Therapeutic Innovation
& Regulatory Science
2016, Vol. 50(2) 195-203

DOI: 10.1177/2168479015604181
tirs.sagepub.com

CBER's Experience With Adaptive Design Clinical Trials

Min Lin, PhD[1], Shiowjen Lee, PhD[1], Boguang Zhen, PhD[1], John Scott, PhD[1], Amelia Horne, DrPH[1], Ghideon Solomon, PhD[1], and Estelle Russek-Cohen, PhD[1]

Abstract

There is considerable interest among pharmaceutical and other medical product developers in adaptive clinical trials, in which knowledge learned during the course of a trial affects ongoing conduct or analysis of the trial. When the FDA released a draft Guidance document on adaptive design clinical trials in early 2010, expectations were high that it would lead to an increase in regulatory submissions involving adaptive design features, particularly for confirmatory trials. A 6-year (2008-2013) retrospective survey was performed within the Center for Biologics Evaluation and Research (CBER) at the FDA to gather information regarding the submission and evaluation of adaptive design trial proposals. We present an up-to-date summary of adaptive design proposals seen in CBER and provide an overview of our experiences. We share our concerns regarding the statistical issues and operational challenges raised during the review process for adaptive design trials. We also provide general recommendations for developing proposals for such trials. Our motivation in writing this paper was to encourage the best study design proposals to be submitted to CBER. Sometimes these can be adaptive, and sometimes a simpler design is most efficient.

Keywords

adaptive design, FDA, clinical trial, survey, biologics

Introduction

Adaptive designs have been used in human clinical trials to advance medical research for many years. Adaptive designs allow researchers to modify the trial procedures and/or statistical methods of ongoing clinical trials using accumulated data at interim analyses. A potential beneficial feature of adaptive designs is that they may speed the development process of investigational products such as drugs, biological products, or medical devices. Therefore, adaptive designs have been attractive to industry researchers who are looking to reduce costs and streamline the product development process.[1] There are also potential ethical advantages of adaptive designs with respect to both efficacy and safety. Adaptive designs can decrease the number of study participants exposed to ineffective or unsafe products, and can lead to earlier availability of products with extraordinary efficacy.[1-4] Many adaptive design features have been proposed in the literature and/or applied to clinical trials,[2,3] including adaptive randomization, adaptive dose finding, and sample size re-estimation. However, for many adaptive designs, there are outstanding statistical and operational issues that can affect the interpretability and integrity of trials.

In early 2010, the United States Food and Drug Administration (FDA) circulated a draft Guidance document,[5] "Adaptive Design Clinical Trials for Drugs and Biologics," for public comments. In this draft Guidance, an *adaptive design clinical study* was 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." By this definition, all planned adaptations should be specified at the design stage. The draft Guidance further classified adaptive designs into two categories: *well-understood* designs and *less well-understood* designs. The terms "well-understood" and "less well-understood" referred to the regulatory review experiences at the FDA rather than the depth of theoretical exposition in the literature. Well-understood designs are those "well-established clinical study designs with planned modifications based on an interim study result analysis ... that either need no statistical correction ... or properly account for the analysis-related multiplicity of choices."[5] For example,

[1] Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

Submitted 27-May-2015; accepted 06-Aug-2015

Corresponding Author:
Min Lin, PhD, Center for Biologics Evaluation and Research, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA.
Email: min.lin@fda.hhs.gov

traditional group sequential designs fall into this category. Less well-understood designs, on the other hand, refer to "study designs with which there is relatively little regulatory experience and whose properties are not fully understood at this time."[5] Three major concerns associated with less well-understood adaptive designs were emphasized in the draft Guidance: (1) control of the study-wise type I error rate; (2) minimization of the impact of any adaptation-associated statistical or operational bias on the estimates of treatment effects; and (3) the interpretability of the results.[5] Generally well-understood adaptive designs are those that the FDA has had extensive experience with in terms of study conduct and statistical properties. While the draft Guidance encouraged the use of adaptive designs in early phase studies, its emphasis was on adequate and well-controlled studies (ie, confirmatory studies).

The Center for Biologics Evaluation and Research (CBER) at FDA regulates a diverse set of biological products and related devices, including preventive and therapeutic vaccines, products to treat or diagnose allergies, blood derivatives and components, gene and cell therapies, human tissue products, medical devices related to the processing of blood and tissues, diagnostic devices for screening blood donors, and xenotransplantation products. The adaptive designs we see in CBER vary by the type of product we evaluate.

In the present paper, we describe the experience with adaptive designs in CBER-regulated products. First we present the findings from a retrospective survey within CBER conducted by its Division of Biostatistics. Then we discuss specific adaptive design issues that have arisen in our regulatory work and present selected examples.

A Retrospective Survey Within CBER

Objectives

The survey to capture our review experience with adaptive designs was conducted over a 6-year period (2008-2013). The objectives of this survey were (1) to summarize the regulatory experience of adaptive designs reviewed at CBER during 2008-2013; (2) to understand the strengths and limitations of the adaptive design proposals in clinical product development, especially in late-phase studies in the context of CBER-regulated products; and (3) to better assist CBER sponsors in preparing clinical studies utilizing adaptive design approaches.

Materials and Methods

We considered Investigational New Drug (IND) and Investigational Device Exemption (IDE) applications for which a statistical evaluation was requested. All studies that elicited formal statistical review memos and involved adaptive design approaches—including more traditional and novel approaches—were identified. The submissions were from 3 CBER product offices: Office of Blood Research and Review (OBRR); Office of Cellular, Tissue, and Gene Therapies (OCTGT); and Office of Vaccine Research and Review (OVRR).

However, we excluded protocols for diagnostic and blood donor screening devices, since adaptive designs have not been seen as particularly advantageous in those areas. Many CBER-regulated diagnostic products are held to a very high performance standard with an objective performance criterion (eg, sensitivity and specificity must exceed some threshold). No alpha penalty is made when sponsors expand the size of a prospective study solely because there are not enough patients with the target condition of interest. In addition, for some products associated with blood and tissue typing, an error made by the diagnostic can result in a mismatch and possibly death or an acute reaction for the patient receiving a transfusion or transplantation. So the performance bar is set quite high and it is not clear that an adaptive design would be particularly helpful.

Phase I studies also were excluded in this survey, as they are not always captured with a formal statistical memo, and many are largely descriptive and do not use formal statistical inference procedures. Adaptive dose-finding studies are common in phase I, especially in oncology, but statistical review is generally requested only for those in which the dose escalation is unclear to the clinical review team. We also did not consider any designs proposed to the Center for Devices and Radiological Health (CDRH) or the Center for Drug Evaluation and Research (CDER), as our medical products are different and therefore our adaptive design experience may differ.

The data for this retrospective survey analysis were collected by searching CBER's Electronic Document Room (EDR) as well as by electronically distributing survey questionnaires to all statistical reviewers within the Division of Biostatistics at CBER. The survey contained open-ended questions regarding common characteristics of study designs, adaptation components, and review comments. All survey analyses were performed with SAS 9.3.1 (SAS Institute Inc., Cary, NC) using descriptive statistics.

We categorized adaptive designs based on their adaptation features. While there are many ways to classify adaptive design methods.[1,4,6] our classification for the purposes of this paper was driven by the frequency with which different features were encountered at CBER. An *adaptive dose finding*[7] design uses accumulating data to determine dose assignment for study participants, usually to establish a maximum tolerated dose (MTD). A *response-adaptive randomization design*,[2,8] which is sometimes referred to as a play-the-winner design, is a design that allows modification of randomization algorithms after the review of the response of previously assigned subjects. *Simon's two-stage designs*[9] are common in oncology and allow early termination of unpromising treatments in phase II efficacy evaluation trials. For a traditional *group sequential design*,[10-12] the study can be terminated prematurely due to futility and/or efficacy (ie, *stop for futility* or *stop for efficacy*) if convincing evidence is found during the planned interim analysis based on prespecified stopping boundaries. *Sample size*

re-estimation[2,13-18] (SSR) adaptation can provide investigators an opportunity to adjust the planned sample size while a trial is ongoing, using information from interim looks to increase power to detect a clinically meaningful treatment effect. An *adaptive seamless phase II/III design*[1,19] is a two-stage design consisting of a learning stage (ie, phase II part) and a confirmatory stage (ie, a phase III part) where data from both phases are combined for a final analysis; trials of this type are reviewed as confirmatory phase III trials. There are also *operationally seamless but inferentially distinct phase II/III studies*, in which two traditional trials (phase II and phase III) are conducted under a single study protocol but analyzed separately; in these trials, we expect the phase III to demonstrate type I error control, whereas type I error control for the phase II component may depend on the objective and context of the study. It should be noted that the terms "seamless" and "phase II/III" were not used in the FDA draft guidance as they have sometimes been adopted to describe various design features.[5]

Results

From 2008 to 2013, a total of 12,095 regulatory submissions were assigned to CBER statisticians for statistical review. Of the 12,095 submissions, 10,638 (88.0%) were IND or IDE applications; 2577 (24.2%) of these contained protocols and/or statistical analysis plans for confirmatory phase III or IV studies. For 1225 (11.5%) IND and IDE proposals, statistical comments were conveyed to the CBER review committees and/or the IND or IDE sponsor via formal statistical review memos. Of those 1225 applications, 140 (11.4%) submissions contained protocols for phase II, III, or IV studies with adaptive design features.

The characteristics of these 140 IND and IDE applications involving adaptive designs are summarized in Table 1. Forty-one (29.3%) were for blood products; 83 (59.3%) were for cellular, tissue, or gene therapies; and 16 (11.4%) were for vaccines or allergenic products. Fifty-three (37.9%) of the 140 submissions were early-phase studies (phase II), 83 (59.3%) were confirmatory (phase III) trials, and 4 (2.9%) were postmarketing (phase IV) studies. Most submissions were IND applications (130 of 140; 92.9%) and most were from industry sponsors (96 of 140; 68.6%). Since we reviewed only 4 phase IV studies with adaptive features and these are reviewed similarly to phase III studies, for subsequent tables we combined these two categories.

The number of adaptive design submissions varied considerably over the 6 years from 2008 through 2013 (Figure 1). The highest number of applications (40 of 140; 28.6%) was recorded in 2011 and the lowest number (10 of 140; 7.1%) was seen in 2008. For confirmatory phase III&IV trials, the number of adaptive design applications increased from 2008 to 2011 and decreased after 2011 (Figure 1). The Drug Information Association's (DIA's) Adaptive Design Scientific Working Group (ADSWG) published the results of a survey on the use of adaptive designs with 2008-2011 data from 18 participating

Table 1. Characteristics of adaptive design submissions by product offices.

	Product Office			
	OBRR (n = 41)	OCTGT (n = 83)	OVRR (n = 16)	Overall (N = 140)
Study phase				
II	10 (24.4%)	39 (47.0%)	4 (25.0%)	53 (37.9%)
III	30 (73.2%)	43 (51.8%)	10 (62.5%)	83 (59.3%)
IV	1 (2.4%)	1 (1.2%)	2 (12.5%)	4 (2.8%)
Submission type				
IND	39 (95.1%)	75 (90.4%)	16 (100%)	130 (92.9%)
IDE	2 (4.9%)	8(9.6%)	0 (0%)	10 (7.1%)
Sponsor type				
Academia	14 (34.2%)	27 (32.5%)	3 (18.7%)	44 (31.4%)
Industry	27 (65.8%)	56 (67.5%)	13 (81.3%)	96 (68.6%)

Abbreviations: IDE, investigational device exemption; IND, investigational new drug; OBRR, Office of Blood Research and Review; OCTGT, Office of Cellular, Tissue, and Gene Therapies; OVRR, Office of Vaccine Research and Review.

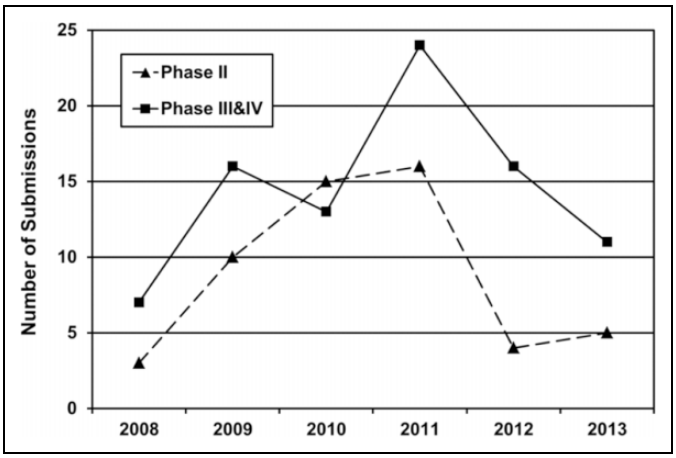

Figure 1. Adaptive design submissions by trial phases in years 2008-2013.

institutions.[20] Their results showed that industry and academia were more willing to propose adaptive designs, including designs classified as less well understood by the FDA draft Guidance, compared to the earlier results of a previous ADSWG survey conducted in 2008.[21] In contrast, our survey showed that, despite the release of the FDA adaptive design draft Guidance,[5] there was no clear trend of increase in regard to the total number of regulatory IND and IDE submissions using adaptive design approaches in the 6-year period. This could be because the term "less well understood" has been interpreted by some to mean not acceptable to FDA when that was not the intent.

Figure 2 displays the frequency of different types of adaptive design proposals seen at CBER by study phase, and Table 2 shows trial characteristics by study phase. Most of the 53 phase II trials (34 of 53; 64.1%) were open-label studies,

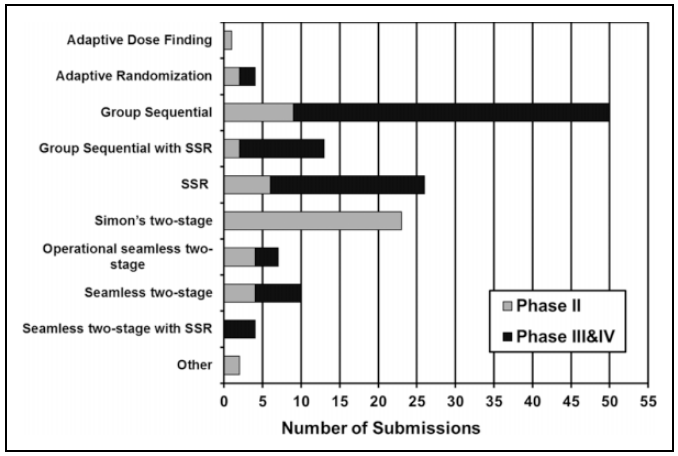

Figure 2. Types of adaptive design submissions by study phases.

approximately half (28 of 53; 52.8%) were single arm trials, and 36 (of 53; 67.9%) used well-understood adaptive design methods. Simon's two-stage designs[7] were employed in 43.4% (23 of 53) of the phase II proposals.

The overwhelming majority of confirmatory (phase III&IV) adaptive design submissions (84 of 87; 96.6%) were parallel controlled studies. This is in large part due to the fact that many of the products CBER regulates are not conducive to crossover designs because they have permanent effects or long carryover effects. Table 2 also shows that 66.7% (58 of 87) of phase III and IV studies were designed as either singly or doubly blinded, 95.4% (83 of 87) were randomized, and 94.3% (82 of 87) were superiority trials. With respect to primary endpoints, 43.7% (38 of 87) of the phase III and IV applications used time-to-event endpoints, and most of these (33 of 38; 86.8%) were from the Office of Cellular, Tissue and Gene Therapies, and includes many oncology products. Thirty-four (39.1%) phase III and IV trials had binary primary endpoints (eg, response or no response); these were most commonly seen in blood product submissions (24 of 34; 70.6%) (Figure 3).

Of the 87 confirmatory studies, 2 (2.3%) used adaptive randomization approaches, 41 (47.1%) employed classical group sequential designs,[10,11] 20 (23.0%) used sample size re-estimation (SSR), and 11 (12.6%) combined group sequential design with SSR. In some submissions, a more conventional group sequential design was modified to consider more than one primary endpoint. Thirteen (14.9%) submissions were seamless phase II/III studies, including 3 (3.4%) operationally seamless but inferentially distinct phase II/III applications and 4 (4.6%) seamless phase II/III trials with SSR incorporated in the confirmatory (phase III) stage (Figure 2). Overall, 56.3% (49 of 87) of the adaptive methods proposed in the phase III and IV trials were categorized as well-understood (Table 2), as defined in the FDA draft Guidance. Of the 35 confirmatory submissions involving SSR methods, 13 (37.1%) used unblinded SSR methods, which are classified as less well-understood approaches in the draft Guidance, 7 (20.0%) proposed blinded SSR methods, which are considered to be well-understood approaches, and more than 40% of the submissions (15 of 35) did not provide adequate information in the study protocol to determine how SSR would be performed (Figure 4). We counted as "Bayesian" any trial in which a Bayesian analysis formed the basis for either an interim or a final decision rule. Although many phase I adaptive submissions include proposed Bayesian adaptive design methods, the majority of phase II, III, and IV studies used frequentist adaptive design approaches only (115 of 140; 82.1%) (Table 2).

Table 2. Characteristics of adaptive design submissions by study phases.

	Study Phases		
	II (n = 53)	III and IV (n = 87)	Overall (N = 140)
Trial design			
Blinded	19 (35.8%)	58 (66.7%)	77 (55.0%)
Open label	34 (64.2%)	29 (33.3%)	63 (45.0%)
Parallel controlled	25 (47.2%)	84 (96.6%)	109 (77.9%)
Single arm	28 (52.8%)	3 (3.4%)	31 (22.1%)
Randomized	25 (47.2%)	83 (95.4%)	108 (77.1%)
Nonrandomized	28 (52.8%)	4 (4.6%)	32 (22.9%)
Method types			
Frequentist	45 (84.9%)	70 (80.5%)	115 (82.1%)
Bayesian	2 (3.8%)	2 (2.3%)	4 (2.9%)
Unclear	6 (11.3%)	15 (17.2%)	21 (15.0%)
Method categories			
Well understood	36 (67.9%)	49 (56.3%)	85 (60.7%)
Less well understood	10 (18.9%)	22 (25.3%)	32 (22.9%)
Unclear	7 (13.2%)	16 (18.4%)	23 (16.4%)
Review decisions			
No comments	29 (54.7%)	23 (26.4%)	52 (37.1%)
Need clarification	24 (45.3%)	64 (73.6%)	88 (62.9%)

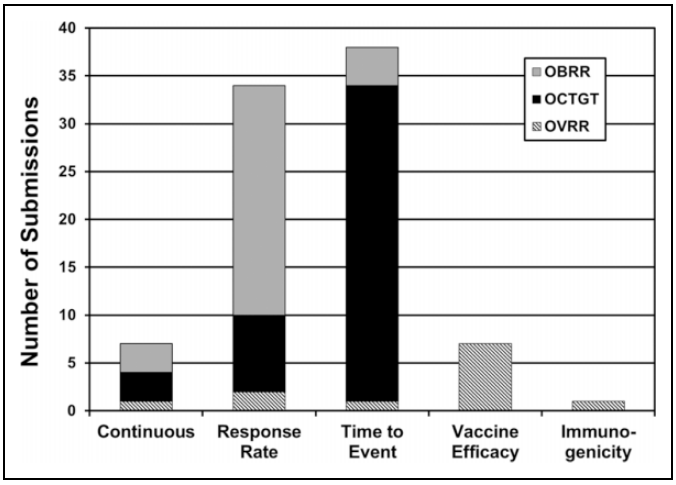

Figure 3. Adaptive design submissions by study endpoints in phase III and IV trials.

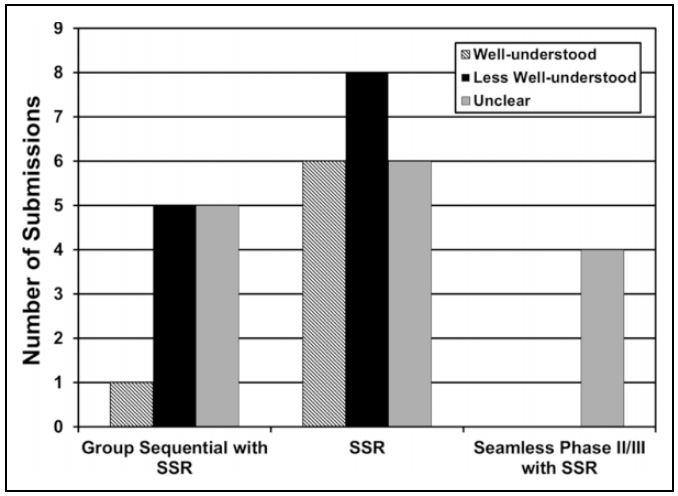

Figure 4. Adaptive design categories for sample size re-estimation methods in Phase III and IV submissions.

Over 50% (29 of 53; 54.7%) of the phase II trials were allowed to proceed without major statistical comments, while further clarifications were requested in most (64 of 87; 73.6%) of the confirmatory adaptive design trials (Table 2). A common issue in these submissions is insufficient detail in protocols and statistical analysis plans about the proposed adaptations to permit the Agency's determination of whether study objectives could be achieved while maintaining the study integrity and type I error control. In these cases, sponsors were always encouraged to submit revised protocols and/or statistical analysis plans.

Lessons from Our Review Experiences

Appropriateness of an Adaptive Design

The use of adaptive design in product development may or may not be appropriate depending upon a number of factors. Adaptive designs require clinical trials to be long enough to learn and adapt. Factors likely to determine a choice of an adaptive approach could include what is known at the planning stage: objectives and populations, proposed disease indications, accrual rate, choice of endpoints, scientific nature of the investigational products, financial resources, and others. The FDA review of a submission with adaptive design features will focus not only on its statistical robustness but also on its clinical appropriateness, and both clinicians and statisticians will weigh in on the appropriateness of a design and analysis plan. We recommend carefully assessing the potential advantages and disadvantages of an adaptive design over a traditional design at the study planning stage, especially for phase III confirmatory trials. Furthermore, sponsors should consider if simple, conventional adaptive designs, such as group sequential methods, will meet their needs. For example, complex adaptive designs may not be cost- or time-efficient for sponsors as they often require detailed design justification, extensive simulation studies, and multiple regulatory review cycles. This process can significantly delay the start of a study.

In practice, an adaptive design with too early an interim look may not be useful because there is too little information to properly adapt. Also, studies that stop too early for success may have inadequate safety data accrued, and both safety and efficacy are considerations for a pre-market approval. In addition, having sufficient sample size to assess secondary endpoints may need to be taken into consideration when planning to stop early, particularly if the secondary endpoint has less frequency (eg, mortality). Another issue arises with international trials: regulators may require a certain amount of in-country data prior to making an approval decision, especially when there are substantial differences in standard of care between countries or regions. Stopping a multiregional trial early without accounting for this issue could adversely affect a development program. Similarly, heterogeneity in the patient population can also raise concerns if one sees particular subgroups that are problematic at the time of regulatory review of premarket applications.

Early-Phase Studies

Early-phase studies often have largely descriptive analyses, but adaptive early-phase studies in which dose escalation is captured in a formula such as with continual reassessment methods (CRM)[22] sometimes require a formal statistical review at CBER. Statistical reviewers need to work with clinicians in evaluating the appropriateness of sponsors' proposals so that the pace of dose escalation is transparent, since phase I studies are largely about safety. However, adaptive designs may be especially appropriate in early-phase studies aimed at choosing among multiple doses, dose schedules, and perhaps other elements of product formulation.

In general, for early-phase adaptive design trials, the statistical reviews will mainly focus on whether or not the safety concerns were appropriately addressed in the study protocol. To this end, issues regarding initial dose selection, dose escalation algorithms, and safety monitoring plans will be carefully evaluated. In CBER, two-stage designs[23-25] are often seen in phase II cancer studies. However, as discussed in the CBER Guidance—Clinical Considerations for Therapeutic Cancer Vaccines,[26] phase II studies with the aim of gaining efficacy signals for the investigational products are suggested to mimic the future phase III confirmatory studies to its utmost extent, such as using clinical endpoints instead of surrogate endpoints. For preventive vaccines, phase II studies often enroll a few hundred participants and may consider multiple treatment groups with varying dose, formulation, and perhaps regimen. Adaptive design approaches may be useful in such trials, and may also provide more preliminary data with regard to both efficacy and safety.

Bayesian Approaches

Our survey indicated Bayesian methods were proposed more often for early-phase than for late-phase adaptive studies. Many

of these proposals were for CRM[22] dose-finding studies or variants thereof. This result supports findings of the DIA Bayesian Scientific Working Group (BSWG), which reported that there are more opportunities for statisticians to implement Bayesian adaptive design approaches in early phase than in late phase clinical development.[27] In very general terms, the chief advantages of Bayesian approaches are the ability to easily incorporate external information into an analysis and mathematical flexibility in fitting certain complicated models that may be intractable in a classical framework. The fact that CBER primarily receives Bayesian proposals for dose-finding models in early-phase trials suggests that it is the latter rather than the former factor spurring Bayesian applications in regulatory submissions to CBER. As discussed in the CDRH/CBER Guidance for the Use of Bayesian Statistics in Medical Device Trials,[28] we recommend that any Bayesian design for a confirmatory trial include an evaluation of type I error. This may limit the usefulness of incorporating external information in analyses in some cases, as incorporating such information can lead to type I error inflation.[29]

Sample Size Re-estimation

Our survey showed that SSR was the second-most frequently used adaptive design method in confirmatory trials, after classical group sequential designs. A similar result was reported based on a survey of scientific advice letters from the European Medicines Agency,[30] in which classical group sequential designs were excluded. The usefulness of interim SSR is mainly due to uncertainty about the true effect size during the planning stage of the study. In general, SSR based on blinded interim analyses of aggregate/overall data are considered not to be problematic, as those approaches have very limited potential to introduce bias or impair the validity and interpretability of study results. SSR based on unblinded knowledge of interim treatment effects, on the other hand, can raise issues related to type I error inflation and/or operational biases, and has to be approached with greater caution. An unblinded SSR may be conducted with or without a group sequential framework, or at the end of the first stage of a seamless phase II/III study. In planning any of the above unblinded SSRs, clear analytical derivations or statistical justifications to demonstrate no type I error inflation are expected, as well as strategies/plans to mitigate operational bias.[5,14,15,17,18] For example, CBER reviewers have agreed to multiple proposals using the conditional power approach described by Chen et al.[14]

It should be noted that SSR approaches can lead to changing the minimum effect size for which a statistically significant result will occur. We recommend clearly pre-specifying the decision rules with respect to sample size increase to ensure statistically significant results are also clinically meaningful. In addition, the timing of an SSR is also critical. An early SSR (eg, 30% of information) may result in an unreliable new sample size because of limited accrued data. A late SSR (eg, 80% information) may be pointless as planned accrual may be completed by this time. Study simulations under different scenarios may help to select an optimal timing for interim analyses. As with any adaptive approach, the timing and the rules for any SSR should be clearly prespecified.

Simulations

Trial simulations play an important role in adaptive design studies, especially in confirmatory trials. For less well-understood adaptive approaches, the major design properties (eg, type I error rate and power) often cannot be assessed through analytical derivations. Trial simulations may greatly facilitate the evaluation of the type I error rate and power, the average sample size that may be required, and timing of interim analysis under various realistic settings. In submissions to CBER, simulations have variously been used to support novel analytical methods, to verify the applicability of asymptotic tests (ie, tests that assume large numbers of patients) to a smaller anticipated study sample size, and to show how the type I error rate is preserved in cases where an analytic solution was not available. We strongly recommend extensive simulation studies for complex adaptive design proposals. However, trial simulations require a variety of assumptions that may be difficult to justify and may well turn out to be wrong in practice. In any case where a final study outcome is out of the range of simulated scenarios at the planning stage, interpreting the results of the trial might be complicated, and this could affect FDA review of a marketing application.

Preventive Vaccine Trials

Preventive vaccine trials based on disease endpoints differ from therapeutic trials in that they are usually conducted in large numbers of healthy subjects, not within a sick population. Therefore, for vaccines to be successful, they need to be regarded as safe, with little tolerance for rare but serious adverse events, and their efficacy usually needs to be considerably better than a placebo.

The most popular adaptive design for phase III vaccine clinical trials that we have seen is the classic group sequential design. Some examples include Prevnar (pneumococcal conjugate vaccine for prevention of invasive pneumococcal disease), Gardasil (human papilloma virus vaccine, types 6, 11, 16, 18), and RotaTeq rotavirus vaccine evaluated in the REST trial. The former vaccine trials used the group-sequential design to evaluate efficacy, while the RotaTeq trial used it mainly to monitor for the serious adverse event (SAE) intussusception, an intestinal obstructive disorder in young children that can be fatal.[31]

However, we have seen some group sequential designs that are not classic in that more than one hypothesis were proposed. For example, for both a therapeutic and a preventive vaccine trial, we have seen proposals for studying a subgroup within a group sequential framework, but the plans entailed continuing to accrue from a subgroup after the primary goal was achieved. Such a design posed an equipoise problem, however, if the overall trial demonstrated efficacy.

Example of a Successful Vaccine Trial

We have seen a successful seamless phase II/III adaptive design trial of a vaccine. This design was implemented in the study of Merck's recently licensed 9-valent human papillomavirus vaccine (HPV-9).[32,33] In this trial, the phase II part randomized subjects in equal numbers to one of 3 HPV-9 vaccine doses or the comparator Gardasil in a 3-dose regimen. The best HPV-9 vaccine dose, among those doses tested, was to be selected for phase III after all post–dose 2 immunogenicity and safety data were analyzed. Following selection of the HPV-9 vaccine dose in phase II, a phase III efficacy study including those subjects enrolled under phase II who received the selected HPV-9 vaccine dose or the control vaccine (Gardasil), and an additional 13,380 subjects enrolled under phase III was to be conducted to assess the efficacy and safety.

Although the FDA Guidance does not encourage seamless phase II/III adaptive designs, partly because they effectively eliminate the end-of-phase-II meeting with the FDA, we believed that there were several characteristics of the setting that supported its use for this trial. First, the HPV-9 vaccine is a higher valent, second-generation version of Gardasil, the 4-valent HPV vaccine licensed in the United States in 2006.[32,33] Thus, limited learning was needed, and there were clear criteria for dose selection. A sponsor internal committee, not involved in the study, was responsible for dose selection based only on immunogenicity data for the 4 original serotypes (6/11/16/18) common to both the HPV-9 vaccine and Gardasil. An external Data Monitoring Committee (DMC) was to confirm the selected dose based on both immunogenicity and safety, and this committee would have no access to efficacy data. Efficacy assessment in phase III was to be based on clinical endpoints related to the 5 new serotypes in HPV-9. Moreover, low statistical correlation was expected between the immune responses for the original 4 serotypes and the efficacy outcomes for the 5 new serotypes. Thus, type I error inflation due to data-driven dose selection was expected to be small. However, to compensate for any minimal type I error inflation that might occur, a conservative statistical approach was to be used to estimate vaccine efficacy (an exact method applied to a discrete distribution).[34] Because of our experience with Gardasil, a high level of efficacy was expected. All relevant details of the trial were prespecified, and blinding and firewalls were in place to protect the integrity of the trial. For all the above reasons, there was little concern about bias or type I error, and CBER concluded that an end-of-phase II meeting was not needed.

After the trial ended, simulation results performed by the applicant confirmed that the empirical type I error rate for the phase II/III adaptive design trial was likely below the nominal 0.025 level, and the immune response and efficacy endpoints indeed appeared to have low statistical correlation. Relative vaccine efficacy of HPV-9 compared to the 4-valent Gardasil was estimated to be 96%-97%, depending on case definition, with high lower confidence limits, which further confirmed the lack of need for concern about type I error.[32]

Although the above HPV-9 trial may be considered a successful implementation of a seamless phase II/III adaptive design, each trial protocol will need to be evaluated on its own merits. In particular, if CBER review committee concludes that an end-of-phase-II meeting is essential prior to the sponsor advancing to phase III because of concerns about type I error, trial integrity, or interpretability of results, then this type of adaptive design would not likely be practical. We would expect the sponsor of a trial to discuss the possibility of a seamless phase II/III study and its advantages and disadvantages when planning a confirmatory trial.

Adaptive designs could be of particular interest in vaccine development when looking at unmet public health needs, for example, HIV, TB, or malaria. In such cases, the high levels of efficacy seen for many preventive vaccines for childhood immunization may not be expected because of the challenges presented by the disease pathogens and natural history of the diseases. Thus, many phase III trials could be unsuccessful. Seamless phase II/III adaptive design trials could be useful in such settings to eliminate unpromising vaccine candidates early, while carrying forward to phase III only those candidates that demonstrate potential public health benefit.

Stopping for Futility

One of the most useful adaptations in clinical trials is consideration of a study termination for futility. CBER regulates many advanced biotechnology products with unknown mechanisms of action and uncertain efficacy. If it is apparent early on that a trial is unlikely to succeed, it may be ethical and financially prudent to stop the trial. There have been various adaptive design proposals over the years that have attempted to ''borrow alpha'' from a binding futility analysis, most often to increase the frequency or nominal significance level of interim efficacy analyses. Because we have encountered multiple cases in which supposedly binding futility boundaries have been crossed and ignored, it has been our practice to ask sponsors to evaluate type I error without accounting for any futility analyses.

Other Considerations

The quality and thoroughness of documentation are critical factors in our evaluation of any adaptive design proposal. A detailed study protocol with careful clinical and statistical considerations/justifications greatly facilitates communication between the FDA and a sponsor. We suggest that submissions provide the rationale for using an adaptive design, the logistics associated with each planned adaptation (eg, SSR, early stopping due to efficacy or futility), the distribution and parameter assumptions made for sample size calculations, the software used for sample size calculations, stopping rules, timing and methods of interim analyses, statistical analysis plan (SAP) for the primary and major secondary endpoints, analytical derivations or sufficient simulation studies along

with well-documented computer programs evaluating study-wise type I error rate, plans to minimize operational bias (eg, who conducts the interim analysis, who becomes aware of the interim results, and how to ensure data quality), Data Monitoring Committee (DMC) charter with details (eg, how DMC conveys the interim findings to the sponsor for decision making), and statistical references cited as part of the documentation.

It is also important to maintain documentation regarding operational compliance, especially in adaptive design trials. For example, autologous cancer vaccine trials are complex, as vaccines are made from each patient's own cells. The dosing time may significantly deviate from the scheduled time frames because cell harvest time and cell amounts may vary from patient to patient. Missing data and protocol compliance are often not scrutinized in detail until after a study is completed. However, adequate documentation during the conduct of a trial may encourage better monitoring of compliance and promote greater study and data quality. So, while we encourage sponsors to pick the best possible design for their objectives, we also caution that study quality and data quality are important factors. With adaptive designs that stop early for success, there may not be a ''cushion'' of excess positive information to overcome problems that are determined once the study is complete.

We realize that our experiences do not span all types of adaptive designs. For example, in spite of a growing literature on adaptive enrichment designs and the release of draft Guidance for Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products,[35] none have been submitted to CBER to date. Incorporating multiple objectives or multiple endpoints into an adaptive design can be particularly challenging and we have not seen many thus far. As noted earlier, we have seen a group sequential design with multiple stated objectives, but the sponsor needed longer term data to answer one of the objectives. So if the first objective of overall treatment efficacy is met and the sponsor decides to extend the study to evaluate some attribute of a particular subgroup, equipoise may no longer hold and it may be unethical to continue to randomize patients to a less effective therapy. So the fact that there is statistical theory to answer a question does not always equate into an appropriate design to meet the objectives. A dialog among an interdisciplinary team including statisticians and clinicians is just as important on the sponsor side as it is on the FDA side.

Conclusions

Our motivation in writing this article was to encourage the best study design proposals to be submitted to CBER. Sometimes these can be adaptive, and sometimes a simpler design is most efficient.

Our retrospective survey summarized our experience with adaptive design submissions in CBER over the past 6 years. The results showed that the number of applications involving adaptive designs did not continue to increase after the release of the FDA draft Guidance.[5] Our experience suggests that whether or not an adaptive feature can actually benefit a study may depend on a number of factors. Early-phase studies with adaptive design elements are generally encouraged, while more deliberation occurs for confirmatory studies. We give submissions with adaptive design elements careful attention because of the potential of inflated type I error rate and, in some cases, the need to verify simulations. Providing a well-written and sufficiently detailed study protocol in the regulatory submission will greatly assist the review process. We particularly suggest the following points for sponsors to consider if they intend to conduct adaptive design clinical trials for confirmatory purposes:

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

We encourage sponsors to propose adaptive design clinical trials when appropriate for their development. It is often productive, however, to discuss such proposals prior to formal submission, using regulatory meeting mechanisms such as pre-IND or Q-submissions[36] to start the dialog.

In summary, the FDA draft guidance offered a taxonomy for clinical trial design that may have been subject to misunderstanding. By labeling some studies as less well-understood, sponsors were discouraged from actually asking if such designs were options. Only through experience can designs move from a less well-understood to well-understood category. Another systematic survey of our submissions could be warranted in the future if our experiences change.

Acknowledgments

The authors thank CBER colleagues who were involved in conducting the internal survey for their assistance with the data collection and interpreting the findings.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Chow SC, Chang M. *Adaptive Design Methods in Clinical Trials*. 2nd ed. New York, NY: Chapman and Hall/CRC, Taylor and Francis; 2011.

2. Chow SC, Chang M. Adaptive design methods in clinical trials—a review. *Orphanet J Rare Dis*. 2008;3:11.
3. Coffey CS, Kairalla JA. Adaptive clinical trials—progress and challenges. *Drugs R&D*. 2008;9:229-242.
4. Kairalla JA, Coffey CS, Thomann MA, Muller KE. Adaptive trial designs: a review of barriers and opportunities. *Trials*. 2012;13:145.
5. FDA. Draft guidance for industry—adaptive design clinical trials for drugs and biologics. http://www.fda.gov/downloads/Drugs/ . . . /Guidances/ucm201790.pdf. Published 2010.
6. Dragalin V. Adaptive designs: terminology and classification. *Drug Inf J*. 2006;40:425-435.
7. Gaydos B, Krams M, Perevozskaya I, et al. Adaptive dose-response studies. *Drug Inf J*. 2006;40:451-461.
8. Hu F, Rosenberger W. *The Theory of Response-Adaptive Randomization in Clinical Trials*. New York, NY: Wiley; 2006.
9. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989;10:1-10.
10. Jennison C, Turnbull BW. *Group Sequential Methods With Applications to Clinical Trials*. London: Chapman & Hall/CRC; 2000.
11. Lan KKG, Demets DL. Discrete sequential boundaries for clinical-trials. *Biometrika*. 1983;70:659-663.
12. Tsiatis AA, Mehta C. On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*. 2003;90:367-378.
13. Bauer P, Koenig F. The reassessment of trial perspectives from interim data—a critical view. *Stat Med*. 2006;25:23-36.
14. Chen YHJ, DeMets DL, Lan KKG. Increasing the sample size when the unblinded interim result is promising. *Stat Med*. 2004; 23:1023-1038.
15. Denne JS. Sample size recalculation using conditional power. *Stat Med*. 2001;20:2645-2660.
16. Gao P, Ware JH, Mehta C. Sample size re-estimation for adaptive sequential design in clinical trials. *J Biopharm Stat*. 2008;18: 1184-1196.
17. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med*. 2011;30:3267-3284.
18. Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics*. 1995;51:1315-1324.
19. Bretz F, Schmidli H, Konig F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: general concepts. *Biom J*. 2006;48:623-634.
20. Morgan CC, Huyck S, Jenkins M, et al. Adaptive design: results of 2012 survey on perception and use. *Ther Innov Regul Sci*. 2014; 48:473-481.
21. Quinlan J, Gaydos B, Maca J, Krams M. Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development. *Clin Trials*. 2010;7:167-173.
22. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*. 1990;46:33-48.
23. Fleming TR. One-sample multiple testing procedure for phase-II clinical-trials. *Biometrics*. 1982;38:143-151.
24. Ivanova A, Monaco J, Stinchcombe T. Efficient designs for phase II oncology trials with ordinal outcome. *Stat Interface*. 2012;5: 463-469.
25. Jung SH, Lee T, Kim K, George SL. Admissible two-stage designs for phase II cancer clinical trials. *Stat Med*. 2004;23: 561-569.
26. FDA. Guidance for industry—clinical considerations for therapeutic cancer vaccines. http://www.fda.gov/downloads/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Vaccines/UCM278673.pdf. Published 2011.
27. Natanegara F, Neuenschwander B, Seaman JW, et al. The current state of Bayesian methods in medical product development: survey results and recommendations from the DIA Bayesian Scientific Working Group. *Pharm Stat*. 2014;13(1):3-12.
28. FDA. Guidance for the use of bayesian statistics in medical device clinical trials. http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf. Published 2010.
29. Viele K, Berry S, Neuenschwander B, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat*. 2014;13:41-54.
30. Elsaesser A, Regnstrom J, Vetter T, et al. Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. *Trials*. 2014;15:383.
31. Heyse JF, Kuter BJ, Dallas MJ, Heaton P, Team RS. Evaluating the safety of a rotavirus vaccine: the REST of the story. *Clin Trials*. 2008;5:131-139.
32. Chen YHJ, Gesser R, Luxembourg A. A seamless phase IIB/III adaptive outcome trial: design rationale and implementation challenges. *Clin Trials*. 2015;12:84-90.
33. Li P, Zhao Y, Sun X, Chan ISF. Multiplicity adjustment in seamless phase II/III adaptive trials using biomarkers for dose selection. In: Chen Z, Liu A, Qu Y, Tang L, Ting N, Tsong Y, eds. *Applied Statistics in Biomedicine and Clinical Trials Design*. New York: Springer; 2015:285-299.
34. Chan ISF, Bohidar NR. Exact power and sample size for vaccine efficacy studies. *Commun Stat-Theor M*. 1998;27: 1305-1322.
35. FDA. Guidance for enrichment strategies for clinical trials to support approval of human drugs and biological products. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf. Published 2012.
36. FDA. Requests for feedback on medical device submissions: the pre-submission program and meetings with food and drug administration staff. http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm311176.pdf. Published 2014.