Adaptive Enrichment Designs for Confirmatory Randomized Trials: Statistical Methods and Software Tools

Michael Rosenblum and Joshua Betz

Johns Hopkins Bloomberg School of Public Health

June 13, 2017

Course Outline

1. Overview of Adaptive Enrichment Designs
2. Optimizing Adaptive Enrichment Designs: Group Sequential Approach
3. Optimizing 2-Stage Adaptive Enrichment Designs using Sparse Linear Programming
4. Optimizing 2-Stage Adaptive Enrichment Designs: 2 Treatments versus Control
5. Existing Software Tools
6. Our New Software Tool (beta version)

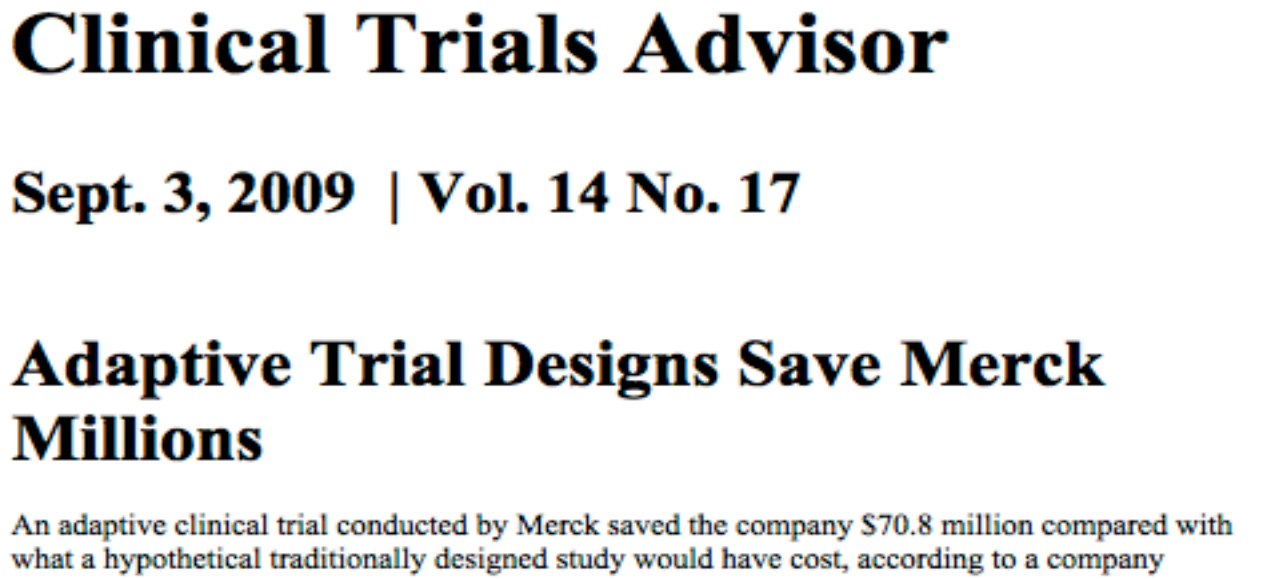
Adaptive Clinical Trial Designs

FDA is Interested:

"A large effort has been under way at FDA during the past several years to encourage the development and use of new trial designs, including enrichment designs."

Adaptive Clinical Trial Designs

Pharmaceutical Companies are Interested:



Clinical Trials Advisor

Sept. 3, 2009 | Vol. 14 No. 17

Adaptive Trial Designs Save Merck Millions

An adaptive clinical trial conducted by Merck saved the company $70.8 million compared with what a hypothetical traditionally designed study would have cost, according to a company

"An adaptive clinical trial conducted by Merck saved the company $70.8 million compared with what a hypothetical traditionally designed study would have cost…"

Why Consider Adaptive Designs?

Potential Benefits:

- Can give More Power to Confirm Effective Treatments/Interventions and Determine Subpopulations who Benefit Most
- Can Reduce Cost, Duration, and Number of Participants
- Caution! adaptive design not always better

Challenge: find the best design tailored to clinical investigator's research question and resource constraints

Adaptive Designs

- Participants Enrolled over Time
- At Interim Analyses, Can Change Sampling in Response to Accrued Data:
  - Adaptive designs could involve changes to:
    - Sample size
    - Enrollment criteria ("enrichment"—my focus)
    - Length of follow-up
    - Randomization probabilities
    - Dose
- SMART designs: If participant fails on initial treatment, randomized to another.

Stroke Trial Application

New Surgical Technique to Treat Intracerebral Hemorrhage (MISTIE, PI: Daniel Hanley)

Subpopulations: intraventricular hemorrhage (IVH) < 10ml vs. not.

Projected proportions: 0.33, 0.67.

Primary outcome: 180 day modified Rankin Scale < 4.

Clinically meaningful, minimum treatment effect: 12% risk difference.

Data set used: MISTIE phase 2 trial data.

Alzheimer's Disease Application

Treatment to reduce progression from mild cognitive impairment to Alzheimer's disease.

Subpopulations: APOE4 carrier or not.

Primary outcome: 2 year change score in Clinical Dementia Rating Sum of Boxes

Clinically meaningful, minimum treatment effect: 30% reduction in mean change score

Data set used: Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort study

General Problem

Two predefined subpopulations that partition overall pop.

Δ_1 = Mean treatment effect for subpopulation 1

Δ_2 = Mean treatment effect for subpopulation 2

Δ_0 = Mean treatment effect for combined population

Goal: construct adaptive enrichment design to test

$$H_{01} : \Delta_1 \leq 0; \quad H_{02} : \Delta_2 \leq 0; \quad H_{00} : \Delta_0 \leq 0$$

that strongly controls familywise Type I error rate,

provides power guarantees, and optimizes expected sample size and/or duration.

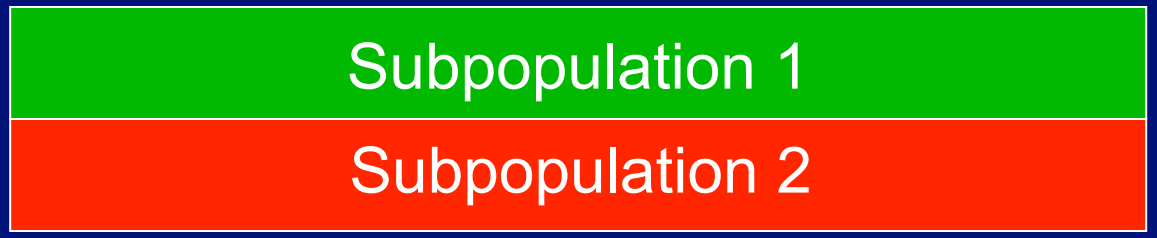
Example of Power and Type I Error Constraints

Power and Type I Error Constraints:

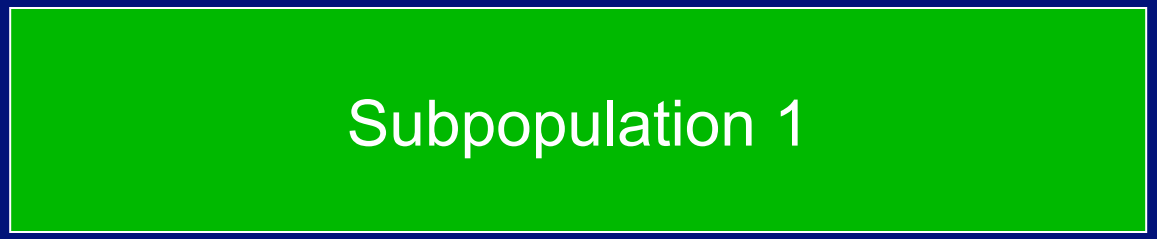
1. If clinically meaningful, minimum effect in both subpopulations, 80% power to reject combined pop. null H_{00}.
2. If clinically meaningful, minimum effect in single subpop., 80% power to reject that null hyp.
3. Strong control of familywise Type I error rate 0.025 (one-sided).

Goal: minimize expected sample size, averaged over scenarios in (1), (2), and global null.

Standard (non-adaptive) Design 1



Standard (non-adaptive) Design 2



2 Stage Adaptive Enrichment Design

Flow of Enrollment and Decision

Stage 1

Enroll Both Subpopulations

- Subpopulation 1
- Subpopulation 2

Decision

- Option 1 → Enroll Both Pop.
  - Subpopulation 1
  - Subpopulation 2
- Option 2 → Enroll Only Subpop.1
  - Subpopulation 1
- Option 3 → Enroll Only Subpop.2
  - Subpopulation 2
- Option 4 → 

Trial

Adaptive Enrichment Design: Group Sequential, Enrollment Modification Rule

- At each analysis k, compute cumulative statistics (e.g., z-statistics) $Z_{0,k}$, $Z_{1,k}$, $Z_{2,k}$ for combined pop., subpop. 1, and subpop. 2, respectively.

- Decision rule based on these statistics to:
  stop entire trial, stop single subpopulation accrual but continue other, continue both. (Cannot restart accrual once stopped.)

- No other adaptive features (e.g., randomization ratio fixed)

Multiple Testing Procedure

$$H_{01} : \Delta_1 \leq 0; \quad H_{02} : \Delta_2 \leq 0; \quad H_{00} : \Delta_0 \leq 0$$

At each analysis k:

1. (Test efficacy) For each population $s \in \{0,1,2\}$, if $Z_{s,k} > u_{s,k,}$ reject H_{0s}.
   Also, if both H_{01} and H_{02} are rejected, reject H_{00}.

2. (Modify Enrollment) Stop enrollment of subpopulation $s \in \{1,2\}$, if any of following occur:
   H_{0s} was rejected, $Z_{s,k} < l_{s,k}$, or $Z_{0,k} < l_{0,k}$.

Boundaries $u_{s,k}$, $l_{s,k}$ set by error-spending functions (Maurer and Bretz, 2013; Rosenblum et al. 2016a,b).

Trial Design Optimization Problem

- Many design parameters to set: number of stages, per-stage sample sizes, efficacy and futility boundaries for each (stage, population) pair

- We developed software tool to automatically optimize over design parameters; goal is to minimize expected sample size under power and Type I error constraints.

  -Algorithm: Simulated Annealing.

  -User-friendly graphical user-interface.

  -Outputs reports comparing optimal designs

Design Optimizer

Design Options

Main Options

Type of Outcome Data: Continuous

Subpopulation 1 proportion	0.33	
Familywise Type I error	0.025	
Maximum total sample size	1600	
Enrollments per Year for Combined Population	420	
Length of Follow-up	0	years

Optimization Target: Minimize Expected — (•) Sample Size ( ) Duration

Advanced Options

Number of Stages	5	
Subpop. 1 Randomization Probability to Treatment Arm	0.5	
Subpop. 2 Randomization Probability to Treatment Arm	0.5	
Number of Different Designs to Search Over	10	
Computational Time Limit	200	minutes

Design Optimizer Outputs

1. Optimized adaptive and standard designs that satisfy all power and Type I error constraints

2. Performance comparisons in terms of: sample size, duration, power, Type I error.

3. Highlight key tradeoffs.

4. Plots of efficacy and futility boundaries

Example of Optimization: Stroke Trial Application

Search over 4 classes designs:

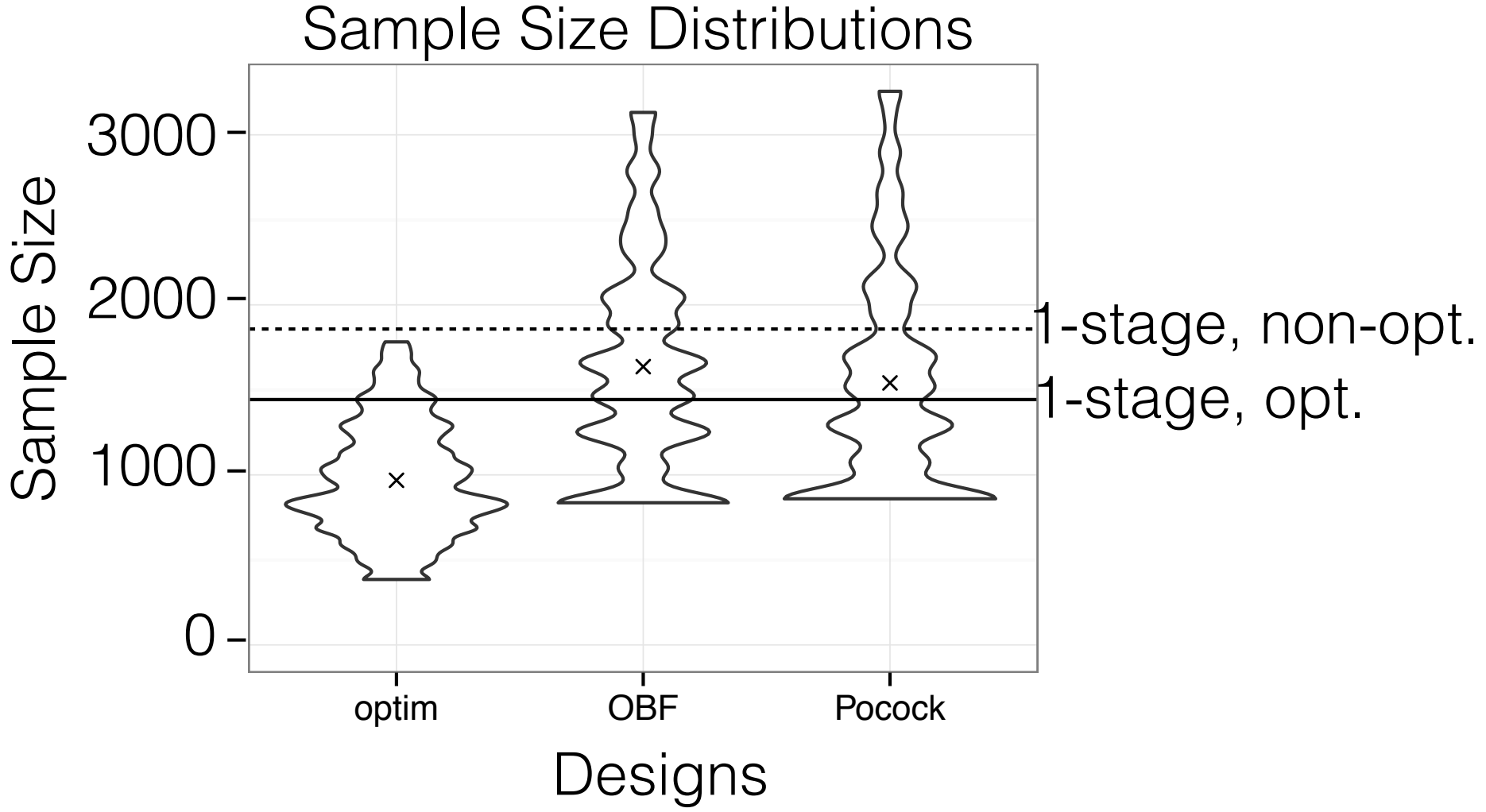
1. Separate error spending functions for efficacy and futility boundaries using power family, unequal per-stage sample sizes, up to 10 stages

2. O'Brien-Fleming boundaries, 5 stages, equal per-stage sample sizes

3. Pocock boundaries, 5 stages, equal per-stage sample sizes

4. Single stage designs

Comparison of Optimized Designs: Stroke Trial Application



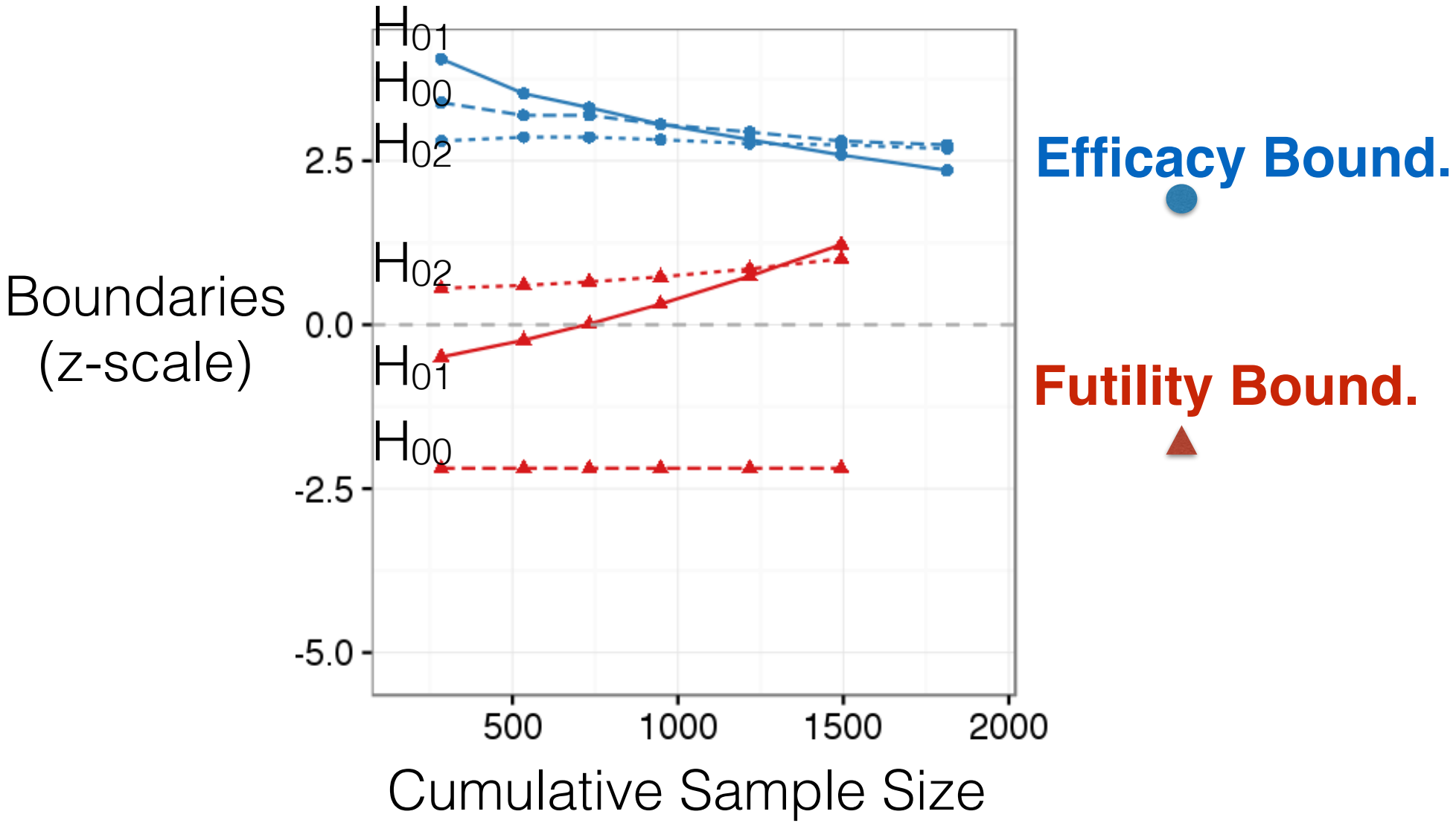
x=Expected Sample Size

Comparison of Optimized Designs: Stroke Trial Application

Performance Tradeoff Summary among Best Designs

	Optimized Adaptive Enrichment Design	Optimized 1-Stage
Expected Sample Size	968	1430
Maximum Sample Size	1787	1430

Optimized Adaptive Design Boundaries: Stroke Trial Application



Alzheimer's Disease Application

Treatment to reduce progression from mild cognitive impairment to Alzheimer's disease.

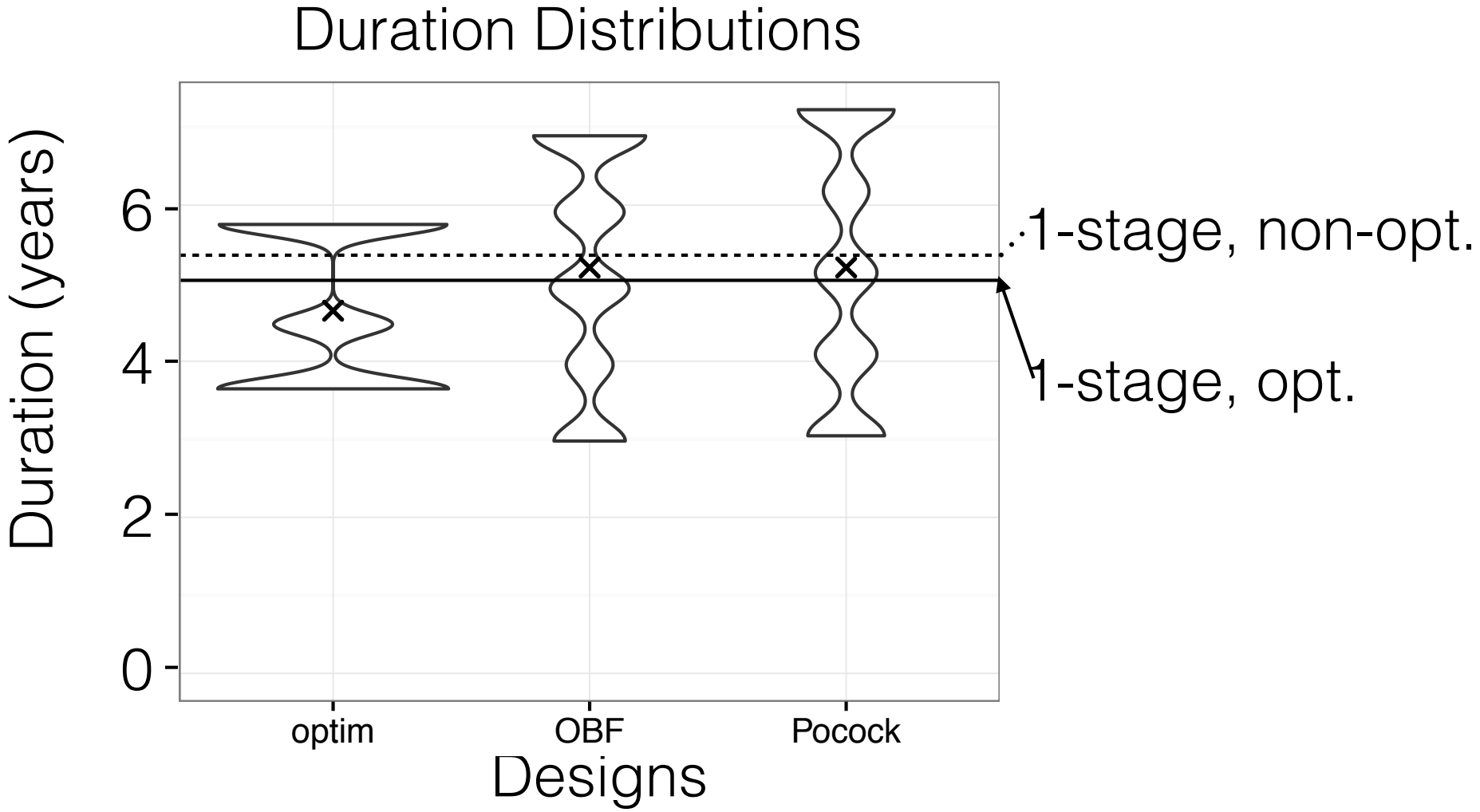
Subpopulations: APOE4 carrier or not.

Primary outcome: 2 year change score in Clinical Dementia Rating Sum of Boxes

Clinically meaningful, minimum treatment effect: 30% reduction in mean change score

Data set used: Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort study

Comparison of Optimized Designs



x=Expected Duration (years)

Performance Tradeoff Summary among Best Designs

	Optimized Adaptive Enrichment Design	Optimized 1-Stage
Expected Duration (Years)	4.65	5.02
Maximum Duration (Years)	5.75	5.02

Acknowledgments

Thank you!

This research was supported by the U.S. Food and Drug Administration (HHSF223201400113C) and the Patient-Centered Outcomes Research Institute (ME-1306-03198). This work is solely the responsibility of the authors and does not represent the views of these agencies.

We thank Michela Gallagher and thank ADNI for providing data.

References:

Fisher, A. and Rosenblum, M. (2016), Stochastic Optimization of Adaptive Enrichment Designs for Two Subpopulations. Johns Hopkins University, Dept. of Biostatistics Working Papers. http://goo.gl/wcQAxP

Rosenblum, M., Qian, T., Du, Y., and Qiu, H., Fisher, A. (2016a) Multiple Testing Procedures for Adaptive Enrichment Designs: Combining Group Sequential and Reallocation Approaches. Biostatistics. (17)4, 650-662. http://goo.gl/extFAl

Rosenblum, M., Thompson, R., Luber, B., Hanley, D. (2016b) Group Sequential Designs with Prospectively Planned Rules for Subpopulation Enrichment. Statistics in Medicine. 35(21), 3776-3791. http://goo.gl/7nHAVn

Maurer, W. and F. Bretz (2013). Multiple testing in group sequential trials using graphical approaches. Statistics in Biopharmaceutical Research 5 (4), 311-320.

Multiple testing procedures for adaptive enrichment designs: combining group sequential and reallocation approaches

Michael Rosenblum
Biostatistics Department
Johns Hopkins Bloomberg School of Public Health (JHBSPH)

Joint work with:
Tianchen Qian, Yu Du, Huitong Qiu, Aaron Fisher

From 2016 paper in *Biostatistics*: http://goo.gl/extFAI

June 13, 2017

Adaptive Enrichment Designs

- Involve preplanned rules for modifying enrollment criteria based on accrued data
- Multiple populations of interest, each with corresponding null hypothesis to test
- Challenge: construct group sequential multiple testing procedure with all of the following properties:
  1. Strong control of familywise Type I error rate (probability of rejecting one or more true null hypotheses)
  2. Leverages correlation among statistics over time and for overlapping populations
  3. Provides strictly greater power than several known methods
  4. Does not require knowing covariance matrix in advance

Related Work

- Methods that leverage covariance among statistics: Tang and Geller (1999), Stallard (2011), Magirr and others (2012), Magnusson and Turnbull (2013)
- Methods that lower rejection thresholds for the remaining null hypotheses after others have been rejected, by reallocating alpha across hypotheses (populations): Holm (1979), Bretz et al. (2009), Maurer and Bretz (2013).
  These don't leverage covariance among statistics.

We combine features from these two types of approaches.

Note: Bretz et al. (2011, Section 3.2) do this, but unlike our method require covariance of future statistics known in advance.

Interleaved Error Spending Functions

- Error spending function approach of Slud and Wei (1982) and Lan and DeMets (1983). Advantage: information accrual rates don't need to be known in advance.
- We use separate error-spending function for each composite population of interest.
- Tests for different populations are interleaved to take advantage of correlations among statistics for different but overlapping populations and statistics for the same population at different times.

Hypotheses and Statistics

- Null Hypotheses: $H_{0j} : \Delta_j \leq 0$, for each $j \in \{0, \ldots, J\}$. Global null hypothesis $H_0 : \Delta_j = 0$ for all $j \leq J$.
- A sequence of analyses $1, \ldots, K$ are preplanned, where analysis k takes place at the end of stage k.
- At analysis k, observe cumulative, Wald statistics $Z_{0,k}, Z_{1,k}, \ldots, Z_{J,k}$.
- Assume $EZ_{j,k} \leq 0$ under H_{0j} for all stages k.
- Covariance matrix of $Z_{j,k}$ fixed but unknown.
- Alpha increments $\alpha_{j,k}$ determined by error spending functions at each stage.

Simple Version Without Interleaved Error Spending Functions

Alpha increments $\alpha_{j,k} \geq 0$ and $\sum_{j \geq 0, k \geq 0} \alpha_{j,k} = 0.05$.

At stage k, consider each null hypothesis $H_{0j}, j = 0, 1, \ldots, J$ and reject H_{0j} if $Z_{j,k} > \mathbf{u_{j,k}}$.

Each efficacy boundary $\mathbf{u_{j,k}}$ is set to be solution to:

$$\alpha_{j,k} = P_{H_0} \left\{ Z_{j,k} > \mathbf{u_{j,k}}; Z_{j,k'} \leq u_{j,k'} \text{ for all } k' \leq k \right\}.$$

This uses covariances for same population across stages, but ignores covariance among populations.
Can improve by interleaving error-spending functions.

Interleaved error spending functions

Alpha increments $\alpha_{j,k} \geq 0$ and $\sum_{j \geq 0, k \geq 0} \alpha_{j,k} = 0.05$.

At stage k, consider each null hypothesis $H_{0j}, j = 0, 1, \ldots, J$ and reject H_{0j} if $Z_{j,k} > \mathbf{u_{j,k}}$.

Each efficacy boundary $\mathbf{u_{j,k}}$ is set to be solution to:

$$\alpha_{j,k} = P_{H_0}\left\{Z_{j,k} > \mathbf{u_{j,k}}; Z_{j',k'} \leq u_{j',k'} \text{ for all } (j', k') \text{ preceding } (j, k)\right\},$$

where (j', k') preceding (j, k) if $k' < k$ or if $(k' = k$ and $j' \leq j)$; and where $\alpha_{j,k} \geq 0$ and $\sum_{j \geq 0, k \geq 0} \alpha_{j,k} = \alpha$ (e.g., 0.05).

This leverages covariances, but does not use alpha reallocation.

Example of Efficacy Boundary Improvements

Efficacy Boundaries (z-scale) for 3 Stage, 2 Hypothesis Trial

Analysis (k)	1	2	3
	Without Interleaving ($\mathcal{M}^{MB}$)		
H_{00} boundaries $u_{0,k}$	2.57	2.32	2.10
H_{01} boundaries $u_{1,k}$	3.45	3.29	3.14
	With Interleaving ($\mathcal{M}^{NEW}$)		
H_{00} boundaries $u_{0,k}$	2.57	2.32	2.09
H_{01} boundaries $u_{1,k}$	3.16	2.95	2.74

$\mathcal{M}^{NEW}$: Our new multiple testing procedure.
$\mathcal{M}^{MB}$: Maurer and Bretz (2013) procedure.

Interleaved error spending functions with alpha reallocation

Closure principle: For each $F \subseteq \{0, \ldots, J\}$, define local test of intersection null hyp. $H_F = \cap_{j \in F} H_{0j}$ with level α. Reject elementary null H_{0j} if every H_F with $j \in F$ rejects.

Local test of H_F: reject if $Z_{j,k} > \mathbf{u}^{\mathbf{F}}_{\mathbf{j},\mathbf{k}}$ for any $j \in F, k \geq 0$, where at analysis k, for each $j \in F$, set $\mathbf{u}^{\mathbf{F}}_{\mathbf{j},\mathbf{k}}$ to be solution to:

$$c_j^F \alpha_{j,k} = P_{H_0}\left\{ Z_{j,k} > \mathbf{u}^{\mathbf{F}}_{\mathbf{j},\mathbf{k}}; \right.$$
$$\left. Z_{j',k'} \leq u^F_{j',k'} \text{ for all } (j',k') \text{ preceding } (j,k) \text{ and } j' \in F \right\},$$

where $c_j^F \geq 1$ and $\sum_{j \in F, k \geq 0} c_j^F \alpha_{j,k} = \alpha$.
Intuitively, c_j^F is alpha inflation factor, reallocating alpha from hypotheses $j \notin F$. It can be set, e.g., using graphical approach of Bretz et al. (2009), but not restricted to this.

Example of Efficacy Boundary Improvements

Efficacy Boundaries (z-scale) for 3 Stage, 2 Hypothesis Trial

Analysis (k)	1	2	3
	Without Interleaving ($\mathcal{M}^{MB}$)		
H_{00} boundaries $u_{0,k}$	2.57	2.32	2.10
H_{01} boundaries $u_{1,k}$	3.45	3.29	3.14
	With Interleaving ($\mathcal{M}^{NEW}$)		
H_{00} boundaries $u_{0,k}$	2.57	2.32	2.09
H_{01} boundaries $u_{1,k}$	3.16	2.95	2.74
	After Alpha Reallocation (both)		
H_{00} boundaries after reject H_{01}	2.55	2.30	2.07
H_{01} boundaries after reject H_{00}	2.55	2.30	2.07

$\mathcal{M}^{NEW}$: Our new multiple testing procedure.
$\mathcal{M}^{MB}$: Maurer and Bretz (2013) procedure.

Main Result

Our new multiple testing procedure: $\mathcal{M}^{NEW}$.
Maurer and Bretz (2013) procedure: $\mathcal{M}^{MB}$.

Theorem

- *For each null hypothesis H_{0j}, $\mathcal{M}^{NEW}$ rejects it by analysis k whenever $\mathcal{M}^{MB}$ does.*
- *If H_{0J} is false, $\mathcal{M}^{NEW}$ has strictly greater power than $\mathcal{M}^{MB}$ to reject H_{0J} by analysis k, under the condition that covariance matrix full rank and each $\alpha_{j,k} > 0$.*

Limitations

- Computation of multivariate normal distribution function restricts total number of stages and hypotheses. For confirmatory trials with 2-4 elementary null hypotheses and 3-5 analysis times, can use Genz et al. (2014).
- Advantages from leveraging covariance only useful if substantial overlap in populations, e.g., a subpopulation that makes up 2/3 of overall population.

References

Rosenblum, M., Qian, T., Du, Y., and Qiu, H., Fisher, A. (2016) Multiple Testing Procedures for Adaptive Enrichment Designs: Combining Group Sequential and Reallocation Approaches. *Biostatistics*. http://goo.gl/extFAI

Bretz, F., W. Maurer, W. Brannath, and M. Posch (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in medicine 28*(4), 586–604.

Bretz, F., M. Posch, E. Glimm, F. Klinglmueller, W. Maurer, and K. Rohmeyer (2011). Graphical approaches for multiple comparison procedures using weighted bonferroni, simes, or parametric tests. *Biometrical Journal 53*(6), 894–913.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn (2014). mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-0. URL http://CRAN.R-project.org/package=mvtnorm.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist. 6*, 65–70.