Optimizing Adaptive Enrichment Designs, and Challenges in Using Data to Construct Realistic Simulations to Evaluate Design Performance

Michael Rosenblum
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Joint work with Aaron Fisher
Working paper: https://goo.gl/JWscpV
October 24, 2016

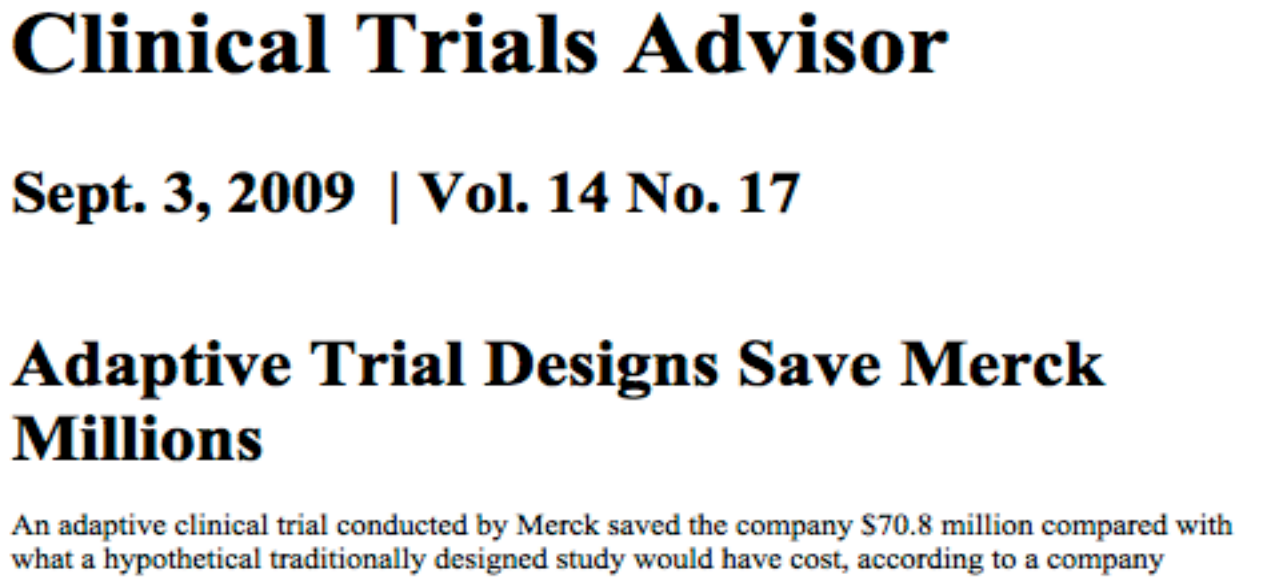
Adaptive Clinical Trial Designs

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.

Overview of My Research on New Adaptive Designs

PI on PCORI funded project: "Innovative Randomized Trial Designs to Generate Stronger Evidence about Subpopulation Benefits and Harms" Specific Aims:

1. Develop and evaluate new adaptive enrichment designs for time-to-event and other delayed outcomes.
2. Conduct extensive simulation studies.
3. Produce user-friendly, free, open-source software to find best design to answer a clinical investigator's research question.

PI on FDA funded project to demonstrate strengths and weaknesses of new adaptive trial designs in the following clinical applications:
stroke treatment (Dan Hanley), slowing progression of Alzheimer's disease (Michela Gallagher), cardiac resynchronization devices (Boston Scientific), and HIV prevention (Craig Hendrix)

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.

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:

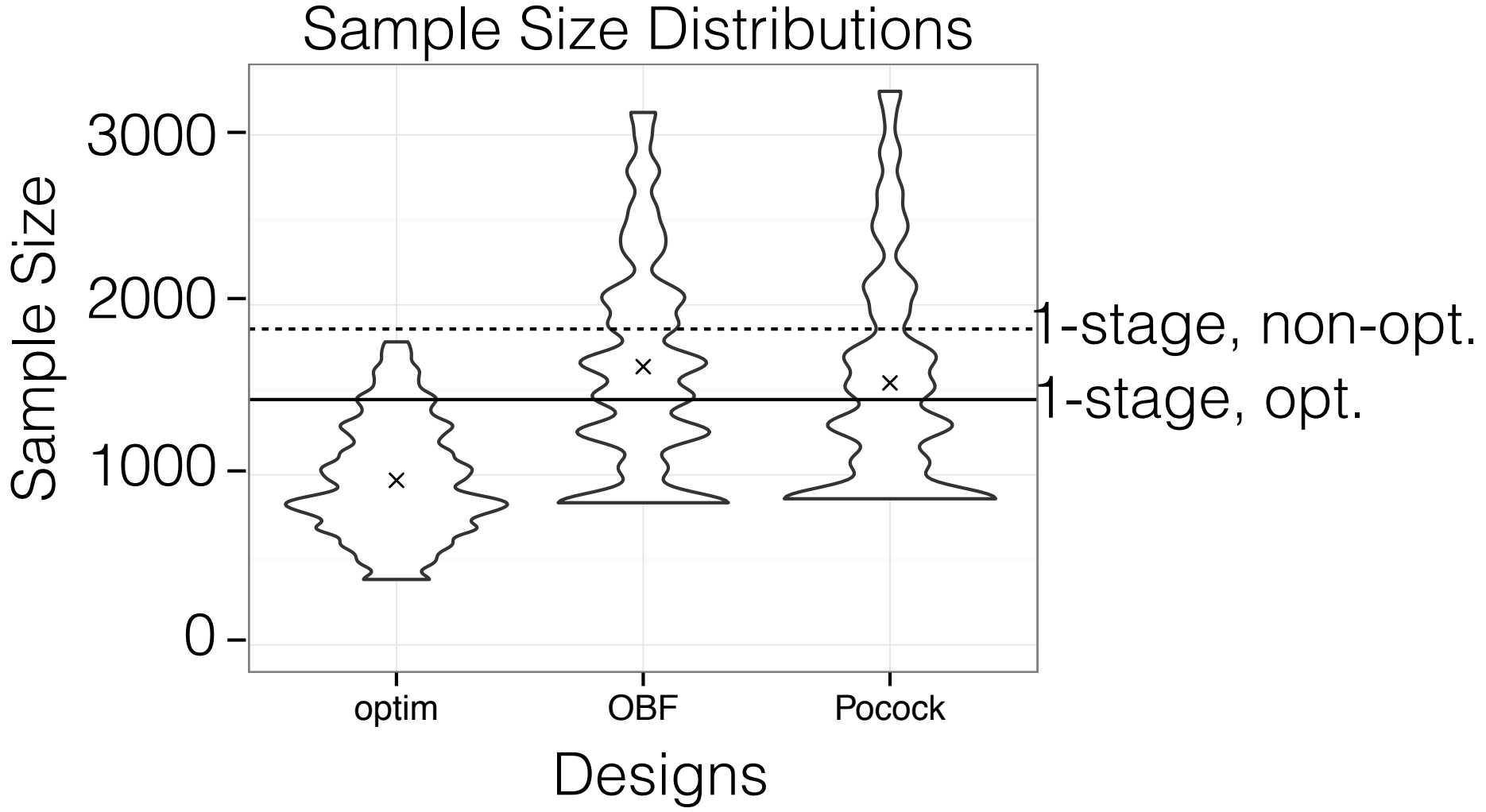
1. Separate error spending functions for efficacy and futility boundaries using power family, unequal per-stage sample sizes, up to 10 stages

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



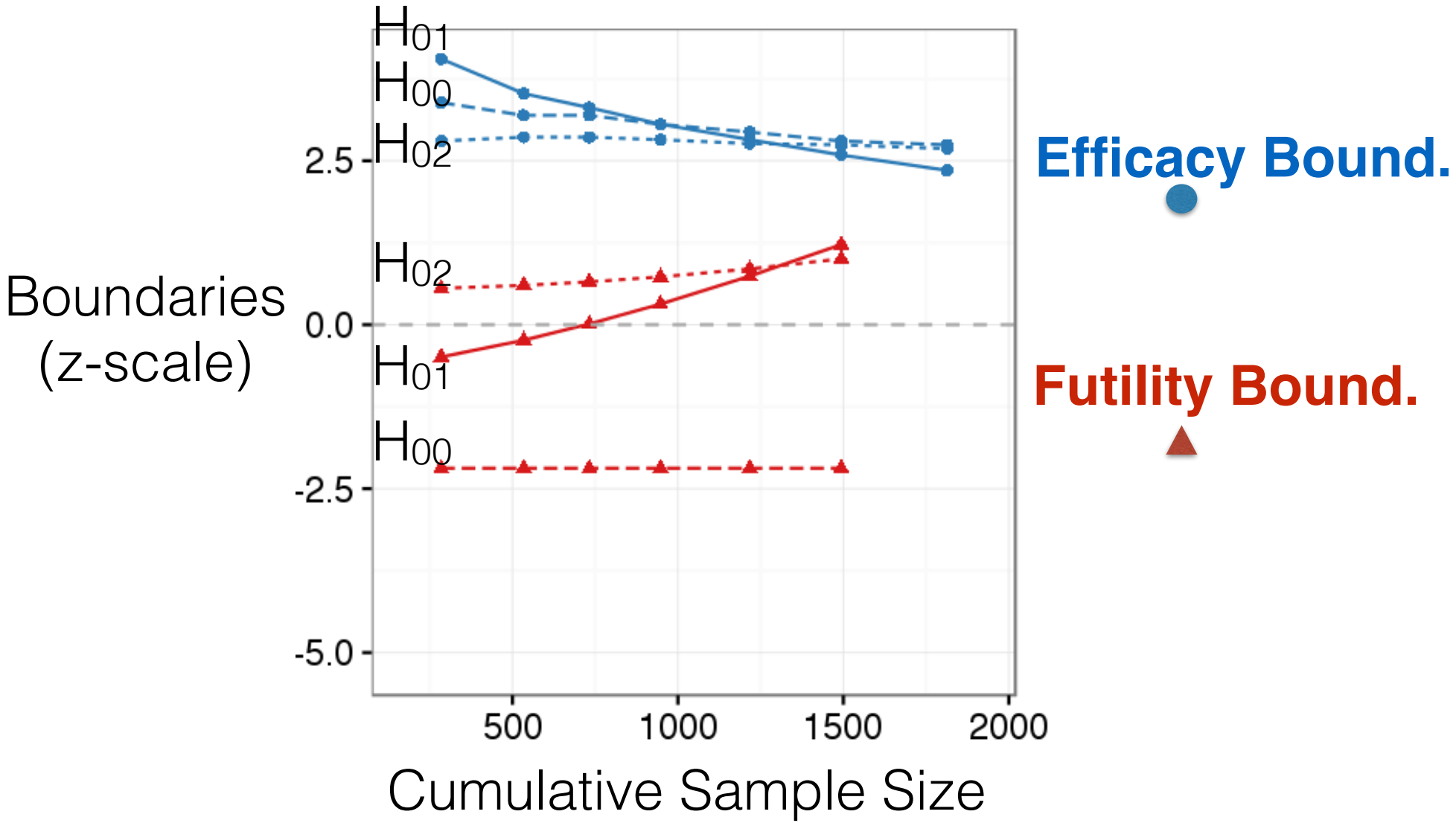
x=Expected Sample Size

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



Challenges in Constructing Realistic Simulations/Scenarios

How to use prior trial data, observational study data, and healthcare data sets to help define realistic scenarios where performance of new designs/estimators can be evaluated.

Want:

a. Mimic key features of real data sets (e.g., correlations between baseline variables and outcome, event rates, outcome variances)

b. Data generating distributions not built using same models that are used in estimators (since could lead to overly optimistic performance evaluation)

Resampling from Datasets

Example Goal: Evaluate designs for future phase 3 trial, by constructing simulation distributions that mimic key features of target populations.

Resample participant data vectors, e.g., **(Baseline variables, Treatment, Outcome)**, from data sets (e.g., ADNI study). Preserves within-patient correlations and outcome variances.

To simulate no treatment effect, can replace each Treatment by independent Bernoulli(1/2). To simulate non-zero effects, can shift outcome distribution in treatment arm.

Caution when extrapolating from phase 2 data to phase 3. Ideally, resample from much larger data set than study being planned. Need e.g., obs. study, healthcare data set, merged data from related RCT's.

Ideas and Open Research Problems

- If multiple data sources, should one mix the populations or do separate simulations (or both)?

- Should one extract summary population features only (e.g., correlations, event rates, dropout) and build data generating distributions using parametric models, or directly resample?

- Related work by Susan Gruber on Observational Medical Dataset Simulator (OSIM2): build simulations with causal structure using DAG.

- Objective third party to create simulation scenarios, to avoid cherry-picking data generating distributions.

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 ADNI and Daniel Hanley 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

Gruber, S. (2015) A Causal Perspective on OSIM2 Data Generation, with Implications for Simulation Study Design and Interpretation. Journal of Causal Inference 3(2): 177-187.

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