.4], \sigma^2 = 1$
- No early stopping and no sample size adaptation
- Exact analytical comparison of group sequential approach vs P-value Combo

Power Comparison with Three Arms

Power Comparison with Three Arms

- Two treatments were compared to a common control
- When $\delta_1 = \delta_2$, the two methods have the same power
- As the $\delta's$ differ, the power gain for GS approach increases
- When $\delta_i = 0$ and $\delta_j = 0.4$, GS approach has 5% more global power than P-val Combo

Power Comparison with Four Arms

- Four-arm trial with normally distributed data –
    - $\delta_1 \in \{0, 0.05, ...0.3\}$
    - $\delta_2 \in \{0, 0.05, ...0.3\}$
    - $\delta_3 = 0.3$
    - $\sigma^2 = 1$
- Dose selection at the end of Stage 1
- Select every dose i for which $\hat{\delta}_{i1} > -0.1$
- Re-allocate available sample size to remaining arms
- No early stopping at Stage 1
- 10,000 simulations at every$(\delta_1 \times \delta_2 \times \delta_3)$ combination

Power Comparison with Four Arms

Power gain at $\delta_3 = 0.3$ and Cutoff = -0.1
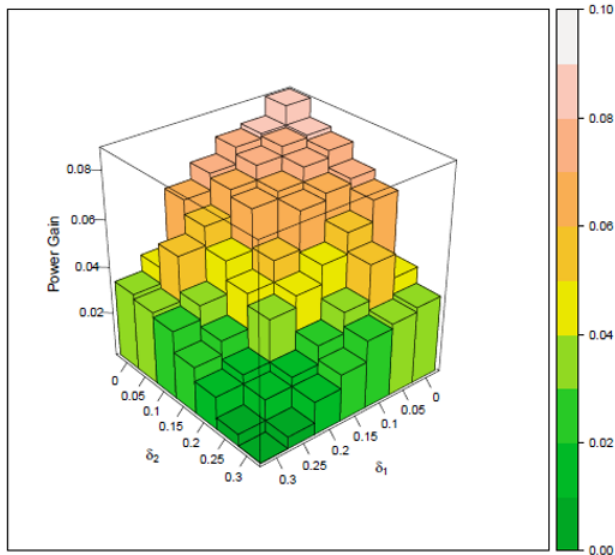
Summary of Comparison

- Three treatments were compared to common control
- Dose selection and sample size re-assessment at Stage 1
- As before, GS approach dominates over P-value Combo
- Power gains increase with increasing heterogeneity of δ
- Up to 12% power gain observed

Power Advantage with Group Sequential Approach

- Group sequential approach requires less closed testing
  - There are two possibilities at the end of Stage 1
    - All doses are selected for Stage 2
    - At least one dose is dropped at Stage 1
  - If all doses are selected, GS approach does not require closed testing but P-value Combo does
- Statistics used by group sequential approach satisfies the sufficiency principle
  - Group sequential test is of the form $\max\left\{\vec{W}_2\right\} \geq b_2$, where $\vec{W}_2$ is based on cumulative data
  - P-value Combo test is of the form $w_1 Z_{p(1)} + w_2 Z_{p(2)} \geq b_2$ where
    $$p(i) = P_0\left(\max\left\{\vec{W}_i \geq \vec{w}_i\right\}\right)$$

Outline

Conclusions and Discussions

- Both are well established methodologies for preserving type I error
- Group sequential approach
  - Boundaries are constructed under global null hypothesis from distribution of the maximum statistic
  - Strong control of type I error is nevertheless guaranteed
  - Natural extension of two arm group sequential trial
  - Exploits the correlation between treatment arms for added efficiency
  - Hypothesis test based on sufficient statistics
  - Straightforward to communicate to clinicians
- P-value combination approach
  - Uses closed testing to preserve type I error
  - Combines p-values from two stages with pre-specified weights
  - Does not utilize correlation between p-values (except Dunnett test)
  - Less transparent to clinicians
  - Slight loss of efficiency

Conclusions and Discussions

- In the context of master protocol
  - Should we control the FWER?
  - Should we control PWER?
  - In what situations, should control FWER vs PWER?
- More complex in the survival setting
  - Can short term readouts (e.g. ORR) be utilized at interim analysis for dose selection?
  - How to monitor such trials in the survival setting?
  - For those patients who are randomized to the arms which are dropped after interim, can they switch to other treatments?