# Standards for the Design, Conduct, and Evaluation of Adaptive Randomized Clinical Trials

Michelle A. Detry, PhD,[1] Roger J. Lewis, MD, PhD,[1-4] Kristine R. Broglio, MS,[1]
Jason T. Connor, PhD,[1,5] Scott M. Berry, PhD,[1] Donald A. Berry, PhD[1,6]

1. Berry Consultants, LLC, Austin, TX
2. Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, CA
3. David Geffen School of Medicine at UCLA, Los Angeles, CA
4. Los Angeles Biomedical Research Institute, Torrance, CA
5. College of Medicine, University of Central Florida, Orlando, FL
6. Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX

Revision Date: March 15, 2012

Contact Information:

Donald A. Berry, PhD
Berry Consultants, LLC,
4301 Westbank Drive,
Bldg B, Suite 140,
Austin, TX 78746
Telephone: 713-817-5586
Email: don@berryconsultants.com

## Table of Contents

# Introduction

Patient-centered outcomes research (PCOR) is intended to help physicians and patients make better treatment decisions using comparative information regarding the effectiveness of commonly used treatments in typical care settings.[1, 2] Comparative effectiveness research (CER) is defined by the Institute of Medicine as "the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care."[3] While CER can include many types of studies[4], including traditional randomized clinical trials (RCTs), *adaptive* RCTs hold particular promise for PCOR.[5-8] An adaptive clinical trial is one in which key trial characteristics (e.g., randomization proportions, sample size, treatment arms, eligibility criteria) evolve according to prespecified rules during the trial, in response to information accruing within the trial itself. Goals of this approach include higher statistical efficiency, improved patient outcomes, or better ethical balance.[9-11] There is a range of roles for adaptive trials, including dose-finding trials, personalized medicine trials in which treatments are developed to be tailored to specific patient traits, and informative comparative effectiveness trials. Adaptive approaches can lead to more effective treatments for participants in the CER trial itself, but also to better treatment of future patients having the disease or condition in question. Adaptive trials are ideal for PCOR because enrollment criteria may be altered to focus on patient subpopulations of particular interest. Appendix 1 lists common terminology related to adaptive trial design.

Adaptive designs are not new to clinical trials. Group sequential designs have

been in use for decades. However, in recent years, increased emphasis has been placed on streamlining drug development, including by regulatory agencies.[12] In 2010, the Center for Drug Evaluation and Research and the Center for Biologics Evaluation and Research of the Food and Drug Administration (FDA) issued a draft guidance for industry regarding the use of adaptive designs for "adequate and well controlled" phase III trials.[13] This guidance supplemented a similarly focused 2007 "Reflection Paper" from the European Medicines Agency.[14] On the medical device side, since the passage of the FDA Modernization Act of 1997[15] the FDA's Center for Devices and Radiologic Health has been encouraging Bayesian adaptive approaches and issued its Bayesian guidance for medical device trials in 2010.[16]

Innovations in adaptive trial design for drug and medical device development have led to highly flexible approaches that extend in a natural way to PCOR. The standard, modern approach to clinical trial design is steeped in hypothesis testing. This tradition is important and it has transformed medical research from case studies and clinical experience into a scientific discipline. But hypothesis testing is not as applicable to CER and PCOR as it is to the medical product registration setting because, estimation of treatment effects is of greater importance than simply determining superiority. Adaptive trial designs do not abandon the tradition of hypothesis testing, but controlling type I and type II errors may be secondary when addressing questions such as, "How do we build a trial with a goal of treating patients as effectively as possible, including those in the trial as well as those who present later?"

Recognizing the need for innovation in clinical trial design, representatives

from Clinical and Translational Science Award programs have identified adaptive clinical trial design as a high priority methodological issue "to increase the efficiency of comparative effectiveness trials."[17] In October 2011, the Patient-Centered Outcomes Research Institute requested proposals to review applicable guidance documents and available literature and to propose minimum standards for the design, conduct, and reporting of adaptive clinical trials. This report was produced in response to that request.

In this report, we describe the results of a comprehensive review of existing literature relative to adaptive RCTs and propose standards for the design, conduct and reporting of adaptive RCTs in PCOR. We augment published information based on our experience in designing and helping to conduct adaptive RCTs. We identify existing gaps in the literature that currently preclude identifying standards in specific areas.

## Literature Review

We conducted a literature search with three goals. The first goal was to identify any applicable guidance or draft guidance documents from regulatory bodies or government agencies. We identified three of these, as mentioned above.[13, 14, 16] Although adaptive design is not the main focus of the Bayesian guidance from FDA's Center for Devices and Radiologic Health,[16] the Bayesian approach is inherently adaptive, and thus the guidance addresses general concepts for adaptive clinical trials.

The second goal of the literature search was to determine a set of best-practice papers. We searched PubMed using the search phrase "adaptive trial design" in combination with the search phrases "recommendations", "best practice", "expert opinion", and "panel discussion". Additionally, we retrieved all white papers from the PhRMA working group website http://www.biopharmnet.com/doc/doc12004-01.html. Our initial search identified 18 articles and 2 special issues on adaptive trial design. Articles narrowly focused on a particular type of adaptation were then excluded. An additional article published since the initial search was encountered and included. A team member abstracted information from the resulting 21 manuscripts[18-38] and a second team member independently verified the abstracted information. Appendix 2 contains a list of the information abstracted. We have incorporated the key content from these best-practices papers into our proposed minimum standards. In addition to the articles we selected, a substantial number of focus pieces and discussion pieces are available and many of these are listed as general references in the guidance documents. However, the moderate number of guidance and review documents we specifically cite provide sufficient and broad support for the proposed minimum standards.

The third goal of the literature search was to identify clinical trials to serve as examples for demonstrating components of the minimum standards. A pool of candidate trials was identified by the searches described in Box 1.

Box 1.

| Search | PubMed Keywords |
|---|---|
| 1 | "comparative effectiveness" and randomized clinical trial and adaptive |
| 2 | "patient centered outcome" and randomized clinical trial |
| 3 | "patient centered outcomes" and randomized clinical trial |
| 4 | adaptive and clinical trial and ("phase IV" or "phase 4") |
| 5 | adaptive and clinical trial and ("phase III" or "phase 3") |
| 6 | adaptive and clinical trial and ("phase II" or "phase 2") |
| 7 | adaptive and clinical trial and seamless |

We reviewed abstracts and selected a subset of potential example trials. We augmented the list with other adaptive clinical trials that were known to study team members. We reviewed each article and completed a structured abstraction form. A selection of relevant example trials, demonstrating compliance with the minimum standards, is described in tables presented later in this report.

We found two publications of adaptive RCTs that clearly qualified as PCOR,[39, 40] there are likely others but the overall number of adaptive RCTs in CER is low. However, adaptive trials in other settings, including those in drug and medical device development, evince properties of importance to PCOR.

## Minimum Standards

We propose eight minimum methodological standards that apply in general to adaptive RCTs (Box 2). These minimum standards were developed based on the information from the guidance documents and selected best-practices articles, along with our experience in adaptive design. Our goal is to provide standards appropriate for adaptive RCTs in PCOR.

Box 2.

| Minimum Standards for Adaptive Randomized Clinical Trials |
| --- |
| 1. Explicit Prospective Specification of Planned Adaptations and Primary Analysis<br>2. Evaluation of Statistical Properties<br>3. Standards for Bayesian Adaptive Randomized Clinical Trial Designs<br>4. Communication and Vetting of Trial Design with Key Stakeholders<br>5. Ensure Clinical Trial Infrastructure is Adequate to Support Planned Adaptation(s)<br>6. Consideration of Operational Bias<br>7. Ensure Proper Oversight of Adaptive Randomized Clinical Trial<br>8. The Reporting of Adaptive Randomized Clinical Trials Should be Consistent with the CONSORT Statement |

We distinguish minimum standards, criteria that are broadly applicable to all adaptive trials, from recommendations. Recommendations generally refer to a higher methodological level that may not be necessary in all settings. In addition, minimum standards, or specific components of minimum standards, may not apply to every adaptive clinical trial. To clarify this issue, we classify adaptive trials designs into two broad categories: simple and complex. “Simple” adaptive trial designs are those in which the design’s operating characteristics, especially type I error and statistical power, can be calculated analytically. From this perspective, traditional group sequential designs, blinded sample size re-estimation, and adaptive randomization to achieve covariate balance are “simple”. “Complex” adaptive trial designs are those for which analytical calculations are not sufficient to define operating characteristics; complex trials often result from the tailoring of a trial to a specific research question. Calculating the operating characteristics for complex trials requires numerical methods or, more typically, computer simulations due to lack of an “off the shelf” tool. Examples of complex trials include those with response-adaptive randomization, arm dropping and adding, and enrichment

designs. The combination of multiple different simple adaptations may result in a complex design.

The following sections describe the minimum standards and include tables that provide information regarding current practice, examples of trials that meet the standards, and the rationale for adopting the proposed standard. These standards apply to the design, execution, analysis, and reporting of results of adaptive RCTs for PCOR.

### *Minimum Standard 1. Explicit Prospective Specification of Planned Adaptations and Primary Analysis*

Prospective specification of all components of both simple and complex adaptive designs is critical. Adaptive designs can be very flexible, but the flexibility is built in during the planning stage.[18, 26, 30] Adaptive trial designs require more planning and detailed prespecification than do standard trials.[18, 30] The extra planning time represents a cost associated with being adaptive, but the benefits often outweigh the costs. Indeed, clearly defining a trial's goals and the likely consequences of different designs, including finding the predictive probabilities of different results, is a worthwhile process even if one does not use an adaptive design in the end.

A common misconception is that adaptive designs allow decisions to be made on an ad hoc basis. On the contrary, all possible adaptations for a design should be described before the trial begins. It should be clear when interim analyses that may lead to an adaptation are to occur, with the "when" defined by the number of participants enrolled, the number of participants evaluated, the number of events

occurring in a trial, or calendar time. The adaptive decision criteria should be specific, e.g. the accrual to a trial will be terminated early if the predictive probability of success is greater than 0.95. The study population whose results might trigger an adaptation should also be clear, e.g. an intent-to-treat (ITT) population including all participants randomized. If the design requires a model, e.g. a time-to-event trial with a longitudinal model relating early and late endpoints, the model structure should be specified before the trial begins.

The analysis for the primary outcome measure or, if applicable, the criteria used to select the primary endpoint, must be completely prespecified. Even in simple adaptive designs, one should prespecify the analysis population for the primary analysis. For example, in a standard group sequential trial with a 90-day endpoint, if the trial stops early for success based on an interim analysis of participants with complete data, the protocol should specify how data from participants recently enrolled but lacking primary outcomes at the stopping analysis are to be analyzed. We recommend the interim and final analyses should be included in a Statistical Analysis Plan (SAP) that is separate from the study protocol, although some detail may also be included in the protocol.

The description of the adaptive design for both simple and complex designs should be adequately detailed so that a sufficiently sophisticated reviewer could implement the adaptations and perform the primary analysis. The prespecified adaptations are to be considered part of the original protocol and the implementation of prespecified adaptations does not constitute a protocol amendment.

| Section | Minimum Standard 1: Explicit prospective specification of planned adaptations and primary analysis |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: **The adaptive clinical trial design should be prospectively planned and the design clearly documented, including:**<br><br>• **All potential adaptations, including timing;**<br>• **Trial results and populations that will be used in determining each adaptation;**<br>• **Statistical models to be used; and**<br>• **Planned analysis of the primary endpoint(s).**<br><br>The description of the design should be sufficiently detailed that it could be implemented without human judgment or intervention (i.e., by an automaton). The specification of the design should be completed and documented in the trial protocol before enrollment begins. This specification should include, in all but the simplest designs, a Statistical Analysis Plan (SAP) that is separate from the trial protocol in which all necessary detail is provided regarding planned interim and final analyses. Prior specification is a prerequisite for valid and meaningful evaluation of an adaptive design (see Minimum Standard 2) and for communication of the design to stakeholders (see Minimum Standard 4). |
| | 2. Current Practice and Examples: Current best practice for adaptive trial design, particularly for Phase III confirmatory trials, is that all potential adaptations and analyses be prospectively defined and ad hoc changes based on subjective observation of trial results are not permissible.[19, 20, 30, 34].<br><br>An example of a trial that included prespecification of the adaptations is the BLISS trial,[41] a randomized placebo-controlled non-inferiority trial to assess the cardiovascular safety of testosterone gel in women with hypoactive sexual desire disorder. A second example is Muss et al,[40] a randomized non-inferiority trial of two adjuvant chemotherapy regimens in older women with primary breast cancer. Both used prespecified adaptive sample size algorithms. A third example is Giles et al;[42] a prespecified adaptive allocation algorithm was used in a phase III trial comparing three chemotherapy regimens in older patients with acute myeloid leukemia or high risk myelodysplastic syndrome. A fourth example is a CER trial comparing two strategies for insulin administration in hospitalized patients.[39] A response adaptive randomization design was implemented and the adaptations and stopping rules were clearly prespecified. |
| | 3. Published Guidance(s): This minimum standard is addressed or supported by the following guidances:<br><br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13]<br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials;[16] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14] |

| | |
|---|---|
| MC Key Criteria: Rationale for and against adoption of the proposed standard | 4. Contribution to Patient Centeredness: The requirement for prior specification enables a thorough consideration of the trial's goals, scientific aims, and the priorities of concerned stakeholders, including patients. |
| | 5. Contribution to Scientific Rigor: Prior specification of the trial design helps to ensure validity and credibility of the trial results. Unplanned or ad hoc adaptations are likely to alter the operating characteristics of the trial and introduce unknown biases in the estimates of treatment effect. |
| | 6. Contribution to Transparency: Prior specification of the design is essential for transparency. A complete and clearly presented prior specification of the clinical trial design, including all planned adaptations and primary analyses, helps stakeholders understand the adaptive design when deciding whether or not to participate in the trial. |
| | 7. Empirical Evidence and Theoretical Basis: Theoretical considerations and simulation studies demonstrate the potential for inflated type I error rates and overestimates of treatment effects when unplanned adaptations or data analyses occur.[43-45] |
| Additional considerations | 8. Degree of Implementation Issues: Detailed, prior specification of the planned adaptive trial design is the accepted standard in registration trials conducted by industry. Prior specification of investigator-initiated adaptive trials is not consistently followed in academic settings. |
| | 9. Other Considerations: None<br>10. Implication of lack of adherence to standard: Unplanned or ad hoc adaptations, especially when based on inspection of unblinded data, increases the risk of obtaining false or misleading conclusions, including false-positive results and bias in estimates of treatment effects. Such trials are also less likely to have sufficient credibility to influence clinical practice. |

### *Minimum Standard 2. Evaluation of Statistical Properties*

The statistical properties of an adaptive trial should be evaluated and understood before a design is implemented. For simple adaptive designs, such as group sequential designs, the statistical properties are well understood and can be determined analytically. Newer complex adaptive designs may require computer simulation to fully understand their statistical properties. Traditionally, the statistical properties to be evaluated, also referred to as operating characteristics, include type I error, power, and sample size. Complex adaptive designs usually have several goals, and their operating characteristics tend to be multifaceted.

The null hypothesis is rarely a single point and is often multidimensional. For example, there are many ways for two treatments to be equally effective. Type I

error rate usually varies over the null hypothesis space of parameters. An adaptive trial can be designed, if necessary, so that its type I error rate is sufficiently low across this space. In registration trials, evaluating and controlling the type I error rate is essential. It is also important to understand the type I error rate for CER and PCOR trials, even though controlling this rate at a traditional 0.05 level may not be required. Statistical power is another important characteristic of an adaptive RCT. The power of a complex adaptive trial should be evaluated by simulation for a variety of assumed true treatment effects. Similar to type I error, adequate power should be determined in the context of the research question. The sample size distribution for a complex adaptive trial should also be evaluated during simulation including its mean or median, probability that the sample sizes reaches its predetermined maximum, and other such measures. Understanding the distribution of possible sample size is especially important for budgeting, assessing feasibility, and determining necessary participant enrollment rates.

We also recommend that simulation be used to assess the effects on trial performance of accrual rate, dropout rate, missing data, possible delays in data updates, and violations of distributional assumptions. In contrast to fixed designs, the rate of execution of a design, affecting the completeness of available information, can have an impact on operating characteristics. Depending on the type of adaptations involved, it may also be important to consider the consequences of drift in participant characteristics over time.

Computer simulation is available in some adaptive trial design software packages, for example AddPlan, Compass, East, FACTS, and Pest (list not

exhaustive). The simulation of complex adaptive trials may require customized code. When used, simulation methods should be thoroughly described and the results summarized in an SAP or other supporting trial document.

Trial designers may choose to produce a separate document, an adaptive design report, as an appendix to the study protocol or SAP. Possible components of the adaptive design report include:

1. Description of the adaptive trial structure.
2. All potential adaptations and the trial results and populations that will be used in determining each adaptation.
3. Statistical models used for decisions and adaptations within the adaptive design, including calculation details or software used.
4. Statistical models and thresholds for the primary analyses and key analyses, including any calculation details or software used.
5. Operating Characteristics for the design (e.g., based on simulation).
6. Example simulated trials to illustrate the behavior of adaptive algorithms.
7. Mode of calculating operating characteristics:
    a. If simulations are used, assumptions used in simulations
    b. Creation of virtual participants
    c. Assumptions regarding accrual rate, dropout rates, and time-to-information
    d. Any other assumptions used for the simulation of trials.

Computer programs used for the simulations should be retained with appropriate version control. We recommend that computer code used for

simulations be provided to appropriate stakeholders (e.g. reviewers from Patient-Centered Outcomes Research Institute, National Institutes of Health, or FDA) when feasible, or that reference be made to proprietary simulation software.

| Section | Minimum Standard 2: Evaluation of statistical properties |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: Adaptive clinical trial designs vary widely in complexity, ranging from some relatively simple designs that have been used for decades and whose statistical performance is well understood (e.g., group sequential methods, blinded sample size re-estimation) to newer, more complex approaches. The performance of innovative, more complex adaptive designs are less well understood and generally require the use of computer simulation to thoroughly elucidate how the design behaves under a range of potential circumstances. **While not necessary for simple designs, the statistical properties of complex adaptive clinical trial designs should be thoroughly investigated over the relevant range of important parameters or clinical scenarios (e.g., treatment effects, accrual rates, delays in the availability of outcome data, dropout rates, missing data, drift in participant characteristics over time, subgroup-treatment interactions, violations of distributional assumptions). Statistical properties to be evaluated should include type I error, power, sample size distributions, as well as the precision and bias in the estimation of treatment effects. Additional performance metrics may also be evaluated (e.g., the frequency with which specific adaptions occur, likelihood of substantial covariate imbalance, likely adequacy of final data for subgroup and safety analyses).**<br><br>The programming code used to create the simulations should be retained with version control. It is recommended that the programming code and software used be made available to stakeholders who have a need to know, including reviewing agencies. |
| | 2. Current Practice and Examples: Traditional operating characteristics, such as type I error rates, power, and sample size distributions, are available analytically for some designs. Results of simulations to determine these characteristics are commonly provided for complex adaptive clinical trial designs.[18, 32, 34, 36] Simulation allows the trial design's performance to be evaluated, the design to be modified if necessary (e.g., altering the criterion for a positive result), and re-evaluated until the desired performance is obtained.[18, 33] Regulatory agencies and other reviewers require thorough characterization of the statistical properties of proposed clinical trials.[13, 14]<br><br>The CER trial comparing two strategies for insulin administration in hospitalized patients described by Fiore et al[39] includes an example of several components of this minimum standard. A response adaptive randomization was implemented to compare the two insulin strategies. Simulation was used to determine the type I error, power, and mean participants assigned to each treatment arm for a variety of trial scenarios and results of the simulations are presented in the manuscript. Type I error is stated to not be as important a metric for this design, compared to designs in different contexts, because in this trial they are investigating two widely used insulin dosing strategies. In addition, the authors state that the R[46] code used to conduct the simulations is available upon request. Cohen et al[47] presented a randomized trial of early or delayed antiretroviral therapy in couples in which one partner was HIV positive and the other was HIV negative. The study used a group sequential design in which the power and the type I |

| | |
|---|---|
| | error could be determined analytically so simulation was not required. The BLISS trial[41] is a randomized placebo-controlled non-inferiority trial to assess the cardiovascular safety of testosterone gel in women with hypoactive sexual desire disorder. This is a Bayesian adaptive trial design that used simulation to assess type I error rate, power, and the expected sample size. Additional examples of published adaptive trials whose performance characteristics were assessed by simulation include Hall et al,[48] Chataway,[49] and the ASTIN trial.[50-52] The trial reported by Hall et al[48] was a group sequential up-and-down design for dose selection in migraine. Simulations were used to investigate the type I error rate, power, trial size, and the dose selected. The trial reported by Chataway et al,[49] was an adaptive seamless design for secondary progressive, multiple sclerosis. Simulation was used to assess the timing of the interim analysis, treatment selection rules, and resource allocation through evaluation of the simulated study power. The programming code was submitted to the comprehensive R[46] archive network (package *asd*) and is publically available. Finally, the ASTIN trial[50-52] was a seamless phase II/III trial, with a phase II dose-finding component, conducted in acute ischemic stroke. Simulations assessed the ability of the design to accurately identify the underlying dose-response curve and whether the drug was effective. |
| | 3. Published Guidance(s): This minimum standard is addressed or supported by the following guidances:<br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13]<br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials;[16] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14] |
| MC Key Criteria: Rationale for and against adoption of the proposed standard | 4. Contribution to Patient Centeredness: A thorough understanding of a trial's statistical properties is critical to evaluating the efficiency and scientific strength of the design. This evaluation leads to better trials, treating participants in a trial more effectively and increasing the likelihood that future patients will receive better treatment. Participants in CER or patients whose treatment is guided by results of adaptive trials must be assured that such trials yield valid conclusions. |
| | 5. Contribution to Scientific Rigor: Some features of modern adaptive clinical trial designs (e.g., frequent interim data analyses, dropping poorly performing arms, early stopping rules) may increase type I error rate, especially if traditional statistical criteria are used. Early termination of a clinical trial, especially when triggered by an observed treatment difference, may result in a biased estimate of the treatment effect and some adaptive methods (e.g., group sequential designs, enrichment designs, seamless phase II/III designs) may result in biased estimates although the magnitude of such bias is generally small. These sources of statistical bias should be assessed quantitatively (e.g., through simulation), especially in settings where the goal is to determine with confidence whether a treatment is effective (as opposed, e.g., to which among a set of treatment strategies is best). Understanding the existence of and magnitude of statistical biases ensures the trial results will be interpreted correctly when applied to the care of patients. It also allows for reducing these biases should that be necessary. |
| | 6. Contribution to Transparency: In addition to allowing quantification and communication of the statistical properties of complex adaptive trial designs, simulation yields examples of specific trials—each representing a realization of the proposed design—that can be used to communicate the design to investigators, trial study team members, DSMB members, regulatory agencies, |

| | |
|---|---|
| | Institutional Review Boards, grant proposal reviewers (see Minimum Standard 3). |
| | 7. Empirical Evidence and Theoretical Basis: The importance of and validity of simulation as a tool to understand the performance of adaptive clinical trial designs is illustrated in many publications, including Fiore et al,[39] White et al,[41] Hall et al,[48] and Chataway et al[49] and the book by Berry, Carlin, Lee, Muller.[11] |
| Additional considerations | 8. Degree of Implementation Issues: The need for customized computer programs to simulate complex designs is a barrier to the implementation of this standard. Recently available software programs specifically written for simulating complex adaptive trial designs are available and their capabilities are being continuously extended (e.g., AddPlan, Compass, East, FACTS, PEST); thus this barrier should be reduced over time. The use of simulation, especially simulation requiring the writing of custom programming, is highly variable. |
| | 9. Other Considerations: Type I error rate is not necessarily the primary concern in PCOR studies. For example, the goal may be to identify as well as possible the best therapy among a set of alternatives, possibly depending on participant characteristics. The evaluation of specific operating characteristics should be considered in the context of the research and clinical questions being investigated.<br><br>When quantifying the expected statistical bias in estimating the treatment effects associated with proposed stopping rules (e.g., simple group sequential designs), one must average over all variability in the resulting data and the outcome of the trial (e.g., whether the trial stops early or does not). There is substantial misinformation and misunderstanding regarding the extent of bias in estimation, and studies purporting to demonstrate empirical evidence of substantial bias based on analyses of highly selected trials have added to this confusion.[53] In many cases, the bias associated with the application of common stopping rules is minimal.[43-45, 54-59]<br><br>10. Implication of lack of adherence to standard: Lack of a thorough elucidation of the trial's statistical properties would make it difficult for reviewers to evaluate whether the proposed design is likely to accomplish its objective. Lack of adherence to this standard can result in adaptive trials being conducted whose results are biased to a clinically important degree or, worse, are qualitatively inaccurate. |

### *Minimum Standard 3. Standards for Bayesian Adaptive Randomized Clinical Trial Designs*

Adaptive clinical trial designs using a Bayesian methodology are typically complex adaptive designs. These designs can be very flexible and the Bayesian approach naturally fits the adaptive paradigm. Because Bayesian designs are becoming more common, and many statisticians were not specifically trained in this approach, we provide a minimum standard specifically for this methodology.

However, having a standard specific for Bayesian designs does not mean they require additional scrutiny; Bayesian designs should be held to the same level of rigor as other designs. In particular, Bayesian adaptive designs must meet each of the other seven minimum standards.

A description of a Bayesian adaptive design should specify the statistical models used in the conduct of the trial and the final analysis, prior probability distributions, and assumptions regarding exchangeability. Prior distributions need to be explicitly and prospectively specified and the rationale for their selection should be reported. For informative priors, the source of the information, justification, and implications of the informative prior should be provided. If prior information is used in the design phase, but not in the final analysis, then this should be made clear. If formal decision analyses are driving adaptation then any utility functions should be clearly defined and their rationale described.

Bayesian adaptive designs allow for the incorporation of prior or external information that may be similar to, but not exchangeable with, information in the proposed trial. Specific details regarding the incorporation of external information should be included and any exchangeability assumptions described. The Bayesian design specification described above may be included within the trial protocol or, preferably, in a separate statistical analysis plan.

| Section | Minimum Standard 3: Standards for Bayesian adaptive randomized clinical trial designs |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: Bayesian approaches to adaptive randomized clinical trials are becoming more common due to their flexibility, potential efficacy, and ease of construction. They are often based on a broad goal, such as delivering good therapy to as many participants as possible, rather than testing a particular hypothesis. If such a design is proposed, **the Bayesian structure and analysis plan for the trial must be clearly and completely specified.** The description of the Bayesian structure should include any statistical models used either during the conduct of the |

| | |
|---|---|
| | trial or for the final analysis, prior probability distributions and their basis, utility functions associated with the trial's goals, and assumptions regarding exchangeability (of participants, of trials, of other levels). Specific details should be provided as to how the prior distribution was determined and if an informative or non-informative prior was chosen. When an informative prior is used, the source of the information should be described. If the prior used during the design phase is different from the one used in the final analysis then the rationale for this approach should be indicated. Utility functions, if employed, should be defined and their source should be described. Computational issues, such as the choice of software, the creation and testing of custom software, and software validation should be addressed as well. Software used for Bayesian calculations during trial design, trial execution, and final analysis must be functionally equivalent. When feasible, software or programs should be made available to relevant stakeholders for evaluation and validation. |
| | 2. Current Practice and Examples: Bayesian trials have been used for the evaluation of both medical devices and drugs. When a Bayesian design is implemented it is recommended that the priors be reported along with the weight attributed to the priors.[30] The FDA published a guidance in 2010 specifically regarding the use of Bayesian statistics in medical device trials.[16] The recommendations contained in the 2010 guidance can appropriately be applied to Bayesian adaptive clinical trials more generally.<br><br>An example trial using a fully Bayesian approach is the ASTIN trial,[50-52] a seamless phase II/III trial in acute ischemic stroke, with a phase II dose-finding component. Berry et al[50] includes details of the statistical models, the adaptive decision algorithms, computational and software issues, simulation, and operating characteristics. An example in the CER setting is a trial comparing two strategies for insulin administration in hospitalized patients using response-adaptive randomization.[39] The model used to simulate the trial data and prior distributions are described. Complete details for the Bayesian adaptive randomization and early stopping considerations are also included. Other examples may be found in the phase II literature. Thall et al[60] present two phase II trial designs that use Bayesian hierarchical models in settings where the disease has multiple subtypes. The authors include a thorough description of how the prior distributions were elicited. Discussion regarding how to elicit information to select a prior, as well as the use of informative and non-informative priors may also be found in Biswas.[61] A phase II trial presented in Maki et al[62] describes the utility function used in that trial and its clinical rationale. |
| | 3. Published Guidance: This minimum standard is addressed or supported by the following guidances:<br><br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials.[16]<br><br>This minimum standard is not addressed by<br><br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14] |
| MC Key Criteria: | 4. Contribution to Patient Centeredness: This standard does not address patient centeredness.<br>5. Contribution to Scientific Rigor: Detailed specification of the components of the Bayesian design, |

| | |
|---|---|
| Rationale for and against adoption of the proposed standard | along with the rationale for the statistical approach, enables effective peer review of the proposed design and of the reported results of the trial.<br>6. Contribution to Transparency: Complete specification of the Bayesian design will help stakeholders understand the trial and evaluate the appropriateness of the design.<br>7. Empirical Evidence: No empirical evidence specifically supports this minimum standard. |
| Additional considerations | 8. Degree of Implementation Issues: Peer review journals vary substantially regarding the level of detail related to the design of Bayesian clinical trials that is required in submissions and included in published reports. |
| | 9. Other Considerations: A strength of the Bayesian approach is the ability to incorporate prior or external information, often derived from trials involving participants who are similar to but not exchangeable with those in the planned trial (e.g., enrolled at different sites, treated at an earlier time). Such information can be incorporated via a hierarchical model—for example assuming exchangeability at the level of the trial—or by down weighting the contribution of the prior information. When such approaches are used, the structure for the incorporation of the external information and any implied exchangeability assumptions should be clearly defined. This information may be contained within the trial protocol or, preferably in a separate statistical analysis plan (SAP).<br>10. Implication of lack of adherence to standard: Non-adherence to this standard may result in uncertainty regarding the proposed or implemented trial design or, further, difficulty in assessing whether the trial was implemented as originally designed (see Minimum Standard 1). |

### *Minimum Standard 4. Communication and Vetting of Trial Design with Key Stakeholders*

Adaptive clinical trials provide flexibility that allows for designs to be better tailored to meet the specific needs of affected communities. Thus, greater importance should be placed on ensuring that key stakeholders understand a proposed adaptive design, whether it is simple or complex. Key stakeholders may include investigators, institutional review boards, DSMBs, patient representatives and patient advocacy groups, potential participants, members of the community from which participants will be drawn, and funding and regulatory agencies. Designs for adaptive RCTs should be vetted with IRBs, funding agencies, investigators, and potential participants or their representatives. Vetting of

regulatory agencies is important if the trial is intended to be a registration trial. Vetting with other interested parties may also be worthwhile; including the multiple perspectives of all key stakeholders may result in a better overall research effort. Acceptance and even enthusiasm for a design from patient representatives and patient advocates may make it easier to meet a trial's accrual goals.

I-SPY 2[10, 63] is an example of a trial in which patient advocates were intimately involved from the inception of the design of the trial and participants have even helped to publicize the trial (http://online.wsj.com/article/SB10001424052748703882404575520190576846812.html). A public website, http://ispy2.org, is available to key stakeholders and others to obtain additional information about the trial.

| Section | Minimum Standard 4: Communication and vetting of trial design with key stakeholders |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: Because adaptive trial designs are more complex and more flexible than traditional designs, there is a greater ability to tailor the design to meet needs of affected communities and a greater need to ensure key stakeholders understand the design. Thus, **adaptive designs should be thoroughly vetted with all key stakeholders with regards to the overall acceptability of the design and to the ability of the design to address important primary and secondary aims**. Key stakeholders may include investigators, institutional review boards, data safety monitoring boards, patient representatives and patient groups, potential participants, members of the community from which participants will be drawn, and funding and regulatory agencies. |
| | 2. Current Practice and Examples: The traditional trial design process has largely been a collaborative effort between scientific and clinical domain experts and statisticians. However, in some areas (e.g., oncology and trials requiring an emergency exception from informed consent) it is common practice to include patient representatives in the design process. Little published literature exists regarding these practices, however. The adaptive design process is ideally a collaborative effort in which statisticians have a responsibility to create designs that address the clinical objectives.[27] All affected stakeholders should understand the benefits and limitations of the trial design.[18] Education may be required for effective vetting of stakeholder groups both within and outside the trial's sponsor (e.g., investigators, ethics committees, health authorities, and journal editors).[30] Those providing oversight of the trial should also be familiar with the adaptive components of the design[35] (see Minimum Standard 7).<br><br>An example trial where the full range of key stakeholders were involved with the design is I-SPY2.[10, 63] "I-SPY 2 is an Innovative public-private collaboration that combines Personalized |

| | |
|---|---|
| | Medicine & Novel Trial Design to develop new cancer treatments much faster and for much less cost" (http://ispy2.org/). It is an adaptively randomized phase II drug screening trial of neoadjuvant therapy for women with high-risk, non-metastatic breast cancer that attempts to match patients with therapies. "I-SPY 2 is sponsored by the Biomarkers Consortium, a unique partnership led by the Foundation for the National Institutes of Health (FNIH), which includes the Food and Drug Administration (FDA), the National Institutes of Health (NIH), and a large number of partners from major pharmaceutical companies, leading academic medical centers, and non-profit and patient advocacy groups" (http://ispy2.org). To enhance communication and facilitate getting the trial activated, the chairs of the Institutional Review Boards from the participating centers attended a special two-day meeting in which the I-SPY 2 trial design was vetted with them. |
| | 3. Published Guidance: This minimum standard is indirectly addressed or supported by the following guidances:<br><br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13]<br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials;[16] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14]<br><br>All three guidances suggest that designs be vetted with appropriate regulatory agencies. The following articles support the broader range of stakeholders indicated in this minimum standard: Gaydos[18] and Quinlan.[30] |
| MC Key Criteria: Rationale for and against adoption of the proposed standard | 4. Contribution to Patient Centeredness: Involving patient advocates, people at risk, and community representatives in the design process, and clearly communicating the final adaptive trial design to those groups, helps to ensure the patient centeredness of the design process and the final design itself. The ability to create a more patient-centered trial is a key strength of the adaptive approach. As suggested above, patient advocates were intimately involved from the inception of I-SPY 2. Indeed, trial participants have been involved in publicizing the trial (http://online.wsj.com/article/SB10001424052748703882404575520190576846812.html). |
| | 5. Contribution to Scientific Rigor: Incorporating multiple perspectives, including those of IRBs, regulatory agencies, and funding or sponsoring organizations, helps to ensure the scientific rigor of the resulting trial design. |
| | 6. Contribution to Transparency: Involvement of key stakeholders in the design process substantially contributes to the overall transparency of the research effort. |
| | 7. Empirical Evidence: There are many examples of trials that were designed from one perspective (e.g., within a subspecialty) and despite showing positive results failed to influence clinical care or patient outcomes to the extent expected because the intervention needed to be implemented in another setting, patient population, or by a different set of providers (e.g., primary care versus emergency department physicians). Effective vetting among those working in the target setting, or with those in the target population, can mitigate this problem. |
| Additional considerations | 8. Degree of Implementation Issues: The degree of involvement of key stakeholders in the trial design process is variable and not well documented in the published literature. Communication of the proposed trial design to potential participants and patient groups is common in some areas (e.g., oncology) and rare in others. |

| | |
|---|---|
| | 9. Other Considerations: None.<br>10. Implication of lack of adherence to standard: Lack of adherence to this standard could make the proposed trial design unacceptable to patient groups, IRBs, or regulatory agencies; fail to address important patient-centered goals (e.g., quality of life); and be difficult to implement or fail to meet enrollment targets. |

### *Minimum Standard 5. Ensure Clinical Trial Infrastructure is Adequate to Support Planned Adaptation(s)*

An often-overlooked component of adaptive trial research proposals, whether simple or complex, is evidence that the trial infrastructure is able to support the planned adaptations. More complex designs may require more complex infrastructures. Reviewers of proposals for adaptive RCTs should consider not only the trial design, but also the adequacy of logistical support and trial infrastructure.

An important requirement for an adaptive design is rapid access to available clinical data, at least relative to the duration of the trial. Proposals should demonstrate that the infrastructure could meet the time lines through electronic data capture, medical monitoring, data transfers, etc. The impact of delays in the availability of outcome data on trial performance should be addressed (e.g., via simulation.) If the trial adaptations involve changing the randomization probabilities, the centralized randomization processes should be described and verification given that the probabilities can be adjusted in a timely fashion.[38] If study arms that contain drugs may be dropped or added, a plan to address changing drug supply requirements should be provided.[18, 38] The infrastructure details should be included in the study protocol.

A common concern, when running any simple or complex adaptive clinical trial, is the cleanliness and completeness of the data used for adaptations. Data management procedures should allow adaptations to be performed in a timely manner.[18, 38] Requiring full data cleaning at the time of an adaptation may result in a lengthy delay or in an adaptation being performed using only a subset of data. Procedures should be in place to query and clean the outcome data needed for adaptations while the trial is ongoing so that, when the adaptation is performed, most of the available data will have been reviewed and validated.[38] However, adaptations generally do not require completely clean datasets or complete observations of all outcomes.[34, 38]

Procedures need to be in place to capture and archive the database used for each adaptation. During the trial, data are continually collected, queried, and cleaned and the database evolves over time. Archiving the database at the time of each adaptation allows for validation of the results by outside parties should that be warranted.

Testing the data management and support infrastructures for complex trials should be performed before the trial begins.[38]

| Section | Minimum Standard 5: Ensure clinical trial infrastructure is adequate to support planned adaptation(s) |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: The benefits of an adaptive clinical trial are tied to having rapid access to available clinical data, including adjudicated primary endpoints where applicable, and the ability to effectively implement the planned adaptations (e.g., changes in randomization proportions, dropping or adding study arms). **The clinical trial infrastructure, including centralized randomization, data collection related to the assessment and recording of key outcomes, data transmission procedures, and processes for implementing the adaptation (e.g., centralized, web based randomization) must be able to support the planned trial.** In simple adaptive trials, qualitative verification of the capabilities of the proposed trial infrastructure may be adequate. Trials with more complicated requirements such as frequent interim analyses require thorough testing prior to trial initiation. Such testing should involve the trial's data collection and data |

| | |
|---|---|
| | management procedures, the implementation of the adaptive algorithm, and methods for implementing the resulting adaptation(s). The impact on the trial's operating characteristics of delays in collecting and analyzing available outcome data should be assessed. |
| | 2. Current Practice and Examples: Industry-sponsored adaptive trials usually implement rapid data collection using electronic data capture.[23, 30, 34, 35, 38] Some designs require that full outcome data be available on all trial participants when making adaptations. More commonly, the trial's adaptations do not require complete outcome data or a completely clean dataset.[34] Current data management procedures are usually adequate to allow adaptations to be performed in a manner that is sufficiently timely to achieve the benefits of the adaptive approach.[18]<br><br>In many adaptive trials, commercial contract research organizations (CROs) provide the infrastructure for data collection, database management, and centralized randomization. While the internal capabilities and validation procedures of commercial CROs are largely proprietary, industry best practices include the testing and validation of both software systems and data collection and management procedures.[38] When adaptive clinical trials are not supported by commercial CROs (e.g. academic investigator-initiated clinical trials), infrastructure requirements are often underappreciated and underfunded.<br><br>The ASTIN trial[50-52] was a seamless phase II/III trial in acute ischemic stroke, with a phase II dose-finding component. Berry[50] presents a detailed account of the trial's data communication interface between the statistical center, the study centers, and the central pharmacy, including the system's requirements, development, testing, and implementation. The procedures developed to implement the adaptive algorithm were validated using artificial data prior to trial initiation. A published description of a CER trial comparing two strategies for insulin administration in hospitalized patients provides detail on the database infrastructure and enrollment procedures that will be used to run the pragmatic trial.[39] The detail provided demonstrates the ability to support the response-adaptive randomization and potential early stopping for superiority required by the design. |
| | 3. Published Guidance: This minimum standard is not addressed by the following guidances:<br><br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13]<br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials;[16] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14]<br><br>This minimum standard is supported by multiple best practices articles, including Gaydos,[18] Benda,[35] Gallo,[34] Meta,[23] He,[38] and Quinlan.[30] |
| MC Key Criteria: Rationale for and against adoption of the proposed standard | 4. Contribution to Patient Centeredness: While this minimum standard does not directly address patient centeredness, it helps to ensure that the trial is successfully implemented as designed. This increases the value of the trial to future patients and helps fulfill an obligation to subjects to maximize the scientific value of their participation. |
| | 5. Contribution to Scientific Rigor: This minimum standard helps to ensure that the trial is successfully implemented as designed. |
| | 6. Contribution to Transparency: Demonstrating the adequacy of the trial's infrastructure and |

| | |
|---|---|
| | testing associated processes will help ensure that the trial can be conducted as designed and without unnecessary delays. |
| | 7. Empirical Evidence: Researchers seldom publish their failures. There are anecdotal accounts of adaptive trials that could not be implemented or completed as designed (e.g., due to delays in data availability preventing the timely application of the adaptive algorithm, errors or delay in implementing changes to randomization tables) due to limitations in the trial infrastructure or associated processes. This minimum standard is intended to prevent such occurrences. |
| Additional considerations | 8. Degree of Implementation Issues: Because many of the known adaptive trials have been conducted by industry sponsors and supported by commercial CROs, proprietary concerns make it difficult to assess the degree of implementation of this minimum standard. In general, the validation of data management and related processes is most complete among CROs that have extensive experience in late-phase registration trials and less consistent when investigator-initiated trials are conducted without CRO support. |
| | 9. Other Considerations: During conduct of the trial, a record or "snapshot" of the database triggering each adaptation should be archived. Because the trial data are continually undergoing queries and revision while the trial is ongoing, these records allow subsequent validation that the adaptations were properly implemented based on the data available at the time.<br>10. Implication of lack of adherence to standard: Irrespective of a trial's adaptive nature, DSMBs have stopped accrual when efficacy or safety data are not sufficiently up to date. Lack of adherence to this standard will increase the probability that the trial cannot be successfully conducted or that there will be unnecessary delays in trial initiation, data availability, the conduct of planned interim analyses, or other aspects of trial implementation. |

### *Minimum Standard 6. Consideration of Operational Bias*

Operational bias results when information from an ongoing trial causes changes to the participant pool, investigator behavior, or other clinical aspects that affect the conduct of the trial in such a way that conclusions about important efficacy or safety parameters are biased. For example, in a 3-armed trial in which one arm will be dropped in midcourse, an investigator might not initially enroll older patients, believing that they are not good candidates for one of the arms. But if that arm is dropped, then the investigator may start to include older patients and the patient population will have changed over the course of the trial. Operational bias is different than statistical bias, the systematic bias caused by modeling and design features. The possibility of operational bias is inevitable during an adaptive

clinical trial—in both complex and simple adaptive designs—and sources of operational bias should be considered.

In an adaptive design there is heightened concern for operational bias because there are more "moving parts" and the knowledge of adaptations may alter the behavior of key stakeholders. For example, simply knowing that a trial continues after an interim analysis or DSMB meeting releases information, presumably limited, to blinded parties.

The extent and effect of operational bias in practice are not well known or well understood. During the design phase, potential sources of operational bias should be identified, their potential effect quantified, and, if possible, mitigated. Any actions proposed to reduce the potential for operational bias should be prespecified and documented. This may include, for example, procedures to disguise allocation codes or drug kit labels to avoid even partial unblinding of adaptive decisions, such as shifts in randomization probabilities or arm dropping/adding.[38]

Some trialists might avoid using adaptive designs to avoid the potential for operational bias. However, designing any clinical trial involves balancing benefits and risks. The decision regarding whether to use an adaptive design should weigh the risks of bias, both operational and statistical, against the possible benefits that adaptive designs have to offer[28] such as reduced trial size and duration, or the reduction of the number of participants assigned to ineffective treatment regimens.

Mitigation of operational bias in an adaptive trial often requires controlling leakage of information beyond the DSMB's sphere of confidentiality.[28, 29] One potential method for controlling leakage is to establish strict communication

structures for the dissemination of adaptation decisions from the individuals who view unblinded data, e.g. members of DSMBs.[38] The details of who is allowed to know what information, at what time, should be considered and decided before the trial begins. A potential strategy to limit the value of information stemming from a trial—such as knowledge that a treatment arm has been dropped—may be to limit the details of the adaptations that are included in protocols and SAPs. As suggested by Gallo,[28] this information could be included in a separate document that has very limited distribution, recognizing that such limitations may inhibit transparency.

Operational bias in adaptive RCTs is a potential problem that is not well understood. The goal of this minimum standard is not to completely remove operational bias, and, indeed, complete removal may not be possible. The goal is to ensure that investigators carefully consider the potential for operational bias, include steps to minimize it when possible, and recognize the possibility of such bias when interpreting trial results.

| Section | Minimum Standard 6: Consideration of operational bias |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: Some information about treatment effects becomes available during the conduct of every clinical trial. The mere fact that a trial is continuing after an interim analysis usually means that the experimental treatment is not performing either exceptionally well or exceptionally poorly. Knowledge of the adaptive algorithm, combined with knowledge of adaptations that have occurred, can provide information regarding the observed treatment effect. Some adaptive designs are particularly susceptible to such "information leakage." For example, if assignment probabilities depend on the current relative efficacy and safety performance of treatments being evaluated, knowledge of those probabilities would be very revealing. Operational bias results when information from an ongoing trial causes changes to the participant pool, investigator behavior, or other clinical aspects that affect the conduct of the trial in such a way that conclusions about important efficacy or safety parameters are biased. For example, investigators might stop offering a trial to patients having particular characteristics because they know the patient would be more likely to get a particular treatment. Or sponsors might limit funding to the trial because they are vested in a treatment that seems to be losing. **Potential sources of operational bias for the proposed trial design should be identified and the consequences of such biases quantified and addressed to the extent possible.** Depending on |

| | |
|---|---|
| | the trial characteristics, some "leakage" may be inevitable and it may not be harmful; however, sources of potential bias and procedures to be used during the trial to mitigate their effects should be documented at the design phase. To enable the evaluation of potential operational bias, standard operating procedures should be developed, documented, and implemented for the conduct of interim analyses and the communication of results. The number of study personnel with access to unblinded data or knowledge of implemented adaptations should be limited and communication between such personnel and investigators and others involved in the trial should be limited or eliminated. |
| | 2. Current Practice and Examples: The effects and extent of operational bias are not well known or well understood. Some investigators will choose to avoid adaptive designs entirely for fear that the perception of possible operational biases will limit the trial's credibility. But concerns for operational bias should be weighed against the advantages that adaptive designs have to offer.[28] Practices to limit operational bias may include establishing communication firewalls for disseminating information about adaptation decisions from groups viewing unblinded data (i.e. DSMBs, independent statisticians) or strategies to disguise allocation codes or drug kit labels.[38] These procedures are put in place at the design stage with the goal of minimizing leakage of information.[28, 29, 38] The potential for operational bias has been reduced in practice by limiting the information provided to study participants and clinicians regarding the adaptations actually implemented. Opinions vary regarding the amount of detail on the adaptations that should be included in protocols, informed consent documents, or other material distributed to investigators and study participants. Some sponsors provide the details of the adaptive algorithm in a document other than the protocol that has limited distribution. Limitations on the amount of detail provided reduce the potential for operational bias but also decrease transparency. Indeed, in some settings, it may not be ethical to withhold such information from participants or other stakeholders.[28]<br><br>Knowledge of decisions resulting from interim analyses of unblinded results conveys information about accumulating data, even to those who may be blinded to treatment assignments. This leakage of information is not limited to newer adaptive designs and also occurs with simple groups sequential designs as well.[19, 28, 29] NSABP C-08, for example, was a randomized trial comparing standard chemotherapy to standard chemotherapy plus bevacizumab in patients with stage II or III colon cancer.[64] Interim analyses were conducted according a group sequential design and it was publicly announced that the trial had not yet stopped at each analysis. Such interim analyses and the associated announcements are typically viewed as acceptable with regard to the amount of information leaked. However, such information is useful in predicting the probability of success of an ongoing clinical trial and these predictions are of interest to the scientific community, the media, pharmaceutical companies, the trial sponsor, and investors. As the NSABP C-08 study was ongoing, Roche was in negotiations to purchase a portion of Genentech, the makers of bevacizumab. Based upon the fact the trial had not yet stopped, both entities estimated the probability of the trial's eventual success when negotiating the price per share (http://www.techzone360.com/news/2009/02/10/3979331.htm). If the study had been successful, Genentech's worth would have increased beyond what Roche paid. (Andrew Pollack "Avastin Falls Short in Test as Colon Cancer Medicine" New York Times April 22, 2009. http://www.nytimes.com/2009/04/23/health/23avastin.html?_r=1&hpw. Last Accessed Nov 4, 2009).<br><br>An example of an adaptive trial with the potential for operational bias is I-SPY 2.[10, 63] Therapeutic assignment probabilities vary depending on patient subtype and on how well the therapies are |

| | |
|---|---|
| | performing for patients with each tumor subtype. While the therapies are assigned randomly, the assigned therapy is known to both investigator and patient, as is the patient's outcome. Because the investigators know the tumor subtypes of their patients and the chronology of therapies these patients have received, as well as the outcomes, the investigator could choose to not offer the trial to a patient based on her subtype. In the extreme, an investigator could encourage a patient to drop from the trial if she is assigned a therapy different from other patients having the same subtype. However, there are 20 centers accruing patients in the trial and no single investigator is likely to have enough information to enable inferences regarding which therapies are better for which patients. In terms of mitigating the potential for operational bias, the investigators participate in monthly conference calls wherein they are periodically reminded that they do not have sufficient patient numbers to draw conclusions about therapeutic effects by patient subtype or for the patient population as a whole. Moreover, in this particular trial, all patients receive standard therapy in addition to any experimental drug, so there is little motivation to seek a more promising alternative to any therapy assigned in the trial.<br><br>Weiss et al[65] describes a two-phase, adaptive, sequential treatment design for patients with prescription opioid dependence. Entry into the second phase was based on failure in the first stage. Thus, if patients knew the entry criteria and desired additional treatment, they could falsify their self-reported measures; masking the entry criteria from patients was essential to ensure the integrity of the study. To ensure blinding, the investigators performed initial and booster training sessions with study staff, including role-playing scenarios with staff, to illustrate how to respond to questions from patients.<br><br>Fiore et al[39] propose a Bayesian adaptive comparative-effectiveness trial comparing two strategies for insulin administration in hospitalized patients. The trial is an open-label trial because blinding of treatment is not feasible. The authors consider that the lack of blinding could result in operational bias and discuss this limitation. |
| | 3. Published Guidance: This minimum standard is addressed or supported by the following guidances:<br><br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13]<br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials;[16] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14] |
| MC Key Criteria: Rationale for and against adoption of the proposed standard | 4. Contribution to Patient Centeredness: This minimum standard does not address patient centeredness. |
| | 5. Contribution to Scientific Rigor: Operational bias due to changes in study conduct or adaptations in response to accumulating data, especially if the adaptations are openly disseminated, may result in inflation of type I error, loss of statistical power, and biased estimates of overall treatment effects or outcomes. |
| | 6. Contribution to Transparency: This minimum standard does not address transparency and, further, approaches intended to limit operational bias may require less transparency (e.g., not providing details of adaptive algorithms to trial participants). |
| | 7. Empirical Evidence and Theoretical Basis: Empirical evidence of operational bias or lack thereof is |

| | |
|---|---|
| | sparse, representing a gap in the available literature. |
| Additional considerations | 8. Degree of Implementation Issues: Careful attention to controlling operational bias is common in phase III trials in the industry setting. Concern regarding the possibility of and the effects of operational biases vary in academic, investigator-initiated adaptive trials. |
| | 9. Other Considerations: The effects of operational bias and methods to mitigate them are not well understood. Indeed, eliminating such biases may be neither possible nor desirable because doing so may also eliminate the advantages of the adaptive design. At present, there is little experience with quantifying or managing operational bias. As opposed to statistical biases, the magnitude and direction of operational bias is difficult or impossible to know and so quantitative adjustments are not possible.<br>10. Implication of lack of adherence to standard: Failure to recognize the possibility of operational bias or to reduce such bias may affect the trial's credibility and impact. |

### *Minimum Standard 7. Ensure Proper Oversight of Adaptive Randomized Clinical Trial*

Oversight of an ongoing complex adaptive RCT is necessary to ensure that the algorithm is functioning properly and that the trial is being conducted as planned. Because such oversight requires unblinding to treatment assignment, a body independent from the sponsor and free from substantial conflicts of interest should be charged with this oversight. The oversight body must be familiar with complex adaptive trial designs and may require special expertise or education.

In trials with a data safety and monitoring board (DSMB), it is common and natural that the DSMB function as the oversight body. In trials without a DSMB, the oversight body may be a single statistician or a small committee. The additional oversight of the algorithm does not affect the usual DSMB responsibilities of ensuring participant safety and trial integrity.

The oversight body should not modify the trial design except to ensure participant safety, because ad hoc design modifications could affect the statistical

properties of the trial. It is critical for the credibility of the trial that the oversight body understands that its role does not involve redesigning the study but, instead, is to ensure the trial is implemented as planned. Sponsor or investigator input into the adaptive decisions may be desired because of commercial, business, or funding implications of adaptations. However, any such involvement of sponsors or investigators requires concordance of regulatory authorities and the processes to be followed should be documented in advance.

| Section | Minimum Standard 7: Ensure proper oversight of adaptive randomized clinical trial |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: During the conduct of an adaptive clinical trial, the proper functioning of the adaptive algorithm(s) should be monitored. **A body with the necessary expertise, access to data, and freedom from substantial conflicts of interest should be charged with ensuring that the trial is being conducted in accord with the adaptive algorithm.** The members of the body should be knowledgeable regarding adaptive design in general and regarding the specifics of the adaptive design used for the current trial; in some cases of complex designs, specific educational efforts will be necessary (see Minimum Standard 4). To preserve the designed operating characteristics of the trial, the body cannot modify the study design except for situations involving participant safety. The overseeing body need not be the data and safety monitoring board (DSMB) for the trial, if one exists, although this option is recommended. If a DSMB is the overseeing body, the DSMB's responsibility for ensuring participant safety and trial integrity is not different from that of a non-adaptive trial. The overseeing body and sponsor should discuss and come to an agreement on responsibilities and authority before any unblinded data are viewed, or the adaptive algorithm is invoked, and this agreement should be clearly documented (e.g., in a DSMB Charter). If oversight of the implementation of the adaptive design requires sponsor input (e.g., because of specialized expertise held only by the sponsor), the number of sponsor personnel participating in this process should be limited, the minimum necessary amount of information should be provided, and unblinded sponsor personnel should be distanced and firewalled from the clinical study team. |
| | 2. Current Practice and Examples: Practices regarding the oversight of clinical trials vary considerably, as does the complexity of the trials themselves. Most potential DSMB or oversight body members are familiar with simple group sequential designs, while more complex adaptive designs represent new territory. At present, few members of DSMBs are adequately prepared to monitor complex adaptive designs. In the worst of scenarios, when charged with monitoring an adaptive trial, DSMBs view their role as devising and making adaptations. This is not an appropriate role for the oversight body,[35] unless the revision relates to safety of the trial participants. The consequences of such ad hoc trial modifications can be disastrous. Even in group sequential designs the current practice is that DSMB members often view their roles as using their own discretion (having seen the totality of the data) in conjunction with observing a protocol defined stopping rule. Failing to stop the trial in such a circumstance may be viewed as a protocol violation and DSMBs should make such recommendations only with strong reasons for |

| | |
|---|---|
| | doing so. When the DSMB is charged with monitoring the function of the adaptive design algorithm they require specialized expertise in adaptive designs.[29, 30] Typically, a statistician who understands the design (and might even be the trial designer) advises the oversight body or DSMB on whether the algorithm is performing as planned, and on whether there are circumstances arising in the actual trial for which the algorithm was unprepared.[18, 28, 34, 35] This statistician reviews unblinded trial information and thus must be firewalled and distanced from the study trial team.[28] Sponsor involvement in oversight bodies is not typical, but the preferences of the sponsor could be relevant in implementing some adaptive decisions, for example, the doses to carry forward into phase III from a seamless phase II/III design.[18] Sponsor involvement should have a convincing rationale and should be limited in terms of scope and the number of representatives.[18, 28, 34]<br><br>An example trial is ASTIN, a seamless phase II/III trial in acute ischemic stroke, with a phase II dose-finding component.[50-52] The independent data monitoring committee (IDMC) served as the oversight body and received weekly updates of the probability that the trial should be terminated for futility or success. The IDMC was responsible for monitoring the performance of the algorithm and confirming decisions to continue or stop the study, in addition to their standard responsibilities of ensuring the safety of the participants and the integrity of the study. A member of the IDMC was an expert independent Bayesian statistician who was intimately involved with monitoring the algorithm's performance.<br><br>Additional examples are found in Barnes et al[66] and Haley et al.[67] The adaptive seamless design reported by Barnes et al[66] required that the data monitoring committee function independently; further, the committee's responsibilities were predefined and their role in evaluating the interim results was clear. The phase IIB/III trial reported by Haley et al[67] included a DSMB that was charged with reviewing the progress of the dose-selection procedure in the Phase IIB component. |
| | 3. Published Guidance: This minimum standard is partially addressed or supported by the following guidances:<br><br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14]<br><br>This minimum standard is not addressed by the following guidance document:<br><br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials.[16]<br><br>The following publications also support this minimum standard: Gaydos,[18] Gallo,[28, 34] and Benda.[35] |
| MC Key Criteria: Rationale for and against adoption of the proposed | 4. Contribution to Patient Centeredness: This standard does not address patient centeredness.<br>5. Contribution to Scientific Rigor: Establishment of an independent oversight body or DSMB to monitor the implementation of the adaptive design will help assure the scientific credibility of the trial. This approach also allows the trial organizers and sponsor to remain blinded to accumulating trial results, allowing them to make necessary trial amendments without potentially biasing knowledge of interim results. |

| | |
|---|---|
| standard | 6. Contribution to Transparency: This minimum standard does not contribute to the transparency of the trial. |
| | 7. Empirical Evidence: There are many examples in which the presence of an independent oversight body such as a DSMB resulted in the protection of human subjects from unnecessary risk. See Lewis et al[68] as an example in the setting of a group sequential trial design that was stopped prior to the first planned interim analysis. Published evidence of the importance of an independent oversight body in the setting of a complex adaptive clinical trial is sparse. However, anecdotal evidence demonstrates that DSMBs find it essential to have someone reporting to them who is knowledgeable about how the design is performing and how it should be performing. In addition, unpublished accounts of adaptive trials describe errors in implementing the adaptive algorithm (e.g., errors in implementing response-adaptive randomization, missed interim analyses). The likelihood of detecting such errors is enhanced when there is a knowledgeable and independent oversight body. |
| Additional considerations | 8. Degree of Implementation Issues: It is common to have at least one adaptive design expert on the oversight body or DSMB of complex adaptive trials, in both industry and academic settings. |
| | 9. Other Considerations: None.<br>10. Implication of lack of adherence to standard: Lack of adherence to this standard, namely conducting an adaptive clinical trial without effective and knowledgeable oversight, may result in the possibility of failing to discover substantial implementation problems or threats to participant safety in a timely manner. For example, even if computer algorithms were failure-free, human error in programming a computer algorithm can occur or issues with implementing the adaptations may occur. The algorithm may be assigning participants to the wrong therapy, or it may be continuing the trial to the detriment of the trial participants. Finally, if there is a malfunction in applying the appropriate adaptations, the trial results may be uninterpretable. |

### *Minimum Standard 8. The Reporting of Adaptive Randomized Clinical Trials Should be Consistent with the CONSORT Statement*

The CONSORT guidelines are intended to improve the reporting of RCTs by establishing a minimum set of reporting standards. The reporting of results from both simple and complex adaptive clinical trials should be consistent with the CONSORT standard. However, the current version, CONSORT 2010,[69] is limited with respect to adaptive design components. We propose several extensions to the CONSORT checklist items to address the following predefined adaptations:

- Response adaptive randomization
- Arm dropping or adding
- Stopping for futility or early success (including continuing follow-up based on prediction of eventual success)

- Sample size re-estimation
- Transitioning of stages (e.g. seamless Phase II/III designs)
- Modification of inclusion/exclusion to focus on responding subpopulations (e.g., enrichment designs).

Minimum standard 8 provides the overall rationale for the proposed extension to the CONSORT statement. Table 9, the supplemental table for minimum standard 8, identifies the corresponding CONSORT sections where the design detail, and resulting adaptations would be described for each of the six selected adaptations. Details regarding each adaptation should be reported, including the populations included in the corresponding interim analysis and the key outcomes that were observed. Reports must make clear which participants were included in all analyses, including both interim and final analyses.

Our list includes the most common adaptations;[10] other adaptations are possible and the trialist should use judgment as to where in the CONSORT structure the design details and resulting adaptations should be described. This minimum standard is specifically for pre-planned adaptations and applies to both simple and complex designs. Ad hoc changes (e.g. unplanned sample size re-estimation) should be clearly identified as not prospectively defined.

| Section | Minimum Standard 8. The reporting of adaptive randomized clinical trials should be consistent with the CONSORT statement |
|---|---|
| Identification and background of the proposed standard | 1. Description of Standard: The CONSORT Statement[69] provides a structure intended to improve the reporting of clinical trials by ensuring the trial's design, results, analysis, and interpretation are reported completely and clearly. An adaptive clinical trial may contain components, both in the trial's design and its results, which are not explicitly addressed in the CONSORT Statement. This standard provides recommendations regarding how these components should be reported within the current CONSORT structure.<br><br>This standard addresses the following potential adaptations and provides suggested "locations" for their reporting within the CONSORT reporting structure:<br><br>• Adaptation of randomization probabilities (CONSORT Sections 8b and 13a);<br>• Dropping or adding study arms (CONSORT Sections 7b and 13a);<br>• Interim stopping for futility and superiority (CONSORT Sections 7b and 14b);<br>• Sample size re-estimation (CONSORT Sections 7a and 7b);<br>• Transitioning of stages (e.g. seamless Phase II/III designs)(CONSORT Sections 3a, 7a, 7b, and 16); and<br>• Modification of inclusion and exclusion criterion (CONSORT Sections 4a, 13a).<br><br>CONSORT Sections 16, 20, and 21 may also be expanded to report additional aspects of an adaptive trial. A supplementary table provides additional details regarding the incorporation of these adaptive trial features into the CONSORT reporting structure.<br><br>If the trial incorporates adaptations other than those listed above, the authors should use their judgment as to where in the CONSORT structure to include both design details and the associated results. All possible adaptations included in the prospective design, even if they did not occur, should be included in the report. |
| | 2. Current Practice and Examples: Currently, many adaptive trials are reported in accordance with the CONSORT guidelines and some adaptations are partially addressed by the current CONSORT statement (e.g., the reporting of interim analyses and stopping guidelines in CONSORT Section 7b). However, there is no detailed published guidance on how best to incorporate other common adaptive features into this reporting structure.<br><br>As an example, Cohen et al[47] closely followed the CONSORT reporting structure in their publication of the results of a group sequential trial examining early antiretroviral therapy for the prevention of HIV infection. |
| | 3. Published Guidance(s): This minimum standard is not addressed or supported by these guidances:<br><br>• Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance);[13]<br>• Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials;[16] and<br>• Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design.[14] |

| | |
|---|---|
| MC Key Criteria: Rationale for and against adoption of the proposed standard | 4. Contribution to Patient Centeredness: This minimum standard does not directly address patient centeredness. However, it helps to ensure that the trial is completely and clearly reported. This increases the likely value of the trial to future patients and helps fulfill an obligation to study participants to maximize their contribution to medical science. |
| | 5. Contribution to Scientific Rigor: Scientific evaluation of trial quality and the clinical impact of a clinical trial both hinge on the clear and complete reporting of the trial's design, implementation, and results. Reporting adaptive clinical trials within the existing and proven CONSORT structure helps to ensure the clear and complete reporting of the trial's design, implementation, and results. |
| | 6. Contribution to Transparency: Following this standard will help to ensure the trial design and all adaptations are well described. |
| | 7. Empirical Evidence and Theoretical Basis: Prior experience has demonstrated that the complete and clear reporting of a clinical trial's design and results is necessary if readers of the report are to be able to accurately judge the trial's quality and clinical importance. This is likely to be particularly important for the reporting of adaptive trials because of their greater potential complexity. |
| Additional considerations | 8. Degree of Implementation Issues: Journal editors and peer reviewers are generally very familiar with the CONSORT reporting standard as applied to non-adaptive clinical trials. Journal editors and peer-reviewers of patient-centered research conducted using adaptive approaches should be made aware of this standard and work with authors to ensure adaptive trials are clearly and completely reported as described above and in the accompanying supplemental table. |
| | 9. Other Considerations: None.<br>10. Implication of lack of adherence to standard: Lack of adherence to this standard will decrease transparency, make it more difficult for readers to judge the quality of and clinical impact of an adaptive trial, and may decrease the likelihood that the trial results influence clinical practice. |

**Table 9. Supplemental Table for Minimum Standard 8**

| CONSORT Section | Topic Area | Standard CONSORT description | Extension for adaptive trials |
|---|---|---|---|
| • **Adaptation of randomization probabilities (e.g. response-adaptive randomization or balance of covariates with adaptive randomization)** | | | |
| 8b | Randomization—sequence generation | Type of randomization; details of any restriction (such as blocking and block size) | Adaptive randomization scheme describing frequency, timing and algorithm for randomization updates. |
| 13a | Participant flow (a diagram is strongly recommended) | For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome | For each arm or relevant population group a summary of how participants were allocated over the course of the trial should be provided. |
| • **Dropping or adding study arms** | | | |
| 7b | Sample Size | When applicable, explanation of any interim analyses and stopping guidelines | Arm dropping/adding algorithm with timing and frequency of analyses. Specify whether sample size determination is overall or by arm. |
| 13a | Participant flow (a diagram is strongly recommended) | For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome | Times at which arms were added or dropped and sample sizes at those times. |
| • **Interim monitoring for futility and/or superiority** | | | |
| 7b | Sample Size | When applicable, explanation of any interim analyses and stopping guidelines | Timing and number of possible interim analyses, stopping thresholds (e.g. group sequential design bounds). In the Bayesian approach, whether stopping based on current probability distributions or predictive probabilities. |
| 14b | Recruitment | Why the trial ended or was stopped | The appropriate statistical summary of the stopping analysis. In the Bayesian approach when stopping accrual is based on a predictive probability, whether that prediction was accurate. |

| | | | |
|---|---|---|---|
| • **Sample size re-estimation** | | | |
| 7a | Sample Size (as planned) | How sample size was determined | Pre-specified plans for sample-size re-estimation, timing of looks and range of potential sample size updates. |
| 7b | Sample Size (after adaptations) | When applicable, explanation of any interim analyses and stopping guidelines | Results of interim analyses giving rise to sample size revisions. Should clearly state if sample size re-estimation was post hoc (not predefined in protocol). |
| • **Transition of Stages (e.g. Seamless phase II/III)** | | | |
| 3a | Trial design | Description of trial design (such as parallel, factorial) including allocation ratio | Decision rules for transitioning from one stage to another including timing and frequency of analyses. |
| 7a | Sample Size | How sample size was determined | Sample size range for each stage and overall. Should clearly state whether participants enrolled during the first stage are included in the primary analysis of the trial. |
| 7b | Interim analyses | When applicable, explanation of any interim analyses and stopping guidelines | Statistic leading to transition, timing of transition if multiple interim analyses. |
| 16 | Numbers analyzed | For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups | The progress of the trial and whether (and why) the transition to subsequent stage occurred. |
| • **Modification of inclusion and exclusion criteria** | | | |
| 4a | Participants | Eligibility criteria for participants | Prespecified rules for modifying eligibility criteria. |
| 13a | Participant flow (a diagram is strongly recommended) | For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome | Number of participants per arm, and population group, randomized before and after updates. |

| | | | |
|---|---|---|---|
| • **General reporting standards that apply to all adaptations** | | | |
| 16 | Numbers analyzed | For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups | Clearly define the number of participants used in each arm for each analysis. For instance if arms are dropped is a dropped arm compared vs. the whole control group, or just the control group at the time of the arm-dropping. |
| 20 | Limitations | Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses | Discuss any planned interim analyses not performed, or performed that led to adaptation that was not implemented (e.g. stopping rule not enforced) and explanation. Discuss any unplanned (ad hoc) protocol revisions. Discuss lessons learned and value of adaptive design in current trial. |
| 21 | Generalizability | Generalizability (external validity, applicability) of the trial findings | Discuss to whom the results should be generalized based upon adaptations or how this study should inform future studies. |

## Gaps in Knowledge

In considering gaps in knowledge related to the planning, conduct, and analysis of adaptive RCTs for PCOR, we distinguish between two types of gaps. We define "gaps in existing guidance" as methodological areas in which existing knowledge, methodological work, or experience is sufficient to reasonably define minimum standards or best practices, but those minimum standards or best practices have not yet been incorporated into published guidance documents from regulatory agencies or other policy-setting bodies. In contrast, we define "gaps in knowledge" as methodological areas in which the existing knowledge, methodological work, or experience is insufficient to reasonably define minimum standards or best practices. Gaps in knowledge can only be addressed by additional work or experience with adaptive RCTs in the PCOR setting.

Gaps in existing guidance were found in the areas of vetting of the proposed trial with key stakeholders (minimum standard 4), verifying sufficient trial infrastructure (minimum standard 5), and ensuring proper trial oversight (minimum standard 6).

Vetting of a design by key stakeholders is only indirectly supported by available guidance documents. These documents generally support vetting only by regulatory agencies, which is not unexpected because regulators wrote the documents. While published literature supports vetting proposed trial designs with additional stakeholders,[18, 30] the development of guidance documents specifically

addressing the vetting of adaptive RCTs for patient-centered outcomes research would be valuable.

None of the identified guidance documents provide recommendations for ensuring that adequate infrastructure is available to support the planned adaptive trial. However, a number of published articles address infrastructure considerations[18, 23, 30, 34, 35, 38] and should be incorporated into formal guidance documents.

Minimum standard 6, related to the proper oversight of adaptive RCTs, is indirectly supported by the FDA adaptive design draft guidance. The guidance indicates that a DSMB should not modify the design of the trial, but it contains no discussion of the role of the DSMB in monitoring the adaptive algorithm. Several best-practices papers address the need for additional monitoring for adaptive trials[18, 28, 34, 35] and could be used as the basis for guidance development.

It would be helpful to have additional published guidance addressing oversight of the adaptive algorithm, membership expertise requirements for the oversight body, and when and how sponsors should be involved in the oversight of an adaptive trial. The numbers of clinicians and statisticians with the necessary knowledge and experience to serve on oversight committees is limited and educational efforts are warranted to expand the population of potential committee members.

Additional research is needed to address knowledge related to operational bias (minimum standard 6) and CONSORT reporting standards (minimum standard 8). The effect and extent of operational bias in simple and complex adaptive clinical

trials is not well understood. However, each of the three guidance documents emphasizes the need to control operational bias. Leakage of information in adaptive clinical trials is inevitable, but it is not clear to what extent trial credibility is affected. Additional work to determine "who should know what and when" and the effects of operational bias should be performed. We were unable to find written standards for CONSORT reporting of the most common adaptations. We propose extensions to the CONSORT 2010 standard in this report, but further experience will be needed to determine if these are sufficient or if further work on reporting standards is necessary.

## Discussion

Adaptive clinical trials should adhere to the same principles of good trial design, conduct, analysis, and reporting as any other clinical trial. However, applying these principles is more involved for adaptive clinical trial designs than for traditional designs. Adaptive designs may require variations in the development process, such as computer simulation to evaluate complex designs, or may require reporting of additional design details and their resulting analyses to fully describe the design and trial outcomes. If each of the minimum standards we propose is met, the resulting trial will have a well justified prespecified design, have undergone a thorough consideration of strengths and weaknesses, have the resources available for implementation, have proper oversight, and the results will be reported clearly and completely.

Guidances and best-practices papers suggest limiting the number of adaptations in a design, especially in confirmatory trials.[13, 14, 18, 22, 32] Some authors argue that confirmatory trials do not need many adaptations because there should be fewer questions at that stage[22] and that, as the number of adaptations increases, the trial becomes an exploratory trial rather than a confirmatory trial. However, if all minimum standards are met, then the number of adaptations need not be limited. Indeed, the benefits of adaptive designs are potentially greater when the trial is more flexible and addresses more questions.[70] Of course, each adaptation should contribute to fulfilling the trial's goals and not just exist for the sake of adapting. The overarching approach should be to use a design that most efficiently answers the research questions and treats trial participants ethically.

As we have indicated, adaptive trials are well suited for patient-centered outcome research. But to be successful, adaptive trials must be carefully constructed, properly vetted, and run according to plan. Being adaptive does not ensure that a design will be best, or even better than a more traditional design. A good adaptive design has one overarching characteristic: it is built on a theme, a set of related trial objectives, and this theme is its focus throughout. Objectives may include efficiently addressing scientific hypotheses (such as identifying a dose-response curve) or delivering good medical care to patients, including both those inside the trial as well as future patients.

Adaptive designs should be prospectively specified. This allows key stakeholders in the research, including potential trial participants, to assess its likely value. All aspects of the design should be transparent unless doing so will

introduce unacceptable risks of operational bias, and the adaptations should be made as planned. Conducting an adaptive trial requires more tools and more attention than does conducting a traditional trial. The risks may be greater as well. A poorly conducted adaptive design may be worthless. The minimum standards describe biases that are inherent to adaptive designs, although their magnitudes may be unknown and are sometimes unknowable. In poorly conducted trials, the biases may be of sufficient magnitude to jeopardize the credibility of the research.

There is a limited number of formal guidance documents related to adaptive trials. The three available guidances, the FDA draft guidance for adaptive designs,[13] the EMA reflection paper on adaptive designs in confirmatory trials,[14] and the FDA guidance on the use of Bayesian statistics[16] provide the foundation for most of our standards. In addition to these guidances, however, there are numerous best-practices and discussion papers on designing and implementing adaptive clinical trials.

We note, however, there are few published adaptive trials conducted in patient-centered outcomes research. When we searched the literature to specifically find examples, we only found two relevant publications.[39, 40] To identify trials to demonstrate compliance with components of the minimum standards, we had to extend our literature search to earlier-phase drug or device trials. Adaptive designs are more commonly used in the drug or device development process, but the same standards are applicable to PCOR as long as trial is designed with the specific goal of best answering the research question. This field remains a very active statistical research area and we anticipate numerous publications in the future.

At the present time, designing a complex adaptive RCT is a time-consuming process. Relatively few statisticians have the experience necessary to build such designs. Because these designs are new to many practitioners, there is substantial potential for mistakes to be made. We encourage clinical trialists to publish their experiences with adaptive designs for PCOR, including any barriers encountered, to improve knowledge in the area generally.

A particular adaptive trial need not be constructed using citable, previously published methodology. Unlike traditional trials whose methodologies have generally been sequentially developed, published, and implemented, more complex adaptive trial designs are often tailored to a specific research question. Therefore, a specific design may have features based upon published methodology; however, the whole of the trial may be unique. In these cases, thorough understanding of the trial's operating characteristics and validation of computer simulation code are important.

The term adaptive clinical trial design covers a wide range of designs that vary greatly in complexity. As new innovative adaptive designs are developed the potential complexity may increase. However, as these approaches become commonplace, what was considered an innovative design in 2012 may become a standard method in 2013. The standards we propose are based on the current state of the art. Our goal is to establish general fundamental principles; adherence to these principles is necessary for good science. As the field of adaptive RCT design continues to evolve these standards will need to be reevaluated and, like modern clinical trial designs, the minimum standards too will need to be adapted.

# Appendix 1: Adaptive Design Terminology

Adding and dropping arms: A trial that has interim analyses to determine whether specific treatment arms are added or removed based on the accruing safety and efficacy information.

Alpha penalty: If a clinical trial designed to test a null hypothesis has interim analyses with the possibility of rejecting the null hypothesis at a particular alpha level (type I error rate), these additional analyses inflate alpha. Controlling the overall alpha (to 0.05, say) requires decreasing the alpha used at each interim analysis. This decrease is the alpha penalty.

Biomarkers: Biomarkers are observations made on trial participants that may be correlated with clinical outcomes. Baseline biomarkers are sometimes used as covariates. Post-therapy biomarkers may be used to predict longer term outcomes based on statistical models that include both variables and that are updated based on accumulating trial data. Post-therapy biomarkers have the promise to make adaptive designs more efficient. When a biomarker is found to be correlated with longer-term endpoints, incorporating it in the design allows for adaptations to be made sooner and more efficiently. When the biomarker turns out to be unrelated or minimally predictive, then a properly modeled trial will not use the biomarker information.

Competitive platform designs: These are trials in which multiple arms are used in a competitive situation to be compared. Response adaptive randomization can be used to favor the better performing arms. Arms can be added to the platform or removed for either success or futility. Economies of scale can allow for the randomization and conclusions to apply to subgroups of participants.

Bayesian statistics: An approach based on applying Bayes theorem for updating probability distributions. This approach is useful in building adaptive designs because its inferential measures are updated naturally as information accumulates. Not all Bayesian designs are adaptive and not all adaptive designs are Bayesian. Bayesian designs that are modified to have prescribed operating characteristics are hybrid Bayesian/frequentist designs.

Complex adaptive design: One for which the operating characteristics cannot be found analytically and are not available in tables.

Covariate adaptive randomization: A technique used to randomize participants depending on a predetermined group of covariates with the goal of balancing participants with regard to those covariates. Randomization probabilities differ depending on the values of the covariates. This technique is not based on unblinded results.

Dose-escalation trials: Initial participants receive a relatively low dose and dosage increases for later participants depending on the results of lower doses. These designs typically focus on the tolerability and safety of doses of therapy rather than on efficacy.

- 3+3 Design: A particular dose-escalation design in which participants are allocated in cohorts of 3 with specific escalation rules as a function of the adverse events observed on the current cohort of 3.
- CRM design: A continual reassessment method that uses result of participants and a dose-response model to determine whether to escalate or deescalate dose in subsequent participants.

Dynamic sample-size selection:

- Stopping for success: Using interim analyses the sample size can be reduced because the trial has met, or is expected to meet, its goals.
- Stopping for futility: Using interim analyses the sample size can be reduced because the trial is unlikely to meet its goals.

Enrichment designs: A trial that has an evolving participant selection criteria based on the interim outcome information. For example, the trial may determine that treatment responders are limited to a biomarker-defined subset of participants and so eligibility is adapted to restrict to that subset.

Endpoint Selection Design: A design that selects the primary endpoint through an adaptive procedure that observes the effect for each treatment for multiple endpoints. This can be done in a seamless phase II/III design, or possibly a confirmatory trial.

Group-sequential design: A design with multiple interim analyses of the primary endpoint with the goal of stopping the trial and announcing that the results have crossed a "stopping boundary."

Interim analysis: An analysis done before the completion of a trial with the possibility of adapting the trial conduct in some fashion.

Operating (performance) characteristics of an adaptive design: Operating characteristics represent the average behavior of a design. Standard operating characteristics of a clinical trial design are power, type I error, and sample size. In an adaptive design there is a much richer set of operating characteristics (or performance characteristics). The probability of each possible adaptation taking place, such as adding an arm, dropping an arm, mean sample size per arm, the probability of making a seamless switch, and the probability of dropping a subset of the population are examples of richer adaptive design operating characteristics. Typically operating characteristics are calculated through trial simulation.

Operational bias: Operational bias results when information from the trial may cause changes to the participant pool, investigator behavior, or other clinical aspects that affect the conduct of the trial in such a way that conclusions about important efficacy and safety parameters may be biased. For example, in a 3-armed trial in which one arm will be dropped in midcourse, an investigator might not enroll older patients feeling that they are not candidates for one of the arms. But if that arm is dropped then the investigator includes older patients and so the patient population will have changed, including when making comparisons of the remaining arms.

Response adaptive randomization: Using the outcome data during the current trial to change the randomization probabilities for future participants.

Seamless-phase trials: Seamless trials are multistage trials that address different questions within each stage, but without pausing accrual between stages. Some aspect of the trial, such as treatment arms involved—randomization probabilities, participant characteristics, and accrual rate—changes between stages.

- Inferentially seamless trial: A seamless trial that includes all participants in each stage in the final analysis at the completion of all stages, usually with adjustments in type I error rate.
- Operationally seamless: A seamless trial in which the final statistical analysis includes outcome data from only the participants during that stage.

Sample size re-estimation: This technique is used at an interim analysis in the trial to modify the sample size. The modification may be based on blinded or unblinded data. The latter usually entails greater alpha penalty.

Shared-control design: A design in which multiple experimental arms are used in a single trial and are compared to a common control arm.

Simple adaptive design: One for which the operating characteristics can be found analytically or in widely available tables.

Simulating clinical trials: A procedure to evaluate performance characteristics of a particular clinical trial design. A computer program generates virtual participants according to the design. The participants' outcomes are generated under a particular scenario, with the participants having particular characteristics, including receiving treatments with a specified distribution of outcomes. Repeating this process 10,000 times, say, provides a comprehensive distribution of the possible outcomes in that scenario when using the design being evaluated. An example scenario is when all treatment arms have the same distribution of outcomes. Then the proportion of simulations in which one arm is declared to be better than others is an estimate of the type I error rate (in that scenario). Varying factors such as treatment effects, accrual rates, drop-out rates, etc., gives a comprehensive picture of the trial design. Selecting a few of the

many thousands of simulated trials and observing how the design behaves when faced with particular outcomes can be informative for the trial designers and for the key stakeholders.

Statistical bias: Statistical bias stems from the trial design and therefore can be evaluated for prospective designs and adjusted if desired. An example in trials that test hypotheses is the "statistical penalty" in the type I error rate resulting from interim analyses because of the increased opportunity for rejecting the null hypothesis.

Null hypothesis (Null space): A hypothesis to be tested within a clinical trial. Typically this is a lack of effect for the experimental agent, such as having an identical mean change to a control arm. In a non-inferiority trial this can be a hypothesis of being worse by the non-inferiority delta. In many trials there is a simple null hypothesis but a range of hypotheses that all qualify as showing "no-effect." This range of null hypotheses is referred to as the null space.

Type I Error: The probability a test concludes a positive result, when the truth is that the experimental regimen is not effective. This is the probability that the trial results in a successful result when a null hypothesis is true. This is usually a big concern in adaptive designs as the adaptations have the potential to inflate the type I error over a nominal level test (see Alpha-Inflation).

Confirmatory Trial: A trial where the goal is to demonstrate to a regulatory body that a treatment is safe and effective and should be approved for use. Typically the design is simple and allows for the confirmation of an earlier finding of a possible efficacy and safety benefit for an experimental regimen. Commonly these trials are referred to as pivotal trials (devices) and phase III trials (drugs and biologics). Additionally the FDA will refer to them as "adequate and well controlled trials."

Exploratory Trial: A trial set up mostly for learning the behavior of an experimental regimen. The goal is typically to understand different dose or exposure on the size of the efficacy of the treatment. Additionally, the tolerability or safety measures of an experimental treatment are sought. The goal is to understand safety and or efficacy. Typically these trials are not strictly type I error controlled. In drugs and biologics these are referred to as phase I or phase II trials and are referred to as pilot trials in medical devices.

# Appendix 2: PCORI Literature Review Abstraction Tool: Items Collected

*Not all items apply to all papers; denote NA where not applicable*

1. Reference title
2. Year
3. Journal
4. Authors
5. Organization of authors (FDA, NIH, academic, working group, etc.)
6. Type of paper (i.e. guidance, best-practices paper, methodology, example of implementation)
7. Peer reviewed?
8. Description of trial design
9. Types of adaptations discussed
10. Bayesian or Frequentist
11. Key recommendations/conclusions
12. Applicable for which general concept(s)
13. Trial design
14. Trial analysis
15. Trial reporting
16. Trial implementation/conduct/operational
17. Applicable for PCOR? If yes, elaborate
18. Ranking of importance/applicability of reference (scale 1-10 where 10 is important)
19. Initials of primary reviewer
20. Initials of confirmation reviewer
21. For confirmation review was there agreement on information abstracted? (Confirmer should revise with suggested changes using track changes. Primary and confirmer should discuss suggestions and come to an agreement.)

# Appendix 3. Table of Guidance Documents

| Description of guidance statements included in the recommended minimum guidelines. | | | | | | | |
|---|---|---|---|---|---|---|---|
| Guideline | Organization or Authors | Year | Program | Country or Region | Guideline subjected to independent external review? | Research Design | Description |
| Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance) | Food and Drug Administration | 2010 | Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) | USA | Yes, currently in distribution for comments | Adaptive clinical trial design for confirmatory and exploratory trials | Document provides discussion of clinical, statistical, and regulatory aspects of adaptive clinical trials; design aspects that deserve special consideration are also described. |
| Guidance for Industry and Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials | Food and Drug Administration | 2010 | Center for Devices and Radiologic Health (CDRH) | USA | Yes | Bayesian statistical methods for medical device clinical trials | Document focuses on guidance for statistical aspects of medical device trials using Bayesian methodology." |
| Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design | European Medicines Agency | 2007 | Committee for Medicinal Products for Human Use (CHMP) | Europe | Yes | Adaptive clinical trial design in confirmatory trials | Document outlines general considerations for trials with interim analyses; minimal requirements that must be fulfilled for confirmatory adaptive trials are presented. |

## Acknowledgments

This project to develop minimum standards for the design, conduct, and evaluation of adaptive randomized clinical trials was funded by the Patient-Centered Outcomes Research Institute.

# References


1. Laine, C. & Davidoff, F. Patient-centered medicine. A professional evolution. *JAMA* **275**, 152-6 (1996).
2. Clancy, C. & Collins, F.S. Patient-Centered Outcomes Research Institute: the intersection of science and health care. *Sci Transl Med* **2**, 37cm18 (2010).
3. Institutes of Medicine (IOM). Initial National Priorities for Comparative Effectiveness Research [online]. Washington, DC: Natl. Acad. Press. http://www.nap.edu/catalog/12648.html (2009).
4. Sox, H.C. & Goodman, S.N. The Methods of Comparative Effectiveness Research. *Annu Rev Public Health* (2011).
5. Berry, D.A. Bayesian approaches for comparative effectiveness research. *Clin Trials* **9**, 37-47 (2012).
6. Luce, B.R. et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med* **151**, 206-9 (2009).
7. Tunis, S.R., Stryer, D.B. & Clancy, C.M. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* **290**, 1624-32 (2003).
8. Zhou, X., Liu, S., Kim, E.S., Herbst, R.S. & Lee, J.J. Bayesian adaptive design for targeted therapy development in lung cancer--a step toward personalized medicine. *Clin Trials* **5**, 181-93 (2008).
9. Berry, D.A. Bayesian clinical trials. *Nat Rev Drug Discov* **5**, 27-36 (2006).
10. Berry, D.A. Adaptive clinical trials in oncology. *Nat Rev Clin Oncol* (2011) [Epub ahead of print 2011 Nov 8].
11. Berry, S.M., Carlin, B.P., Lee, J.J. & Muller, P. Bayesian Adaptive Methods for Clinical Trials (ed. Hall, C.) (Taylor & Francis Group, Boca Raton, FL, 2010).
12. FDA. Critical Path Opportunities Report [online], http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/UCM077254. pdf (2006).
13. Food and Drug Administration. Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Guidance). Available from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790. pdf. Accessed November 21, 2011.
14. Committee For Medicinal Products For Human Use (CHMP). Reflection Paper On Methodological Issues In Confirmatory Clinical Trials Planned With An Adaptive Design [online], http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf (2007).
15. FDA. Food And Drug Administration Modernization Act Of 1997 [online], http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FDAMA/FullTextofFDAMAlaw/default.htm (1997).

16. Food and Drug Administration. Guidance for Industry and FDA Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. Available from http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm . Accessed November 27, 2011.
17. Helfand, M. et al. A CTSA agenda to advance methods for comparative effectiveness research. *Clin Transl Sci* **4**, 188-98 (2011).
18. Gaydos, B. et al. Good practices for adaptive clinical trials in pharmaceutical product development. *Drug Inf J* **43**, 539-556 (2009).
19. Bauer, P. Methodological developments vs. regulatory requirements. *Biom J* **48**, 609-12; discussion 613-22 (2006).
20. Gallo, P. & Maurer, W. Challenges in implementing adaptive designs: comments on the viewpoints expressed by regulatory statisticians. *Biom J* **48**, 591-7; discussion 613-22 (2006).
21. Hung, H.M., O'Neill, R.T., Wang, S.J. & Lawrence, J. A regulatory view on adaptive/flexible clinical trial design. *Biom J* **48**, 565-73 (2006).
22. Koch, A. Confirmatory clinical trials with an adaptive design. *Biom J* **48**, 574-85 (2006).
23. Mehta, C.R. & Jemiai, Y. A consultant's perspective on the regulatory hurdles to adaptive trials. *Biom J* **48**, 604-8; discussion 613-22 (2006).
24. O'Neill, R.T. FDA's critical path initiative: a perspective on contributions of biostatistics. *Biom J* **48**, 559-64 (2006).
25. Schafer, H. Adaptive designs from the viewpoint of an academic biostatistician. *Biom J* **48**, 586-90; discussion 613-22 (2006).
26. Wittes, J. & Lachenbruch, P.A. Opening the adaptive toolbox. *Biom J* **48**, 598-603; discussion 613-22 (2006).
27. Dragalin, V. Adaptive designs: terminology and classification. *Drug Inf J* **40**, 425-435 (2006).
28. Gallo, P. Confidentiality and trial integrity issues for adaptive designs. *Drug Inf J* **40**, 445-450 (2006).
29. Gallo, P. & Krams, M. PhRMA working group on adaptive designs: introduction to the full white paper. *Drug Inf J* **40**, 421-423 (2006).
30. Quinlan, J. & Krams, M. Implementing adaptive designs: logistical and operational considerations. *Drug Inf J* **40**, 437-444 (2006).
31. Hung, H.M., Wang, S.J. & O'Neill, R.T. Methodological issues with adaptation of clinical trial design. *Pharm Stat* **5**, 99-107 (2006).
32. Phillips, A.J. & Keene, O.N. Adaptive designs for pivotal trials: discussion points from the PSI Adaptive Design Expert Group. *Pharm Stat* **5**, 61-6 (2006).
33. Chang, M., Chow, S.C. & Pong, A. Adaptive design in clinical research: issues, opportunities, and recommendations. *J Biopharm Stat* **16**, 299-309; discussion 311-2 (2006).
34. Gallo, P. et al. Adaptive designs in clinical drug development--an Executive Summary of the PhRMA Working Group. *J Biopharm Stat* **16**, 275-83; discussion 285-91, 293-8, 311-2 (2006).

35. Benda, N. et al. Perspectives on the use of adaptive designs in clinical trials. Part II. Panel discussion. *J Biopharm Stat* **20**, 1098-112 (2010).
36. Wang, S.J. Perspectives on the use of adaptive designs in clinical trials. Part I. Statistical considerations and issues. *J Biopharm Stat* **20**, 1090-7 (2010).
37. PhRMA Adaptive Design Working Group, Adaptive designs software requirements subgroup. Adaptive Trials - User Requirements for Statistical Simulation and Analysis Software. (2008)
38. He, W. et al. Practical considerations and strategies for executing adaptive clinical trials. *Drug Inf J* **44**, 160-174 (2012).
39. Fiore, L.D. et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin Trials* **8**, 183-95 (2011).
40. Muss, H.B. et al. Adjuvant chemotherapy in older women with early-stage breast cancer. *N Engl J Med* **360**, 2055-65 (2009).
41. White, W.B. et al. A cardiovascular safety study of LibiGel (testosterone gel) in postmenopausal women with elevated cardiovascular risk and hypoactive sexual desire disorder. *Am Heart J* **163**, 27-32 (2012).
42. Giles, F.J. et al. Adaptive randomized study of idarubicin and cytarabine alone or with interleukin-11 as induction therapy in patients aged 50 or above with acute myeloid leukemia or high-risk myelodysplastic syndromes. *Leuk Res* **29**, 649-52 (2005).
43. Goodman, S.N. Stopping at nothing? Some dilemmas of data monitoring in clinical trials. *Ann Intern Med* **146**, 882-7 (2007).
44. Goodman, S.N. Stopping trials for efficacy: an almost unbiased view. *Clin Trials* **6**, 133-5 (2009).
45. Freidlin, B. & Korn, E.L. Stopping clinical trials early for benefit: impact on estimation. *Clin Trials* **6**, 119-25 (2009).
46. Team., R.D.C. (ed. Computing, R.F.f.S.) (Vienna, Austria, 2011).
47. Cohen, M.S. et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* **365**, 493-505 (2011).
48. Hall, D.B., Meier, U. & Diener, H.C. A group sequential adaptive treatment assignment design for proof of concept and dose selection in headache trials. *Contemp Clin Trials* **26**, 349-64 (2005).
49. Chataway, J. et al. A novel adaptive design strategy increases the efficiency of clinical trials in secondary progressive multiple sclerosis. *Mult Scler* **17**, 81-8 (2011).
50. Berry, D.A. et al. in Case Studies in Bayesian Statistics V (eds. Gatsonis, C.A. et al.) 99-181 (Springer-Verlag, New York, New York, 2002).
51. Krams, M. et al. Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN): an adaptive dose-response study of UK-279,276 in acute ischemic stroke. *Stroke* **34**, 2543-8 (2003).
52. Grieve, A.P. & Krams, M. ASTIN: a Bayesian adaptive dose-response trial in acute stroke. *Clin Trials* **2**, 340-51; discussion 352-8, 364-78 (2005).
53. Bassler, D. et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* **303**, 1180-7 (2010).

54. Berry, S.M., Carlin, B.P. & Connor, J. Bias and trials stopped early for benefit. *JAMA* **304**, 156; author reply 158-9 (2010).
55. Ellenberg, S.S., DeMets, D.L. & Fleming, T.R. Bias and trials stopped early for benefit. *JAMA* **304**, 158; author reply 158-9 (2010).
56. Goodman, S., Berry, D. & Wittes, J. Bias and trials stopped early for benefit. *JAMA* **304**, 157; author reply 158-9 (2010).
57. Korn, E.L., Freidlin, B. & Mooney, M. Bias and trials stopped early for benefit. *JAMA* **304**, 157-8; author reply 158-9 (2010).
58. Schomer, D.L. & Lewis, R.J. Stopping seizures early and the surgical epilepsy trial that stopped even earlier. *JAMA* **307**, 966-8 (2012).
59. Goodman, S.N. Systematic reviews are not biased by results from trials stopped early for benefit. *J Clin Epidemiol* **61**, 95-6; author reply 96-8 (2008).
60. Thall, P.F. et al. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat Med* **22**, 763-80 (2003).
61. Biswas, S., Liu, D.D., Lee, J.J. & Berry, D.A. Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center. *Clin Trials* **6**, 205-16 (2009).
62. Maki, R.G. et al. Randomized phase II study of gemcitabine and docetaxel compared with gemcitabine alone in patients with metastatic soft tissue sarcomas: results of sarcoma alliance for research through collaboration study 002 [corrected]. *J Clin Oncol* **25**, 2755-63 (2007).
63. Barker, A.D. et al. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* **86**, 97-100 (2009).
64. Allegra, C.J. et al. Phase III trial assessing bevacizumab in stages II and III carcinoma of the colon: results of NSABP protocol C-08. *J Clin Oncol* **29**, 11-6 (2011).
65. Weiss, R.D. et al. A multi-site, two-phase, Prescription Opioid Addiction Treatment Study (POATS): rationale, design, and methodology. *Contemp Clin Trials* **31**, 189-99 (2010).
66. Barnes, P.J. et al. Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design. *Pulm Pharmacol Ther* **23**, 165-71 (2010).
67. Haley, E.C., Jr. et al. Phase IIB/III trial of tenecteplase in acute ischemic stroke: results of a prematurely terminated randomized clinical trial. *Stroke* **41**, 707-11 (2010).
68. Lewis, R.J. et al. Monitoring a clinical trial conducted under the Food and Drug Administration regulations allowing a waiver of prospective informed consent: the diaspirin cross-linked hemoglobin traumatic hemorrhagic shock efficacy trial. *Ann Emerg Med* **38**, 397-404 (2001).
69. Schulz, K.F., Altman, D.G. & Moher, D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* **152**, 726-32 (2010).
70. Berry, D.A. Adaptive clinical trials: the promise and the caution. *J Clin Oncol* **29**, 606-9 (2011).