Practical Recommendations Document: Planning Adaptive Enrichment Designs for Confirmatory Randomized Clinical Trials

Michael Rosenblum, mrosen@jhu.edu
January 26, 2018

This document presents an overview of adaptive enrichment designs intended for clinical investigators and statisticians who are considering using these designs. Case studies are used to compare these adaptive designs to standard designs in simulations that mimic key features from 4 clinical applications:
Case Study 1: Superiority trial of surgical treatment for severe stroke.[1]
Case Study 2: Superiority trial of cardiac resynchronization device.[2]
Case Study 3: Non-inferiority trial of HIV treatments.[3]
Case Study 4: Superiority trial of drug for Alzheimer's disease prevention.[1]

Outline:

1. Overview of Adaptive Enrichment Designs for Confirmatory Trials:
    a. What is a confirmatory trial?
    b. What is an adaptive design?
    c. What is an adaptive enrichment design?
    d. Goals of adaptive enrichment designs.
    e. Three recently conducted/ongoing adaptive enrichment designs
2. Case Studies Overview
    a. Summary of Each Clinical Application
    b. Trial Design Requirements for Each Case Study
3. Key Requirements for Adaptive Enrichment Designs to be Useful:
    a. Need pre-defined subpopulations where effect of new treatment hypothesized to differ by subpopulation
    b. Subpopulation size can't be too small as fraction of total population
    c. Long delay times between patient enrollment and measurement of her/his primary outcome can limit the value added by adaptive enrichment
4. Key Tradeoffs: Adaptive Enrichment Designs Versus Standard Group Sequential Designs
    a. What is likely to be learned at end of trial
    b. Sample size and duration required (both average and maximum)
    c. Expected number of participants assigned to inferior/superior arm
5. Regulator Perspectives on Adaptive Enrichment Designs
    a. What does the U.S. Food and Drug Administration recommend?
    b. What does the European Medicines Agency recommend?
6. Software Tool for Tailoring Adaptive Enrichment Design to One's Scientific Goals and Logistical Constraints, and Compare to Standard Designs
    a. Overview: target audience
    b. Inputs from the user specific to their trial design problem
    c. What information is output by the software tool (automatically generated report comparing performance of different designs)

Acknowledgments: This work was funded by U.S. Food and Drug Administration (HHSF223201400113C) and the Patient-Centered Outcomes Research Institute (ME-1306-03198). The views expressed in this document are those of the author and not necessarily those of FDA, PCORI, or any other organization mentioned below.

1. Overview of Adaptive Enrichment Designs for Confirmatory Trials:

a. What is a confirmatory trial?

Our focus is on confirmatory randomized clinical trials, often referred to as adequate and well-controlled trials. These terms mean that the trial is designed to demonstrate efficacy and safety of an intervention (such as a drug, medical device, surgical procedure, etc.).

b. What is an adaptive design?

Adaptive designs involve preplanned rules for modifying the conduct of a trial based on accruing data. We use the definition of adaptive designs from the FDA Draft Guidance on Adaptive Designs for Drugs and Biologics[4], "an *adaptive design clinical study* is defined as a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study."

Examples of modifications that could be made in an adaptive trial include the following, excerpted from[4]:

• study eligibility criteria (either for subsequent study enrollment or for a subset selection of an analytic population)
• randomization procedure
• treatment regimens of the different study groups (e.g., dose level, schedule, duration)
• total sample size of the study (including early termination)
• concomitant treatments used
• planned schedule of patient evaluations for data collection (e.g., number of intermediate timepoints, timing of last patient observation and duration of patient study participation)
• primary endpoint (e.g., which of several types of outcome assessments, which timepoint of assessment, use of a unitary versus composite endpoint or the components included in a composite endpoint)
• selection and/or order of secondary endpoints
• analytic methods to evaluate the endpoints (e.g., covariates of final analysis, statistical methodology, Type I error control)

c. What is an adaptive enrichment design?

This document focuses on a single type of adaptation called "adaptive enrichment", defined by[5] to be a preplanned rule for modifying enrollment criteria based on accruing data in an ongoing trial. Adaptive enrichment may be considered when there is suggestive evidence (available in the planning phase of a trial) that a subpopulation may be more likely to benefit from treatment. The subpopulation could be defined by severity of disease, a risk score, or a biomarker measured at baseline (i.e., prior to randomization). Adaptive enrichment designs may be useful if the goals of the trial include not only

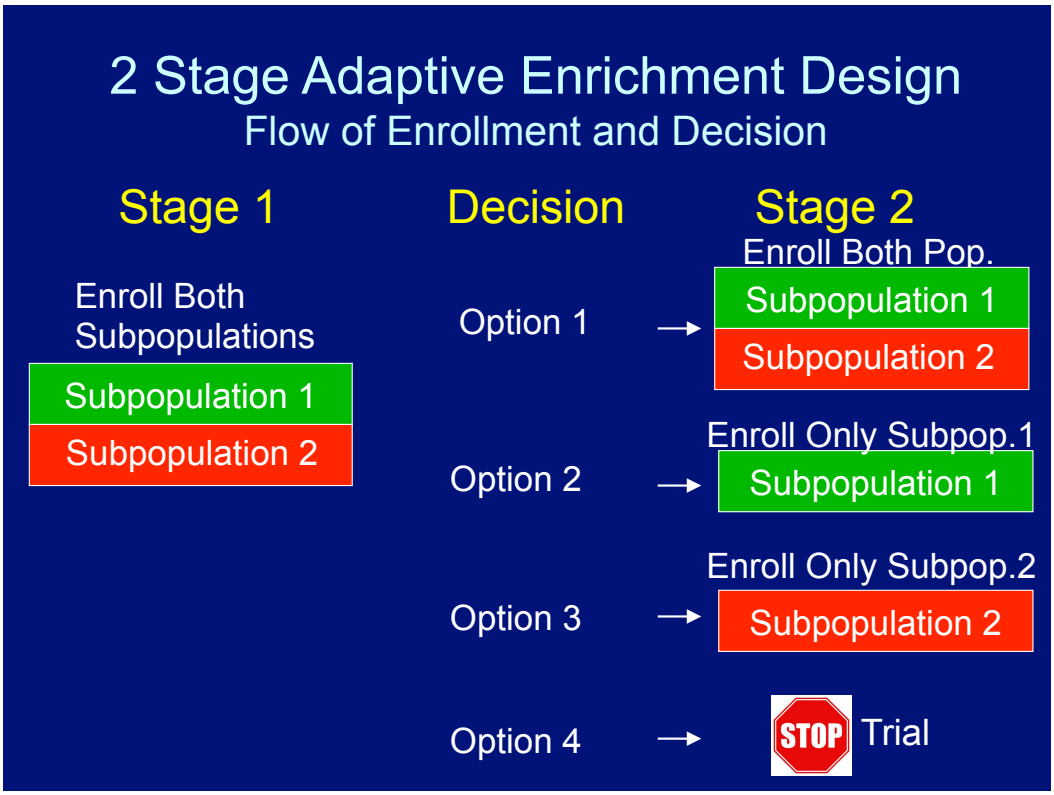
determining whether a treatment benefits the overall target population, but also to learn which subpopulations benefit.

Enrichment refers to selection of a subset of the overall population that is conjectured to be more likely to benefit (or may benefit more) from treatment. An enrichment design (FDA 2012) refers to a design that only enrolls such a subpopulation, from the very start. Adaptive enrichment refers to designs that start by enrolling from the overall target population, but may (based on a preplanned rule applied to accruing data) restrict enrollment and follow-up to a subpopulation.

A motivation for considering adaptive enrichment designs is that a standard design (which enrolls the overall population) may have low power to detect a treatment effect if only a subpopulation benefits. Adaptive enrichment designs can provide more power to detect overall population and subpopulation benefits[6]. However, there is typically a tradeoff in that greater sample size is required to do this, compared to a standard design that is only powered to detect a benefit in the overall population. Our trial optimization software tool presents this tradeoff by comparing performance of different designs, allowing the trial designer to decide which design is the best match to her/his research goals and constraints.

Figure 1 below presents a schematic of a 2 stage adaptive enrichment design that is similar to Figure 2 in the manuscript[7]. A decision is made after stage 1 to: (1) continue enrolling both subpopulations; (2) enroll only subpopulation 1; (3) enroll only subpopulation 2; (4) stop the trial. This type of design is considered in Case Studies 1-4, but involving multiple stages after which such enrollment decisions can be made. Once a subpopulation has accrual stopped, it cannot be restarted.

Figure 1:



d. Goals of adaptive enrichment designs.
Adaptive enrichment designs typically involve power requirements for detecting treatment benefits in each subpopulation (and possibly in the overall population) and Type I error requirements. Under these requirements, the goal is to minimize the trial cost in terms of sample size and duration. Just as for standard group sequential designs (i.e., designs for a single population but with multiple stages where the trial can be stopped early for efficacy or futility), there is a tradeoff between the average sample size and the maximum possible sample size.

e. Three recently conducted or ongoing adaptive enrichment designs
These are presented only to illustrate that such designs are implemented in practice; we do not discuss details of these particular designs and they are not related to our case studies.
i. The DAWN trial[8] compares mechanical thrombectomy versus standard medical care for treating acute stroke. It uses a Bayesian adaptive enrichment design involving interim analyses where the inclusion criteria could be modified to restrict enrollment to subpopulations with smaller infarct sizes at baseline.
ii. The DEFUSE 3 trial[9] involves, “A novel adaptive design will identify, at interim analyses, the group with the best prospect for showing benefit from endovascular treatment, based on baseline core lesion volumes and the times since stroke onset. Interim analyses will be conducted at 200 and 340 patients, at which time

the study may stop for efficacy/futility, or the inclusion criteria may be adjusted in the case of futility."

iii. Tappas:[10] "An adaptive enrichment phase 3 trial of TRC105 and pazopanib versus pazopanib alone in patients with advanced angiosarcoma." The trial involves an interim analysis where the population may be enriched (restricted) and sample size increased.

2. Case Studies Overview

a. Summary of Each Clinical Application

The four case studies involve simulations comparing adaptive versus standard designs that use key features from clinical applications in four disease areas. Table 1 gives key features of each clinical application and the data sources we drew from to create our simulation studies.

Table 1: Summary of Clinical Applications Used in Case Studies

Disease	**Treatments**	**Data Sources**	**Subpopulations of Interest**	**Primary Outcome**
Stroke	Surgery versus standard of care; rtPA.	MISTIE[11] and CLEAR[12] trials	Intraventricular hemorrhage (IVH) volume <10ml vs. ≥10ml (33%, 67%)*	180-day functional disability score (modified Rankin Scale)
Heart disease	Atrioventric-ular delay used in CRT device	SMART-AV trial[13]	QRS duration <150ms vs. ≥150ms (49%, 51%)*	Left ventricular end systolic volume at 180 days.
HIV	Antiretroviral combination therapies	PEARLS trial[14]	Women, Men (47%,53%)*	Time to virologic failure, AIDS defining illness, or death
Alzheimer's Disease	Drug for mild cognitive impairment	ADNI[15] cohort study	ApoE4 genotype carrier versus non-carrier (47%,53%)*	2 year Clinical Dementia Rating Scale (CDR)

*Percentages indicate proportions of study population in each subpopulation.

For each clinical application, it was conjectured by clinical investigators that a subpopulation (see column 4) may be more likely to benefit from treatment than the overall target population. This motivated the consideration of adaptive enrichment designs to help learn about treatment effects in subpopulations.

b. Trial Design Requirements for Each Case Study

The trial design requirements for each of the four clinical applications differed, but had some common features. All of the designs are required to control the study-wide Type I error rate (probability of at least one false positive result) at the desired level (e.g., 0.05), asymptotically. Each application involved two

subpopulations that partition the overall target population (column 4 of Table 1). Interest was in testing null hypotheses for these subpopulations (and sometimes for the combined population). We next describe the null hypotheses tested, the scenarios of interest, and the power requirements in each case study.

Case Studies 1 (stroke) and 4 (Alzheimer's disease) have the same structure.[1] A single treatment was compared to control in two subpopulations. The 3 null hypotheses are no average treatment benefit for subpopulation 1, for subpopulation 2, and for the combined population, respectively. There were 4 scenarios of interest: (i) only subpopulation 1 benefits, (ii) only subpopulation 2 benefits, (iii) both subpopulations benefit, and (iv) neither subpopulation benefits; benefiting means that the average treatment effect equals the minimum, clinically meaningful level, and not benefiting means the average treatment effect is zero. The power requirements were the following: 80% power to reject the null hypothesis for subpopulation 1 in scenario (i); 80% power to reject the null hypothesis for subpopulation 2 in scenario (ii); and, 80% power to reject the null hypothesis for the combined population in scenario (iii).

Case study 2 (heart disease) involves testing 2 treatments (A and B) versus control in two subpopulations. There are four null hypotheses, one for each treatment by subpopulation combination, each of the type: no average treatment benefit (compared to control) in the given subpopulation. There were six scenarios of interest, defined by the treatment effects in each treatment by subpopulation combination, which could be either 0 (no effect) or 15ml (the minimum, clinically meaningful treatment effect). The six scenarios are defined as follows: (i) neither subpopulation benefits from either treatment; (ii) treatment A benefits only subpopulation 1 and treatment B benefits no one; (iii) both treatments benefit subpopulation 1 but not subpopulation 2; (iv) treatment A benefits both subpopulations and treatment B benefits no one; (v) treatment A benefits both subpopulations and treatment B benefits only subpopulation 1; (vi) treatment A and treatment B benefit both subpopulations. The power requirement in each scenario is 80% power to reject each null hypothesis corresponding to a treatment benefit. For example, in scenario (iii), the requirement is 80% power to reject the null hypothesis for treatment A and subpopulation 1, and also for treatment B and subpopulation 1; in scenario (vi), the power requirement is 80% for each treatment by subpopulation combination.

Case study 3 (HIV treatment) involves testing whether treatment A is non-inferior to treatment B in two subpopulations (women and men), where the outcome is time to the first of virologic failure, AIDS defining illness, or death. The non-inferiority margin is a hazard ratio of 1.35 comparing treatment A to B. There are two null hypotheses (one per subpopulation) of the form: treatment A is inferior to treatment B (hazard ratio at least 1.35) in the given subpopulation. Four scenarios were considered: (i) A is equivalent to B (hazard ratio 1) for each subpopulation; A is equivalent to B for subpopulation 1 but inferior to B (hazard ratio 1.35) for subpopulation 2; (iii) A is equivalent to B (hazard ratio 1) for

subpopulation 1 but highly inferior to B (hazard ratio 2.14) for subpopulation 2; A is inferior to B (hazard ratio 1.35) for each subpopulation. In each scenario, it was required to have 80% power to reject each inferiority null hypothesis when the corresponding hazard ratio was 1 (equivalence of A and B) or smaller (A superior to B).

3. Key Requirements for Adaptive Enrichment Designs to be Useful:

a. Need predefined subpopulations where effect of new treatment hypothesized to differ by subpopulation

For confirmatory trials (which is the focus of this document), the subpopulations of interest must be defined in the study protocol before the trial starts. In each of our case studies, we consider two subpopulations that partition the overall population. However, more subpopulations could be considered, e.g., a nested sequence of populations as in[5] and some of the real trials in Section 1.e above. Throughout, we assume that the subpopulation definitions are specified in the study protocol.

Selection of the subpopulations should be based on biological knowledge. It was conjectured that subpopulations defined by smaller hemorrhage volume, shorter QRS duration, sex=female, and APOE4 carrier would be more likely to benefit from treatment in the trial corresponding to each of Case Studies 1-4, respectively. Ideally, the subpopulation definition would also be corroborated with data from previous studies (e.g., phase 2 trials or observational studies) if such studies have been conducted. It must be clinically important to learn which subpopulations benefit from treatment in order to justify use of an adaptive enrichment design.

b. Subpopulation size can't be too small as fraction of overall population

If a subpopulation makes up only a small fraction of the overall target population, then an adaptive enrichment design involving such a subpopulation may be infeasible. This is because of the potentially large time required to enroll a sufficient number from such a subpopulation in the case that enrollment rate is proportional to subpopulation fraction. This would be the case for subpopulations defined, e.g., by a biomarker, where biomarker status is first determined at the time of screening a patient into the trial. However, for subpopulations defined by characteristics that are known to would-be participants before screening such as sex, age (and possibly disease severity), it may be possible to include such subpopulations if the projected enrollment rate would be fast enough. Another argument against considering subpopulations that make up a small fraction of the overall population is that even if a treatment benefit were demonstrated in the subpopulation, the public health impact would be limited to this small population (and inappropriate off-label use may end up wasting healthcare resources and/or causing side-effects).

Subpopulation sizes in Case Studies 1-4 are, respectively, 33% with small hemorrhage volume, 49% with short QRS duration, 47% female, and 47% ApoE4 carriers.

c. Long delay times between patient enrollment and measurement of her/his primary outcome can limit the value added by adaptive enrichment.

Another key factor that determines the added value of adaptive enrichment designs is whether sufficient information will have accrued in time to make a useful decision about changing enrollment criteria. The information accrual rate depends on the enrollment rate and the delay time between enrollment and measurement of the primary outcome. As an extreme example, if all enrollment is completed before any primary outcomes have been measured, then there is no opportunity to change enrollment criteria based on accrued primary outcome data. In such a case, a decision could still be made to terminate patient follow-up early (for the overall population or a subpopulation among those in the pipeline) after sufficiently many primary outcomes have been observed. But this will not impact sample size (i.e., number enrolled), only study duration.

It is an open question as to what constitutes sufficient information to make a useful decision about changing enrollment or terminating follow-up. The answer depends on the relative importance assigned to the trial's multiple goals, which include power, precision (of estimators), cost, and minimizing exposure to ineffective/harmful treatments. The situation is more complex in adaptive enrichment designs than in standard designs, since power and precision apply to multiple populations instead of a single population. We recommend running simulations involving different analysis times to determine how much information is sufficient for making a useful decision about enrollment or follow-up modification. Our software tool optimizes the interim analysis times, so may be useful for this purpose. The impact of the enrollment rate and the time to observe participant outcomes is further discussed in Section 4.b.

Some research has been done where interim decisions are based on outcomes measured earlier than the primary endpoint. We recommend caution in using such an approach, since earlier outcomes may fail to be surrogates for the primary outcome; in such cases, a decision rule that depends on the assumption of surrogacy may lead to poor trial performance.

4. Key Tradeoffs: Adaptive Enrichment Designs Versus Standard Group Sequential Designs

a. What is likely to be learned at end of trial

Standard designs have high power to detect an effect in the overall population, but may have low power to detect any effects if the treatment only benefits a subpopulation. It was shown in[1] that an advantage of adaptive enrichment designs compared to standard designs is that the former have high power (e.g., 80%) to detect an effect in a subpopulation if only that subpopulation benefits,

and also have high power to detect and overall effect if the treatment benefits both subpopulations. However, this added power typically comes at the cost of higher sample size (both maximum sample size and average sample size).

b. Sample size and duration required (both expected and maximum)
The previous paragraph compared standard designs that only have high power to detect a treatment benefit in the overall population versus adaptive designs that have high power for both the overall population and for the case where only a subpopulation benefits. A major tradeoff between optimized standard designs and optimized adaptive enrichment designs is average sample size versus maximum possible sample size (and similarly expected duration versus maximum possible duration). The adaptive designs typically have lower average sample size but higher maximum sample size. This is the same type of tradeoff encountered when comparing (non-adaptive) group sequential designs versus single stage designs.

We conducted four simulation studies that mimic key features of data from completed randomized trials (Case Studies 1-4). In each simulation study, we optimized adaptive and standard designs under the power and Type I error constraints from Section 2.b, and compared their performance. The results are summarized in Table 2.

Table 2: Summary of Tradeoffs Comparing Standard (1-Stage) Versus Adaptive Enrichment Designs.

Case Study	**Design Type**	**Average Performance**	**Worst-case Performance**
1. Stroke	 Standard: Adaptive:	Average Sample Size 1443 981	Maximum Sample Size 1443 1762
2. Heart disease	 Standard: Adaptive:	Average Sample Size 1779 1341	Maximum Sample Size 1779 1917
3. HIV	 Standard: Adaptive:	Average Sample Size 1426 No improvement	Maximum Sample Size 1426 No improvement
4. Alzheimer's Disease	 Standard: Adaptive:	Average Duration 5.0 years 4.6 years	Maximum Duration 5.0 years 5.8 years

In Case Study 1 (stroke treatment), the optimized adaptive design reduced average sample size by 32% compared to standard designs; the tradeoff is that the worst-case (maximum) sample size was 22% larger for the adaptive design. For Case Study 2 (cardiac resynchronization device for treating heart disease), the optimized adaptive design reduced average sample size by 25% compared to standard designs; the tradeoff is that the worst-case (maximum) sample size

was 8% larger for the adaptive design. For Case Study 3 (HIV treatments), the adaptive designs provided no added benefits. For Case Study 4 (Alzheimer's disease prevention), the adaptive designs provided no benefits in terms of reducing the average sample size, but did reduce the average trial duration at the cost of greater maximum duration.

Case Studies 1 and 2 saw substantial reductions in average sample size while Case Studies 3 and 4 did not see any such reductions. This is likely due to differences in the relationships between participant enrollment rates and the time (called delay time) between enrollment and observation of each participant's outcome. In Case Studies 1 and 2, the enrollment rates for the combined population were 420 and 240 per year, respectively, while the delay time was 0.5 years in both studies. The standard design maximum durations were 3.9 years and 7.9 years, respectively, for Case Studies 1 and 2.

As a rough benchmark for how much information has accrued in time to be useful for modifying future enrollment, for the standard design we computed the fraction of the total sample size has not yet been enrolled at the time point when half of the total outcomes are observed. We call this the remaining sample size fraction. When the outcome is a time-to-event, as in Case Study 3, we replace "time when half of the total outcomes are observed" by "expected time when half of the total events have occurred", since the total number of events determines the information available for this outcome type.

The remaining sample size fraction is 33% and 43% for case studies 1 and 2, respectively. This suggests that adaptive enrichment may be able to reduce the average sample size, since 33-43% of the sample size remains to be enrolled when half the statistical information has accrued. For example, such designs could have rules for stopping a subpopulation's accrual early for futility at such a time point. For case studies 3 and 4, where the enrollment rates for the combined population were 724 and 500 per year, respectively, the remaining sample size fraction is 0% in both cases. That is, all enrollment has been completed before half the information has accrued. In such cases, one may expect that adaptive enrichment designs are unlikely to lead to reductions in average sample size, since too little information is available to make a useful decision before enrollment has been completed. However, in such cases it is still possible to use adaptive enrichment to reduce the average trial duration, since follow-up for a subpopulation can be stopped early for efficacy or futility. In case study 4, this led to a decrease of 0.4 years in average duration compared to a standard design, but at the cost of 0.8 years for maximum duration; unfortunately, this is not a particularly favorable tradeoff.

c. Expected number of participants assigned to inferior/superior arm

Another potential advantage adaptive enrichment designs is reducing the number of participants assigned to an ineffective or inferior treatment arm. This may be

especially relevant in trials where the treatment has known toxicity or can cause severe adverse events. It may be ideal for cases where it is suspected that the treatment benefits one subpopulation but harms the complementary subpopulation. In[16], simulation studies showed that substantial reductions in the number assigned to the inferior treatment are possible through the use of adaptive enrichment designs, when there is such a qualitative interaction.

5. Regulator Perspectives on Adaptive Enrichment Designs

a. What does the U.S. Food and Drug Administration recommend?

The 2010 FDA draft guidance on adaptive designs for drugs and biologics[4] defines adaptive designs as having preplanned rules for modifying features of the trial, such as sample size, enrollment criteria, or dose of treatment. Their focus is on adequate and well-controlled trials (also called confirmatory trials). They classify different types of adaptations as either "well understood" or "less well understood". Adaptive enrichment designs are in the latter category. Being in the "less well understood" category means that additional justification needs to be provided for why the adaptive design is being proposed. According to a manuscript written by statisticians at the FDA's Center for Biologics Evaluation and Research[17]:

"Providing a well-written and sufficiently detailed study protocol in the regulatory submission will greatly assist the review process. We particularly suggest the following points for sponsors to consider if they intend to conduct adaptive design clinical trials for confirmatory purposes:

1. Why use an adaptive design instead of a conventional one?
2. What features does the proposed design have and are details related to timing and execution clearly spelled out in the protocol?
3. Is the study-wise type I error rate controlled?
4. Are the included simulation studies adequate?
5. Are steps being taken to avoid or minimize operational bias?
6. Are study success criteria and stopping rules explicitly specified?"

We encourage sponsors to propose adaptive design clinical trials when appropriate for their development. It is often productive, however, to discuss such proposals prior to formal submission, using regulatory meeting mechanisms such as pre-IND or Q-submissions to start the dialog."

Similar recommendations as above apply to adaptive designs for medical devices.[18]

b. What does the European Medicines Agency recommend?

According to the European Medicines Agency's Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design[19], in confirmatory trials the use of adaptive designs (compared to standard designs) requires "further justification: adaptive designs should not be seen as a means to alleviate the burden of rigorous planning of clinical trials." For

adaptive enrichment designs, which involve preplanned rules for modifying enrollment at the end of each stage, "the applicant will need to carefully argue why a combination of the results from different stages is capable of substantiating a final treatment recommendation and why the confirmatory nature of the trial is not damaged."[19]

6. Software Tool for Tailoring Adaptive Enrichment Design to One's Scientific Goals and Logistical Constraints, and Compare to Standard Designs, with our Software Tool

a. Overview: target audience

The open-source software tool is available, free, here: http://rosenblum.jhu.edu
It is intended for investigators planning a confirmatory trial where it's suspected that a subpopulation may benefit more than the overall population. The subpopulation could be defined by a risk score or biomarker measured at baseline. The subpopulation must be defined in advance, e.g., based on prior data or medical knowledge. Adaptive enrichment designs have potential to provide stronger evidence than standard designs about treatment benefits for the subpopulation, its complement, and the combined population.

Adaptive enrichment designs have a preplanned rule for changing enrollment criteria based on accrued data in an ongoing trial; for example, future enrollment may be restricted to a subpopulation if the complementary subpopulation is not benefiting. This software tool can help in planning such a trial, by tailoring an adaptive enrichment design to the scientific goals and logistical constraints of the investigator, and comparing performance of the adaptive design to more traditional designs. The software searches over hundreds of candidate adaptive designs with the aim of finding one that satisfies the user's requirements for Type I and II error at the minimum cost. This requires substantial computation and is typically completed within 24 hours, at which time a summary report is emailed to the user.

b. Inputs from the user specific to their trial design problem

i. Type of Primary Outcome (binary, continuous, or time-to-event)
ii. Subpopulation 1 proportion
iii. Familywise Type I error
iv. Maximum total sample size
v. Enrollment Rate per Year for Combined Population
vi. Length of Follow-up for Each Participant
vii. Optimization Target: minimize average sample size or expected duration.
viii. Option to Incorporate Precision Gain from Adjustment for Prognostic Baseline Variables
ix. If outcome is time-to-event, additional inputs include: selection of superiority versus non-inferiority trial, censoring probability.
x. Power requirements
xi. Scenarios over which average sample size (or duration) is to be calculated, e.g., the scenarios described in (1.d) above.

c. What information is output by the software tool (automatically generated report comparing performance of different designs)

The software outputs a report comparing the performance of the best designs found in its search. Performance criteria reported include the following:

i. Sample size
ii. Duration
iii. Power
iv. Type I error
v. Estimator bias, variance, and mean squared error
vi. Confidence interval performance

---

[1] Fisher, A. and Rosenblum, M., Stochastic Optimization of Adaptive Enrichment Designs for Two Subpopulations. Journal of Biopharmaceutical Statistics (Submitted December 5, 2016, revision requested March 19, 2017) Working paper: http://biostats.bepress.com/jhubiostat/paper279

[2] Steingrimsson, Jon Arni; Betz, Joshua; Qian, Tiachen; and Rosenblum, Michael, Optimized Adaptive Enrichment Designs For Multi-Arm Trials: Learning Which Subpopulations Benefit From Different Treatments (2017). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 288. http://biostats.bepress.com/jhubiostat/paper288

[3] Betz, Josh; Steingrimsson, Jon Arni; Qian, Tianchen; and Rosenblum, Michael. Comparison Of Adaptive Randomized Trial Designs For Time-To-Event Outcomes That Expand Versus Restrict Enrollment Criteria, To Test Non-Inferiority (2017). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 289. http://biostats.bepress.com/jhubiostat/paper289

[4] FDA. (2010). Draft guidance for industry. Adaptive design clinical trials for drugs and biologics. http://goo.gl/sND1mU

[5] Wang, S. J., Hung, H. & O'neill, R. T. (2009). Adaptive patient enrichment designs in therapeutic trials. Biomet. J. 51, 358–74.

[6] Rosenblum, M. (2015), Adaptive Randomized Trial Designs that Cannot be Dominated by Any Standard Design at the Same Total Sample Size. Biometrika. 102(1). 191-202. http://goo.gl/ovNu1S

[7] Rosenblum, M., and Hanley, D.F. (2017) Topical Review: Adaptive Enrichment Designs for Stroke Clinical Trials. Stroke. 48(6). https://doi.org/10.1161/STROKEAHA.116.015342

[8] Tudor G Jovin, Jeffrey L Saver, Marc Ribo, Vitor Pereira, Anthony Furlan, Alain Bonafe, Blaise Baxter, Rishi Gupta, Demetrius Lopes, Olav Jansen, Wade Smith, Daryl Gress, Steven Hetts, Roger J Lewis, Ryan Shields, Scott M Berry, Todd L Graves, Tim Malisch, Ansaar Rai, Kevin N Sheth, David S Liebeskind, Raul G Nogueira. (2017) Diffusion-weighted imaging or computerized tomography perfusion assessment with clinical mismatch in the triage of wake up and late presenting strokes undergoing neurointervention with Trevo (DAWN) trial methods. International Journal of Stroke. Vol 12, Issue 6, pp. 641-652. DOI 10.1177/1747493017710341

[9] Gregory W Albers. Endovascular Therapy Following Imaging Evaluation for Ischemic Stroke 3 (DEFUSE 3) ClinicalTrials.gov Identifier: NCT02586415. https://www.clinicaltrials.gov/ct2/show/NCT02586415?term=DEFUSE&rank=1

[10] Robin Lewis Jones, Steven Attia, Cyrus R. Mehta, Lingyun Liu, Kamalesh Kumar Sankhala, Steven Ian Robinson, Vinod Ravi, Nicolas Penel, Silvia Stacchiotti, William D. Tap, Delia Alvarez, Richard Yocum, Charles P. Theuer, and Robert G. Maki. Tappas: An adaptive enrichment phase 3 trial of TRC105 and pazopanib versus pazopanib alone in patients with advanced angiosarcoma (AAS). Journal of Clinical Oncology 2017 35:15_suppl, TPS11081-TPS11081 NCT02979899.
http://ascopubs.org/doi/abs/10.1200/JCO.2017.35.15_suppl.TPS11081

[11] Hanley, D. F., Thompson, R. E., Muschelli, J., Rosenblum, M., McBee, N., Lane, K., Bistran-Hall, A. J., Mayo, S. W., Keyl, P., Gandhi, D., Morgan, T. C., Ullman, N., Mould, W. A., Carhuapoma, J. R., Kase, C., Ziai, W., Thompson, C. B., Yenokyan, G., Huang, E., Broaddus, W. C., Graham, R. S., Aldrich, E. F., Dodd, R., Wijman, C., Caron, J-L., Huang, J., Camarata, P., Mendelow, A. D., Gregson, B., Janis, S., Vespa, P., Martin, N., Awad, I., Zuccarello, M., for the MISTIE II Investigators. (2016) Safety and efficacy of minimally invasive surgery plus alteplase in intracerebral haemorrhage evacuation (MISTIE): a randomized, controlled, open-label, phase 2 trial. Lancet Neurology. 15(12). 1228-1237. http://dx.doi.org/10.1016/S1474-4422(16)30234-4

[12] Hanley, D.F., Lane, K., McBee, N., Ziai, W, Tuhrim, S., Lees, K.R., Dawson, J., Gandhi, D., Ullman, N., Mould, W.A., Mayo, S.W., Mendelow, A.D., Gregson, B., Butcher, K., Vespa, P., Wright, D.W., Kase, C.S., Carhuapoma, J.R., Keyl, P.M., Diener-West, M., Betz, J.F., Thompson, C., Sugar, E.A., Yenokyan, G., Janis, S., John, S., Harnof, S., Lopez, G., Aldrich, E.F., Harrigan, M.R., Ansari, S., Jallo, J., Caron, J-L., LeDoux, D., Adeoye, O., Zuccarello, M., Adams, H.P., Rosenblum, M., Thompson, R.E., Awad, I.A., for the CLEAR III Investigators. (2017) Thrombolytic removal of intraventricular haemorrhage in treatment of severe stroke: results of the randomised, multicentre, multiregion, placebo-controlled CLEAR III trial. The Lancet. 389(10069), 603–611. http://dx.doi.org/10.1016/S0140-6736(16)32410-2

[13] Kenneth A. Ellenbogen, Michael R. Gold, Timothy E. Meyer, Ignacio Fernndez Lozano, Suneet Mittal, Alan D. Waggoner, Bernd Lemke, Jagmeet P. Singh, Francis G. Spinale, Jennifer E. Van Eyk, Jeffrey Whitehill, Stanislav Weiner, Maninder Bedi, Joshua Rapkin and Kenneth M. Stein. (2010) Primary Results From the SmartDelay Determined AV Optimization: A Comparison to Other AV Delay Methods Used in Cardiac Resynchronization Therapy (SMART-AV) Trial. Circulation. 122:2660-2668. https://doi.org/10.1161/CIRCULATIONAHA.110.992552

[14] Campbell TB, Smeaton LM, Kumarasamy N, Flanigan T, Klingman KL, Firnhaber C, Grinsztejn B, Hosseinipour MC, Kumwenda J, Lalloo U, Riviere C, Sanchez J, Melo M, Supparatpinyo K, Tripathy S, Martinez AI, Nair A, Walawander A, Moran L, Chen Y, Snowden W, Rooney JF, Uy J, Schooley RT, De Gruttola V, Hakim JG; PEARLS study team of the ACTG. Efficacy and safety

of three antiretroviral regimens for initial treatment of HIV-1: a randomized clinical trial in diverse multinational settings. PLoS Med. 2012;9(8):e1001290. Epub 2012 Aug 14.

[15] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

[16] Luber, Brandon S.; Rosenblum, Michael; and Chambaz, Antoine (2013) Trial Designs That Simultaneously Optimize The Population Enrolled And The Treatment Allocation Probabilities. Johns Hopkins University, Dept. of Biostatistics Working Papers. http://biostats.bepress.com/jhubiostat/paper256

[17] Min Lin, Shiowjen Lee, Boguang Zhen, John Scott, Amelia Horne, DrPH, Ghideon Solomon, Estelle Russek-Cohen. CBER's Experience With Adaptive Design Clinical Trials. Therapeutic Innovation & Regulatory Science. Vol 50, Issue 2, pp. 195 - 203. DOI 10.1177/2168479015604181

[18] U.S. Food and Drug Administration (2016), Adaptive Designs for Medical Device Clinical Studies. Guidance for Industry and Food and Drug Administration Staff. http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf accessed February 23, 2017.

[19] European Medicines Agency. 2007. Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design (Chmp/Ewp/2459/02) http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf Accessed 8/28/17.