Adaptive Designs for Clinical Trials of Drugs and Biologics

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

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be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance.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

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

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

September 2018
Clinical/Medical

Adaptive Designs for Clinical Trials of Drugs and Biologics

Guidance for Industry

Additional copies are available from:
Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

and/or

Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
https://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

September 2018
Clinical/Medical

Contains Nonbinding Recommendations

Draft — Not for Implementation

TABLE OF CONTENTS

I. INTRODUCTION AND SCOPE 1

II. DESCRIPTION OF AND MOTIVATION FOR ADAPTIVE DESIGNS 2

A. Definition 2

B. Important Concepts 2

C. Motivation and Examples 3

D. Limitations 6

E. Choosing to Adapt 6

III. PRINCIPLES FOR ADAPTIVE DESIGNS 7

A. Controlling the Chance of Erroneous Conclusions 7

B. Estimating Treatment Effects 8

C. Trial Planning 8

D. Maintaining Trial Conduct and Integrity 9

IV. ADAPTIVE DESIGNS BASED ON NON-COMPARATIVE DATA 9

V. ADAPTIVE DESIGNS BASED ON COMPARATIVE DATA 10

A. Group Sequential Designs 11

B. Adaptations to the Sample Size 13

C. Adaptations to the Patient Population (e.g., Adaptive Enrichment) 14

D. Adaptations to Treatment Arm Selection 15

E. Adaptations to Patient Allocation 16

F. Adaptations to Endpoint Selection 17

G. Adaptations to Multiple Design Features 17

VI. SPECIAL CONSIDERATIONS AND TOPICS 17

A. Simulations in Adaptive Design Planning 17

B. Bayesian Adaptive Designs 20

C. Adaptations in Time-to-Event Settings 21

D. Adaptations Based on a Potential Surrogate or Intermediate Endpoint 21

E. Secondary Endpoints 22

F. Safety Considerations 22

G. Adaptive Design in Early-Phase Exploratory Trials 23

H. Unplanned Design Changes Based on Comparative Interim Results 23

I. Design Changes Based on Information From a Source External to the Trial 23

VII. MAINTAINING TRIAL INTEGRITY 24

VIII. REGULATORY CONSIDERATIONS 26

A. Interactions With FDA 26

B. Documentation Prior to Conducting an Adaptive Trial 27

C. Evaluating and Reporting a Completed Trial 29

IX. REFERENCES 30

Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry[1]

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. INTRODUCTION AND SCOPE

This document provides guidance to sponsors and applicants submitting investigational new drug applications (INDs), new drug applications (NDAs), biologics licensing applications (BLAs), or supplemental applications on the appropriate use of adaptive designs for clinical trials to provide evidence of the effectiveness and safety of a drug or biologic.[2] The guidance describes important principles for designing, conducting, and reporting the results from an adaptive clinical trial. The guidance also advises sponsors on the types of information FDA needs to evaluate the results from clinical trials with adaptive designs, including Bayesian adaptive and complex trials that rely on computer simulations for their design.

The primary focus of this guidance is on adaptive designs for clinical trials intended to support the effectiveness and safety of drugs. The concepts contained in this guidance are also useful for early-phase or exploratory clinical trials as well as trials conducted to satisfy post-marketing commitments or requirements.

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

This guidance will replace the 2010 draft guidance for industry *Adaptive Design Clinical Trials for Drugs and Biologics*.

[1] This guidance has been prepared by the Office of Biostatistics in the Center for Drug Evaluation and Research (CDER) and the Division of Biostatistics in the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration.

[2] The term *drug* as used in this guidance refers to both human drugs and biological products unless otherwise specified.

II. DESCRIPTION OF AND MOTIVATION FOR ADAPTIVE DESIGNS

A. Definition

For the purposes of this guidance, an *adaptive design* is defined as a clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulating data from subjects in the trial.

B. Important Concepts

The following are descriptions of important concepts used in this guidance:

- An *interim analysis* is any examination of data obtained from subjects in a trial while that trial is ongoing, and is not restricted to cases in which there are formal between-group comparisons.[3] The observed data used in the interim analysis can include one or more types, such as baseline data, safety outcome data, pharmacokinetic, pharmacodynamic or other biomarker data, or efficacy outcome data.

- A *non-comparative analysis* is an examination of accumulating trial data in which the treatment group assignments of subjects are not used in any manner in the analysis. A *comparative analysis* is an examination of accumulating trial data in which treatment groups are identified, either with the actual assigned treatments or with codes (e.g., labeled as A and B, without divulging which treatment is investigational). The terms *unblinded analysis* and *blinded analysis* are also sometimes used to make the distinction between analyses in which treatment assignments are and are not identified, respectively.

- An interim analysis can be comparative or non-comparative regardless of whether trial subjects, investigators, and other personnel such as the sponsor and data monitoring committee (DMC) remain blinded to comparative results. The importance of limiting access to comparative interim results is discussed in detail in section VII of this guidance.

- The term *prospective*, for the purposes of this guidance, means that the adaptation is planned and details specified before any comparative analyses of accumulating trial data are conducted. In nearly all situations, potential adaptive design modifications should be

[3] This definition is different from the definition in the FDA International Council for Harmonization (ICH) guidance for industry *E9 Statistical Principles for Clinical Trials* (ICH E9), which defines an *interim analysis* as "any analysis intended to compare treatment arms with respect to efficacy or safety . . . ." This guidance uses a broader meaning for *interim analysis* to accommodate the wide range of analyses of accumulating data that can be used to determine trial adaptations. We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA Drugs or Biologics guidance web pages at

https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

https://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

planned and described in the clinical trial protocol (and a separate statistical analysis plan, if used) prior to initiation of the trial.

- This guidance distinguishes between those trials that are intended to provide substantial evidence of effectiveness and other trials, termed *exploratory trials.*[4] This distinction depends on multiple features of a clinical trial, such as the clinical relevance of the primary endpoint, quality of trial conduct, rigor of control of the chance of erroneous conclusions, and reliability of estimation.

- A *fixed sample trial* is a clinical trial with a targeted total sample size, or a targeted total number of events,[5] that is specified at the design stage and not subject to prospectively planned adaptation.

- A *non-adaptive trial* is a clinical trial without any prospectively planned opportunities for modifications to the design.

- *Bias* is a systematic tendency for the estimate of treatment effect to deviate from its true value.

- *Reliability* is the extent to which statistical inference from the clinical trial accurately and precisely evaluates the treatment effect.

- *Generalizability* is the degree to which inference, based on the clinical trial or trials, is applicable to real clinical practice.

- A critical component of the demonstration of the effectiveness and, in some cases, safety of a drug is the test of a null hypothesis in a clinical trial. If the null hypothesis is rejected at a specified level of significance (typically a one-sided level equal to .025), with demonstration of a clinically meaningful effect of the drug, the evidence generally supports a conclusion of effectiveness. Sometimes, however, the null hypothesis is rejected even though the drug is ineffective. This is called a *Type I error*. Typically, there are multiple scenarios for which the null hypothesis is true. We will use the term *Type I error probability* to refer to the maximum probability of rejecting the null hypothesis across these scenarios.

C. Motivation and Examples

Adaptive designs can provide a variety of advantages over non-adaptive designs. These advantages arise from the fundamental property of clinical trials with an adaptive design: they

[4] A variety of terms have been used to describe different kinds of clinical trials, such as phase 1, phase 2, and phase 3 (as in 21 CFR 312.21); pivotal; registration; and confirmatory (as in the ICH E9 guidance). These terms will not be used in this guidance.

[5] In settings where the primary outcome of interest is the time to event (such as death), the statistical power of the trial is determined by the total number of observed events rather than the sample size.

allow the trial to adjust to information that was not available when the trial began. The specific nature of the advantages depends on the scientific context and type or types of adaptation considered, with potential advantages falling into the following major categories:

- Statistical efficiency: In some cases, an adaptive design can provide a greater chance to detect a true drug effect (i.e., greater statistical power) than a comparable non-adaptive design.[6] This is often true, for example, of group sequential designs (section V.A) and designs with adaptive modifications to the sample size (section V.B). Alternatively, an adaptive design may provide the same statistical power with a smaller expected sample size[7] or shorter expected calendar time than a comparable non-adaptive design.

- Ethical considerations: There are many ways in which an adaptive design can provide ethical advantages over a non-adaptive design. For example, the ability to stop a trial early if it becomes clear that the trial is unlikely to demonstrate effectiveness can reduce the number of patients exposed to the unnecessary risk of an ineffective investigational treatment and allow subjects the opportunity to explore more promising therapeutic alternatives.

- Advantages in generalizability and improved understanding of drug effects: An adaptive design can make it possible to answer broader questions than would normally be feasible with a non-adaptive design. For example, an adaptive enrichment design (section V.C) may make it possible to demonstrate effectiveness in either a given population of patients or a targeted subgroup of that population, where a non-adaptive alternative might require infeasibly large sample sizes. An adaptive design can also yield improved understanding of the effect of the experimental treatment. For example, a design with adaptive dose selection (section V.D) may yield better estimates of the dose-response relationship, which may also lead to more efficient subsequent trials.

- Acceptability to stakeholders: An adaptive design may be considered more acceptable to stakeholders than a comparable non-adaptive design because of the added flexibility. For example, sponsors may be more willing to commit to a trial that allows planned design modifications based on accumulating information. Patients may be more willing to enroll in trials that use response-adaptive randomization (section V.E) because these trials can increase the probability that subjects will be assigned to the more effective treatment.

The following examples of clinical trials with adaptive designs illustrate some of the potential advantages:

- A clinical trial was conducted to evaluate Eliprodil for treatment of patients suffering from severe head injury (Bolland et al. 1998). The primary efficacy endpoint was a three-category outcome defining the functional status of the patient after six months of treatment. There was considerable uncertainty at the design stage about the proportions of

[6] An example of a comparable non-adaptive design is a fixed sample design with sample size equal to the expected sample size of the adaptive design.

[7] The expected sample size is the average sample size if the trial were repeated many times.

patients in the placebo control group who would be expected to experience each of the three different functional outcomes. An interim analysis was prespecified to update estimates of these proportions based on pooled, non-comparative data in order to potentially increase the sample size. This approach was chosen to avoid a trial with inadequate statistical power and therefore helped ensure that the trial would efficiently and reliably achieve its objective. The interim analysis ultimately led to a sample size increase from 400 to 450 patients.

- PARADIGM-HF was a clinical trial in patients with chronic heart failure with reduced-ejection fraction designed to compare LCZ696, a combination of the neprilysin inhibitor sacubitril and the renin-angiotensin system (RAS) inhibitor valsartan, with the RAS inhibitor enalapril with respect to risk of the composite endpoint of cardiovascular death or hospitalization for heart failure (McMurray et al. 2014). The trial design included three interim analyses to occur after accrual of one-third, one-half, and two-thirds of the total planned number of events to potentially stop the trial for superior efficacy of LCZ696 over enalapril based on comparative results. The addition of interim analyses with stopping rules for efficacy reduced the expected sample size and expected duration of the trial while maintaining a similar probability of trial success, relative to a trial with a single analysis after observation of a fixed total number of events. PARADIGM-HF was stopped after the third interim analysis because the prespecified stopping boundary for compelling superiority of LCZ696 over enalapril had been crossed. The group sequential design therefore facilitated a more rapid determination of benefit than would have been possible with a fixed sample design.

- To evaluate the safety and effectiveness of a nine-valent human papillomavirus (HPV) vaccine, a clinical trial with adaptive dose selection was carried out (Chen et al. 2015). The trial randomized subjects to one of three dose formulations of the nine-valent HPV vaccine or an active control, the four-valent HPV vaccine. An interim analysis was carried out to select one of the three dose formulations to carry forward into the second stage of the trial. The goal of the trial was to select an appropriate dose and confirm the safety and effectiveness of that dose in a timely manner.

- STAMPEDE was a clinical trial designed to inform the practice of medicine and simultaneously evaluate multiple treatments in prostate cancer by comparing standard androgen deprivation therapy (ADT) with several different treatment regimens that combined ADT with one or more approved therapies (Sydes et al. 2012). The trial design included multiple interim analyses to potentially drop treatment arms that were not performing well based on comparative results. The use of a common control group, along with sequential analyses to potentially terminate treatment arms, allowed the simultaneous evaluation of several treatments more efficiently than could be achieved in multiple individual trials.

- PREVAIL II was a clinical trial conducted to evaluate ZMapp plus the current standard of care as compared to the current standard of care alone for treatment of patients with Ebola virus disease (Davey et al. 2016; Dodd et al. 2016). The trial utilized a novel Bayesian adaptive design in which decision rules for concluding effectiveness at interim

and final analyses were based on the Bayesian posterior probability that the addition of ZMapp to standard of care reduces 28-day mortality. Interim analyses were planned after every 2 patients completed, with no potential action taken until a minimum number of patients (12 per group) were enrolled. The design also allowed the potential to add experimental agents as new treatment arms and the potential to supplement or replace the current standard of care arm with any agents determined to be efficacious during the conduct of the trial.

D. Limitations

The following are some of the possible limitations associated with a clinical trial employing an adaptive design:

- Adaptive designs require certain analytical methods to avoid increasing the chance of erroneous conclusions and introducing bias in estimates. For complex adaptive designs, such methods may not be readily available.

- Gains in efficiency in some respects may be offset by losses in other respects. For example, an adaptive design may have a reduced minimum and expected sample size, but an increased maximum sample size,[8] relative to a comparable non-adaptive fixed sample design. In addition, preplanning adaptive design modifications can require more effort at the design stage, leading to longer lead times between planning and starting the trial. The use of an adaptive design also adds logistical challenges in ensuring appropriate trial conduct and trial integrity. In particular, approaches to appropriately limit access to comparative interim results may be complex and add costs to the trial.

- The opportunity for efficiency gains through adaptation may be limited by important scientific constraints or in certain clinical settings. For example, a minimum sample size may be expected for a reliable evaluation of safety. There also may be limited utility in certain types of adaptations if the primary outcome of interest is ascertained over a longer period of time than the time it takes to enroll most or all patients in the trial.

- An adaptive change to a trial design may lead to results after the adaptation that are not sufficiently similar to those before the adaptation. This may lead to challenges in interpretability and limit the generalizability of results.

E. Choosing to Adapt

In general, the decision to use or not use adaptive elements in a clinical trial design will depend on a large number of factors, including the potential advantages and disadvantages described in the preceding sections. There may also be a variety of non-scientific considerations. In short, designing a clinical trial is a complex process, and it is not the intent of this guidance to require or restrict the use of adaptive designs in general or in specific settings. However, FDA

[8] The minimum and maximum sample sizes are the smallest and largest sample sizes, respectively, that could be selected under the adaptive design if the trial were repeated many times.

encourages sponsors to explore a variety of design options in planning, and to discuss their considerations with the appropriate FDA review division at regulatory meetings such as End-of-Phase-2 (EOP2) or Type C meetings.

III. PRINCIPLES FOR ADAPTIVE DESIGNS

In general, the design, conduct, and analysis of a proposed adaptive clinical trial intended to provide substantial evidence of effectiveness should satisfy four key principles: the chance of erroneous conclusions should be adequately controlled, estimation of treatment effects should be sufficiently reliable, details of the design should be completely prespecified, and trial integrity should be appropriately maintained. While all clinical trials intended to provide substantial evidence of effectiveness should satisfy these four principles, the following sections outline considerations specific to adaptive designs.

A. Controlling the Chance of Erroneous Conclusions

Because clinical trials play a central role in premarket decision-making, it is critical to assess the probability that any trial design under consideration will lead to incorrect conclusions of safety or effectiveness, incorrect conclusions of lack of safety or effectiveness, or misleading estimates of the clinical parameters that contribute to an overall assessment of benefit-risk. For example, there are a number of ways in which adaptive features can inflate the Type I error probability of a trial. The most obvious examples of this are cases in which multiple statistical hypothesis tests are performed. Consider a group sequential design, in which a preliminary test to potentially stop the trial for efficacy is performed after 50 percent of planned subjects have completed the trial. If the trial is not stopped early, a final test is performed once 100 percent of the planned subjects have completed the trial. If each of these two tests were performed at the conventional .025 one-sided significance level and the drug were not effective, the overall chance of the trial yielding a Type I error would exceed 2.5 percent. This is a well-known problem, and a variety of methods exist to determine appropriate significance levels for interim and final analyses that together ensure the overall Type I error probability of the trial is controlled at 2.5 percent (Jennison and Turnbull 1999).

Explicit multiple hypothesis tests are not the only way adaptive design features can lead to erroneous conclusions. Consider a naive approach to adaptive patient population selection, in which data in the overall trial population and in a subpopulation are examined halfway through a trial, and the population with the larger treatment effect at that point is chosen for continued study. If the final analysis is performed in the selected population at a .025 significance level and includes the same data that were used to choose the patient population, the Type I error probability would exceed 2.5 percent. Other adaptive design features may introduce still more subtle Type I error probability inflation.

Adaptive design proposals for trials incorporating null hypothesis testing should therefore address the possibility of Type I error probability inflation. In some cases, such as simple group sequential designs (section V.A), statistical theory can be used to derive significance levels that ensure Type I error probability is controlled at the desired level. In other cases, such as sample

size reestimation based on non-comparative interim results (section IV), it can be shown that performing analyses at the conventional .025 significance level has a negligible effect on the Type I error probability (Kieser and Friede 2003). In still other cases, such as many Bayesian adaptive designs (section VI.B), it may be necessary to use simulations (section VI.A) to evaluate the chance of an erroneous conclusion.

B. Estimating Treatment Effects

It is important that clinical trials produce sufficiently reliable treatment effect estimates to facilitate an evaluation of benefit-risk and to appropriately label new drugs, enabling the practice of evidence-based medicine. Some adaptive design features can lead to statistical bias in the estimation of treatment effects and related quantities. For example, each of the two cases of Type I error probability inflation mentioned in section III.A above has a potential for biased estimates. Specifically, a conventional end-of-trial treatment effect estimate such as a sample mean that does not take the adaptations into account would tend to overestimate the true population treatment effect. This is true not only for the primary endpoint which formed the basis of the adaptations, but also for secondary endpoints correlated with the primary endpoint. Furthermore, confidence intervals for the primary and secondary endpoints may not have correct coverage probabilities for the true treatment effects.

For some designs there are known methods for adjusting estimates to reduce or remove bias associated with adaptations (Jennison and Turnbull 1999; Wassmer and Brannath 2016). Such methods should be prospectively planned and used for reporting results when they are available. Biased estimation in adaptive design is currently a less well-studied phenomenon than Type I error probability inflation, however, and methods may not be available for other designs. For these other designs, the extent of bias in estimates should be evaluated, and treatment effect estimates and associated confidence intervals should be presented with appropriate cautions regarding their interpretation.

C. Trial Planning

In general, as with any clinical trial,[9] it is expected that the details of the adaptive design are completely specified prior to initiation of the trial and documented accordingly (see section VIII.B). Prospective planning should include prespecification of the anticipated number and timing of interim analyses, the type of adaptation, the statistical inferential methods to be used, and the specific algorithm governing the adaptation decision. Complete prespecification is important for a variety of reasons. First, for many types of adaptations, if aspects of the adaptive decision-making are not planned, appropriate statistical methods to control the chance of erroneous conclusions and to produce reliable estimates may not be feasible once data have been collected. Second, complete prespecification helps increase confidence that adaptation decisions were not based on accumulating knowledge in an unplanned way. For example, consider a trial with planned sample size reestimation based on pooled, non-comparative interim estimates of the variance (section IV) in which personnel involved in the adaptive decision-making (e.g., a monitoring committee) have access to comparative interim results. Prespecification that includes

[9] ICH E9 recommends prespecification of the design and analysis plan for all clinical trials.

the exact rule for modifying the sample size reduces concern that the adaptation may have been influenced by knowledge of comparative results and precludes the need for a statistical adjustment to account for modifications based on comparative interim results (section V.B). Finally, complete prespecification can motivate careful planning at the design stage, reduce unecessary sponsor access to comparative interim data, and help ensure that the DMC, if involved in implementing the adaptive design, effectively focuses on its primary responsibilities of maintaining patient safety and trial integrity (see section VII for further discussion).

D. Maintaining Trial Conduct and Integrity

Adaptive designs can create additional trial operational complications. Knowledge of accumulating data can affect the course and conduct of a trial, and the behavior of its sponsor, investigators, and participants, in ways that are difficult to predict and impossible to adjust for. Therefore, it is generally recommended that access to comparative interim results be limited to individuals with relevant expertise who are independent of the personnel involved in conducting or managing the trial for all clinical trials (not only adaptive ones).[10] Maintaining confidentiality of comparative interim results is especially challenging when the trial design includes adaptive features. Two examples of issues that could arise in adaptive trials are:

- If investigators are provided access to comparative results from an early interim analysis, knowledge of a small or unfavorable estimated treatment effect based on unreliable data could be misinterpreted as reliable evidence of no effect, leading to decreased adherence and decreased efforts to retain patients, increasing the amount of missing data in the remainder of the trial.

- After an interim analysis in a design with sample size reestimation based on comparative results (section V.B), knowledge that the targeted sample size has been increased could be interpreted by investigators and potential trial subjects as indicative of a less-than-expected interim treatment effect, potentially depressing future enrollment and endangering the success of the trial.

As these and other similar issues are generally impossible to adjust for once data have been collected, planning for an adaptive design trial should include a consideration of possible sources and consequences of trial conduct issues and plans to avoid these issues. Plans should describe the processes intended to control access to information and to document that access throughout the trial. This is discussed in more detail in section VII.

IV. ADAPTIVE DESIGNS BASED ON NON-COMPARATIVE DATA

This section addresses adaptive clinical trial designs in which adaptations are based entirely on analyses of non-comparative data, that is, without incorporating information about treatment assignment. Such analyses are sometimes called *blinded* or *masked* analyses. We avoid these terms in this guidance because they can misleadingly conflate knowledge of treatment

[10] This recommendation is conveyed, for example, in ICH E9.

assignment with the use of treatment assignment in adaptation algorithms. It is possible to include adaptations based on non-comparative data even in open-label trials, but there are added challenges in those cases to ensuring that the adaptations are completely unaffected by knowledge of comparative data. In general, adequately prespecified adaptations based on non-comparative data have a negligible effect on the Type I error probability. This makes them an attractive choice in many settings, particularly when uncertainty about event probabilities or endpoint variability is high.

Accumulating outcome data can provide a useful basis for trial adaptations. The analysis of outcome data without using treatment assignment is called *pooled analysis*. The most widely used category of adaptive design based on pooled outcome data involves sample size adaptations (sometimes called *blinded sample size reestimation*). Sample size calculations in clinical trials depend on several factors: the desired significance level, the desired power, the assumed or targeted difference in outcome due to treatment assignment, and additional *nuisance parameters*—values that are not of primary interest, but which may affect the statistical comparisons. In trials with binary outcomes such as response, the probability of response in the control group is commonly considered a nuisance parameter. In trials with continuous outcomes such as symptom scores, the variance of the scores is a nuisance parameter. By using accumulating information about nuisance parameters, sample sizes can be adjusted according to prespecified algorithms to ensure the desired power is maintained. In some cases, these techniques involve statistical modeling to estimate the value of the nuisance parameter, because the parameter itself depends on knowledge of treatment assignment (Gould and Shih 1992).

Another example of adapting based on pooled outcome data is the planned interim reevaluation of the prognostic strength of a biomarker or other baseline characteristic in a prognostic enrichment strategy.[11] For example, a trial may be targeting heavy enrollment among patients with a certain biomarker to increase the number of endpoint events, but interim pooled outcome data may suggest the biomarker in question does not have the anticipated effect on the pooled event rate, perhaps leading to a change in recruitment strategies.

V. ADAPTIVE DESIGNS BASED ON COMPARATIVE DATA

This section will discuss different types of clinical trial designs in which there are prespecified rules for stopping the trial or modifying the design based on interim analyses of comparative data. Such analyses are sometimes called *unblinded* or *unmasked analyses*. There are a few important concepts that are generally applicable to the sections that follow. First, in contrast to adaptations based on non-comparative data, adaptations based on comparative data generally do directly increase the Type I error probability and induce bias in treatment effect estimates. Therefore, statistical methods should take into account the adaptive trial design. Second, when adaptations are based on comparative interim analyses, additional steps may need to be taken to ensure appropriate trial conduct. This is discussed in more detail in section VII. Finally, stopping or adaptation rules can be specified on a variety of different scales, such as the estimate of

[11] See additional discussion in the FDA draft guidance for industry *Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products*.

treatment effect, fixed sample p-value, conditional probability of trial success, Bayesian posterior probability that the drug is effective, or Bayesian predictive probability of trial success. The choice of scale is relatively unimportant as long as the operating characteristics of the designs are adequately evaluated.

A. Group Sequential Designs

Group sequential trials allow for one or more prospectively planned interim analyses of comparative data with prespecified criteria for stopping the trial. The inclusion of sequential analyses can provide ethical and efficiency advantages by reducing the expected sample size and calendar time of clinical trials and by accelerating the approval of effective new treatments. For example, a group sequential design with a single interim analysis and a commonly used stopping boundary for efficacy can reduce the expected sample size of the trial by roughly 15 percent relative to a comparable non-adaptive fixed sample trial.[12]

Group sequential designs may include rules for stopping the trial when there is sufficient evidence of efficacy to support regulatory decision-making or when there is evidence that the trial is unlikely to demonstrate efficacy, which is often called stopping for futility. Performing each of the multiple statistical hypothesis tests for efficacy in a group sequential trial at the conventional .025 one-sided significance level would inflate the Type I error probability and therefore increase the chance of erroneous conclusions. A variety of methods exist to determine appropriate stopping boundaries for the interim and final analyses such that the Type I error probability is appropriately controlled. For example, the O'Brien-Fleming approach tends to require very persuasive early results to stop the trial for efficacy (O'Brien and Fleming 1979). Alternative approaches such as that proposed by Pocock require less persuasive early results and have higher probabilities of early stopping (Pocock 1977). These and other approaches rely on prospective planning of both the number of interim analyses and the specific sample size or number of event targets at which those analyses will occur.

The Lan-DeMets alpha-spending[13] approach accommodates varying degrees of required evidence for early stopping by specifying a function for how the Type I error probability is spent throughout the trial, while also allowing for flexibility in determining the number and timing of interim analyses (Lan and DeMets 1983). The flexibility in timing helps accommodate scheduling of monitoring meetings at specific calendar times rather than at specific interim sample sizes or number of event targets. The flexibility in the number of analyses can help accommodate faster- or slower-than-expected enrollment rates. If, however, interim analysis times are chosen based on accumulating comparative results, the Type I error probability can be inflated. For example, adjusting the next interim analysis to occur sooner than originally planned because the current interim analysis result is close to the stopping boundary would not be

[12] A group sequential design with an interim analysis that occurs when outcome information is available on half of the maximum number of patients and that utilizes an O'Brien-Fleming stopping boundary for efficacy, reduces the expected sample size of the trial by roughly 15 percent under the alternative hypothesis (at which there is 90 percent power), as compared to a design with a single analysis planned when all patients have been enrolled and had their outcomes ascertained.

[13] The Type I error probability of a clinical trial is often denoted by the Greek letter α (alpha).

appropriate. Because of this potential issue with the Lan-DeMets alpha-spending approach, sponsors should put in place additional safeguards such as a targeted number of interim analyses and an approximate schedule for their occurrence, as well as a decision framework for changing the number or timing of analyses after the trial has begun. The decision framework should be based on information that is statistically independent of the estimated treatment effect (e.g., enrollment rate or scheduling logistics). For example, the decision framework could specify semi-annual interim analyses, with additional analyses planned if enrollment is considerably slower than a prespecified target.

There are a number of additional considerations for ensuring the appropriate design, conduct, and analysis of a group sequential trial. First, for group sequential methods to be valid, it is important to adhere to the prospective analytic plan and terminate the trial for efficacy only if the stopping criteria are met. Second, guidelines for stopping the trial early for futility should be implemented appropriately. Trial designs often employ *nonbinding* futility rules, in that the futility stopping criteria are guidelines that may or may not be followed, depending on the totality of the available interim results. The addition of such nonbinding futility guidelines to a fixed sample trial, or to a trial with appropriate group sequential stopping rules for efficacy, does not increase the Type I error probability and is often appropriate. Alternatively, a group sequential design may include *binding* futility rules, in that the trial should always stop if the futility criteria are met. Binding futility rules can provide some advantages in efficacy analyses (e.g., a relaxed threshold for a determination of efficacy), but the Type I error probability is controlled only if the stopping rules are followed. Therefore, if a trial continues despite meeting prespecified binding futility rules, the Agency will likely consider that trial to have failed to provide evidence of efficacy, regardless of the outcome at the final analysis. Note also that some DMCs may prefer the flexibility of nonbinding futility guidelines.

Third, a trial terminated early for efficacy will have a smaller sample size for the evaluation of safety and potentially important secondary efficacy endpoints. Therefore, early stopping for efficacy is typically reserved for circumstances where there are compelling ethical reasons (e.g., the primary endpoint is survival or irreversible morbidity) or where the stopping rules require highly persuasive results in terms of both the magnitude of estimated treatment effect and the degree of evidence of an effect. In some cases, there may be a limit on how early group sequential interim analyses should occur or whether they should occur at all because of a minimum sample size expected for a reliable evaluation of safety. This is often true, for example, in preventive vaccine trials.

Finally, conventional fixed sample estimates of the treatment effect such as the sample mean tend to be biased toward greater effects than the true value when a group sequential design is used. Similarly, confidence intervals do not have the desired nominal coverage probabilities. Therefore, a variety of methods exist to compute estimates and confidence intervals that appropriately adjust for the group sequential stopping rules (Jennison and Turnbull 1999). To ensure the scientific and statistical credibility of trial results and facilitate important benefit-risk considerations, an approach for calculating estimates and confidence intervals that appropriately accounts for the group sequential design should be prospectively planned and used for reporting results.

B. Adaptations to the Sample Size

One adaptive approach is to prospectively plan modifications to the sample size based on comparative interim results, i.e., interim estimates of the treatment effect. This is often called *unblinded sample size adaptation* or *unblinded sample size reestimation.* Sample size determination for a fixed sample design depends on many factors, such as feasibility, the event rate in the control arm or the variability of the primary outcome, the Type I error probability, and the desired power to detect a hypothesized treatment effect size. In section IV, we described potential adaptations based on non-comparative interim results to address uncertainty at the design stage in the variability of the outcome or the event rate on the control arm. In contrast, designs with sample size adaptations based on comparative interim results might be used when there is considerable uncertainty about the true treatment effect size. Similar to a group sequential trial, a design with sample size adaptations based on comparative interim results can provide adequate power under a range of plausible effect sizes, and therefore, can help ensure that a trial will have high power if the true magnitude of treatment effect is less than what was hypothesized, but still clinically meaningful. Furthermore, the addition of prespecified rules for modifying the sample size can provide efficiency advantages with respect to certain operating characteristics in some settings.

Indiscriminately modifying the sample size of a trial without proper adjustment can inflate the Type I error probability. Consider a design with one interim analysis at which the interim estimate of treatment effect is used to modify the final sample size. If one carries out a hypothesis test at the end of the trial at the conventional .025 significance level, the Type I error probability can be more than doubled (Proschan and Hunsberger 1995).[14] Therefore, one of a variety of existing methods should be used to appropriately control the Type I error probability in the presence of this type of adaptive design. For example, hypothesis testing approaches have been developed based on combining test statistics or p-values from the different stages of the trial in a preplanned manner or through preservation of the conditional Type I error probability (Bauer and Kohne 1994; Fisher 1998; Cui et al. 1999; Denne 2001; Müller and Schäfer 2001; Chow and Chang 2011). These approaches also accommodate adaptations to aspects of the sampling plan other than the maximum sample size, such as the number and spacing of future interim analyses.

The additional considerations regarding adherence to the adaptation plan, the evaluation of safety, and the estimation of treatment effects that were discussed in section V.A on group sequential designs also apply to designs with sample size adaptations based on comparative data. Of note, prospective planning should include prespecification of not only the statistical hypothesis testing method that will be used, but also the specific rule governing the sample size modification. It is also critical that the adaptation rule and analysis plan are followed. Finally, there are additional challenges in maintaining trial integrity in the presence of sample size adaptations. For example, sample size modification rules are often based on maintaining the conditional probability of a statistically significant treatment effect at the end of the trial (often called the conditional power) at or near some desired level. In this scenario, knowledge of the

[14] This means that even use of the Bonferroni method to adjust for the two analyses conducted would not be adequate.

adaptation rule and the adaptively chosen sample size allows a relatively straightforward back-calculation of the interim estimate of treatment effect. Therefore, additional steps should be taken to limit personnel with this detailed knowledge so that trial integrity can be maintained. See section VII for additional discussion.

The principles discussed in this section also apply to trials with time-to-event endpoints where the adaptive design allows prospectively planned modifications to the total number of events based on comparative interim results. However, there are some special additional considerations in such settings that are discussed further in section VI.C.

C. Adaptations to the Patient Population (e.g., Adaptive Enrichment)

In many settings, it may be expected that the treatment effect will be greater in a certain targeted subset of the trial population. This subpopulation could be defined, for example, by a demographic characteristic or by a genetic or pathophysiologic marker that is thought to be related to the drug's mechanism of action. In such a setting, consideration may be given to a design that allows adaptive modify cations to the patient population based on comparative interim results. For example, a trial might enroll subjects from the overall trial population up through an interim analysis, at which time a decision will be made based on prespecified criteria whether to continue enrollment in the overall population or to restrict future enrollment to the targeted subpopulation. Data accumulated both before and after the interim analysis may be combined to draw inference on the treatment effect in the targeted group. This type of design, often called an *adaptive enrichment*[15] design, can provide advantages over alternative non-adaptive designs. In particular, such an adaptive design can provide greater power[16] at the same sample size as a fixed sample design in the overall population. Furthermore, unlike a trial restricting enrollment to the targeted subpopulation, the adaptive design allows an evaluation of the experimental treatment in the non-targeted (complementary) subpopulation.

A design that allows adaptive modifications to the patient population often involves both (1) modification of design features, such as the enrolled population and the population evaluated in the primary analysis, based on comparative interim results; and (2) hypothesis tests in multiple populations, such as a targeted subpopulation and the overall population. Therefore, statistical hypothesis testing methods should account for both sources of multiplicity. For example, one approach is to combine test statistics or p-values from the different stages of the trial in a preplanned manner, while also using an appropriate multiple testing procedure (Wassmer and Brannath 2016). Such an approach could potentially also accommodate adaptations to the sample size or to the proportion of patients enrolled from a particular subpopulation (e.g., increasing the proportion in a subset rather than completely restricting enrollment to that subset).

[15] This terminology is used, for example, in the FDA draft guidance for industry *Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products*. When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

[16] Power in this context could be defined, for example, as the probability of successfully identifying a true treatment effect in either the targeted subpopulation or the overall population.

There are a number of important considerations beyond those previously discussed for group sequential designs and designs with adaptive modifications to the sample size. First, in the case of an adaptive enrichment design, the proposed adaptive modifications to the patient population should be motivated by results from previous (e.g., early-phase) trials and/or strong biologic plausibility that the benefit-risk profile will be most favorable in a particular subpopulation. Second, if the baseline characteristic that is thought to modify the treatment effect is not binary in nature, any threshold or thresholds used to define subpopulations should be appropriately justified. Third, the identification of the targeted subpopulation may depend on the use of an in vitro companion diagnostic device or test. In this scenario, the diagnostic device or test should have adequate performance characteristics.[17] Finally, the extent to which the trial should be designed to characterize the treatment effect in the complementary subpopulation may depend on a number of factors, such as the pathophysiologic or empirical rationale for enrichment, the toxicities of the drug, the distribution of the baseline marker defining the subpopulations, the justification for a threshold defining subpopulations, and the potential for off-label use in the complementary subpopulation if approval is limited to the targeted subpopulation.

D. Adaptations to Treatment Arm Selection

Another adaptive approach is to prospectively plan modifications to the treatment arms included in the clinical trial based on comparative interim results. Modifications could include adding or terminating arms. This kind of design has often been used in early-phase exploratory dose-ranging trials. An adaptive dose-ranging trial might begin with several doses and incorporate interim analyses based on comparative data to select doses for continued evaluation, with the goal of providing improved characterization of the dose-response relationship relative to a non-adaptive design and allowing selection of an optimal dose or doses for evaluation in future confirmatory trials. For example, the continual reassessment method (CRM) is an approach to adaptively escalate the doses evaluated in early-phase trials based on observed toxicities in order to reliably and efficiently estimate the maximum tolerated dose for a new drug (Le Tourneau et al. 2009). Adaptive treatment arm selection is also possible in trials intended to provide substantial evidence of effectiveness. For example, in a setting where it is plausible that either or both of two doses might have a favorable benefit-risk profile, an adaptive design with sequential analyses allowing early termination of one of the dose arms can meet its scientific objective in a more efficient manner than alternative non-adaptive designs. Such an adaptive design could in principle allow interim modifications to additional aspects of the design, such as the number of additional patients that will be enrolled (the sample size) and the randomization ratio for treatment arms carried forward.

For trials intended to provide substantial evidence of effectiveness, statistical hypothesis testing methods should account for the adaptive selection of a best dose or doses from among the multiple doses evaluated in the trial, as well as any additional adaptive modifications, such as the potential to stop the trial early or to modify future sample sizes. In the simple case of a design with more than one dose that includes interim analyses to potentially stop enrollment for a

[17] See the FDA draft guidance for industry and FDA staff *In Vitro Companion Diagnostic Devices.* When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

particular dose for efficacy or futility, typical group sequential testing methods can be used, along with some multiple testing approach to control the Type I error probability across the multiple doses evaluated. If the design allows for additional adaptations such as modifications to the sample size, methods such as those described for sample size and population adaptations should be used. As with other adaptive designs, prospective planning is important and should include prespecification of not only the testing method, but also the specific adaptation rule for selecting treatment arms and for any other potential interim modifications. In general, seamless designs that incorporate both dose selection and confirmation of efficacy of a selected dose (based on data from the entire trial) can be considered if the principles outlined in section III are followed.

A special case of adaptive treatment arm selection occurs in the context of an adaptive *platform trial* designed to compare more than one experimental treatment against an appropriate control for a disease (e.g., Woodcock and LaVange 2017). Two features of these trials often incorporated for efficiency gains are use of a common control arm and use of prospectively planned adaptations to select promising treatments at interim analyses for continued study. Because these trials may involve investigational agents from more than one sponsor, may be conducted for an unstated length of time, and often involve complex adaptations, they should generally involve extensive discussion between all stakeholders and FDA.

E. Adaptations to Patient Allocation

This section considers two types of adaptations to patient allocation: adaptations based on comparative baseline characteristic data and adaptations based on comparative outcome data. The first type is *covariate-adaptive treatment assignment*, a technique in which a patient's treatment assignment depends in part or entirely on his or her baseline characteristics and the baseline characteristics and treatment assignments of previously enrolled patients. Such an approach is used to promote balance between treatment groups on baseline covariates. One well-known example of covariate-adaptive randomization is *minimization* (Pocock and Simon 1975), which involves assigning each consecutive patient to treatment in such a way that differences between treatment groups on potentially prognostic covariates are minimized. Covariate-adaptive treatment assignment techniques do not directly increase the Type I error probability when analyzed with the appropriate methodologies (generally randomization or permutation tests). These techniques can increase the predictability of treatment assignment relative to simple randomization, but this predictability can be mitigated with an additional random component to prevent perfectly deterministic treatment assignment.

The second type is response-adaptive randomization, an adaptive feature in which the chance of a newly-enrolled subject being assigned to a treatment arm varies over the course of the trial based on accumulating outcome data for subjects previously enrolled. There are a variety of response-adaptive randomization techniques, some of which go by names such as *play the winner* designs. Statistical, ethical, and pragmatic rationales are all sometimes given for using response-adaptive randomization. In statistical terms, response-adaptive techniques can in some circumstances minimize the variance of the test statistics, leading to shorter trials, smaller sample sizes, and/or greater statistical power. The ethical argument for response-adaptive randomization is that this design feature can lead to more trial subjects being assigned to the more promising of

the treatment arms. Finally, a pragmatic argument is that clinical trials with this design feature can be appealing to potential participants, thereby increasing speed and ease of accrual. Note that the arguments for response-adaptive randomization are controversial, and some researchers feel that inconclusive interim results should not be used to alter randomization in an ongoing trial and/or that statistical efficiency is not substantially improved in two-arm trials to justify adjusting randomization ratios (Hey and Kimmelman 2015, and accompanying commentaries).

Response-adaptive randomization alone does not generally increase the Type I error probability of a trial when used with appropriate statistical analysis techniques. It is important to ensure that the analysis methods appropriately take the design of the trial into account. Finally, as with many other adaptive techniques based on outcome data, response-adaptive randomization works best in trials with relatively short-term ascertainment of outcomes.

F. Adaptations to Endpoint Selection

This is a design that allows adaptive modification to the choice of primary endpoint based on comparative interim results. Such a design might be motivated by uncertainty about the treatment effect sizes on multiple patient outcomes that would be considered acceptable primary endpoints by FDA. As with other adaptive designs, the adaptation rule should be prespecified, and statistical hypothesis testing methods should account for the adaptive endpoint selection. Because endpoint selection involves important clinical considerations, early discussion with the FDA review division is recommended when such designs are being considered.

G. Adaptations to Multiple Design Features

It is possible for a clinical trial to be more complex by combining two or more of the adaptive design features discussed in this guidance. The same general principles apply to these complex designs as to simpler adaptive designs. It may be particularly difficult to estimate Type I error probability and other operating characteristics for designs that incorporate multiple adaptive features. Clinical trial simulations (section VI.A) will often be necessary to evaluate the trial design.

VI. SPECIAL CONSIDERATIONS AND TOPICS

A. Simulations in Adaptive Design Planning

Clinical trial simulations often play a critical role in planning and designing clinical trials in general, and are particularly important for adaptive trials. Simulations can be used, for example, to select the number and timing of interim analyses, or to determine the appropriate critical value of a test statistic for declaring efficacy or futility. Simulations can also be useful for comparing the performance of alternative designs. Finally, a major use of simulations in adaptive trial

design is to estimate trial operating characteristics[18] and to demonstrate that these operating characteristics meet desired levels.

Traditional non-adaptive clinical trials have generally relied on statistical theory to ensure that Type I error probability is controlled at a desired level and to obtain estimates of the power of the trial. In the simplest case, when testing a single endpoint in a fixed-sample size clinical trial design, it can typically be shown that the final test statistic has a certain asymptotic probability distribution,[19] and inference and operating characteristics can then be based on the properties of this distribution. For many adaptive designs, such as traditional group sequential designs, it is similarly possible to derive asymptotic probability distributions mathematically and base inference and planning on those distributions.

For some adaptive designs, however, it is either not possible to derive relevant distributions of test statistics, or the distributions themselves are not computationally tractable. This tends to be the case for more complex adaptive designs, such as designs that adapt several elements or designs that use predictive probability models to determine analysis time points. In these cases, trial operating characteristics can often be estimated by means of clinical trial simulations. For example, for Type I error probability and power, the basic logic of this approach is to simulate many instances of the trial based on various assumptions and evaluate the proportion of simulations which would have met the predetermined bar for supporting a conclusion of effectiveness under each set of assumptions.

For simulations intended to estimate Type I error probability, hypothetical clinical trials would be simulated under a series of assumptions compatible with the null hypothesis. For each set of such assumptions, the proportion of simulated trials that led to a false positive conclusion would be taken as an estimate of Type I error probability under those assumptions. In almost all cases, there are an infinite number of scenarios potentially compatible with the null hypothesis. Identifying which scenarios should be considered when estimating Type I error probability can be challenging, and may rely on a combination of medical and mathematical considerations.

These scenarios may include varying assumptions about nuisance parameters. These nuisance parameters can include statistical parameters, such as the variance of a symptom scale or the probability of response in the control group, and also operational parameters, such as the speed of subject accrual to a trial. For example, consider a trial comparing 2-year mortality rates between an experimental therapy and placebo in an oncology indication with very low (for example, median 6-month) survival. The null hypothesis is equal mortality rates in the two arms. Possible scenarios consistent with this null hypothesis would include equal mortality rates of 5 percent, of 50 percent, of 99 percent, of 99.01 percent, and so on. While it is impossible to simulate every scenario compatible with the null hypothesis, it may be possible to determine a

[18] Trial operating characteristics are properties of the trial with a given design. For example, properties of interest might include Type I error probability; power; expected, minimum, and maximum sample size; bias of treatment effect estimates; and coverage of confidence intervals (the probability the confidence interval would include the true treatment effect if the clinical trial were repeated many times).

[19] The asymptotic distribution of a test statistic is the approximate probability distribution of that statistic when the sample size gets large.

limited set of scenarios that adequately represent the plausible range of potential false positives. In this example, medical experts may feel comfortable ruling out any scenario with a 2-year placebo mortality rate below 75 percent, for instance, based on literature and clinical experience with the disease. Mathematical considerations can also play a role in determining which scenarios need to be simulated to estimate Type I error probability. It may be possible to argue that certain scenarios necessarily have lower Type I error probability than other scenarios based on monotonicity.

In many cases, it will not be possible to estimate Type I error probability for every set of null assumptions even after taking clinical and mathematical considerations into account. It is common to perform simulations on a grid of plausible values and argue based on the totality of the evidence from the simulations that maximal Type I error probability likely does not exceed a desired level across the range covered by the grid. In the example above, simulations might be performed at placebo and experimental treatment mortality rates equal to 75, 80, 85, 90, 95, and 99 percent. If, in each of these scenarios, estimated Type I error probability was below .025, that could be considered sufficient evidence that Type I error probability was adequately controlled for all scenarios with placebo mortality between 75 and 99 percent. However, with any approach, the evaluation at the end of the trial should consider whether the statistical inference is appropriate and the conclusions are justified in light of the accumulated information about the nuisance parameters. In the example, if the observed placebo mortality rate was unexpectedly 50 percent, additional simulations would be required.

Another complicating factor is the presence of multiple endpoints. When it is desired to test multiple clinical endpoints and control the familywise Type I error probability across all of these endpoints, null hypothesis scenarios require simulating all endpoints for each subject, which may in turn require knowledge of the correlational structure of the multiple endpoints. Typically, this is too complex an issue to address in clinical trial simulation. In some cases, however, it can be argued that assuming independence among multiple endpoints will provide an upper bound on the Type I error probability. This is true, for instance, when using the Bonferroni or Holm approach to control for multiple testing.[20]

It is important to consider the precision of simulated operating characteristics, which depends on the number of simulated trials (iterations). The number of iterations should be sufficient to facilitate an understanding and review of the proposed clinical trial design. Using 100,000 iterations per scenario, for instance, ensures a 95% confidence interval for estimated Type I error probability with a width of approximately ± 0.1%, which would be sufficient in most cases. This will allow very small differences in estimated Type I error probability to be identified, which may be important in some cases. In general, it is also preferable to use different random seeds for different simulation scenarios; this helps avoid consistently atypical results across scenarios. In some cases, fewer iterations may suffice to evaluate Type I error probability. For example, it may be sufficient to use 10,000 iterations if a particularly fine grid of scenarios is explored and every scenario has an estimated Type I error probability below the desired level. Also, a smaller

[20] Additional discussion on the Bonferroni, Holm, and other multiple testing approaches can be found in the FDA draft guidance for industry *Multiple Endpoints in Clinical Trials.* When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/RegulatoryInformation/Guidances/default.htm.

number of simulations can generally be used if the upper bound of the 95% confidence interval for the Type I error probability estimate is below the desired level.

Clinical trial simulations can also be used to estimate power and other relevant operating characteristics, such as expected sample size, expected calendar time, and bias in treatment effect estimates, for complex adaptive designs. Similar considerations apply to these estimates as to Type I error probability estimates. The level of precision expected for Type I error probability estimates, however, is generally not needed for other operating characteristics, and so it is usually appropriate to investigate a sparser set of scenarios using smaller numbers of iterations for power and other operating characteristics.

B. Bayesian Adaptive Designs

The term *Bayesian adaptive design* has been used to refer to a wide variety of clinical trial designs that use Bayesian statistical reasoning and/or calculations in various ways (Berry, et al. 2010). Some examples of Bayesian adaptive design features are:

- Use of predictive statistical modeling, possibly incorporating information external to a trial, to govern the timing and decision rules for interim analyses
- Use of assumed dose-response relationships to govern dose escalation and selection
- Explicit borrowing of information from external sources, e.g., previous trials, natural history studies, and registries, via informative prior distributions to improve the efficiency of a trial
- Use of posterior probability distributions to form trial success criteria

In general, the same principles apply to Bayesian adaptive designs as to adaptive designs without Bayesian features. One common feature of most Bayesian adaptive designs is the need to use simulations (section VI.A) to estimate trial operating characteristics.[21] Because many Bayesian methods themselves rely on extensive computations (Markov chain Monte Carlo (MCMC) and other techniques), trial simulations can be particularly resource-intensive for Bayesian adaptive designs. It will often be advisable to use conjugate priors or computationally less burdensome Bayesian estimation techniques such as variational methods rather than MCMC to overcome this limitation (Tanner 1996).

Special considerations apply to Type I error probability estimation when a sponsor and FDA have agreed that a trial can explicitly borrow external information via informative prior distributions. Type I error probability simulations need to assume that the prior data were generated under the null hypothesis. This is usually not a sensible assumption, as the prior data

[21] Note that Type I error probability and power are, by definition, frequentist concepts. As such, any clinical trial whose design is governed by Type I error probability and power considerations is inherently a frequentist trial, regardless of whether Bayesian methods are used in the trial design or analysis. Nevertheless, it is common to use the term "Bayesian adaptive design," to distinguish designs that use Bayesian methods in any way from those that do not.

are typically being used specifically because they are not compatible with the null hypothesis. Furthermore, controlling Type I error probability at a conventional level in cases where formal borrowing is being used generally limits or completely eliminates the benefits of borrowing. It may still be useful to perform simulations in these cases, but it should be understood that estimated Type I error probabilities represent a worst-case scenario in the event that the prior data (which are typically fixed at the time of trial design) were generated under the null hypothesis. A comprehensive discussion of Bayesian approaches is beyond the scope of this document. As with any complex adaptive design proposal, early discussion with the appropriate FDA review division is recommended for adaptive designs that formally borrow information from external sources.

C. Adaptations in Time-to-Event Settings

There are certain additional considerations specific to adaptive trials in which the primary endpoint is the time to occurrence of a certain event, such as time to death or time to tumor response. In these trials, power is dependent on the number of events rather than the number of subjects. As such, it is common to target a fixed number of events rather than a fixed number of subjects. Sample size adjustment in these trials has the purpose of modifying the number of events and, therefore, may take the form of increasing the number of subjects, the length of the follow-up period for each subject, or both. In addition, interim analyses in time-to-event settings may utilize information on surrogate or intermediate outcomes, and use of such approaches should be appropriately accounted for in the analysis (see next section for further discussion).

D. Adaptations Based on a Potential Surrogate or Intermediate Endpoint

Most adaptive designs rely on ongoing monitoring of the primary endpoint or endpoints. However, in cases where a potential surrogate or intermediate endpoint exists that is correlated with the primary endpoint, and the primary endpoint itself is difficult or slow to ascertain, an adaptive design can be based on the potential surrogate or intermediate endpoint. For example, consider a trial of a neoadjuvant treatment for high-risk early-stage breast cancer, where the primary endpoint is overall survival, median survival time is well over 2 years, and pathological complete response (pCR) may be reasonably likely to predict clinical benefit.[22] In this case, it may be sensible to base sample size reassessment or other adaptive features on pCR rather than mortality. The final evaluation of efficacy would still be based on the primary endpoint (overall survival in this example). Similarly, an adaptive design could be based on a 3-month measurement of patient symptoms when the primary endpoint is the assessment of the same symptom outcome at 1 year. These approaches involve assumptions about the relationship between the potential surrogate or intermediate endpoint and the primary endpoint, and any evaluation of Type I error probability or other trial operating characteristics should consider the possible effects of misspecification of this relationship.

In many trials with adaptive designs in time-to-event and longitudinal outcome settings, the plan is to adapt based on only primary endpoint information. In such cases, it would be inappropriate

[22] See the FDA guidance for industry *Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval.*

to use surrogate or intermediate outcome information at the interim analysis in an unplanned manner. For example, it has been noted (Bauer and Posch 2004) that in time-to-event settings, using surrogate information at the time of an interim analysis from subjects for whom events have not been observed to help predict future event times can lead to Type I error probability inflation. Additional safeguards such as limitation of access to comparative interim results and prespecification of the adaptation rule can help increase confidence that such unplanned approaches were not carried out (see section VII for additional discussion).

E. Secondary Endpoints

Most clinical trials have one or more secondary endpoints specified in addition to the primary endpoint,[23] and adaptive designs can have consequences for the analysis of these secondary endpoints. Consider group sequential designs: It is widely understood that multiple analyses of the primary endpoint can inflate the Type I error probability and lead to biased estimation of treatment effects on that endpoint. Less well appreciated, however, is that Type I error probability inflation and biased estimation can also apply to any endpoint correlated with the primary endpoint (Hung et al. 2007). Most secondary endpoints in clinical trials are correlated with the primary endpoint, often very highly correlated. For some designs such as group sequential approaches, methods exist to adjust secondary endpoint analyses for the adaptation (Glimm et al. 2009). Without such adjustment, appropriate caution should be applied in interpreting secondary endpoint results.

F. Safety Considerations

Although adaptive design clinical trial planning often focuses on outcomes intended to demonstrate effectiveness, safety objectives also play a critical role. First, there are cases where adaptations are planned on safety rather than efficacy endpoints. One example is early-phase dose-ranging trials in oncology that attempt to identify a maximum tolerated dose using the CRM or other adaptive techniques. Another example is the Rotavirus Efficacy and Safety Trial (REST) that formed a primary basis for the 2006 approval of a rotavirus vaccine, RotaTeq (Heyse et al. 2008). REST was a group sequential trial designed to evaluate the risk of intussusception, a serious gastrointestinal condition, in up to 100,000 infants, of whom a subset was used for an efficacy evaluation.

Second, the acquisition of sufficient safety information to support product approval is usually a major concern in trials that adapt on efficacy endpoints. Trials with early stopping for strong evidence of effectiveness still need to collect sufficient safety data to allow for a reliable benefit-risk evaluation of the investigational drug. For this reason, the size of a safety database should be taken into account when planning the number, timing, and stopping boundaries of interim analyses. In particular, the timing of interim analyses may be restricted by the expectation for a minimum number of patients studied and a minimum length of exposure to ensure a reliable safety evaluation.

[23] See the FDA draft guidance for industry *Multiple Endpoints in Clinical Trials* for a discussion of general considerations in the evaluation of multiple endpoints in clinical trials.

Finally, it is important to consider whether adaptations can potentially put trial subjects at excessive risk. This would be a concern in particular in early-phase dose-escalation trials. Adaptation rules that allow for successive cohorts of subjects to receive quickly escalating doses could lead to subjects receiving unsafe high doses that would have been avoided by a design with more gradual dose-escalation. This is particularly true when there is a possibility for serious adverse events with a delayed onset of action of the investigational drug.

G. Adaptive Design in Early-Phase Exploratory Trials

Exploratory trials in drug development are intended to obtain information on a wide range of aspects of drug use that guide later decisions on how best to study a drug (e.g., choices of dose, regimen, population, concomitant treatments, or endpoints). There can be a series of separate early trials in which different aspects of the drug's effect are sequentially examined or a more complex trial attempting to evaluate multiple different aspects simultaneously. The flexibilities offered by adaptive designs may be particularly useful in this exploratory period of development by allowing initial evaluation of a broad range of choices. Using adaptive designs in early development trials to learn about various aspects of dosing, exposure, pharmacodynamics, variability in patient response, or response modifiers offers sponsors opportunities that can improve the designs and possibly the chances of success of later-phase trials.

Although exploratory trials do not generally have the same statistical expectations as trials intended to provide substantial evidence of effectiveness, it is still important to be aware of the potential for erroneous conclusions to be made in exploratory trials. For example, flaws in an exploratory multiple-dose comparison trial could lead to suboptimal dose selection for a subsequent confirmatory trial, with a resultant failure to show effectiveness or a finding of unnecessarily excessive toxicity. Thus, exploratory trials that incorporate adaptations should still follow good principles of adaptive trial design so that the risk of adversely affecting the development program is minimized.

H. Unplanned Design Changes Based on Comparative Interim Results

When trial data are examined in a comparative interim analysis, data analyses that were not prospectively planned as the basis for adaptations may unexpectedly appear to indicate that some specific design change (e.g., restricting analyses to some population subset, dropping a treatment arm, adjusting sample size, modifying the primary endpoint, or changing analysis methods) is ethically important or might increase the potential for a statistically significant final trial result. For example, unexpected toxicity in one arm of a multiple-arm trial might motivate dropping that treatment arm. Such revisions based on non-prospectively planned analyses can create difficulty in controlling the Type I error probability and in interpreting the trial results. Sponsors are strongly encouraged not to implement such changes without first meeting with FDA to discuss the changes being considered, provided patient safety is not compromised.

I. Design Changes Based on Information From a Source External to the Trial

Unpredictable events that occur outside of an ongoing trial during the course of drug development programs may provide important new information relevant to the ongoing trial and

may motivate revisions to the trial design. For example, there may be unexpected safety information arising from a different study (perhaps in a different patient population), new information regarding the disease pathophysiology or patient characterization that identifies disease subtypes, new information on pharmacokinetics or pharmacodynamic responses to the drug, or other information that might have led to a different trial design had the information been known when the ongoing trial was being designed. When this occurs, there may be reason to revise the trial design in some manner rather than, for example, terminating the existing trial and starting a new trial with a modified design. In cases of serious safety concerns, and particularly in large trials, revising the trial design may be critical to allowing the trial to continue. Well-motivated design changes based on only information external to the trial do not affect the validity of statistical inference and will often be considered acceptable to the Agency. Practically, it is very challenging to ensure that a decision to modify a trial was based entirely on external information except in cases where the sponsor is completely blinded to comparative interim results. This is one reason why limitation of access to comparative interim results is so important (see section VII).

VII. MAINTAINING TRIAL INTEGRITY

In general, it is strongly recommended that access to comparative interim results is limited to individuals with relevant expertise who are independent from the personnel involved in conducting or managing the trial. Ensuring that patients, investigators and their staff, and sponsor personnel do not have access to comparative interim results serves two important purposes. First, it provides the greatest confidence that potential unplanned design modifications are not motivated in any way by accumulating data. For example, knowledge of comparative interim results by trial management personnel may make it difficult for regulators to determine whether a protocol amendment seemingly well-motivated by information external to the trial was influenced, in any way, by access to accumulating comparative data. If it is thought that design changes may have been influenced by comparative interim results, appropriate statistical methods to control the chance of erroneous conclusions and to produce reliable estimates may not be known, may be challenging to implement, or may greatly reduce the efficiency of the trial.

Second, limitation of access to comparative interim results provides the greatest assurance of quality trial conduct. Knowledge of accumulating data by trial investigators can adversely affect patient accrual, adherence, retention, or endpoint assessment, compromising the ability of the trial to reliably achieve its objective in a timely manner (Fleming et al. 2008). Issues with trial conduct are difficult to predict and generally impossible to adjust for in statistical analyses. Therefore, a clinical trial with an adaptive design should include rigorous planning, careful implementation, and comprehensive documentation of approaches taken to maintain confidentiality of comparative interim results and to preserve trial integrity.

There are multiple potential models for implementing a plan to maintain confidentiality in an adaptive design trial. A dedicated independent adaptation body could be established, exclusive of a DMC, if one exists. Alternatively, the adaptive decision-making role could be assigned to the DMC, although its primary responsibility should remain to ensure patient safety and trial

integrity.[24] This latter model might best be reserved for group sequential designs and other straightforward adaptive designs with simple adaptation algorithms. There are arguments favoring both approaches. For example, use of separate bodies might facilitate the inclusion of more relevant expertise on each committee and allow the DMC to most effectively focus on its primary responsibilities. On the other hand, use of a single body such as a DMC for both purposes avoids the logistical challenges of determining information sharing with and interactions between multiple monitoring groups.

Regardless of the chosen approach, the committee tasked with making adaptation recommendations should have members with the proper expertise, including a statistician or statisticians who are knowledgeable about the adaptation methodology, the monitoring plan, and the decision rules. Furthermore, the responsibility of this committee should be to make adaptation recommendations or decisions based on appropriately implementing a carefully designed and prespecified adaptation plan, not to identify potential design aspects to adapt after reviewing comparative interim results. Therefore, it is important for the DMC and/or adaptation committee to be involved at the design stage in extensive discussions with the sponsor about hypothetical scenarios and whether actions dictated by the adaptation plan would be considered reasonable by all involved parties.

Safeguards should be in place to ensure that the persons responsible for preparing and reporting interim analysis results to the DMC or the adaptation committee are physically and logistically separated from the personnel tasked with managing and conducting the trial, whether those personnel reside within the sponsor organization, another organization such as a contract research organization (CRO), or both. This practice will help ensure that persons involved in the day-to-day management and conduct of the trial do not have access to treatment assignments or comparative results, even inadvertently. Similarly, recommendations from the DMC or adaptation committee back to the sponsor should generally exclude any details of the interim analysis results, for the reasons cited above.

Although it is generally recommended that no sponsor representatives have access to comparative interim results, there are specific situations where limited access for specific sponsor personnel may be justified. For example, some adaptive trials may involve decisions, such as dose selection, that are typically the responsibility of the sponsor in non-adaptive settings and have important long-term implications for the drug development program. Limited access by sponsor personnel might be justifiable in such circumstances, for example, if a small number of sponsor representatives are involved, the individuals allowed access are not otherwise involved in trial conduct or management, and appropriate procedures are put in place to ensure that comparative interim results remain unknown to other key parties, such as patients, investigators, and the trial steering committee. However, risks to trial integrity are most easily minimized by completely restricting sponsor access to comparative interim results, and this is likely achievable in most circumstances through extensive planning and discussion between the sponsor and the DMC or adaptation committee at the design stage.

[24] See the FDA guidance for clinical trial sponsors *Establishment and Operation of Clinical Trial Data Monitoring Committees* for a detailed discussion of the roles, responsibilities, and operating procedures of DMCs in clinical trials.

Appropriate limitation of access entails carefully planned procedures to maintain and verify confidentiality, as well as documentation of monitoring and adherence to the operating procedures. Approaches typically include the use of confidentiality agreements for persons with access to interim data; the use of logistical or physical firewalls that prevent access by trial personnel to any data that include information that might allow one to infer treatment assignment; and development and use of a data access plan that identifies who has access to confidential data, when that access occurs, and what types of data and results are involved. Important documentation is discussed in more detail in section VIII.

There is also potential in adaptive trials for knowledge of the adaptation decision to convey information about the interim results. Knowledge of a sample size modification algorithm and the adaptively chosen sample size, for example, can allow back-calculation of the interim estimate of the treatment effect. Therefore, steps should be taken where possible to minimize the information that can be inferred by observers. Prespecification of the specific adaptation rule remains critical, although the protocol could perhaps outline only the general approach, with details on the specific algorithm reserved for documents such as the DMC or adaptive design charter that are made available to fewer individuals. Careful consideration and planning about the degree of information that is disseminated following an interim analysis is also important. In general, investigators and trial participants should be shielded as much as possible from knowledge of adaptive changes. For example, if the sample size is increased after an interim analysis, trial sites could be informed that the targeted enrollment number has not been reached rather than being notified of the specific targeted final sample size. The use of a discretized rather than a continuous adaptation decision threshold is another possible approach to limit the knowledge that can be inferred to help minimize risks to trial integrity.

VIII. REGULATORY CONSIDERATIONS

A. Interactions With FDA

The purpose and nature of interactions between a trial sponsor and FDA vary depending on the stage of development. The increased complexity of some adaptive trials and uncertainties regarding their operating characteristics may warrant earlier and more extensive interactions than usual. Early in the development of a drug, FDA's review of a trial protocol typically focuses on the safety of trial participants rather than the validity of inference about pharmacologic activity or efficacy. However, as resources allow, FDA might review exploratory protocols to consider the relevance of the information being gathered to guide the design of later trials. Sponsors who have questions about adaptive design elements in an early-phase exploratory trial should seek FDA feedback by requesting a meeting (or written responses only) addressing those questions. Discussion of the plans for an adaptive trial can be the basis for requesting a Type C meeting. FDA's ability to address such requests early in development may be limited and will depend on competing workload priorities and on the specifics of the development program.

At later phases of development, FDA will have a more extensive role in evaluating the design and analysis plan to ensure that the trial will provide sufficiently reliable results to inform a

regulatory decision. Regulatory mechanisms for obtaining formal, substantive feedback from FDA on later stage clinical trials are well-established and include, for example, EOP2 meetings. Depending on the preexisting knowledge regarding the drug and its intended use, and the nature of the adaptive features, an EOP2 meeting may be the appropriate setting for a sponsor to obtain feedback, or earlier interactions with FDA may be advisable (e.g., at a Type C or EOP2A meeting). Earlier interactions can help allow time for iterative discussions without slowing product development.

FDA's review of complex adaptive designs often involves challenging evaluations of design operating characteristics, usually requiring extensive computer simulations, as well as increased discussion across disciplines and FDA offices about the evaluations. This may make it difficult for FDA to adequately review such designs under short timelines. Given the timelines (45-day responses) and commitments involved with special protocol assessments (SPAs), we recommend the submission of SPAs for trials with complex adaptive designs only if there has been extensive previous discussion between FDA and the sponsor regarding the proposed trial and design.

FDA's review of proposed late-phase adaptive clinical trials will include considerations about whether the design and analysis plan satisfy the key principles outlined in this guidance. In particular, the sponsor should prespecify the details of the adaptive design and justify that the chance of erroneous conclusions will be adequately controlled, estimation of treatment effects will be sufficiently reliable, and trial integrity will be appropriately maintained. Furthermore, it is good practice for a sponsor to have explored a variety of adaptive and non-adaptive design options in planning, and to discuss its considerations in choosing the proposed adaptive design with the Agency.

Although FDA should be advised during the course of a trial of any proposed changes to the trial design (usually through protocol amendments), the Agency will generally not be involved in the prospectively planned adaptive decision-making. This is the responsibility of the sponsor, typically through the use of a committee (such as a DMC) designated to implement the adaptive design. Minutes from open sessions of a monitoring committee may be requested by the Agency during an ongoing trial, but minutes of closed sessions or any other communication or information about comparative interim results should be kept confidential until the trial concludes, except in unusual circumstances where patients' safety is at risk.

B. Documentation Prior to Conducting an Adaptive Trial

To allow for a thorough FDA evaluation, the documented plan for a clinical trial with an adaptive design will necessarily be more complex than for a trial with a non-adaptive design. In addition to the typical components of a non-adaptive clinical trial protocol and statistical analysis plan, such as those discussed in the ICH guidance *E9 Statistical Principles for Clinical Trials*, documentation submitted to the Agency prior to initiation of an adaptive design trial should include:

- A rationale for the selected design. As discussed in other sections, it is good practice to evaluate the important operating characteristics of the proposed design as compared to alternative adaptive and non-adaptive designs, and it can be useful to submit such

information to FDA. However, the ultimate choice of design is the sponsor's responsibility.

- A detailed description of the monitoring and adaptation plan, including the anticipated number and timing of interim analyses, the specific aspects of the design that may be modified, and the specific rule that will be used to make adaptation decisions.

- Information on the roles of the bodies responsible for implementing the adaptive design, such as the DMC and/or the dedicated adaptation committee, if applicable.

- Prespecification of the statistical methods that will be used to produce interim results and guide adaptation decisions, and to carry out hypothesis tests, estimate treatment effects, and estimate uncertainty in the treatment effect estimates at the end of the trial. Software to carry out interim and final analyses should be prespecified. If software for adaptation algorithms and testing and estimation methods is not commercially available, computer code should be programmed and submitted to FDA before the trial.

- Evaluation and discussion of the design operating characteristics, which should typically include Type I error probability; power; expected, minimum, and maximum sample size; bias of treatment effect estimates; and coverage of confidence intervals. Such evaluations might be achieved through analytical calculations and/or computer simulations. If operating characteristics are evaluated analytically, appropriate details (e.g., literature references or proofs) for the methodology should be submitted.

- In cases where simulations are the primary or sole technique for evaluating trial operating characteristics as defined above, a detailed simulation report should be submitted, including:

    - An overall description of the trial design.

    - Example trials, in which a small number of hypothetical trials are described with different conclusions, such as a positive trial with the original sample size, a trial stopped for futility after the first interim look, a positive trial after increasing the sample size, etc.

    - A description of the set of parameter configurations used for the simulation scenarios, including a justification of the adequacy of the choices.

    - Simulation results detailing the estimated Type I error probability and power under the various scenarios.

    - Simulation code. Since FDA reviewers will need to verify simulation studies used to evaluate trial operating characteristics, it is important to document the software package used for simulations and, if custom software was used, to provide the code used for the simulations. When code is provided, it should be readable and adequately commented. The code should include the random seeds used to generate the

simulation results. It is also helpful to provide code written in widely-used statistical programming languages. Even in cases where another language has been used to generate simulation results (typically for reasons of computational efficiency), it can be helpful to provide a runnable version of the code in a widely-used statistical programming language to facilitate the simulation review. In some cases, it will be important to include additional detailed information, such as formulas and instructions for use of simulation code.

- A summary providing overall conclusions.

- A comprehensive written data access plan defining how trial integrity will be maintained in the presence of the planned adaptations. This documentation should include information regarding: (1) the personnel who will perform the interim analyses; (2) the personnel who will have access to interim results; (3) how that access will be controlled; (4) how adaptive decisions will be made; and (5) what type of information will be disseminated following adaptive decisions, and to whom it will be disseminated. The data access plan should describe what information, under what circumstances, is permitted to be passed on to the sponsor or investigators. In addition, it is recommended that sponsors establish procedures to evaluate compliance with the data access plan and to document all interim meetings of the committee tasked with making adaptation decisions, i.e., the DMC or adaptation committee (e.g., with written minutes describing what was reviewed, discussed, and decided).

This written documentation could be included in the clinical trial protocol and/or in separate documents such as a statistical analysis plan, a DMC charter, or an adaptation committee charter. Although different types of information might be included in different documents, all important information described above should be submitted to FDA during the design stage so that the review division has sufficient time to provide feedback prior to initiation of the trial.

C. Evaluating and Reporting a Completed Trial

A marketing application to FDA that relies on a trial with an adaptive design should include sufficient information and documentation to allow FDA to thoroughly review the results. In particular, in addition to the typical content of an NDA or a BLA,[25] the application should include:

- All prospective plans, any relevant committee charters (e.g., the DMC or adaptation committee charter), and any supporting documentation, as described above (e.g., literature references, programming code, and a simulation report).

- Information on compliance with the planned adaptation rule and with the procedures outlined in the data access plan to maintain trial integrity.

[25] See, for example, the FDA guidance for industry *Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications*.

- Records of deliberations and participants for any interim discussions by any committees involved in the adaptive process (e.g., minutes from closed and open DMC or adaptation committee meetings, minutes from steering or executive committee meetings).

- Results of the interim analysis or analyses used for the adaptation decisions.

- Appropriate reporting of the adaptive design and trial results in section 14 of the proposed package insert. For example, the trial summary should describe the adaptive design utilized. In addition, treatment effect estimates should adequately take the design into account, or if naive estimates such as unadjusted sample means are used, the extent of bias should be evaluated and estimates should be presented with appropriate cautions regarding their interpretation.

More limited information (e.g., reports without the database copies and less detailed information on other aspects) may be sufficient for trial summaries provided to FDA during the course of development to support ongoing discussions within an IND.

IX. REFERENCES

Bauer, P and K Kohne, 1994, Evaluation of Experiments with Adaptive Interim Analyses. *Biometrics,* Dec 1:1029–1041.

Bauer, P and M Posch, 2004, Letter to the Editor. *Statistics in Medicine,* Apr 30; 23(8):1333–1334.

Berry, S, BP Carlin, JJ Lee, and P Muller, 2010, Bayesian Adaptive Methods for Clinical Trials, *CRC Press*; July 19.

Bolland, K, MR Sooriyarachchi, and J Whitehead, 1998, Sample Size Review in a Head Injury Trial with Ordered Categorical Responses, *Statistics in Medicine,* Dec 30; 17(24):2835–2847.

Chen, YJ, R Gesser, and A Luxembourg, 2015, A Seamless Phase IIB/III Adaptive Outcome Trial: Design Rationale and Implementation Challenges, *Clinical Trials,* Feb 1; 12(1):84–90.

Chow, SC and M Chang, 2011, Adaptive Design Methods in Clinical Trials, *CRC Press*; Dec 1.

Cui, L, HM Hung, and SJ Wang, 1999, Modification of Sample Size in Group Sequential Clinical Trials, *Biometrics*, Sep 1; 55(3):853–857.

Davey Jr, RT, L Dodd, MA Proschan, J Neaton, JN Nordwall, JS Koopmeiners, J Beigel, J Tierney, HC Lane, AS Fauci, MB Massaquoi, F Sahr, and D Malvy, 2016, A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection, *New England Journal of Medicine,* Oct 13; 375(15):1448–1456.

Denne, JS, 2001, Sample Size Recalculation Using Conditional Power, *Statistics in Medicine,* Sep 15; 20(17-18):2645–2660.

Dodd, LE, MA Proschan, J Neuhaus, JS Koopmeiners, J Neaton, JD Beigel, K Barrett, HC Lane, and RT Davey, 2016, Design of a Randomized Controlled Trial for Ebola Virus Disease Medical Countermeasures: PREVAIL II, the Ebola MCM Study, *Journal of Infectious Diseases*, Jun 15; 213(12):1906–1913.

Fisher, LD, 2008, Self-designing Clinical Trials, *Statistics in Medicine,* Jul 30; 17(14):1551–1562.

Fleming, TR, K Sharples, J McCall, A Moore, A Rodgers, and R Stewart, 2008, Maintaining Confidentiality of Interim Data to Enhance Trial Integrity and Credibility, *Clinical Trials,* April; 5(2):157–167.

Glimm, E, W Maurer, and F Bretz, 2010, Hierarchical Testing of Multiple Endpoints in Group-sequential Trials, *Statistics in Medicine,* Jan 30; 29(2):219–228.

Gould, AL and WJ Shih, 1992, Sample Size Re-estimation Without Unblinding for Normally Distributed Outcomes with Unknown Variance, *Communications in Statistics – Theory and Methods,* Oct 1; 21(10): 2833–2853.

Hey, SP and J Kimmelman, 2015, Are Outcome-adaptive Allocation Trials Ethical? *Clinical Trials,* Feb 3; 12(2):102-106.

Heyse, JF, BJ Kuter, MJ Dallas, and P Heaton,2008, Evaluating the Safety of a Rotavirus Vaccine: The REST of the story, *Clinical Trials,* 5(2):131–139.

Hung, HMJ, S-J Wang, and R O'Neill, 2007, Statistical Considerations for Testing Multiple Endpoints in Group Sequential or Adaptive Clinical Trials, *Journal of Biopharmaceutical Statistics*, Nov 14; 17(6):1201–1210.

Jennison, C and BW Turnbull, 1999, Group Sequential Methods with Applications to Clinical Trials. *CRC Press,* Sep 15.

Kieser, M and T Friede, 2003, Simple Procedures for Blinded Sample Size Adjustment that do not Affect the Type I Error Rate, 2003, *Statistics in Medicine*, Dec 15; 22(23):3571–3581.

Lan, KG and DL DeMets, 1983, Discrete Sequential Boundaries for Clinical Trials, *Biometrika*, Dec 1; 70(3):659–663.

Le Tourneau, C, JJ Lee, and LL Siu, 2009, Dose Escalation Methods in Phase I Cancer Clinical Trials, *Journal of the National Cancer Institute*, May 20; 101(10):708–720.

McMurray, JJ, M Packer, AS Desai, J Gong, MP Lefkowitz, AR Rizkala, JL Rouleau, VC Shi, SD Solomon, K Swedberg, and MR Zile, 2014, Angiotensin–Neprilysin Inhibition Versus Enalapril in Heart Failure, 2014, *New England Journal of Medicine*, Sep 11; 371(11):993–1004.

Müller, HH and H Schäfer, 2001, Adaptive Group Sequential Designs for Clinical Trials: Combining the Advantages of Adaptive and of Classical Group Sequential Approaches, *Biometrics*, Sep 1; 57(3):886–891.

O'Brien, PC and TR Fleming. A Multiple Testing Procedure for Clinical Trials, 1979, *Biometrics*, Sep 1; 549–556.

Pocock, SJ, 1977, Group Sequential Methods in the Design and Analysis of Clinical Trials, *Biometrika*, Aug 1; 64(2):191–199.

Pocock, SJ and R Simon, 1975, Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial, *Biometrics*, Mar 1; 31(1):103–115.

Proschan, MA and SA Hunsberger, 1995, Designed Extension of Studies based on Conditional Power, *Biometrics*, Dec 1:1315–1324.

Sydes, MR, MK Parmar, MD Mason, NW Clarke, C Amos, J Anderson, J de Bono, DP Dearnaley, J Dwyer, C Green, and G Jovic, 2012, Flexible Trial Design in Practice-stopping Arms for Lack-of-benefit and Adding Research Arms Mid-trial in STAMPEDE: A Multi-arm Multi-stage Randomized Controlled Trial, *Trials*, Sep 15; 13(1):1.

Tanner, MA, 1996, *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, New York: Springer.

Wassmer, G and W Brannath, 2016, *Group Sequential and Confirmatory Adaptive Designs in Clinical Trials*, Springer series in pharmaceutical statistics, New York: Springer.

Woodcock, J and LM LaVange, 2017, Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both, *New England Journal of Medicine*, Jul 6; 377(1):62–70.