Guidance for Industry

Adaptive Design Clinical Trials for Drugs and Biologics

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (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 Robert O'Neill or Sue-Jane Wang at 301-796-1700, Marc Walton at 301-796-2600 (CDER), or the Office of Communication, Outreach and Development (CBER) at 800-835-4709 or 301-827-1800.

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)

February 2010
Clinical/Medical

Guidance for Industry

Adaptive Design Clinical Trials for Drugs and Biologics

Additional copies are available from:

Office of Communication
Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 51, rm.2201
Silver Spring, MD 20993-0002
(Tel) 301-796-3400
http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

or

Office of Communication, Outreach and Development, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
(Tel) 800-835-4709 or 301-827-1800
http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/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

February 2010
Clinical/Medical

Contains Nonbinding Recommendations

Draft — Not for Implementation

TABLE OF CONTENTS

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

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I. INTRODUCTION

This guidance provides sponsors and the review staff in the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration (FDA) with information regarding adaptive design clinical trials when used in drug development programs.[2] This guidance gives advice on topics such as (1) what aspects of adaptive design trials (i.e., clinical, statistical, regulatory) call for special consideration, (2) when to interact with FDA while planning and conducting adaptive design studies, (3) what information to include in the adaptive design for FDA review, and (4) issues to consider in the evaluation of a completed adaptive design study. This guidance is intended to assist sponsors in planning and conducting adaptive design clinical studies, and to facilitate an efficient FDA review.

FDA's guidance documents, including this guidance, 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.

II. BACKGROUND

There is great interest in the possibility that clinical trials can be designed with *adaptive* features (i.e., changes in design or analyses guided by examination of the accumulated data at an interim point in the trial) that may make the studies more efficient (e.g., shorter duration, fewer patients), more likely to demonstrate an effect of the drug if one exists, or more informative (e.g., by

[1] This guidance has been prepared by the Office of Biostatistics and the Office of New Drugs in the Center for Drug Evaluation and Research (CDER) in cooperation with 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.

providing broader dose-response information). This guidance discusses clinical, statistical, and regulatory aspects of a wide range of adaptive design clinical studies that can be proposed as part of a drug development program, including both familiar and less familiar approaches. The familiar design methods are included because they represent, in many cases, well-established and relatively low-risk means of enhancing study efficiency and informativeness that may deserve wider use. The less familiar design methods incorporate methodological features with which there is little experience in drug development at this time. As more experience is obtained with the less familiar designs, the understanding of circumstances where these designs are most useful and where they may pose risks to study validity and interpretation can improve. This guidance describes aspects of adaptive design trials that deserve special consideration and provides advice on the information that should be provided to FDA and how best to interact with FDA to facilitate an efficient review.

The greatest interest in adaptive design clinical trials has been in the adequate and well-controlled study setting intended to support marketing a drug. Because these studies have the greatest regulatory impact, this guidance is generally oriented toward the use of adaptive design methods in adequate and well-controlled studies, where avoiding increased rates of false positive study results (increased Type I error rate) is critical, and introducing bias should be minimized. Many adaptive methods, however, are also applicable to exploratory studies. This guidance encourages sponsors to gain experience with the less well-understood methods in the exploratory study setting (see section IV.D).

III. DESCRIPTION OF AND MOTIVATION FOR ADAPTIVE DESIGNS

A. Definition and Concept of an Adaptive Design Clinical Trial

For the purposes of this guidance, an *adaptive design clinical study* is defined as a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study. Analyses of the accumulating study data are performed at prospectively planned timepoints within the study, can be performed in a fully blinded manner or in an unblinded manner, and can occur with or without formal statistical hypothesis testing.

The term *prospective* here means that the adaptation was planned (and details specified) before data were examined in an unblinded manner by any personnel involved in planning the revision. This can include plans that are introduced or made final after the study has started if the blinded state of the personnel involved is unequivocally maintained when the modification plan is proposed. It may be important to discuss with FDA the documentation that will provide unequivocal assurance of blinding for the pertinent personnel while a study is ongoing. Changes in study design occurring after an interim analysis of unblinded study data and that were not prospectively planned are not within the scope of this guidance.

There is a critical distinction between adaptations based on an interim analysis of unblinded results of the controlled trial (generally involving comparative analyses of study endpoints or outcomes potentially correlated with these endpoints) and adaptations based on interim noncomparative analysis of blinded data (including study endpoint data but also including data

such as discontinuation rates and baseline characteristics). Revisions not previously planned and made or proposed after an unblinded interim analysis raise major concerns about study integrity (i.e., potential introduction of bias). Protocol revisions intended to occur after any unblinded analysis should be prospectively defined and carefully implemented to avoid risking irresolvable uncertainty in the interpretation of study results. In contrast, revisions based on blinded interim evaluations of data (e.g., aggregate event rates, variance, discontinuation rates, baseline characteristics) do not introduce statistical bias to the study or into subsequent study revisions made by the same personnel. Certain blinded-analysis-based changes, such as sample size revisions based on aggregate event rates or variance of the endpoint, are advisable procedures that can be considered and planned at the protocol design stage, but can also be applied when not planned from the study outset if the study has remained unequivocally blinded.

The range of possible study design modifications that can be planned in the prospectively written protocol (or a separate, but also prospective, statistical analytic plan (SAP), if used) is broad. Examples include changes in the following:

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

For the purposes of this guidance, study design aspects that are revised based on information obtained **entirely** from sources outside of the specific study are not considered adaptive design clinical trials. Such study revisions can be prospectively planned or a response to unanticipated external events. For example, a study might be initiated before availability of expected additional information (e.g., dose response or pharmacokinetic information from a separate study) with the intent of revising the study when the external information becomes available. Revisions might also occur when additional information arises in an unexpected manner (e.g., new safety or effectiveness findings from some other source) and leads to a decision that study revision is warranted (see section IV.E for further discussion of this situation). Prospective study revisions based on information obtained from both a study-external and a study-internal source are considered adaptive designs and within the framework of this guidance, because study-internal information is used.

Study oversight responsibilities of sponsors include study monitoring for various purposes, such as to assess and ensure the quality of the study conduct and data, to project overall duration of

study enrollment, to aid study supply logistics. These important processes have been enhanced by modern technology that can facilitate frequently (and perhaps nearly continuously) updated summaries of relevant, but blinded, study information. These procedures are important to timely completion of quality studies (that provide high quality data) and are not considered adaptive features of a study. We encourage using these procedures.

B. Other Concepts and Terminology

Other concepts and terminology used in this guidance are described here:

- *Interim analysis*, for purposes of this guidance, is any examination of the data obtained in a study while that study is still 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 data elements of any data type, such as baseline data, safety outcome data, pharmacodynamic or other biomarker data, and efficacy outcome data. Analyses of outcome data can use data elements such as the observed value of a patient assessment at a specific study timepoint, event rates, or the timepoint in the study when a specific *event* occurs for the patient. Any examination of the study data, even without an intent to modify the study (sometimes called an *administrative look*), is nonetheless an interim analysis. The implications of interim analyses, as discussed below, are very different depending on whether the data examined are unblinded as to treatment group and on the particular data involved.

- Blinded analyses are those in which the treatment group assignments of study subjects are not known and are therefore not used in any manner in the analysis.

- Unblinded analyses are those in which the treatment group assignments of subjects are known and used in some manner in the analysis, usually (but not always) as a formal comparison between treatment groups. By-group results presented to decision-makers with treatment groups openly identified or with the actual identification of the group masked are both considered an unblinded analysis, and introduce the same concerns as unblinded analyses where the groups are fully identified.

- *Conventional study design* is used in this guidance to mean clinical studies of a fixed sample size that do not use adaptive elements.

- Bias in general is a systematic tendency for the estimate of treatment effect to deviate from its true value or for a statistical analysis to lead to an increased rate of Type I error. The biases of particular concern for this guidance are (1) those related to changes in study design or (2) analyses based on interim study information that have the effect of making a treatment-

[3] This definition is different from the definition in FDA's International Conference on Harmonization (ICH) guidance, *E9 Statistical Principles for Clinical Trials* (ICH E9 guidance), 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* than the ICH E9 guidance to accommodate the broad range of analyses of accumulated data that can be used to determine study adaptations at an intermediate point in the study. We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA Drugs page at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

favorable study conclusion more likely when there is in fact no treatment effect, or that lead to overestimation of the magnitude of a true treatment effect. Bias can be introduced by knowing the results associated with various choices of endpoints, subject subsets, or analyses, and choosing the most favorable. In some cases bias can be minimized by adjusting the study alpha levels (e.g., to correct for the multiplicity of analyses). In general, bias from analyses can be introduced when there are choices made based on unblinded analyses of data, whether of study endpoints or other information (e.g., pharmacodynamic or other biomarker endpoints) that correlates with study endpoints.

- The major focus of this guidance is adequate and well-controlled effectiveness (A&WC) studies intended to provide substantial evidence of effectiveness required by law to support a conclusion that a drug is effective (see 21 CFR 314.126) . A variety of terms have been used to describe different kinds of clinical trials, but a critical distinction relates chiefly to the purpose and planned use of the study results in the drug development process. The terms commonly used include *phase 1*, *phase 2*, and *phase 3* (21 CFR 312.21), and *confirmatory study* (as in the ICH E9 guidance). These terms will not be used in this guidance. The important distinction for this guidance is between A&WC studies (used here to refer only to effectiveness studies) and other studies, termed *exploratory studies*. This distinction depends on multiple features of a clinical study design, and is not necessarily determined by any single aspect of study design. For example, a multiple parallel group study evaluating a range of dose levels may have as the primary hypothesis a test of dose response. Dose-response studies may be either A&WC or exploratory, depending on features such as the nature of the primary endpoint (e.g., a clinical efficacy versus a pharmacodynamic endpoint) or the rigor of control of the Type I error rate. Because A&WC studies are used to support drug marketing, adaptive features should be used only when doing so will not increase the Type I error rate.

- The term *exploratory study*, as used in this guidance, includes studies that are not A&WC, often because they do not rigorously control the Type I error rate. Exploratory studies can be designed from the outset to allow multiple changes to the study design during the study based on interim examinations of study data, and can have multiple endpoints to be considered in the results. The term *exploratory study* in this guidance also includes studies designed to be controlled studies using an endpoint that is not suitable to be a basis of marketing approval. Exploratory studies are generally conducted earlier in the drug development program than the A&WC studies and have an important informative role in drug development. Care should be taken in their design and interpretation so that the limited amount of data, adaptive design elements, or multiple endpoints of an exploratory study do not give rise to unwarranted certainty that can lead to poor choices in areas such as dose, patient population, study endpoints.

- An A&WC study can have exploratory *elements* without becoming an exploratory *study*. The prospectively planned analyses that will support an effectiveness claim should be treated with care and rigor. A wide variety of other analyses (e.g., prospective secondary and tertiary endpoints, post hoc analyses) may be examined with less assurance of control of Type I error rate and can suggest directions for subsequent studies.

- The terms *seamless* and *phase 2/3 study* have sometimes been used in describing an adaptive design A&WC study that includes an interim analysis and an adaptation that changes the study design from having features common in an exploratory study (e.g., multiple-dose groups, multiple endpoints) to a design similar to a simple A&WC study (e.g., a single comparison with a single-dose group, a single primary endpoint). However, these terms do not add to understanding of the design beyond the already inclusive term *adaptive*. *Phase 2/3* can also lead to confusion regarding whether the study was initially designed to be A&WC, and ultimately demonstrate effectiveness. The term *seamless*, indicating that there is no long pause after the interim analysis (e.g., as between two independent studies, or between stages of single study) and that data collected from both before and after the interim analysis are used in the final analysis, describes the process of combining data in the final analysis, and is an element of all adaptive designs. Because these terms provide no additional meaning beyond the term *adaptive*, they are not used in this guidance.

C. Motivation for Using Adaptive Design in Drug Development

Interest in adaptive design study methods arises from the belief that these methods hold promise for improving drug development compared to conventional study design (i.e., non-adaptive) methods. Compared to non-adaptive studies, adaptive design approaches may lead to a study that (1) more efficiently provides the same information, (2) increases the likelihood of success on the study objective, or (3) yields improved understanding of the treatment's effect (e.g., better estimates of the dose-response relationship or subgroup effects, which may also lead to more efficient subsequent studies). FDA shares the interest of drug developers in these advantages, but is also concerned with several aspects of such approaches, notably the possible introduction of bias and the increased possibility of an incorrect conclusion.

In many drug development programs, adequate knowledge regarding all the important parameters needed for planning study design may not be present at the time the study is designed. A conventionally designed study is planned using assumptions about, and *best estimate* values for, critical elements of study design (e.g., population means or event rates, variance, dose-response effect size and location, discontinuation rates) that are not precisely known. Because a study may fail to achieve its goal when the prestudy estimates or assumptions are substantially inaccurate, conventional study designs may take the uncertainty into consideration to increase the likelihood of study success. For example, a conventional design to show an effect might use a dose-response design with multiple fixed-size randomized groups to ensure that an optimal dose level is included in the study. This design accepts the likelihood that several groups with suboptimal doses will be studied, with an attendant decrease in study efficiency. The accumulating study data, however, can provide improved knowledge of the dose-response (or other parameters) during the course of the study, if those data can be examined. An adaptive design that can ascertain when further data collection for a particular group is not useful (because that group has already been shown to represent a suboptimal dose choice), and thereby lead to discontinuation of data collection for that group, may decrease cost or time without decreasing the informativeness of the study. Similarly, an adaptive design approach that can adjust the study sample size to avoid both an underpowered study (because of an overly optimistic parameter estimate such as low variance or large treatment-effect size) and

an excessively large study (because of an overly conservative estimate of variance or effect size) might increase the study efficiency and the ability to achieve the study goal.

A potential benefit of adaptive design studies might be to yield more informative data than would otherwise be feasible given the constraints on time and resources that are allocated to a development program. Reducing the time and resources needed to assess each specific choice within a range of parameter values allows more choices to be studied using the same time frame and resources. This reduction may permit exploring a broader range of options (e.g., wider range of doses or schedules, or broader population) or more finely exploring choices within the range (e.g., narrower steps between adjacent dose levels). The resulting better optimization of the drug's use from the more extensive data may lead to an improved balance of benefit and risk or a successful drug development program that might have failed because of inadequate optimization, two obvious benefits.

A component of the potential value of adaptive design methods relates to eliminating the time period that occurs between separate exploratory and A&WC studies in conventional drug development programs. Although the efficiency gain from this elimination of time is apparent, the approach entails risks (see section IV.B) and the apparent time advantage may be less valuable if a greater period of reflection and data exploration would have allowed the design of better studies.

IV. GENERAL CONCERNS ASSOCIATED WITH USING ADAPTIVE DESIGN IN DRUG DEVELOPMENT

A. Potential to Increase the Chance of Erroneous Positive Conclusions and of Positive Study Results That Are Difficult to Interpret

Two principal issues raised by adaptive design methods are as follows:

- whether the adaptation process has led to design, analysis, or conduct flaws that have introduced bias that increases the chance of a false conclusion that the treatment is effective (a Type I error)

- whether the adaptation process has led to positive study results that are difficult to interpret irrespective of having control of Type I error

Bias can affect the validity of the statistical conclusions reached for a study and can arise from problems with the study conduct or subjective decision-making during the course of the study (called operational bias). In the case of some of the more recently developed adaptive methods, the magnitude of the risk of bias and the size of the potential bias, and how to eliminate these effects, are not yet well understood. The level of concern is greatest in an A&WC study setting but is also important in an exploratory study, where bias can adversely affect development decisions, such as choice of dose, population or study endpoints in subsequent studies. The risk of bias is greatly reduced or entirely absent when adaptations rely only on blinded analyses and the blinding is strictly maintained.

1. Bias Associated with the Multiplicity of Options

Design of a clinical study calls for the selection of design features (e.g., dose, population, endpoint, timing of the endpoint, analysis method) from among multiple possibilities. For a conventional design study, the choices are usually made before enrolling the first study subject and before any study results are seen, which contributes to avoiding bias. Where there is the opportunity to choose a study result from among the results on many endpoints, study groups, or data time points, it is well recognized that bias is introduced because of the opportunity to choose the successful result from among the multiplicity of options. In this circumstance an approach to controlling the Type I error rate should always be used.

For a situation in which multiple sequential statistical analyses of a single primary hypothesis are conducted at successive interim stages of a clinical trial, group sequential methods have been developed (see section V.D) that maintain control of the Type I error rate. Inherent in most adaptive designs are choices made from among multiple candidates (e.g., doses, population subsets, endpoints) after the study begins and at one or multiple time points during the study. Often the decisions are based on unblinded examination of interim study results. These adaptation choices create multiple opportunities to *succeed* in showing a treatment effect, with greater likelihood of doing so than when there are no adaptation opportunities. This bias inherent in this multiplicity may be readily recognized, but in complex cases may be difficult to understand and account for with statistical adjustments.

Related to statistical multiplicity, but distinct because it is not possible to enumerate the universe from which choices are made, is the situation in which a sponsor chooses a particular analysis (e.g., time point, subset, covariates, endpoint) after an unblinded, not prospectively specified exploration of the study data to identify the analysis that provides the most favorable result. A study where this occurs cannot be regarded as an A&WC study and is outside the scope of adaptive design studies discussed in this document, where all adaptive choice options are prospectively specified.

2. Difficulty in Interpreting Results When a Treatment Effect Is Shown

Adaptive designs that select the best observed interim treatment effect among the options (and especially when this occurs multiple times within a study) have the potential to select the option with an interim result that is, by random chance, more favorable than the true value. This selection process introduces a bias that will tend to provide final estimates of treatment effect that overestimate the true effect. Adjustments that appropriately control the Type I error rate are not directed at controlling the bias introduced into the effect estimate. Although in all clinical studies the uncertainty about the size of a treatment effect is captured in the confidence intervals around the point estimate, the bias in the point estimate introduced by adaptive designs could be important in decisions related to weighing benefits and risks. Because there is limited experience with the less well-understood adaptive design methods, the size of this bias and the conditions that may influence the size are not yet generally well understood. It is very important to consider this potential when planning and analyzing adaptive design studies.

When the study design includes adaptations that, during the course of the study, change the nature or type of data used in the primary analysis (e.g., changing the endpoint or study population between study stages), interpreting the study results could become more difficult. There may be, for example, uncertainty relating to which types of events are affected by the treatment or for what patient population an effect has been demonstrated. This uncertainty can be increasingly problematic when multiple adaptations are made during conduct of the study (see also section VI.F). To address this problem, analysts usually examine the overall study result and results within the relevant patient, event type, or other subsets, as well as results between the successive study portions, although it is recognized that there are limitations to detecting relevant within-study differences in treatment effect. Some of the newer methods of adaptive design offer the possibility of multiple and more complex study revisions. It is not yet known, however, whether increasingly complex designs could lead to increasingly limited amounts of data on a subset of interest, making subset examination even less informative and study interpretation excessively dependent upon judgment.

3. Operational Bias

Many study adaptations call for unblinding of the analysts charged with implementing the planned design revisions. Access by these analysts to the interim unblinded results raises concern about the possibility that the analysts might influence investigators in how they manage the trial, manage individual study patients, or make study assessments, bringing into question whether trial personnel have remained unequivocally objective. In contrast, if the personnel involved in managing study conduct, interacting with investigators, and addressing unexpected study issues remain unequivocally blinded, it is unlikely that operational bias could be introduced. Because operational bias is a nonstatistical source of bias, statistical methods cannot correct or adjust for this bias.

Shielding the investigators as much as possible from knowledge of the chosen adaptive changes is important because knowledge of the interim unblinded data used to make the adaptation decision, or even knowledge only of the specific adaptive choice, has the potential to introduce operational bias into the treatment-effect estimates. This can occur if investigators, because of their knowledge of the specific adaptation decisions, treat, manage, or evaluate patients differently. Inaccurate estimates can be produced if, for example, knowing what adaptation was selected influences investigators to identify either more or fewer endpoint events in all groups. This inaccuracy could contribute to false positive conclusions in non-inferiority trials and false negative conclusions in superiority trials. If there were some element of patient-level unblinding because of *side effects* of the treatment observable in the patient or laboratory results, the bias could also include a differential influence between treatment groups.

The role of managing study conduct and addressing unexpected study issues is a responsibility that is separate and distinct from the role a Data Monitoring Committee (DMC) will have if it is used to implement a prospective adaptation plan. Because a DMC is unblinded to interim study results, it can help implement the adaptation decision according to the prospective adaptation algorithm, but it should not be in a position to otherwise change the study design except for serious safety-related concerns that are the usual responsibility of a DMC. Indeed, FDA's guidance for clinical trial sponsors on *Establishment and Operation of Clinical Trial Data*

Monitoring Committees (DMC guidance)[4] makes the point strongly that a steering committee or other group that could possibly decide to alter study design (in a partially or fully nonprospectively specified manner) should be blinded to any interim treatment results. It is therefore critical to limit the number of study personnel who have access to unblinded data. All plans for the conduct of the unblinded interim analysis, dissemination of interim results, study modification decisions (of any kind), and distribution of detailed knowledge of the decisions should be carefully considered and documented.

B. Potential for Counterproductive Impacts of Adaptive Design

Adaptive design studies are intended to be part of an overall development program that has the intermediate goal of advancing to the next step of the program and the ultimate goal of obtaining the A&WC study data important for marketing approval. The complete program ideally should include characterizing the dose-response relationship for favorable and unfavorable effects and identifying, where possible, patient subsets that respond particularly well or poorly. Typical development programs consist of a sequence of independent studies that build upon the available information to design the next study. Completed studies are analyzed and evaluated, allowing thoughtful use of the knowledge obtained from the study to inform the choices of design and goals for the next study. A concern is that an adaptive study design will limit the opportunity to reflect on data and design a thoughtful, complete program.

1. Potential to Limit Identifying Gaps in Knowledge

An adaptive study design that is practical and interpretable can modify only a limited number of design aspects, so that only those areas of design uncertainty considered the most critical and least understood (e.g., from among dose, schedule, population, endpoint, concomitant therapies, and others) are incorporated into the adaptive features of the study design. Other aspects of the drug's use might be assumed adequately known and therefore not in need of further investigation. Using adaptive design approaches and the limited number of variables they can feasibly address, particularly for A&WC studies, may increase the pressure to make assumptions so that it would not be impractical to carry out the adaptive study, even if there is only limited prior information to support these assumptions. Avoiding the acknowledgement of uncertainties and the critical importance of actively investigating them might increase the potential for a development program to fail to demonstrate effectiveness or a favorable benefit-risk comparison because of poor choices regarding how to use the drug.

2. Elimination of Time to Thoughtfully Explore Study Results

One of the proposed advantages of an adaptive design is elimination of the time between completing exploratory studies and initiating the subsequent A&WC studies. Particularly when an exploratory study is expected to be followed by an A&WC study, only a limited number of areas of uncertainty (e.g., choice of dose, patient population, endpoint selection, sample size) might be thought to remain before the design of the A&WC study. These few areas are usually the focus of the exploratory study, and it is often hoped that the study results can be rapidly

[4] Available on the Internet at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

analyzed and applied to making the final design choices of the A&WC study. In comparison, incorporating limited exploratory goals within an adaptive design A&WC study and eliminating the independent exploratory study allows the expectation of a decrease in duration of the development program.

An often overlooked value of the time period between studies is the opportunity to thoughtfully examine the data from the exploratory study in ways not identified in the prospective analytic plan but that may reveal an unexpected aspect of the data (e.g., a substantial response difference between patient characteristic-based subsets, interactions with concomitant therapies, difficulties in adhering to a particular study procedure or other study conduct aspect, or other significant findings). This examination of the data may be important to improving the design of the A&WC study, leading to a more informative study and to one more likely to be successful. Such unexpected results are unlikely to be identified by the limited, rapid, interim data analysis of the adaptive design study. Lack of time allocated to fully explore the data may also lead to inadequate recognition of safety issues that should be assessed in A&WC studies (see section VIII), potentially lengthening the overall development program.

In light of these possibilities, using adaptive design approaches to eliminate a separate exploratory study may be less risky in situations where there are substantial amounts of relevant, well-considered, prior experience that may minimize the likelihood that there will prove to be any such important, but unrecognized, issues in the use of the drug.

3. Cautious Use of Adaptive Design Can Advance the Overall Development Program

Careful use of adaptive design methods may aid the orderly, thoughtful accumulation of data needed to optimize, establish, and adequately describe a drug's usefulness, and help avoid the negative impacts. Adaptive design studies may work best, and with least risk, when there truly are just a few issues (e.g., dose, population subsets, endpoints) that need to be examined and are built into an adaptive design.

C. Complex Adaptive Designs — Potential for Increased Planning and More Advanced Time Frame for Planning

The complexity of many adaptive study designs will call for more advance planning by sponsors, with longer lead times between initiating planning and starting the study. Interaction with FDA during study planning is particularly important for the more complex adaptive design studies, especially at the point that the A&WC studies are about to be designed. Modifying the sponsor-FDA interactions may be important to provide opportunity to obtain the comprehensive regulatory advice that may help lead to a successful study (see section X).

It has been suggested that because an adaptive design study can incorporate a planned exploration stage into an A&WC study with examination of the data in the interim analysis, followed later by analysis of the full study data in the final result, the two stages of the study can be viewed as the independent replication that is typically expected in considering whether there is substantial evidence of effectiveness that is needed for marketing approval (see 21 CFR

314.125(b)(5)). That is not the case, however, as the goal of a single, adaptive design A&WC study is to use data from all stages of the study to test one (or more) primary hypotheses. The study remains a single-study source of evidence.

D. Adaptive Design in Exploratory Studies

Exploratory studies 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, endpoints). There can be a series of separate studies in which different aspects of the drug are sequentially examined, or a more complex study attempting to evaluate multiple different aspects simultaneously. The flexibilities offered by adaptive design trials may be particularly useful in this exploratory period of development by allowing initial evaluation of a broad range of choices in drug use and more efficient recognition, as well as discontinuing evaluation of the options that are suboptimal. An adaptive design trial might allow multiple aspects of use to be optimized by sequential adaptations within a single study. Using adaptive designs in early development studies to learn about various aspects of dosing, exposure, differential patient response, response modifiers, or biomarker responses offers sponsors opportunities that can improve later studies. In particular, some of the adaptive methods whose practical properties are as yet less understood (see section VI) have been proposed in the literature to allow a more vigorous examination of certain aspects of drug use than has typically occurred in drug development programs. For example, in some circumstances both dose-group selection and response-adaptive randomization appear to have the potential to obtain a more precise description of the dose-response relationship by starting with a broader range of doses, closer spacing of doses, or both, in a study of approximately the same sample size as is generally used in a conventional exploratory study where only coarser knowledge of the relationship is obtained.

Because exploratory studies have less impact on regulatory approval decisions (than do the A&WC studies), they may be a suitable setting for gaining increased experience with the adaptive design methods discussed in section VI that so far have been infrequently used in actual studies. FDA encourages sponsors to gain experience with these adaptive design methods in this setting.

Although exploratory studies can have less rigor than A&WC studies, it is still important to be aware that inflation of the Type I error rate or biased estimates may occur in the results of exploratory studies. When unrecognized, these flaws can lead to counterproductive design decisions for subsequent studies. For example, flaws in an exploratory multiple-dose comparison study could lead to suboptimal dose selection for the subsequent A&WC study, with a resultant failure to show effectiveness or a finding of unnecessarily excessive toxicity. Thus, although unrecognized flaws in an exploratory study raise less concern regarding regulatory decisions than when similar flaws occur in an A&WC study, exploratory study design should still follow good principles of study design and consider the risk of adversely affecting the development program.

Adaptive design exploratory studies are usually different in multiple aspects of design rigor from A&WC studies so that design revisions while the study is underway will usually not be sufficient

to convert the study into an A&WC study. Studies that are intended to provide substantial evidence of effectiveness should not be designed as exploratory studies, but rather as A&WC studies at initial planning.

E. Study Design Changes That Are Not Considered Adaptive Design

1. Revisions After Unplanned Findings in an Interim Analysis

When study data are examined in an interim analysis, there may be data analyses that were not prospectively planned as the basis for adaptations, but that unexpectedly appear to indicate that some specific design change (e.g., restricting analysis to some population subset, adjusting sample size, changing between primary and secondary endpoints, changing specific methods of endpoint analysis) might increase the potential for a statistically successful final study result. As stated earlier in section III.A, such revisions based on nonprospectively planned analyses and decision paths are not regarded as adaptive design for the purposes of this guidance and will usually create difficulty in controlling the Type I error rate and difficulty in interpreting the study results.

2. Revisions Based on Information From a Study External Source

Unpredictable events that occur outside of an ongoing study during the course of drug development programs may provide important new information relevant to the ongoing study and may motivate revising the study 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 study design had the information been known when the ongoing study was designed. When this occurs, there may be reason to revise the study design in some manner (we call this a *reactive revision*) without terminating the existing study (i.e., starting an entirely new study with a modified design). In cases of serious safety concerns, and particularly in large studies, revising the study design may be critical to allowing the study to continue.

When important unexpected information arises, personnel who are (or become) familiar with both the new information and the design of the ongoing study should be given responsibility for determining revisions to the study design. If the new information is derived from sources entirely outside of the study under consideration, then the revision does not fall into the category of adaptive design. If the personnel who are determining the study revisions have no knowledge of any unblinded data or other information obtained during the study, then their decision-making cannot be influenced by study internal information to consciously or unconsciously introduce a study bias. Therefore, when contemplating a reactive study revision, study sponsors should ensure that the personnel determining the revision have no knowledge of unblinded results from the ongoing study. Importantly, the DMC of a study is not the appropriate group to determine the study revisions because they are aware of results from within the study and this could influence their decisions (see the DMC guidance).

Although carefully performed reactive revisions should not introduce a bias into the study, it is important to pay close attention to maintaining (and documenting maintenance of) the study blind. Reactive revisions, however, can lead to interpretive problems. When an important revision in study design is made midway in a study, it may not be fully clear how the data from before the revision and after the revision should be combined, and how to interpret the study results. Resolution of these interpretive difficulties when the overall study result is statistically significant will inevitably depend on judgment.

V. GENERALLY WELL-UNDERSTOOD ADAPTIVE DESIGNS WITH VALID APPROACHES TO IMPLEMENTATION

There are well-established clinical study designs that have planned modifications based on an interim study result analysis (perhaps multiple times within a single study) that either need no statistical correction related to the interim analysis or properly account for the analysis-related multiplicity of choices. A considerable experience in modern drug development provides confidence that these design features and procedures will enhance efficiency while limiting risk of introducing bias or impairing interpretability.

Many of the best-understood adaptive design methods do not involve examining unblinded study outcome data and examine only aggregate study outcome data, baseline data, or data not related to the effectiveness outcome (see sections V.A, B, and C). Other adaptive methods use the well-understood group sequential design (see section V.D and the ICH E9 guidance). In group sequential designs, unblinded interim analyses of accruing study data are used in a planned and confidential manner (i.e., by a DMC) that controls Type I error and maintains study integrity.

This section will describe some of the approaches that are well-understood, emphasizing the principles that explain why they are well understood. The descriptions and discussion in the following subsections are intended to aid in determining whether other existing or future-developed methods share the same principles.

A. Adaptation of Study Eligibility Criteria Based on Analyses of Pretreatment (Baseline) Data

Clinical studies are generally planned with expectations about the patient population characteristics and the rate at which eligible patients will be identified and enrolled. For example, the study designers may have tried to enroll patients with a broad distribution in certain identified characteristics to maximize a study's utility. Examination of baseline characteristics of the accumulating study population might show that the expected population is not being enrolled and that by modifying eligibility criteria, subsequent subject enrollment may be shifted towards a population with greater numbers of patients with the desired characteristics. Similarly, if the study enrollment rate is substantially slower than expected, the screening log can be examined for noncritical entry criteria that might be modified to allow greater numbers of screened patients to qualify.

Such examination of baseline information and modification of study eligibility criteria can contribute to timely completion of informative studies. Knowing the baseline characteristics of

the overall study population at any time during the study does not generate concerns of introducing statistical bias as long as the treatment assignment remains blinded.

A possible risk of such an approach is the potential to impair the interpretation of the study result when the study population changes mid-way and an important relationship of treatment effect to the changed patient characteristic exists (i.e., a treatment-patient factor interaction). Exploratory analyses of the data obtained before and after the eligibility change can help to identify such problems.

Because post-baseline patient data are not involved in the analyses, the study sponsor or investigator steering committee can review the baseline-data summaries and make design changes to the eligibility criteria without risk to the integrity of the study.

B. Adaptations to Maintain Study Power Based on Blinded Interim Analyses of Aggregate Data

One of the important challenges in planning A&WC studies is deciding on the sample size at the study design stage. In general, the estimated power of a study to detect a treatment effect is dependent upon the study sample size, the targeted (e.g., the sponsor's assumed actual or minimum acceptable) treatment-effect size, the assumed population variance of the patient measure being studied, or the expected control group event rate for event-driven studies. If any of the assumptions used to calculate the sample size are incorrect, the study may be underpowered and fail to show an effect. There are several approaches to maintaining study power.

In studies using a discrete outcome (event) endpoint, a blinded examination of the study overall event rate can be compared to the assumptions used in planning the study. Examining the data in this blinded analysis does not introduce statistical bias, and no statistical adjustments are required. If this comparison suggests the actual event rate is well below the initial assumption, the study will be underpowered. The study sample size can be increased to maintain the desired study power or, alternatively, study duration might be increased to obtain additional endpoint events. Study resizing based on a revised estimate of event rate should be used cautiously early in the study, as variability of the estimated event rate can be substantial. Consequently, this adaptive approach may be best applied later in the study when population estimates of the event rate are more stable.

For studies using a time-to-event analysis, another approach is not to plan a specific study sample size in the protocol, but rather to continue patient enrollment until a prospectively specified number of events has occurred (an *event-driven* study). The interim data analyses are of the overall number of study endpoint events, rather than the overall rate of events.

Similarly, when a continuous outcome measure is the study endpoint, a blinded examination of the variance of the study endpoint can be made and compared to the assumption used in planning the study. If this comparison suggests the initial assumption was substantially too low and the study is consequently underpowered, an increase in the study sample size can maintain the desired study power. As with event endpoints, study resizing based on a revised estimate of

variance should be used cautiously early in the study, as variability of the estimated variance can be substantial.

In some studies with continuous outcome measures the duration of patient participation and time of last evaluation may be the preferred design feature to modify. A study of a chronic, progressive disease with a treatment intended to stabilize the clinical status is dependent upon the control group demonstrating a worsening of the condition, but there may have been only limited prior data upon which the design-assumed rate of progression was based. An interim analysis of the aggregate rate of progression can be useful to assess whether the duration of the study should be adjusted to allow for sufficient time for the group responses to be distinguished, given the assumed treatment-effect size. A combination of sample size and duration modification can also be applied in this case to maintain the desired study power.

Alternatively, if it is thought that patients can be stratified at baseline (e.g., by a genetic or disease-phenotype characteristic) into subsets expected to differ in an important aspect related to the endpoint (e.g., event rate, variance, rate of disease progression), the blinded interim analysis of the event rate (or, e.g., variance) can be done by subset and study eligibility criteria modified to focus the remainder of the study on the subset(s) with the advantageous tendency (e.g., greater event rate, lower variance). A sample size readjustment could be considered at the same time.

Usually, the blinded interim analyses considered here are used to make decisions to increase the sample size, but not to decrease the study size. Decreasing sample size is not advisable because of the chance of making a poor choice caused by the high variability of the effect size and event rate or variance estimates early in the study.

The ability of these procedures to increase the potential for a successful study while maintaining Type I error control has been recognized and discussed in the ICH E9 guidance. Sample size adjustment using blinded methods to maintain desired study power should generally be considered for most studies.

Because these methods avoid introducing bias by using only blinded interim analyses, all study summaries should not contain any information potentially revealing the between-group differences. For example, even a data display showing the distribution of aggregate interim results might reveal the presence, and suggest a size, of a treatment effect (e.g., a histogram showing a bimodal distribution of the endpoint data), and might influence the personnel making these adaptations.

C. Adaptations Based on Interim Results of an Outcome Unrelated to Efficacy

There are some circumstances where study modifications are based on an interim analysis of outcomes that are independent of, and uninformative about, the treatment-related efficacy effect. Concerns about statistical and operational bias usually are not raised by such interim analyses and modifications if there has been no unblinded analysis of any effectiveness-related data. Control of Type I error rate is thus maintained without a statistical adjustment for such adaptations.

At the time that a study is being designed it is not uncommon to be uncertain about how patients may respond to the treatment in ways not measured by the efficacy outcome. For example, there may be a known or potential adverse reaction with an incidence too low to have been accurately estimated from prior experience, but of a severity so substantial that it could outweigh the possible benefits from the treatment. Randomized, parallel, dose-response studies are generally most informative when a broad range of doses are studied. When this is done, however, some doses might cause an unacceptable rate of a serious adverse effect or a less serious adverse effect sufficient to make the treatment unattractive (e.g., causing a high treatment discontinuation rate). It is therefore important to look for these events at an interim stage of the study and discontinue a dose group with unacceptable observed toxicity. If the adverse effect is completely independent of the treatment's benefit, then an unblinded analysis of the rate of the adverse effect provides no knowledge of the efficacy results and the Type I error rate remains controlled without an adjustment. Similarly, if an unexpected serious toxicity is observed in safety monitoring, dropping the dose groups with excessive toxicity is usually appropriate.

It is common to have study designs that initiate testing with several dose or regimen groups, with the intent of dropping dose groups that are poorly tolerated and enrolling subsequent patients into the remaining groups. To ensure full awareness of the process and avoid missteps that could compromise the study integrity, the design and analysis plan should specify the number of groups to be terminated, how they will be selected, and the appropriate analysis procedures for testing the final data (e.g., adjustment for multiplicity when more than one dose is planned to be carried to completion). A design of this type may be particularly useful in long duration studies where the adverse event of concern occurs at a low rate (and therefore cannot be precisely assessed in small exploratory studies) and occurs relatively early after initiating treatment. For example, studies of platelet inhibiting drugs have sought to demonstrate long-term efficacy using the highest dose not causing excessive rates of early bleeding.

It is important to emphasize that this approach may be undesirable if there might be greater effectiveness associated with the more toxic dose that could outweigh the increased toxicity in a risk-benefit comparison. The nature and implications of the possibly greater toxicity should be carefully considered and this approach used only when there is confidence the greater toxicity will outweigh greater effectiveness.

If there are no efficacy-related interim analyses performed, the interpretability of the final study result is not impaired by concerns of statistical bias or operational bias in study conduct. Study planning should assure that the personnel who make the modification decision (e.g., a steering committee) have not previously seen any unblinded efficacy analyses. As emphasized, the outcome examined must not be the efficacy outcome, nor an outcome related to efficacy in any way that allows inferences to be formed regarding the efficacy outcome. Thus, secondary or tertiary efficacy endpoints, or biomarkers thought to have some relationship to efficacy, should not be used in this approach. A design modification based on an efficacy-related endpoint or biomarker will call for an appropriate statistical adjustment (see section VI.C).

Situations where a drug-induced serious or fatal outcome is an event to be avoided (thus monitored for treatment-related increase) and is also an important component of a composite efficacy outcome cannot be considered in this paradigm. Other approaches (e.g., group

sequential designs) should be used in these situations to protect the integrity of the study. The concern is that because the interim results are related to efficacy, the DMC might be biased in making any subsequent decisions about study modification.

D. Adaptations Using Group Sequential Methods and Unblinded Analyses for Early Study Termination Because of Either Lack of Benefit or Demonstrated Efficacy

Group sequential statistical design and analysis methods have been developed that allow valid analyses of interim data and provide well-recognized alpha spending approaches to address the control of the Type I error rate (e.g., O'Brien-Fleming, Lan-DeMets, Peto methods) to enable termination of a study early when either no beneficial treatment effect is seen or a statistically robust demonstration of efficacy is observed. Aspects of group sequential monitoring are discussed in the ICH E9 guidance.

In circumstances where the drug has little or no benefit, the data accumulated before planned completion of the study might provide sufficient evidence to conclude that the study is unlikely to succeed on its primary objective, even if it were carried to completion. Discontinuing the study for these reasons at this interim point, often called *futility*, might save resources and avoid exposure of more patients to a treatment of no value.

Studies with multiple groups (e.g., multiple-dose levels) can be designed to carry only one or two groups to completion out of the several initiated, based on this type of futility analysis done by group. One or more unblinded interim analyses of the apparent treatment effect in each group is examined, and groups that meet the prospective futility criterion are terminated. However, because of the multiplicity arising from the several sequential interim analyses over time with multiple between-group analyses done to select groups to discontinue (see section VI.A), statistical adjustments and the usual group sequential alpha spending adjustments need to be made in this case to control Type I error rates.

For the group sequential methods to be valid, it is important to adhere to the prospective analytic plan, terminating the group if a futility criterion is met, and not terminating the study for efficacy unless the prospective efficacy criterion is achieved. Failure to follow the prospective plan in either manner risks confounding interpretation of the study results.

If the drug is more effective than expected, the accumulating data can offer strong statistical evidence of the therapy's success well in advance of the planned completion of the study. If the study outcome is one of great clinical importance, such as survival or avoidance of irreversible disability, ethical considerations may warrant early termination of the study and earlier advancement of the product towards widespread availability in medical practice. It is important to bear in mind that early termination for efficacy should generally be reserved for circumstances where there is the combination of compelling ethical concern and robust statistical evidence. A study terminated early will have a smaller size than the initially planned study size. It will therefore provide less safety data than planned. A potential also exists for more difficulty with the efficacy analysis and interpretation related to issues that become apparent only during the later detailed analysis (e.g., related to loss to follow-up or debatable endpoint assessments) and decreased power to assess patient subsets of interest.

Group sequential designs offer a method for early termination of a study as an adaptive design element, allowing the study sample size to be reduced to the size accumulated at the time of an interim data analysis. Most of the commonly used methods employ conservative (small p-value) criteria for terminating on the basis of demonstrated efficacy.

Implementation of group sequential design methods involves unblinded analyses of the treatment effect, thereby raising significant concerns for potentially introducing bias into the conduct of the study or into subsequent decisions regarding the conduct of the study. Protocols using group sequential designs have addressed this concern by using a committee independent of the study's conduct and sponsor to examine these analyses in a secure and confidential manner. An independent, nonsponsor-controlled Data Monitoring Committee (DMC) (see the DMC guidance) is an inherent part of the group sequential method's protection of study integrity. These well-established DMC procedures more recently have led to using DMCs to implement other adaptive procedures as well. Less well settled, however, is which parties prepare the analyses for the DMC to consider and the independence of the statistician preparing these analyses. The DMC guidance does not reach firm conclusions on this, but it is critical that the analyses be carried out either externally to the study sponsor or by a group within the sponsor that is unequivocally separated from all other parties to the study.

An unblinded interim analysis exposes the DMC (or other involved committee) to confidential information. Any subsequent decisions or recommendations by the DMC related to any aspect of study design, conduct or analysis can be influenced by the knowledge of interim results, even if the decision is intended to be unrelated to the prior interim analysis. For example, if new information should become available from a source outside the study, but relevant to the ongoing study, the DMC will no longer be the appropriate group to consider and recommend study design changes in response to the new information. This task will usually fall to a blinded steering committee. This issue is emphasized in the DMC guidance.

E. Adaptations in the Data Analysis Plan Not Dependent on Within Study, Between-Group Outcome Differences

The statistical analytic plan (SAP) for the clinical trial often makes assumptions regarding the distribution of the outcome data. Analytic methods may also be sensitive to the amount of, or approach to, various types of observed data (e.g., distribution of values, missing data). When study data do not conform to the assumptions of the planned analytic methods or are overly sensitive to other data behavior, the validity of conclusions drawn from the study analysis can be affected.

Generally, the prospective SAP should be written carefully and completely, and implemented without further changes once the study has started. However, if blinding has been unequivocally maintained, limited changes to the SAP late in the study can be considered. The ICH E9 guidance suggests that after a blinded inspection of the data, the SAP can be updated regarding the appropriate data transformations, adding covariates identified from other research sources or reconsideration of parametric versus nonparametric analysis methods. In some cases, with unequivocal assurance that unblinding has not occurred, this approach can also be applied to

changes in the primary endpoint, composition of the defined endpoint-event, or endpoint analytic sequence ordering.

In certain situations, the optimal statistical analysis plan may be difficult to specify fully before completing the study and examining the relevant characteristics of the final outcome data. If these characteristics are examined for the entire study population in a blinded manner, analytic plan modifications based on these characteristics do not introduce bias. The prospective analysis plan should clearly specify the characteristics and the procedure for selecting the analysis methodology based on these data characteristics.

Examples of where this may be useful include situations in which the observed data violate prospective assumptions regarding the distribution of the data, or where data transformations or use of a covariate is called for in the analysis to achieve adequate conformity with the method's assumptions.

Adaptation of the primary endpoint according to prospectively specified rules may also be useful in some circumstances. For example, when an outcome assessment that is preferred as the primary endpoint proves difficult to obtain, a substantial amount of missing data may occur for this assessment. An analytic plan might direct that if the amount of missing data in the preferred outcome assessment exceeds some prospectively stated criterion, a specified alternative outcome would be used as the primary efficacy endpoint. Similarly, when a composite event endpoint is used but there is uncertainty regarding the event rates to expect for the possible components, an analytic plan accommodating inclusion of one or two specific additional types of events might be appropriate if an insufficient number of events within the initial composite were observed in the overall study. In a similar manner, selection or sequential order of secondary endpoints might also be adapted.

VI. ADAPTIVE STUDY DESIGNS WHOSE PROPERTIES ARE LESS WELL UNDERSTOOD

This section provides an overview of adaptive study designs with which there is relatively little regulatory experience and whose properties are not fully understood at this time. These clinical trial design and statistical analysis methods are primarily intended for circumstances where the primary study objective(s) cannot be achieved by other study designs, such as those described in section V. The study design and analysis methods discussed in this section are limited to parallel group randomized study designs, and they can have several adaptive stages. The chief concerns with these designs are control of the study-wide Type I error rate, minimization of the impact of any adaptation-associated statistical (see section VII.B) or operational bias on the estimates of treatment effects, and the interpretability of trial results. This section does not discuss sequential group dose escalation study designs or dose de-escalation study designs, which are non-comparative designs that can be conducted in early drug development.

The less well-understood adaptive design methods are all based on unblinded interim analyses that estimate the treatment effect(s). The focus of the discussions in this section is primarily on

specific categories of adaptation methods, whereas the more general implementation issues that the methods raise are discussed in section VII.

A. Adaptations for Dose Selection Studies

A critical component of drug development is to estimate the shape and location of the dose-response relationship for effectiveness and adverse effects, which can have different dose relationships. Understanding these relationships facilitates selecting doses for more definitive effectiveness and safety evaluation in the A&WC studies of late clinical development (see FDA's ICH E4 guidance *on Dose-Response Information to Support Drug Registration*[5]), and in some cases can provide labeling guidance on starting and maximum doses for patient management. Too often, however, the A&WC studies evaluate only a single dose or two doses spanning a narrow dose range based on a tenuously understood dose-response relationship developed from very limited data. Unsuccessful development can result from focusing on a single dose in the A&WC studies if the single selected dose does not demonstrate effectiveness or if very important but less common adverse effects are identified in the larger A&WC studies, whereas a different dose could have provided an improved benefit to risk comparison. It is also possible that the selected dose is needlessly large and the excessive dose causes a serious but uncommon adverse effect that will be discovered only in the postmarketing period. Consequently, there is considerable interest in whether adaptive design techniques based on unblinded interim analysis of efficacy data can enable improved understanding of the dose-response relationship.

The term *dose* refers not only to a specific chosen dose level, but also includes the schedule (i.e., administration frequency) and in some cases the duration of use. The different doses evaluated in a dose-response study can be distinguished by any of these aspects of a regimen. Typically, a dose exploration study randomizes patients among placebo and several dose groups. The resulting data can be analyzed to identify the one or several groups with best response (i.e., the existence of a dose-response relationship for effectiveness or safety) or for the therapeutic window (by balancing safety, including tolerability, and efficacy).

An adaptive exploratory dose-response study is intended to begin with multiple doses (sometimes many) across a range. The number of dose groups is adaptively decreased during the course of the study, using the accruing efficacy or safety data in a prospectively specified plan for design modification at one or more unblinded interim analyses. The response evaluated at the interim analyses is often the clinical efficacy endpoint, but could also be a biomarker. Many adaptive study designs only eliminate unsuitable or uninformative doses, but addition of new, potentially more preferable doses is also possible. Some adaptive designs can also adjust the sample size of the overall study or of the individual dose groups to obtain response estimates of a particular desired precision. In some situations, an exposure-response relationship for effectiveness or safety may be used in place of dose-response. These prospectively planned study designs offer flexibility that can allow many potential modifications.

[5] Available on the Internet at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.

A particularly interesting exploratory approach is to use an adaptive design exploratory dose-response study with a moderate number of doses (five to seven) with the objective of identifying the shape (from among several different potential modeled-shape relationships) and location of the dose-response relationship, as well as optimizing selection of two or three doses (which might be the same as, or between, doses that were tested in the exploratory study) for evaluation in subsequent A&WC studies. Irrespective of whether this particular approach is used, fully evaluating more than one dose in the larger A&WC studies is almost always advisable whenever feasible.

Highly flexible modifications should generally be limited to an exploratory study, but some of these approaches, when used with rigorous protection of the Type I error rate, might have a role in A&WC studies. For example, a common design for an A&WC study is to evaluate two doses thought likely to offer a favorable benefit-risk comparison. If there was significant residual uncertainty in selecting the two doses, the study design might also include a third dose to begin the study (higher or lower than the two doses thought likely). An interim analysis of the treatment effect in each dose group would enable terminating the dose that appeared least likely to be useful, allowing the study to continue thorough evaluation of two doses with improved chances for success. Using this approach in an A&WC study will call for careful statistical adjustment to control the Type I error rate and should be limited to modest pruning of the number of dose groups.

In some development programs a biomarker (or an endpoint other than a clinical effectiveness endpoint) might be used for the interim analysis to determine the adaptive modification. If there is limited or uncertain predictiveness of the biomarker for the clinical outcome, however, there may be uncertainty regarding how well such a design will optimize the drug's clinical effects. Sponsors should consider the level of uncertainty in that relationship and the potential consequences when planning to employ this approach. In addition, because of the correlation between the biomarker and the ultimate clinical endpoint, introduction of bias is a concern and statistical adjustments are needed to control the Type I error rate.

B. Adaptive Randomization Based on Relative Treatment Group Responses

Adaptive randomization is a form of treatment allocation in which the probability of patient assignment to any particular treatment group of the study is adjusted based on repeated comparative analyses of the accumulated outcome responses of patients previously enrolled (often called *outcome dependent randomization*, for example, the *play the winner* approach). The randomization schedule across the study groups can change frequently or continuously over the duration of the study. This design is facilitated when the subjects' outcomes are observed soon after initial exposure relative to the rate at which study enrollment occurs. Previously, this randomization method had been used in placebo controlled studies chiefly to place more patients into the group with better outcomes. More recently the approach has been revised to suit the objective of dose-response evaluation. The method allocates fewer subjects to doses that appear to have a low probability of a treatment-related efficacy response, to have a high probability of an adverse event, or to be unlikely to contribute additional information on the shape of the dose-response profile. Outcome dependent adaptive randomization is particularly valuable for exploratory studies because it can make practical an increase in the number of tested treatment

options (increased breadth to the range of doses tested and/or decreased step size between doses) explored for the drug's activity and facilitate estimation of the dose-response relationship, and hypothesis testing is not the study objective. Adaptive randomization should be used cautiously in A&WC studies, as the analysis is not as easily interpretable as when fixed randomization probabilities are used. Particular attention should be paid to avoiding bias and controlling the Type I error rate.

The expectation in clinical studies of balance among the treatment groups with regard to important baseline characteristics relies upon the use of randomization, and provides a valid basis for statistical comparisons. When patient outcome is a function of covariates and treatment group assignment, changing randomization probabilities over the course of the study raises a concern regarding the balance of patient characteristics among the treatment groups. If patients enrolled into the study change in the relevant baseline characteristics (either measured or unmeasured) over the time course of the study, the changing allocation probabilities could lead to poor balance in patient characteristics between the groups at the end of the study. If the characteristics in poor balance have an influence on outcome, inaccuracy is introduced into the estimated treatment effect between groups. A dose-response profile obtained from an exploratory study with this approach could lead to poor dose selection for subsequent studies; this issue should be considered for such studies. Such poor balance in important characteristics could be a very significant problem for an A&WC study.

To address the concern regarding patient characteristics, we recommend that sponsors maintain randomization to the placebo group to ensure that sufficient patients are enrolled into the placebo group along the entire duration of the study. Examining an exploratory analysis of response over time within the placebo group, and examining exploratory comparisons of response in the placebo group to drug-treated groups by dividing the study into periods of enrollment, may help evaluate this concern for a completed study. Maintaining the placebo group will also best maintain the power of the study to show a treatment effect. It is also prudent to consider the treatment-effect estimate obtained from an adaptive randomization exploratory study cautiously, and this estimate should probably be used more conservatively in setting the sample size of a subsequent A&WC study to offset the potential over-estimate of effect size.

C. Adaptation of Sample Size Based on Interim-Effect Size Estimates

In a fixed sample size A&WC study design, planning for the sample size involves consideration of the following: a postulated treatment effect size, an assumption about the placebo event rate in event outcome studies or the variability of the primary outcome endpoint in other studies, the desired Type I error rate, and the desired power to detect the posited treatment-effect size. Other factors (e.g., stratification and dropout rates) can also be considered. Usually, the sample size (or total event count) is prospectively determined and fixed in advance using this information; however, study designs with group sequential methodology (see section V.D) might stop the study early with a smaller than planned sample size (or event count) for either lack of effect or overwhelming evidence of an effect larger than expected.

Section V.B describes a number of adaptations of sample size or event count (or study duration in certain circumstances) based on blinded analyses. In contrast, one adaptive design approach is

to allow an increase in the initially planned study sample size based on knowledge of the unblinded treatment-effect sizes at an interim stage of the study if the interim-observed treatment effect size is smaller than had been anticipated but still clinically relevant. In general, using this approach late in the study is not advisable because a large percentage increase in sample size at that point is inefficient. In some designs, other study features that affect the estimated power of the study might be changed at the same time, such as modifying the components of a composite primary endpoint (see section VI.E). In other cases, an adaptation that focuses on another aspect of study design (e.g., dose, population, study endpoint) could alter the study power, warranting reestimation of study sample size to maintain study power. There are several methods for modifying the sample size of the trial, and these methods frequently are based on conditional power or predictive power. Adaptive designs employing these methods should be used only for increases in the sample size, not for decreases. The potential to decrease the sample size is best achieved through group sequential designs with well-understood alpha spending rules structured to accommodate the opportunity to decrease the study size by early termination at the time of the interim analysis.

A change in study sample size related to an unblinded data analysis (as opposed to one based on blinded analyses discussed in section V.B) can cause an increase in the Type I error rate. To protect against such an increase, a statistical adjustment is necessary for the final study analysis. Some methods for this adjustment decrease the alpha level at which statistical significance is determined, whereas other methods will perform the hypothesis test at the usual alpha level but weight the data from the successive portions of the study unequally. Another method combines aspects of both alpha adjustment and weighting adjustment, and generally results in reasonable sample size increases. The weights for each study portion should be selected prospectively and not determined after the unblinded interim analysis. The selected balance of weights should be carefully considered because they can affect the statistical efficiency of the design. Differential weighting, however, can lead to some difficulties in interpreting the final analysis. When the weighting is not proportional to the patient numbers in each stage, individual patient data from the different stages do not have equal contribution to the overall treatment-effect estimate. This could lead to an estimate of the treatment effect that is different from the estimate when all patients are given equal weight, with resulting confusion regarding the amount of benefit demonstrated.

Estimates of treatment effect observed early in the study, when there are relatively fewer patient data, are generally variable and can be misleadingly large or small. Thus, those responsible for monitoring the study should act conservatively when deciding upon study changes using the early estimates. This is similar in spirit to the approach used in group sequential design alpha spending functions, where more conservative alpha spending is used early in the study.

D. Adaptation of Patient Population Based on Treatment-Effect Estimates

As previously noted (for blinded analysis methods discussed in section V.B), modification of the patient population enrolled (i.e., enrichment modification designs) into a study can sometimes improve the power of a study to detect a treatment effect. The blinded-analysis methods are useful when the purpose of the modification is to increase the ability to show a treatment effect

when the treatment effect is not expected to substantially differ among the various population subsets. These methods do not raise concern about increasing the Type I error rate.

In some circumstances, however, genetic, physiologic, or other baseline characteristics are thought to potentially distinguish patient subsets that have differing responsiveness to the drug treatment. Identifying these characteristics is typically done as part of exploratory studies, and is important to selecting the patient population for study in the A&WC studies. Adaptive design studies using unblinded interim analyses (of either clinical or biomarker data) for each subset of interest have been proposed as another method for identifying population subsets with relatively greater treatment responsiveness. Adaptive methods might, for example, be used within a traditional dose response exploratory study so that the study results guide optimal design for dose and population selections for the subsequent A&WC study. In some cases where the data from exploratory studies are suggestive of population subset-response differences, but inadequate to confidently select a fixed patient population for the A&WC study, these methods might be cautiously applied in an A&WC study to modify eligibility criteria after the interim analysis. These designs are less well understood, pose challenges in avoiding introduction of bias, and generally call for statistical adjustment to avoid increasing the Type I error rate.

Adaptive methods that have been proposed include (1) changing only the eligibility criteria, with no change in the study overall sample size and with the final analysis including the entire study population, or (2) modifying the plan for the final analysis to include only patients with the preferred characteristic. Other methods can increase the sample size for the population subset with the desired characteristic. The prospective study plan should ensure control of the Type I error rate for all hypotheses tested. Each method will involve different approaches to statistical adjustment. There may be no statistical adjustment necessary if there are no changes in the hypotheses tested. Caution should be exercised in planning studies where an interim analysis and eligibility modification are performed multiple times, because when multiple revisions to the study population are made it may be challenging to obtain adequate estimates of the treatment effect in the populations of interest, or to interpret to what patient population the results apply.

E. Adaptation for Endpoint Selection Based on Interim Estimate of Treatment Effect

Planning a clinical trial involves careful selection of the primary and secondary effectiveness endpoints. At the planning stage, the optimal endpoints for assessing the disorder or the disease aspects that best exhibit the particular drug's effects may not be well understood. Choosing endpoints in this circumstance may be difficult at the time of study design. Changing the ordering of endpoints (including switching primary and secondary endpoints) based on an unblinded interim analysis of treatment effect might have value in such cases. Endpoint adaptation should have appropriate statistical procedures to control the Type I error rate for the multiplicity of possible endpoint selections. If the size of the interim dataset is insufficient to provide a stable assessment of the effect-sensitivity differences between endpoints, however, this approach risks selecting a poor endpoint.

Primary endpoint revision usually takes one of two forms, replacement of the designated primary endpoint with an entirely new endpoint, or modification of the primary endpoint by adding or removing data elements to the endpoint (e.g., the discrete event types included in a composite

event endpoint). In addition to prospectively stating all possible endpoint modifications study designers should ensure that all possible choices are appropriate for the objective of the study (e.g., all possible primary endpoints in an A&WC study are clinical efficacy endpoints). This adaptive design approach is an alternative to a fixed design with two (or more) primary endpoints and appropriate multiplicity adjustment. Study planners should ensure the adaptive design provides advantages over the fixed design before adopting it.

A general concern with endpoint modification involves the quality of the data on each endpoint. For example, knowledge of which endpoint has been designated the primary endpoint and/or the chief secondary endpoint could influence the study conduct at some sites in the evaluations for endpoints (or endpoint event components) designated less important (i.e., as *only* backup endpoints) and lead to lower quality data than for those initially designated most important. An interim analysis that includes these lower quality endpoint data can result in misleading effect-size comparisons between endpoints and a counterproductive change in the endpoint. Sponsors conducting an endpoint-adaptive study should be particularly alert to ensuring that the data on each endpoint are collected in a uniform manner with good quality, both before and after the interim analysis and design modification.

F. Adaptation of Multiple-Study Design Features in a Single Study

In theory, adaptive design methods allow more than one design feature to be modified during a study. The study design should prospectively account for the multiple adaptations and maintain control of the study-wide Type I error rate. An adaptive design study could include interim analyses for any of a number of adaptations, such as modification of treatment dose, efficacy endpoint, patient subset, study duration, or study sample size. These revisions could be made at one time or divided across several times during a study.

When multiple adaptations are planned within a single study, the study will become increasingly complex and difficult to plan, with increased difficulty in interpreting the study result. In addition, if there are interactions between the changes in study features, multiple adaptations can be counterproductive and lead to failure of the study to meet its goals.

Because of these concerns, an A&WC study should limit the number of adaptations. Exploratory studies may be better suited to circumstances when multiple adaptations are warranted.

G. Adaptations in Non-Inferiority Studies[6]

Non-inferiority studies rely on many of the same types of assumptions in determining the study design features that are used to design superiority-comparison studies. Accuracy of these assumptions similarly affects whether the study is adequately powered to achieve the study objective. When there is uncertainty in these assumptions, non-inferiority studies also have the potential to be strengthened by interim analyses that examine the accuracy of some of these assumptions and readjust the study size, if appropriate. A blinded interim analysis (e.g., of overall event rate, variance, demographic features of the study population) can often be entirely

[6] A draft guidance is under development and will publish soon. When finalized, this guidance will provide additional information on non-inferiority studies.

sufficient to enable reconsideration of study sample size (see section V.B), and might pose fewer difficulties and risks than methods that rely on an unblinded analysis.

When blinded interim analyses of non-inferiority studies are conducted, a larger sample size might improve the statistical power to meet the prospective non-inferiority margin, and can also increase the potential to demonstrate superiority of the test agent over the comparator in the case where this is true. If the superiority demonstration is also a (secondary) goal of the study sponsor, but the extent of the superiority could not be estimated at the time of study design so that the feasibility of the sample size was uncertain, an adaptive design to modify the study size based on an unblinded interim analysis could be considered. The methods discussed previously are suitable for this adaptive modification if the non-inferiority objective is met at the interim analysis point, and may call for a statistical adjustment to control the Type I error rate for the superiority comparison.

Many design features of a non-inferiority study may not be suitable for adaptation. Chief among these features is the non-inferiority margin. The non-inferiority margin should be carefully determined during study design, is based largely on historical evidence that does not change, and should not be part of a modification plan for a study. The patient population enrolled in the study may also be difficult to change. The non-inferiority margin is based on historical studies that had enrolled patients meeting specified criteria, and may apply only to a study population that is similar in important characteristics. Changing the enrolled patient population (e.g., to increase the rate of enrollment) to a population substantially different from that enrolled in the historical studies may compromise the validity of the non-inferiority comparison. Similarly, adequate historical data on which to base a non-inferiority margin is often available for only one endpoint, so that endpoint selection cannot be adaptively modified in the study.

VII. STATISTICAL CONSIDERATIONS FOR LESS WELL-UNDERSTOOD ADAPTIVE DESIGN METHODS

This section deals with statistical considerations for an adaptive design study that incorporates the more complex approaches described in section VI and that is intended to be an A&WC trial. This section discusses the concern for statistical bias as defined in the ICH E9 guidance. The primary statistical concern of an A&WC study is to control the overall study-wide Type I error rate for all hypotheses tested. This rate can increase in adaptive design studies because of multiplicity related to the multiple adaptation options (and the associated multiple potential hypotheses) or by using biased estimates of the treatment effect. Another concern is avoiding inflation of the Type II error rate (i.e., increased chances of failing to demonstrate a treatment effect when one exists) for the important hypotheses of the study.

A. Controlling Study-wide Type I Error Rate

The Type I error rate for the entire study may be increased if inadequate adjustment is made for the many possible choices for adaptation and the many opportunities to demonstrate nominally statistically significant differences. At each stage of interim analysis and adaptation, there can be opportunities for early rejection of some of the several null hypotheses being tested, the

possibility of increasing sample sizes, or the selection of the final hypothesis from among several initial hypothesis options. These many choices based on unblinded analyses represent multiplicity that may inflate the Type I error rate that needs to be controlled in A&WC studies.

Avoiding problems with study interpretation and controlling the Type I error rate for all involved hypotheses is best accomplished by prospectively specifying and including in the SAP all possible adaptations that may be considered during the course of the trial. Determining the appropriate statistical correction by taking into account the relative amount of data available at the time of the interim analysis, as well as correlation of the multiple endpoints, is challenging and should be addressed at the protocol design stage. Under some limited circumstances, adaptations not envisioned at the time of protocol design may be feasible, but ensuring control of the Type I error rate remains critical. The flexibility to apply such late changes should be reserved for situations where the change is limited in scope and is particularly important, and should not to be proposed repeatedly during a study.

Statistical bias can be introduced into adaptive design studies that make modifications based on interim analyses of a biomarker or an intermediate clinical endpoint thought to be related to the study final endpoint, even though the final study analysis uses a clinical efficacy endpoint. This is because of the correlation between the biomarker and final study endpoint. This potential source of bias should be considered and addressed when the protocol is designed, including appropriate control of the Type I error rate.

One type of adaptation based on an unblinded interim analysis of treatment effects is an increase in the study sample size to maintain study power when the observed effect size is smaller than that initially planned in the protocol. When a statistical bias in the estimate of treatment effect exists, an increase in the sample size does not eliminate the bias. Instead, if flaws in the design (or conduct) of a study introduce a small bias, the increase in sample size can result in the bias increasing the Type I error rate more than would occur without the sample size increase. Thus, the impact of small biases can be magnified when sample size increases are enabled.

B. Statistical Bias in Estimates of Treatment Effect Associated with Study Design Adaptations

Estimates of the treatment effect are used to make decisions at each stage of an adaptive design study. Because these estimates can be based on a relatively small amount of data, they can be very variable or unstable. The effect estimates for the selected adaptations have the potential to overstate the true effect size because the adaptive choice is usually selected based on the largest of the observed interim treatment effects among the design choice options, which can reflect an unusual distribution of patient observations (often called *random highs* in group sequential designs). This could also lead to selecting a wrong adaptation choice and thus miss detecting a true treatment effect (i.e., lead to a Type II error).

In an adaptive design study, the overall treatment effect is obtained by combining in some manner the treatment effect observed in each stage, and this overall effect estimate should be used for hypothesis testing. How the combining of each stage's data is accomplished can affect the validity of the overall treatment-effect estimate. Of particular concern are situations in which

the estimates of the treatment effect obtained before and after the design modification differ substantially. Inconsistent treatment effect estimates among the stages of the study can make the overall treatment effect estimate difficult to interpret. The estimate of treatment effect(s) for an adaptive design A&WC study should be critically assessed at the completion of the study.

C. Potential for Increased Type II Error Rate

Adaptive design trials should be planned not only to control the Type I error rate for all involved hypotheses, but also to avoid increasing the chance of failing to demonstrate a treatment effect when one exists (the Type II error rate). Type II errors may occur because of suboptimal adaptive selection of design modifications or because of insufficient power to detect a real treatment effect on an endpoint. In general, one of the postulated benefits of adaptive designs is the potential to improve the power of the study to demonstrate a treatment effect through sample size increases or other design modifications. Adaptive design methods, however, also have the potential to inflate the Type II error rate for one or more hypotheses. An example of this is a study that begins with multiple doses (or populations or other study features) and that early in the study is adaptively modified to eliminate all but one or two doses to be continued to the study's end. This study risks failing to demonstrate treatment effects by making erroneous choices based on interim results that are very variable because of the limited amount of early study data. If this risk is not considered by study planners, an apparently efficient adaptive design study can mislead the drug development program and result in program failure, when it might have succeeded had there been better adaptation choices made. Another example is stopping for futility reasons where a liberal futility stopping criterion may substantially increase the Type II error rate.

D. Role of Clinical Trial Simulation in Adaptive Design Planning and Evaluation

Many of the less well-understood and complex adaptive designs involve several adaptation decision points and many potential adaptations. For study designs that have multiple factors to be simultaneously considered in the adaptive process, it is difficult to assess design performance characteristics and guide sample size planning or optimal design choices because these characteristics might depend upon the adaptations that actually occur. In these cases, trial simulations performed before conducting the study can help evaluate the multiple-trial design options and the clinical scenarios that might occur when the study is actually conducted, and can be an important planning tool in assessing the statistical properties of a trial design and the inferential statistics used in the data analysis. Section IX provides guidance for the format and content for reporting of clinical trial simulation studies to be included in the adaptive design protocol and the SAP.

In general, clinical trial simulations rely on a statistical model of recognized important design features and other factors, including the posited rate of occurrence of clinical events or endpoint distribution, the variability of these factors among patient subsets, postulated relationships between outcomes and prognostic factors, correlation among endpoints, the time course of endpoint occurrence or disease progression, and the postulated patient withdrawal or dropout patterns, among others. More complex disease models or drug models might attempt to account for changing doses, changing exposure duration, or variability in bioavailability. The multiple

ways to adapt and the multiple ways to declare a study as positive can be simulated as part of study planning.

Some modeling and simulation strategies lend themselves to a Bayesian approach that might be useful. The Bayesian framework provides a way to posit models (i.e., *priors*) for the study design and the adaptive choices as they might probabilistically occur, and may aid in evaluating the impact of different assumed distributions for the parameters of the model and modeled sources of uncertainty. The Bayesian approach can be a useful planning tool at the study design stage to accommodate a range of plausible scenarios. Using Bayesian predictive probability, which depends upon probabilities of outcomes conditional on what has been observed up to an interim point in the adaptive study, may aid in deciding which adaptation should be selected, while the study design is still able to maintain statistical control of the Type I error rate in the frequentist design.

Trial simulations can also be helpful in comparing the performance characteristics among several competing designs under different scenarios (e.g., assumptions about drug effect such as the shape and location of the dose-response relationship, the magnitude of the response, differing responses in subgroups, the distribution of the subgroups in the enrolled population, the clinical course of the comparison group (usually the placebo group), and study dropout rate and pattern). The simulations will allow between-design comparisons of the probability of success of the trial for the objective (e.g., to lead to correct dose selection, to identify a response above a specific threshold, to identify the correct subgroup), and comparisons of the potential size of bias in the treatment-effect estimates. For drug development programs where there is little prior experience with the product, drug class, patient population, or other critical characteristics, clinical trial simulations can be performed with a range of potential values for relevant parameters encompassing the uncertainty in current knowledge.

In general, every adaptation may create a new hypothesis whose Type I error rate should be controlled. There have been suggestions that because of the complexity resulting from multiple adaptations and the difficulty in forming an analytical evaluation, modeling and simulation provide a solution for demonstrating control of the Type I error rate for these multiple hypotheses. Using simulations to demonstrate control of the Type I error rate, however, is controversial and not fully understood.

E. Role of the Prospective Statistical Analysis Plan in Adaptive Design Studies

The importance of prospective specification of study design and analysis is well recognized for conventional study designs, but it is of even greater importance for many of the types of adaptive designs discussed in sections V and VI, particularly where unblinded interim analyses are planned. As a general practice, it is best that adaptive design studies have a SAP that is developed by the time the protocol is finalized. The SAP should specify all the changes prospectively planned and included in the protocol, describe the statistical methods to implement the adaptations, describe how the analysis of the data from each adaptive stage will be incorporated into the overall study results, and include the justification for the method of control of the Type I error rate and the approach to appropriately estimating treatment effects. The SAP for an adaptive trial is likely to be more detailed and complex than for a non-adaptive trial.

Any design or analysis modification proposed after any unblinded interim analysis raises a concern that access to the unblinded data used in the adaptations may have influenced the decision to implement the specific change selected and thereby raises questions about the study integrity. Therefore, such modifications are generally discouraged. Nonetheless, circumstances can occur that call for the SAP to be updated or for some other flexibility for an unanticipated adaptation. The later in the study these changes or updates are made, the more a concern will arise about the revision's impact. Generally, the justifiable reasons to do so are related to failure of the data to satisfy the statistical assumptions regarding the data (e.g., distribution, proportionality, fit of data to a model).

In general, it is best that any SAP updates occur before any unblinded analyses are performed, and that there is unequivocal assurance that the blinding of the personnel determining the modification has not been compromised. A blinded steering committee can make such protocol and SAP changes, as suggested in the ICH E9 guidance and in the DMC guidance, but adaptive designs open the possibility of unintended sharing of unblinded data after the first interim analysis. Any design or analysis modifications made after an unblinded analysis, especially late in the study, may be problematic and should be accompanied by a clear, detailed description of the data firewall between the personnel with access to the unblinded analyses and those personnel making the SAP changes, along with documentation of adherence to these plans. Formal amendments to the protocol and SAP need to be made at the time of such changes (see 21 CFR 312.30).

VIII. SAFETY CONSIDERATIONS IN ADAPTIVE DESIGN TRIALS

A. Safety of Patients in Adaptive Design Dose Escalation Studies Early in Drug Development

Studies designed with sequential cohorts of subjects that plan to escalate the dose for each successive cohort are a common design in first-in-human and other early drug development safety studies, and are a form of adaptive design studies. A concern regarding subject safety may arise in some of these studies. In traditional dose escalation studies, results from each fixed-size cohort determine the dose for the subsequent cohort based on planned rules (e.g., escalate the dose, repeat the same dose, or repeat the adjacent lower dose). Such studies commonly start at a dose well below a dose with observed animal toxicity, and it is intended that each cohort provide reasonable confidence regarding the safety of a dose level before the study proceeds to the next higher dose level. A common occurrence is that the lowest dose (or several of the lowest doses) have little to no effect and are not studied further in drug development. This traditional design is intended to provide safety for subjects in the study when the drug's safety profile is not known, but it is not intended to reach higher doses rapidly.

Some newer adaptive design algorithms permit a change in dose level after each patient is treated based on the accumulated responses of previously enrolled subjects. These algorithms lead to more dose-level changes, both increases and decreases of the dose, as the algorithm selects an exposure for each subject to the dose that will contribute the greatest amount of information towards the ultimate conclusion. By permitting escalation after each individual subject if that

subject did not have an unacceptable adverse response, it is possible to reach the middle or higher end of the dose-response curve with fewer subjects at each of the prior levels. This design emphasizes completing the study more rapidly than the traditional sequential fixed-size cohort design. Where there is little to no prior safety experience with a drug (or related drugs) and the known or hypothetical adverse effects can be serious, however, an adaptive study aggressively designed for most rapidly reaching a decision on the highest tolerable dose might be inappropriate. Study designs that call for a specified minimum number of subjects at a dose level prior to escalation, or designs that allow for smaller cohorts when physiologic activity markers do not show a response, can be appropriate in some circumstances.

Sponsors should explore the features of different study designs with regard to the balance of efficiency (study size) and subject safety. Study simulations with multiple combinations of escalation criteria, dose-step size, and hypothetical-assumed relationships of exposure to severity and frequency of adverse events may be useful in evaluating different designs. These simulations can assist in assessing the risks and selecting a design that offers improved efficiency without increasing risk excessively. Depending on the rapidity of dose escalation in the design, it may be important to submit these simulations and analyses to FDA when the selected design is submitted.

B. Earlier Design and Conduct of Adequate and Well-Controlled Studies with Major Expansion in the Number of Treatment-Exposed Subjects

In drug development programs, the safety-related data of each completed study are commonly examined before finalizing the design and starting the subsequent study. This opportunity is often not available in conventional development programs when the A&WC studies are initiated with little delay between them, so one study is not completed before the next is initiated. Development programs using adaptive design methods are sometimes intended to condense the development program into fewer fully independent studies, with more rapid advancement from small early studies into the large A&WC studies. This approach may lead to having only a limited amount of safety data available at the time that a large adaptive study is being planned that will entail a great increase in the number of patients exposed to the drug. This circumstance is in contrast to a typical non-adaptive development program where a large A&WC study would be preceded by shorter, moderate sized exploratory studies and the safety data analyzed and considered to inform design of the larger study.

There are advantages to the usual sequential approach that should be considered in selecting a study design. If there is a significant adverse effect that is inadequately understood or unrecognized because of the limited safety data of the very early studies, evaluating the data from a moderate-sized study might indicate that effect and lead to design changes to the large A&WC study to improve safety for patients within the A&WC study. Although it is important to monitor for serious adverse effects in any large clinical study, the adaptive design study that is initiated when there is only limited prior patient safety experience has greater uncertainty regarding the potential drug-associated risks, and thus patient safety protection may call for more frequent and/or extensive patient assessment for safety parameters during the study (or at least the earlier portion of it). Increasing the safety data monitoring may not fully resolve this concern, and it may be important to take other steps, such as enrolling limited numbers of

patients until sufficient safety data are accumulated and examined to support expansion of the study to larger numbers of patients being enrolled more rapidly.

Another safety-related concern relates to the adequacy of the safety database attained in the overall development program. A safety concern that becomes recognized in the data of a moderate-sized study can lead to planning for better evaluation in the A&WC study designed subsequently. The more comprehensive evaluation thus obtained may be necessary to ensure an adequate safety assessment for regulatory review. An adaptive design development program that eliminates the independent mid-sized study and initiates the large adaptive A&WC study before recognizing the safety issue will not have included such additional safety assessments. It may then be necessary to carry out further safety studies, leading in the end to a less efficient drug development program rather than the more efficient program that was sought.

IX. CONTENT OF AN ADAPTIVE DESIGN PROTOCOL

A. A&WC Adaptive Design Studies

Although FDA's ICH E3 guidance on *Structure and Content of Clinical Study Reports* (ICH E3 guidance)[7] describes the documentation that should be included in the protocol of an A&WC study, the added complexities introduced by adaptive design methods usually call for more detailed documentation, especially for the less-familiar adaptive design methods where significant design modifications are planned based on unblinded interim analyses.

When FDA is asked to evaluate an adaptive design study (see also section X), the process is more challenging because of the complex decision criteria and processes inherent in some of these designs. The protocol and supporting documentation should contain all the information critical to allow a thorough FDA evaluation of the planned study. This documentation should include the rationale for the design, justification of design features, evaluation of the performance characteristics of the selected design (particularly less well-understood features), and plans to assure study integrity when unblinded analyses are involved. Documentation of the rules of operation of the DMC (or other involved groups) should usually be more extensive than for conventional studies, and should include a description of the responsibilities of each entity involved in the process.

B. Adequate Documentation in a Protocol for an Adaptive Design Study

FDA review of a complex adaptive design protocol cannot be carried out without an adequately detailed protocol, SAP, and supportive information. Protocols for adaptive design studies intended to be A&WC should include a detailed description of all of the important design and decision features of the proposed trial, such as the study's planned endpoints, design, criteria for success, hypotheses to be tested, conduct procedures, data management and quality control. The SAP is an important part of that documentation because it states in detail the prospective hypotheses and statistical methods of analysis. The documentation for an adaptive design A&WC study should include the following:

[7] Available on the Internet at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.

- A summary of the relevant information about the drug product, including what is known at the present stage of development about the drug from other studies, and why an adaptive study design, in contrast to a non-adaptive design, has been chosen in this situation. The role of the chosen adaptive study design in the overall development strategy should also be discussed.

- A complete description of all the objectives and design features of the adaptive design, including each of the possible adaptations envisioned, the assumptions made in the study design with regard to these adaptations, the statistical analytical approaches to be used and/or evaluated, the clinical outcomes and quantitative decision models for assessing the outcomes, the relevant calculations that describe treatment effects, and the quantitative justifications for the conclusions reached in planning the trial.

- A summary of each adaptation and its impact upon critical statistical issues such as hypotheses tested, Type I errors, power for each of the hypotheses, parameter estimates and confidence intervals, sample size. In general, the study design should be planned in a frequentist framework to control the overall study Type I error rate. A Bayesian framework that incorporates uncertainty into planning parameters in a quantitative manner (i.e., prior distributions on parameters) can also be useful for planning purposes to evaluate model assumptions and decision criteria. If models are used to characterize the event rates, disease progression, multiplicity of outcomes, or patient withdrawal rates, these models should be summarized clearly to allow evaluation of their underlying assumptions. Summary tables and figures should be included that incorporate all the important quantitative characteristics and metrics that inform about the adaptive design.

- Computer simulations intended to characterize and quantify the level of statistical uncertainty in each adaptation and its impact on the Type I error, study power (conditional, unconditional) or bias (in hypothesis testing and estimates of the treatment effect). The simulations should consider the impact of changes in a single design feature (e.g., the number of dose groups to be dropped), as well as the combination of all proposed adaptive features.

  The computer programs used in the simulations should be included in the documentation, as should graphical flowcharts depicting the different adaptive pathways that might occur, the probabilities of their occurrence, and the various choices for combining information from the choices. For example, the following quantitative models can be used to reflect various study features considered in evaluating the stages of an adaptive design and the impact of combining information from each of the stages:

  – Models for study endpoints or outcomes.

  – Models for the withdrawal or dropout of subjects (e.g., for lack of compliance, toxicity, or lack of benefit).

  – Models of the procedure for selecting among multiple study endpoints (e.g., selection of the types of events included in a composite endpoint).

For each design evaluated with simulations, the documentation should clearly describe the following:

- A listing of all branching options possible at each stage of adaptation along with the chances of selection of each option.
- Various design features and assumptions.
  - Event rate background
  - Entrance criteria and event rate association with such criteria
  - Subgroup differences or heterogeneity in response
- Procedure for combining data on treatment effects from different stages of the study, including any weightings.
- Statistical methods for estimation of treatment effects at each study stage, and at final study completion along with the statistical bias in the estimate.
- Statistical calculations of the Type I error properties of the design at each study stage and at final study completion, and the calculations of study power.

- Full detail of the analytic derivations, if appropriate. For some adaptations, statistical calculations of the Type I error and/or statistical bias in treatment-effect estimates can be performed analytically without using simulations. If the analytic approaches are based on published literature, the portions of the analytic approaches specifically relevant to the adaptive design employed should be provided in detail.

- The composition, written charter, and operating procedures for the personnel assigned responsibility for carrying out the interim analyses, adaptation selection, and any other forms of study monitoring. This information should include all the written agreements that the sponsor has in place and written assurances from the involved parties for the protection of information that should not be shared outside of the limited team with access to the unblinded data. A description of whether a sponsor-involved statistician will perform the unblinded analysis and/or whether sponsor-involved personnel (e.g., sponsor employees or contract research organization (CRO) staff) will make recommendations for the adaptation should be included. A well-trusted firewall established for trial conduct beyond those established for conventional group sequential clinical trials can help provide assurance that statistical and operational biases have not been introduced.

X. INTERACTIONS WITH FDA WHEN PLANNING AND CONDUCTING AN ADAPTIVE DESIGN

The purpose and nature of the interactions between a study sponsor and FDA varies with the study's location (stage) within the drug development program. The increased complexity of some adaptive design studies and uncertainties regarding their performance characteristics may warrant earlier and more extensive interactions than usual. This section discusses general

principles on interactions between sponsors and FDA with regard to the use of adaptive designs in a development program.

A. Early and Middle Period of Drug Development

FDA's review of an exploratory study protocol is usually focused upon the safety of the study participants, and does not typically scrutinize the protocol as closely for design elements related to assessment of pharmacologic activity, efficacy, and strength of inferences. As resources allow, however, FDA might review exploratory protocols to consider the relevance of the information being gathered to guide the design of later studies (e.g., do the doses being examined seem reasonable for early efficacy evaluations; are the endpoints or biomarkers being examined reasonable for the stage of drug development).

Review comments from the FDA on the adaptive design features in exploratory protocols will generally be less formal than for late stage drug development studies. Sponsors who have specific questions about the adaptive design elements in an exploratory study should seek FDA feedback either by identifying the specific issues, questions, and the requested feedback in the submission containing the protocol, or by requesting a meeting to discuss specific questions. Discussion of the plans for an adaptive design study can be the basis for requesting a Type C meeting. FDA's ability to address such requests for studies in early phases of drug development, however, may be limited and will depend on competing workload priorities and on the particulars of the drug and use under development. Innovative therapeutics for an area of unmet medical need are likely to garner more review attention than other products FDA believes do not fall into this category.

B. Late Stages of Drug Development

FDA has a more extensive role in assessing the design of studies that contribute to substantial evidence of effectiveness. FDA's review focus in later stages of drug development continues to include safety of study subjects, but also includes assuring that studies performed at this stage contain plans for assessment of safety and efficacy that will result in data of sufficient quality and quantity to inform a regulatory decision. Regulatory mechanisms for obtaining formal, substantive, feedback from FDA on design of the later stage trials and their place in the drug development program are well established (e.g., the End-of-Phase 2 (EOP2) meeting and Special Protocol Assessments (SPA)).

Depending on the preexisting breadth and depth of information regarding the drug, its specific use, and the nature of the adaptive features, an EOP2 meeting may be the appropriate place in development for initial discussion of an adaptive design A&WC study. However, if there is only limited knowledge of certain critical aspects of the drug's use before conducting the adaptive study, and the study is intended to obtain such knowledge using the study's adaptive features (particularly less well-understood methods), discussion with FDA earlier than usual is advisable (e.g., at a Type C or End-of-Phase 2A meeting). An early meeting for A&WC study protocols with complex adaptive features allows time to carefully consider the plan and to revise and reevaluate it as appropriate, without slowing the clinical development program. This early discussion should specifically address the adaptive methodology in general and the suitability of

the selected approach to achieve the study's goals. This early, focused adaptive design discussion may not eliminate the value of a subsequent EOP2 meeting.

FDA's review of proposed A&WC studies in a drug development program includes considering whether the totality of the existing information combined with the expected information from the proposed studies will likely be adequate to enable a review of a marketing application for approval. This analysis is often enhanced by an EOP2 meeting that includes assessing the adequacy of plans for evaluating the drug's dose-response, treatment-regimen selection, choice of patient population, and other important aspects of the therapy's use. It is important to recognize that use of less well-understood adaptive methods may limit FDA's ability to offer such an assessment. FDA may be unable to assess in advance whether the adaptively selected aspects of drug use (e.g., dose, regimen, population) will be sufficiently justified by the study results. As usual, FDA will review and comment to the extent possible on aspects of the drug's use that the sponsor considers well defined, as well as non-adaptive aspects of the study.

As previously discussed, FDA will generally not be involved in examining the interim data used for the adaptive decision-making and will not provide comments on the adaptive decisions while the study is ongoing. FDA's review and acceptance at the protocol design stage of the methodology for the adaptation process does not imply its advance concurrence that the adaptively selected choices will be the optimal choices. For example, if for feasibility of design, the adaptive selection of dose is based on one aspect of a drug's effect, but the optimal choice depends on the interplay between two aspects of drug effect, the data resulting from the study will be evaluated to judge whether adequate dose selection has been made.

C. Special Protocol Assessments

Special protocol assessments (SPA) entail timelines (45-day responses) and commitments that may not be best suited for adaptive design studies. The full review and assessment of a study using less well-understood adaptive design methods can be complex, will involve a multidisciplinary evaluation team, and might involve extended discussions among individuals within different FDA offices before reaching a conclusion. If there has been little or no prior discussion between FDA and the study sponsor regarding the proposed study and its adaptive design features, other information requests following initial FDA evaluation are likely and full completion of study assessment within the SPA 45-day time frame is unlikely. Sponsors are therefore encouraged to have thorough discussions with FDA (as noted in section X.B above) regarding the study design and the study's place within the development program before considering submitting an SPA request.

Even when adequate advance discussion has occurred, the nature of a full protocol assessment of an adaptive design study may not be the same as for an SPA request for a conventional study, as one or more critical final decisions regarding study design are made after the study has started. FDA cannot realistically commit to accepting aspects of study design yet to be determined. Thus, although an adaptive design SPA request that had been preceded by adequate advance discussion, enabling a complete protocol review, the FDA response may have certain limitations that an SPA regarding a non-adaptive study would not require.

XI. DOCUMENTATION AND PRACTICES TO PROTECT STUDY BLINDING AND INFORMATION SHARING FOR ADAPTIVE DESIGNS

Protecting study blinding is important in all clinical trials, but in the case of an adaptive design study, where the design is modified after examination of unblinded interim data, protecting study blinding is particularly important to avoid the introduction of bias in the study conduct and to maintain confidence in the validity of the study's result.

In addition to the full documentation required for a study protocol (21 CFR 312.23(a)), there should be comprehensive and prospective, written standard operating procedures (SOPs) that define who will implement the interim analysis and adaptation plan, and all monitoring and related procedures for accomplishing the implementation, providing for the strict control of access to unblinded data (see the DMC guidance). SOPs for an adaptive design study with an unblinded interim analysis are likely to be more complex than SOPs for non-adaptive studies to ensure that there is no possibility of bias introduction. This written documentation should include (1) identification of the personnel who will perform the interim analyses, and who will have access to the interim results, (2) how that access will be controlled and verified, and how the interim analyses will be performed, including how any potential irregularities in the data (e.g., withdrawals, missing values) will be managed, and (3) how adaptation decisions will be made. Other issues that should be addressed in these SOPs are (1) whether there are any foreseeable impediments to complying with the SOPs, (2) how compliance with the SOPs will be documented and monitored, and (3) what information, under what circumstances, is permitted to be passed from the DMC to the sponsor or investigators. It is likely that the measures defined by the SOPs will be related to the type of adaptation and the potential for impairing study integrity.

In general, a person or group that is independent of the personnel involved with conducting or potentially modifying (e.g., a steering committee) the study should be used for the review of an interim analysis of unblinded data and adaptive decision-making. This process should be based on the study management structure set in place by the study sponsor, steering committee, or other group responsible for the study, and in accordance with the well-specified adaptation plans. This role could be assigned to an independent DMC when a DMC is established for other study monitoring purposes. DMCs typically will be provided certain kinds of information, of which some might be unblinded analyses, and procedures are usually in place to ensure that this information does not become available outside of the committee. Alternatively, a DMC might be delegated only the more standard roles (e.g., ongoing assessment of critical safety information) and a separate adaptation committee used to examine the interim analysis and make adaptation recommendations. In either case, the specific duties and procedures of the committees should be fully and prospectively documented.

The planned operating procedures should call for written minutes of all committee meetings that describe what was reviewed, discussed, and decided. Sponsors should plan for procedures to maintain these records in a secure manner with restricted access to enable post-study review of adherence to the prospective process. For the same purpose, the actual interim analysis results and a *snapshot* of the databases used for that interim analysis and adaptation decision should also be retained in a secure manner.

In recent years there has been greatly increased use of contract research organizations (CRO) for many tasks previously performed by direct employees of the study sponsor. In particular, this has included assigning to CROs the task of performing the interim analysis and making study decisions based on the interim results. Many CROs do not have long histories of carrying out these responsibilities. Study sponsors should have assurance that the personnel performing these roles have appropriate expertise, and that there are clear and adequate written SOPs to ensure compliance with the precautions needed to maintain study integrity. The CRO should be able to maintain confidentiality of the information examined in the interim analysis and it should establish that it has the ability to do so. A failure either to make the appropriate decisions as directed in the prospective SAP or to maintain confidentiality of the interim results might have an adverse impact on the interpretation of the study results. The processes established, as well as how they were performed, should be well documented in the final study report. The ability for FDA to verify compliance, potentially by on-site auditing, may be critical.

XII. EVALUATING AND REPORTING A COMPLETED STUDY

Sponsors often seek to communicate to FDA the results of a completed adaptive design study before undertaking a subsequent study within an investigational new drug application (IND). Marketing applications should always include study reports for completed studies. To allow FDA to thoroughly review the results of adaptive design studies, complete and detailed documentation should be supplied in addition to the detailed information for the prospective FDA review of the protocol. All prospective plans and planning support information, detailed description of the study conduct in all aspects, and comprehensive analysis of results should be included in a marketing application submission. More limited information (e.g., reports without the database copies, less-detailed information on other aspects) may be sufficient for study summaries provided to FDA during the course of development to support ongoing discussions within the IND.

In addition to the guidance provided by the ICH E3 guidance regarding the format and content of a clinical study report, there are some unique features to reporting the conduct and analysis of an adaptive design study to FDA. Information submitted regarding the prospective plans should be complete. This information should include the study protocol and study procedure documents, including DMC or other committee charters. The submission should also include the supportive information that was developed to assist the sponsor in the prospective planning and FDA in the prospective review of the study. This information can include the rationale for using an adaptive design, the role of the study within the overall drug development program, and the simulations and other statistical evaluations performed prospectively. Submissions should include copies of published articles critical to assessing the methodology.

Complete information describing the study conduct should include the following:

- information on compliance with the planned adaptive process and procedures for maintaining study integrity
- description of the processes and procedures actually carried out when there were any deviations from those planned,
- records of the deliberations and participants in the internal discussions by any committees (e.g., DMC meeting minutes, steering or executive committee meeting minutes) involved in the adaptive process,
- results of the interim analysis used for the adaptation decisions (including estimates of treatment effects, uncertainty of the estimates, and hypothesis tests at that time),
- assessment of adequacy of any firewalls established to limit dissemination of information

Sponsors should consider using a diagrammatic display of the course of the study, illustrating the adaptive plan and the actual decisions made at each juncture. A copy of the study databases that were used for the interim analyses and adaptation decisions should be maintained in a data-locked manner and also submitted. If there were multiple stages for adaptation with multiple interim analyses, each stage should be fully represented in the report, both as cumulative information and as information acquired during each stage separately. It is important to include all of this information for FDA evaluation of the study conduct, analysis, and interpretation.

The analysis of study final results should be complete and should adhere to the prospective analytic plan. Any deviations from the prospective plan should be detailed and discussed, and the sponsor should assess any potential bias in the results these deviations might have introduced. It may be important to include relevant exploratory analyses of the study data in this assessment.

Exploration of the study data should include examining the consistency of treatment effects and other relevant results between study stages. Statistical tests for differences in treatment-effect estimates between stages of the trial will generally have poor statistical power and are not by themselves a sufficient approach to this issue. Comparability between patients recruited before and after the adaptation can be examined, for instance, by baseline characteristics as well as clinical outcome. If these evaluations suggest a potential shift in population, outcome, or other parameters, more detailed evaluation will be warranted.

GENERAL REFERENCES

Bauer P, Köhne K (1994). Evaluations of experiments with adaptive interim analyses. *Biometrics* 50, 1029-1041.

Bauer P, Römel J (1995). An adaptive method for establishing a dose response relationship. *Statistics in Medicine* 14, 1595-1607.

Bauer P, Kieser M (1999). Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine* 18, 1833-1848.

Bauer P, Brannath W, Posch M (2001). Flexible two stage designs: an overview. *Methods of Information in Medicine* 40, 117-121.

Bauer P, Koenig, F (2006). The reassessment of trial perspectives from interim data—a critical view. *Statistics in Medicine* 25, 23-36.

Bauer P, Koenig F, Brannath W, Posch M (2009). Selection and bias – two hostile brothers. *Statistics in Medicine* (under review).

Berry DA, Eick SG (1995). Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine* 14, 231-246.

Biometrical Journal 48(4) (August 2006). Special issue: adaptive designs in clinical trials.

Bornkamp B, Bretz F, Dmitrienko A, Enas G, Gaydos B, Hsu CH, Konig F, Krams M, Liu Q, Neuenschwander B, Parke T, Pinheiro J, Roy A, Sax R, Shen F (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics* 17, 965-995.

Brannath W, Posch M, Bauer P (2002). Recursive combination tests. *Journal of the American Statistical Association* 97, 236-244.

Brannath W, Bauer P, Maurer W, Posch M (2003). Sequential tests for noninferiority and superiority. *Biometrics* 59**,** 106-114.

Brannath W, Konig F, Bauer P (2006). Estimation in flexible two stage designs. *Statistics in Medicine* 25, 3366-3381.

Brannath W, Mehta CR, Posch M (2008). Exact confidence bounds following adaptive group sequential tests. *Biometrics* 65, 539-546.

Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine* 28, 1445-1463.

Bretz F, Schmidli H, Konig F, Racine A, Maurer W (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: general concepts. *Biometrical Journal* 48(4), 623-634.

Bretz F, Koenig F, Brannath W, Glimm E, Posch M (2009). Tutorial in biostatistics: adaptive designs for confirmatory clinical trials. *Statistics in Medicine* 28, 1181-1217.

Bowden J, Glimm E (2008). Unbiased estimation of selected treatment means in two-stage trials. *Biometrical Journal* 50, 515-527.

Burman C, Sonesson C (2006). Are flexible designs sound? (with discussion). *Biometrics* 62, 664-683.

Chen YHJ, DeMets DL, Lan KKG (2004). Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine* 23, 1023-1038.

Cheng Y, Berry DA (2007). Optimal adaptive randomized designs for clinical trials. *Biometrika* 94, 673-689.

Coburger S, Wassmer G (2001). Conditional point estimate in adaptive group sequential test designs. *Biometrical Journal* 7, 821-833.

Coburger S, Wassmer G (2003). Sample size reassessment in adaptive clinical trials using a bias corrected estimate. *Biometrical Journal* 7, 812-825.

Committee for Medicinal Products for Human Use (CHMP) (2007). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. Available at http://www.emea.europa.eu/pdfs/human/ewp/245902enadopted.pdf.

Cui L, Hung HMJ, Wang SJ (1999). Modification of sample size in group sequential clinical trials. *Biometrics* 55, 321-324.

Dahiya RC (1974). Estimation of the mean of the selected population. *Journal of the American Statistical Association* 69, 226-230.

Denne JS (2001). Sample size recalculation using conditional power. *Statistics in Medicine* 2, 2645-2660.

Fisher LD (1998). Self-designing clinical trials. *Statistics in Medicine* 17, 1551-1562.

Freidlin B, Simon R (2005). Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clinical Cancer Research* 11, 7872-7878.

Friede T, Kieser M (2001). A comparison of methods for adaptive sample size adjustment. *Statistics in Medicine* 20, 3861-3874.

Gould AL, Shih WJ (1992). Sample size reestimation without unblinding for normally distributed outcomes with unknown variance. *Communications in Statistics (A)—Theory and Methods* 21(10), 2833-2853.

Hommel G (2001). Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal* 43(5), 581-589.

Hommel G, Lindig V, Faldum A (2005). Two-stage adaptive designs with correlated test statistics. *Journal of Biopharmaceutical Statistics* 15, 613-623.

Hung HMJ, Wang SJ, O'Neill R (2006). Methodological issues with adaptation of clinical trial design. *Pharmaceutical Statistics* 5(2), 99-107.

Hung HMJ, O'Neill R, Wang SJ, Lawrence J (2006). A regulatory view on adaptive/flexible clinical trial design (with rejoinder). *Biometrical Journal* 48, 565-573, 613-615.

Inoue LY, Thall PF, Berry DA (2002). Seamlessly expanding a randomized phase II trial to phase III. *Biometrics* 58, 823-831.

Jennison C, Turnbull BW (2000). Group sequential methods with applications to clinical trials. Chapman & Hall/CRC, Boca, Raton, FL.

Jennison C, Turnbull BW (2003). Mid-course sample size modification in clinical trials. *Statistics in Medicine* 22, 971-993.

Jennison C, Turnbull BW (2005). Meta-analysis and adaptive group sequential designs in the clinical development process. *Journal of Biopharmaceutical Statistics* 15, 537-558.

Jennison C, Turnbull BW (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: opportunities and limitations. *Biometrical Journal* 48, 650-655.

Journal of Biopharmaceutical Statistics 15(4) (2005). Special issue: adaptive design in clinical research.

Kelly PJ, Stallard N, Todd S (2005). An adaptive group sequential design for phase II/III clinical trials that select a single treatment from several treatments. *Journal of Biopharmaceutical Statistics* 15, 461-658.

Kieser M, Bauer P, Lemacher W (1999). Inference on multiple endpoints in clinical trials with adaptive interim analyses. *Biometrical Journal* 41, 261-177.

Kieser M, Schneider B, Friede T (2002). A bootstrap procedure for adaptive selection of the test statistic in flexible two-stage designs. *Biometrical Journal* 44, 641-652.

Kimani PK, Stallard N, Hutton JL (2009). Dose selection in seamless phase II/III clinical trials based on efficacy and safety. *Statistics in Medicine* 28, 917-936.

Lawrence J (2002). Design of clinical trials using an adaptive test statistic. *Journal of Biopharmaceutical Statistics* 12, 193-205.

Lawrence J, Hung HMJ (2003). Estimation and confidence intervals after adjusting the maximum information. *Biometrical Journal* 45, 143-152.

Lehmacher W, Wassmer G (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* 55, 1286-1290.

Li G, Shih WJ, Xie T, Lu J (2002). A sample size adjustment procedure for clinical trials based on conditional power. *Biostatistics* 3, 277-287.

Liu GF, Zhu GR, Cui L (2008). Evaluating the adaptive performance of flexible sample size designs with treatment difference in an interval. *Statistics in Medicine* 27, 584-596.

Liu Q, Chi G (2001). On sample size and inference for two-stage adaptive designs. *Biometrics* 57, 172-177.

Liu Q, Proschan MA, Pledger G (2002). A unified theory of two-stage adaptive designs. *Journal of the American Statistical Association* 97, 1034-1041.

Mehta C, Gao P, Bhatt DL, Harrington RA, Skerjanec S, Ware JH (2009). Optimizing trial design: sequential, adaptive, and enrichment strategies. *Circulation* 119, 597-605.

Miller F, Guilbaud O, Dette H (2007). Optimal designs for estimating the interesting part of a dose-effect curve. *Journal of Biopharmaceutical Statistics* 17, 1097-1115.

Müller H, Schäfer H (2001). Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* 57, 886-891.

Pharmaceutical Statistics 5(2) (April/June 2006). Special issue on adaptive design.

Posch M, Bauer P (1999). Adaptive two stage designs and the conditional error function. *Biometrical Journal* 41, 689-696.

Posch M, Bauer P (2000). Interim analysis and sample size reassessment. *Biometrics* 56, 1170-1176.

Posch M, Koenig F, Branson M, Brannath W, Dunger-Baldauf C, Bauer P (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine* 24, 3697-3714.

Proschan MA, Hunsberger SA (1995). Designed extension of studies based on conditional power. *Biometrics* 51, 1315-1324.

Schmidli H, Bretz F, Racine-Poon A, Maurer W (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: applications and practical considerations. *Biometrical Journal* 48, 635-643.

Schmidli H, Bretz F., Racine-Poon A (2007). Bayesian predictive power for interim adaptation in seamless phase II/III trials where the endpoint is survival up to some specified timepoint. *Statistics in Medicine* 26, 4925-4938.

Shen Y, Fisher LD (1999). Statistical inference for self-designing clinical trials with a one-sided hypothesis. *Biometrics* 55, 190-197.

Stallard N, Todd S (2005). Point estimates and confidence regions for sequential trials involving selection. *Journal of Statistical Planning and Inference* 135, 402-419.

Tsiatis AA, Mehta C (2003). On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika* 90, 367-378.

Tsong Y, Hung HMJ, Wang SJ, Cui L, Nuri WA (1997). Dropping a treatment arm in clinical trial with multiple arms. Proceedings of the American Statistical Association, Biopharmaceutical Section [CD-ROM]. American Statistical Association, Alexandria, VA.

Wassmer G (2000). Basic concepts of group sequential and adaptive group sequential procedures. *Statistical Papers* 41, 253-279.

Wang L, Cui L (2007). Seamless phase II/III combination study through response adaptive randomization. *Journal of Biopharmaceutical Statistics* 17, 1177-1187.

Wang SJ, Hung HMJ, Tsong Y, Cui L (2001). Group sequential test strategies for superiority and non-inferiority hypotheses in active controlled clinical trials. *Statistics in Medicine* 20, 1903-1912.

Wang SJ, Hung HMJ, O'Neill R (2004). Uncertainty in planning phase III trials based on phase II data: sample size. Proceedings of the American Statistical Association, Biopharmaceutical Section [CD-ROM]. American Statistical Association, Alexandria, VA,

Wang SJ, Hung HMJ (2005). Adaptive covariate adjustment in trial design. *Journal of Biopharmaceutical Statistics* 15, 605-611.

Wang SJ, Hung HMJ (2005). Trials in trials: alpha allocation strategy and sub-trial planning. Proceedings of the American Statistical Association, Biopharmaceutical Section [CD-ROM]. American Statistical Association, Alexandria, VA.

Wang SJ, Hung HMJ, O'Neill R (2006). Adapting the sample size planning of a phase III trial based on phase II data. *Pharmaceutical Statistics* 5, 85-97.

Wang SJ, O'Neill RT, Hung HMJ (2007). Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharmaceutical Statistics* 6, 227-244.

Wang SJ, Hung HMJ, O'Neill R (2007). Stagewise planning for clinical trials from phase II to phase III. Proceedings of the American Statistical Association, Biopharmaceutical Section [CD-ROM]. American Statistical Association, Alexandria, VA**.**

Wang SJ (2008). Utility of adaptive strategy and adaptive design for biomarker-facilitated patient selection in pharmacogenomic or pharmacogenetic clinical development program. *Journal of Formosan Medical Association* 107(12S), 18-26.

Wang SJ, Hung HMJ, O'Neill RT (2009). Adaptive patient enrichment designs in therapeutic trials. *Biometrical Journal* 51(2), 358-374.

Wang SJ, Hung HMJ, O'Neill RT (2009). Impacts of type I error rate with inappropriate use of learn for confirm in adaptive designs. *Biometrical Journal* (under review).

Yao Q, Wei LJ (1996). Play the winner for phase II/III clinical trials. *Statistics in Medicine* 15, 2413-2423.