Chow and Lin, Pharm Anal Acta 2015, 6:3
DOI: 10.4172/2153-2435.1000341

Pharmaceutica Analytica Acta

Review Article
Open Access

Analysis of Two-Stage Adaptive Seamless Trial Design

Chow SC[1]* and Lin M[2]

[1] *Duke University School of Medicine, Durham, North Carolina, USA*
[2] *Food and Drug Administration, Silver Spring, Maryland, USA*

Abstract

In the past decade, adaptive design methods in clinical research have attracted much attention because it offers the principal investigators (1) potential flexibility for identifying clinical benefit of a test treatment under investigation, but efficiency for speeding up the development process. One of the most commonly considered adaptive designs is probably a two-stage seamless (e.g., phase I/II or phase II/III) adaptive design. The two-stage seamless adaptive designs can be classified into four categories depending upon study objectives and study endpoints at different stages. These categories include (I) design with same study objectives and study endpoints at different stages, (II) designs with same study objectives but different study endpoints at different stages, (III) designs with different study objectives but same study endpoints at different stages, and (IV) designs with different study objectives and different study endpoints at different stages. In this article, an overview of statistical methods for analysis of these different types of two-stage designs is provided. In addition, a case study concerning the evaluation of a test treatment for treating hepatitis C infected patients utilizing type (IV) trial design is presented.

Keywords: Adaptive designs; Seamless phase II/III; Efficiency; Flexibility; Validity; Integrity

Introduction

In the past decade, adaptive design methods in clinical research have attracted much attention because it offers the principal investigators (1) potential flexibility for identifying clinical benefit of a test treatment under investigation, but efficiency for speeding up the development process. The FDA adaptive design draft guidance defines an adaptive design as a clinical study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study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 [1]. As it is recognized by many investigators/researchers, the use of adaptive design methods in clinical trials may allow the researchers to correct assumptions used at the planning stage and select the most promising option early. In addition, adaptive designs make use of cumulative information of the on-going trial, which provide the investigator an opportunity to react earlier to surprises regardless of positive or negative results Thus, the adaptive design approaches may speed up the drug development process.

Despite the possible benefits for having a second chance to modify the trial at interim when utilizing an adaptive design, it can be more problematic operationally due to bias that may have introduced to the conduct of the trial. As indicated by the FDA draft guidance, operational biases may occur when adaptations in trial and/or statistical procedures are applied after the review of interim (unblinded) data. As a result, it is a concern whether scientific integrity and validity of trial are warranted. Chow and Chang [2] indicated that trial procedures include, but not limited to, inclusion/exclusion criteria, dose/dose regimen and treatment duration, endpoint selection and assessment and/or laboratory testing procedures employed. On the other hand, statistical procedures are referred to as study design, statistical hypotheses (which can reflect study objectives), endpoint selection, power analysis for sample size calculation, sample size re-estimation, and/or sample size adjustment, randomization schedules and statistical analysis plan (SAP). With respect to these trial and statistical procedures, commonly employed adaptations at interim include sample size re-estimation at interim analysis, adaptive randomization with unequal treatment allocation (e.g., change from 1:1 ratio to 2:1 ratio), deleting, adding, or modifying treatment arms after the review of interim data, (4) shifting in patient population due to protocol amendment, different statistical methods, (6) changing study endpoints (e.g., change response rate and/or survival to time-to-disease progression in cancer trials), and changing hypotheses/objectives (e.g., switch a superiority hypothesis to a non-inferiority hypothesis). Therefore, the use of the adaptive design methods in clinical trials seems promising because of its potential flexibility for identifying any possible clinical benefit, signal, and/or trend regarding efficacy and safety of the test treatment under investigation. However, major adaptations may have an impact on the integrity and validity of the clinical trials, which may raise some critical concerns to the accurate and reliable evaluation of the test treatment under investigation. These concerns include (1) that the control of the overall type I error rate at a pre-specified level of significance, (2) that the correctness of the obtained p-values, and (3) that the reliability of the obtained confidence interval. Most importantly, major (significant) adaptations may have resulted in a totally different trial that is unable to address the scientific/medical questions the original study intended to answer.

As indicated by Chow [3], a seamless trial design is defined as a trial design that combines two independent trials into a single study that can addresses study objectives from individual studies. An adaptive seamless design is referred to as a seamless trial design that would use data collected before and after the adaptation in the final analysis. In practice, a two-stage seamless adaptive design typically consists of two stages (phases): a learning (or exploratory) phase (stage 1) and a confirmatory phase (stage 2). The objective of the learning phase is

***Corresponding author:** Chow SC, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, 2424 Erwin Road, Hock Suite 1102, Room 11068, Durham, NC, USA, Tel: 919-668-7523; E-mail: sheinchung.chow@duke.edu

Received January 22, 2015; **Accepted** February 21, 2015; **Published** February 28, 2015

Citation: Chow SC, Lin M (2015) Analysis of Two-Stage Adaptive Seamless Trial Design. Pharm Anal Acta 6: 341. doi: 10.4172/2153-2435.1000341

not only to obtain information regarding the uncertainty of the test treatment under investigation but also to provide the investigator the opportunity to stop the trial early due to safety and/or futility/efficacy based on accrued data or to apply some adaptations such as adaptive randomization at the end of Stage 1. The objective of the second stage is to confirm the findings observed from the first stage. A two-stage seamless adaptive trial design has the following advantages that (1) it may reduce lead time between studies (the traditional approach); (2) it provides the investigator the second chance to re-design the trial after the review of accumulated date at the end of Stage 1. Most importantly, data collected from both stages are combined for a final analysis in order to fully utilize all data collected from the trial for a more accurate and reliable assessment of the test treatment under investigation.

As indicated in Chow [3] and Chow and Tu [4], in practice, two-stage seamless adaptive trial designs can be classified into the following four categories depending upon study objectives and study endpoints at different stage.

Table 1 indicates that there are four different types of two-stage seamless adaptive designs depending upon whether study objectives and/or study endpoints at different stages are the same. For example, Category I designs (i.e., SS designs) include those designs with same study objectives and same study endpoints, while Category II and Category III designs (i.e., SD and DS designs) are referred to those designs with same study objectives but different study endpoints and different study objectives but same study endpoints, respectively. Category IV designs (i.e., DD designs) are the study designs with different study objectives and different study endpoints. In practice, different study objectives could be treatment selection for Stage 1and efficacy confirmation for Stage 2. On the other hand, different study endpoints could be biomarker, surrogate endpoints, or a clinical endpoint with a shorter duration at the first stage versus clinical endpoint at the second stage. Note that a group sequential design with one planned interim analysis is often considered an SS design.

In practice, typical examples for a two-stage adaptive seamless design include a two-stage adaptive seamless phase I/II design and a two-stage adaptive seamless phase II/III design. For the two-stage adaptive seamless phase I/II design, the objective at the first stage may be for biomarker development and the study objective for the second stage is usually to establish early efficacy. For a two-stage adaptive seamless phase II/III design, the study objective is often for treatment selection (or dose finding) while the study objective at the second stage is for efficacy confirmation. In this article, our focus will be placed on Category II designs. The results can be similarly applied to Category III and Category IV designs.

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 [1]. In this article, a two-stage adaptive seamless phase II/III design only refers to a study containing an exploratory phase II stage (stage 1) and a confirmatory phase III stage (stage 2) while data collected at both phases (stages) will be used for final analysis.

Study Objectives	Study	Endpoint
	Same (S)	Different (D)
Same (S)	I=SS	II=SD
Different (D)	III=DS	IV=DD

Source: Chow [2]

Table 1: Types of Two-stage seamless Adaptive Designs.

One of the questions that are commonly asked when applying a two-stage adaptive seamless design in clinical trials is sample size calculation/allocation. For the first kind (i.e. Category I, SS) of two-stage seamless designs, the methods based on individual p-values as described in Chow and Chang [2] can be applied. However, for other kinds (i.e. Category II to Category IV) of two-stage seamless trial designs, standard statistical methods for group sequential design are not appropriate and hence should not be applied directly. For Category II-IV trial designs, power analysis and/or statistical methods for data analysis are challenging to the biostatistician. For example, a commonly asked question is "How do we control the overall type I error rate at a pre-specified level of significance?" in the interest of stopping trial early, "How to determine stopping boundaries?" is a challenge to the investigator and the biostatistician. In practice, it is often of interest to determine whether the typical O'Brien-Fleming type of boundaries is feasible. Another challenge is "How to perform a valid analysis that combines data collected from different stages?" to address these questions, Cheng and Chow [5] proposed the concept of a multiple-stage transitional seamless adaptive design which takes into consideration of different study objectives and study endpoints.

Properties of Two-Stage Adaptive Design

As compared to the traditional approach (i.e., having two separate studies), a two-stage seamless adaptive design is preferred in terms of controlling type I error rate and power. For comparison of controlling the overall type I error rate, consider a two-stage adaptive trial design that combines a phase II trial and a phase III study. Let α_{II} and α_{III} be the corresponding type I error rate for the phase II trial and the phase III study, respectively. Thus, for the traditional approach, the overall type I error rate is given by $\alpha = \alpha_{II}\alpha_{III}$. In the two-stage adaptive seamless phase II/III design, on the other hand, the actual desired alpha is given by $\alpha = \alpha_{III}$. Thus, as compared to the traditional approach, the a for a two-stage adaptive phase II/III design is actually $1/\alpha_{II}$ times larger. Similarly, let $Power_{II}$ and $Power_{III}$ be the power for the phase II trial and the phase III study, respectively. Then, the power for the traditional approach is $Power = Power_{II} * Power_{III}$. In the two-stage phase II/III adaptive design, the power is given by $Power = Power_{III}$. Thus, as compared to the traditional approach, the power for a two-stage phase II/III adaptive design is $1/Power_{II}$ times larger.

A two-stage seamless adaptive trial design has the following advantages. First, it may help in reducing lead time between studies for the traditional approach. In practice, the lead time between end of the phase II trial and kick-off the phase III study is estimated about 6-12 months. This is because that usually the phase III study will not be initiated until the final clinical report of the phase II trial is completed. After the completion of a clinical study, it will usually take about 4-6 months to clean and lock the database, programming and data analysis, and final report. Besides, before we kick-off the phase III trial, protocol development, site selection/initiation, and IRB review/approval will also take some time. Thus, the use of a two-stage phase II/III adaptive trial design will definitely reduce the lead time between studies. In addition, the nature of adaptive trial design will also allow the investigator to make a go/no-go decision early (i.e., at the end of the first stage). In terms of sample size required, a two-stage phase II/III adaptive design may require a smaller sample size as compared to the traditional approach. Most importantly, a two-stage phase II/III adaptive trial design allows us to fully utilize data collected from both stages for a combined analysis which will provide a more accurate and reliable assessment of the test treatment under investigation.

In what follows, an overview of statistical methods for analysis of different types (i.e. Category I to IV) of two-stage designs is provided. In addition, a case study concerning the evaluation of a test treatment for treating patient with hepatitis C infection of a clinical study utilizing a Category IV adaptive design is presented.

Analysis for Category I Adaptive Designs

Category I design with same study objectives and same study endpoints at different stages is considered similar to a typical group sequential design with one planned interim analysis. Thus, standard statistical methods for group sequential design are often employed. It, however, should be noted that with various adaptations that applied, these standard statistical methods may not be appropriate. In practice, many interesting methods for Category I designs are available in the literature. These methods include (1) Fisher's criterion for combining independent p-values [6-8], (2) weighted test statistics [9] (3) the conditional error function approach [10,11] and (4) conditional power approaches [12].

Among these methods, Fisher's method for combining p-values provides great flexibility in selecting statistical tests for individual hypotheses based on sub-samples. Fisher's method, however, lacks flexibility in the choice of boundaries [13]. For Category I adaptive designs, many related issues have been studied. For example, Rosenberger and Lachin [14] explored the potential use of response-adaptive randomization. Chow, Chang, and Pong [15] examined the impact of population shift due to protocol amendments. Li et al., [12] studied a two-stage adaptive design with a survival endpoint, while Hommel et al. [16] studied a two-stage adaptive design with correlated data. An adaptive design with a bivariate-endpoint was studied by Todd [17] Tsiatis and Mehta [18] showed that there exists a more powerful group sequential design for any adaptive design with sample size adjustment,

For illustration purpose, in what follows, we will introduce the method based on sum of p-values (MSP) by Chang [2,19]. The MSP follows the idea of considering a linear combination of the p-values from different stages.

Theoretical framework

Consider a clinical trial utilizing a K-stage design. This is similar to a clinical trial with K interim analyses, while the final analysis is the K th interim (final) analysis. Suppose that at each interim analysis, a hypothesis test is performed. The objective of the trial can be formulated as the following intersection of the individual hypothesis tests from the interim analyses

$$H_0 : H_{01} \cap \cdots \cap H_{0K},$$

where $H_{0i}, i=1,\ldots,K$ is the null hypothesis to be tested at the i th interim analysis. Note that there are some restrictions on H_{0i}, that is, rejection of any $H_{0i}, i=1,\ldots,K$ will lead to the same clinical implication (e.g. drug is efficacious); hence all $H_{0i}, i=1,\ldots,K$ are constructed for testing the same endpoint within a trial. Otherwise the global hypothesis cannot be interpreted.

In practice, H_{0i} is tested based on a sub-sample from each stage, and without loss of generality, assume H_{0i} is a test for the efficacy of a test treatment under investigation, which can be written as,

$$H_{0i} : \eta_{i1} \geq \eta_{i2} \quad \text{versus} \quad H_{ai} : \eta_{i1} < \eta_{i2},$$

where η_{i1} and η_{i2} are the responses of the two treatment groups at the i th stage and we assume bigger values are better. It is often the case that when $\eta_{i1} = \eta_{i2}$, the p-value p_i for the sub-sample at the i th stage is uniformly distributed on (0, 1) under H_0. Under the null hypothesis, Bauer and Kohne [6] used Fisher's combination of the p-values to construct a test statistic for multiple-stage adaptive designs. Following similar idea, Chang [19] considered a linear combination of the p-values as follows,

$$T_k = \sum_{i=1}^{K} w_{ki} p_i, \, i = 1,\ldots,K, \tag{1}$$

Where $w_{ki} > 0$ and K is the number of interim analyses planned. If $w_{ki} = 1$, this leads to

$$T_k = \sum_{i=1}^{K} p_i, \, i = 1,\ldots,K. \tag{2}$$

T_k can be viewed as cumulative evidence against H_0. Thus, the smaller the T_k is, the stronger the evidence is. Alternatively, we can consider $T_k = \sum_{i=1}^{K} p_i / K$, which an average of the evidence is against H_0. Intuitively, one may consider the stopping rules

$$\begin{cases} \text{Stop for efficacy} & \text{if } T_k \leq \alpha_k \\ \text{Stop for futility} & \text{if } T_k \geq \beta_k \quad , \\ \text{Continue} & \text{otherwise} \end{cases} \tag{3}$$

Where T_k, α_k, and β_k are monotonic increasing functions of k, $\alpha_k < \beta_k, k = 1,\ldots,K-1$, and $\alpha_K = \beta_K$. Note that α_k and β_k are referred to as the efficacy and futility boundaries, respectively. To reach the k th stage, a trial has to pass 1 to $(k-1)$ th stages. Therefore, a so-called proceeding probability can be defined as the following unconditional probability:

$$\begin{aligned} \psi_k(t) &= P\left(T_k < t, \alpha_1 < T_1 < \beta_1, \ldots, \alpha_{k-1} < T_{k-1} < \beta_{k-1}\right) \\ &= \int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_{k-1}}^{\beta_{k-1}} \int_{-\infty}^{t} f_{T_1 \cdots T_k}(t_1,\ldots,t_k) dt_k dt_{k-1} \cdots dt_1, \end{aligned} \tag{4}$$

Where $t \geq 0, t_i, i = 1,\ldots,k$ is the test statistic at the i th stage, and $f_{T_1 \cdots T_k}$ is the joint probability density function. Thus, the error rate at the k th stage can be obtained as

$$\pi_k = \psi_k(\alpha_k). \tag{5}$$

Since the typeI error rates at different stages are mutually exclusive, the experiment-wise typeI error rate is sum ofπ_k, k=1,...K. Thus, we have

$$\alpha = \sum_{k=1}^{K} \pi_k. \tag{6}$$

Note that stopping boundaries can be determined with appropriate choices of a_k. The adjusted p-value calculation is the same as the one in a classic group sequential design[20]. The key idea is that when the test statistic at the kth stage $T_k = t = \alpha_k$ (i.e.just on the efficacy stopping boundary), the p-value is equal to alpha spent $\sum_{i=1}^{k} \pi_i$. This is true regardless of which error spending function is used and consistent with the p-value definition of the traditional design. As indicated in Chang [19], the adjusted p-value corresponding to an observed test statistic $T_k = t$ at the k th stage can be defined as

$$p(t;k) = \sum_{i=1}^{k-1} \pi_i + \psi_k(t), \, k = 1,..,K. \tag{7}$$

Note that p_i in equation (1) is the stage-wise (unadjusted) p-value from a sub-sample at the i th stage, while $p(t;k)$ are adjusted p-values calculated from the test statistic, which are based on the cumulative sample up to the k th stage where the trial stops, equations (6) and (7) are valid regardless how p_i are calculated.

Two-stage design

In this section, for simplicity, we will consider the method of sum of p-values (MSP) and apply the general framework to the two-stage designs as outlined in Chang [19] and Chow and Chang [2] which are suitable for the following adaptive designs that allow (1) early efficacy stopping, (2) early stopping for both efficacy and futility; and(3) early futility stopping. These adaptive designs are briefly described below.

Early efficacy stopping – For simplicity, consider $K = 2$ (i.e., a two-stage design) which allows for early efficacy stopping (i.e., $\beta_1 = 1$). By (5), the typeI error rates to spend at Stage 1 and Stage 2 are given by

$$\pi_1 = \psi_1(\alpha_1) = \int_0^{\alpha_1} dt_1 = \alpha_1, \tag{8}$$

and

$$\pi_2 = \psi_2(\alpha_2) = \int_{\alpha_1}^{\alpha_2}\int_t^{\alpha_1} dt_2 dt_1 = \frac{1}{2}(\alpha_2 - \alpha_1)^2, \tag{9}$$

respectively. Using equations (8) and (9), (6) becomes

$$\alpha = \alpha_1 + \frac{1}{2}(\alpha_2 - \alpha_1)^2. \tag{10}$$

Solving for α_2, we obtain

$$\alpha_2 = \sqrt{2(\alpha - \alpha_1)} + \alpha_1. \tag{11}$$

α_1 is the stopping probability (error spent) at the first stage under the null hypothesis condition and $\alpha - \alpha_1$ is the error spent at the second stage. As a result, if the test statistic $t_1 = p_1 > \alpha_2$, it is certain that $t_2 = p_1 + p_2 > \alpha_2$. Therefore, the trial should stop when $p_1 > \alpha_2$ for futility.

Based on relationship among α_1, α_2 and α as given in (10), various stopping boundaries can be considered with appropriate choices of α_1, α_2 and α For illustration purpose, Table 2 provides some examples of the stopping boundaries from equations (10, 11).

By (7)-(11), the adjusted p-value is given by

$$p(t;k) = \begin{cases} t & \text{if } k = 1 \\ \alpha_1 + \frac{1}{2}(t - \alpha_1)^2 & \text{if } k = 2 \end{cases}, \tag{12}$$

Where $t = p_1$ if the trial stops at Stage 1 and $t = p_1 + p_2$ if the trial stops at Stage 2.

Early efficacy or futility stopping – For this case, it is obvious that if $\beta_1 \geq \alpha_2$, the stopping boundary is the same as it is for the design with early efficacy stopping. However, futility boundary β_1 when $\beta_1 \geq \alpha_2$ is expected to affect the power of the hypothesis testing. Therefore,

$$\pi_1 = \int_0^{\alpha_1} dt_1 = \alpha_1, \tag{13}$$

and

$$\pi_2 = \begin{cases} \int_{\alpha_1}^{\beta_1}\int_{t_1}^{\alpha_2} dt_2 dt_1 & \text{for } \beta_1 \leq \alpha_2 \\ \int_{\alpha_1}^{\alpha_2}\int_{t_1}^{\alpha_2} dt_2 dt_1 & \text{for } \beta_1 > \alpha_2 \end{cases} \tag{14}$$

One-sided α	α_1	0.005	0.010	0.015	0.020	0.025	0.030
0.025	α_2	0.2050	0.1832	0.1564	0.1200	0.0250	-
0.05	α_2	0.3050	0.2928	0.2796	0.2649	0.2486	0.2300

Source: Chang [19] Statistics in Medicine, 26, 2772-2784.

Table 2: Stopping boundaries for two-stage efficacy designs.

One-sided α				$\beta_1 = 0.15$		
0.025	α_1	0.005	0.010	0.015	0.020	0.025
	α_2	0.2154	0.1871	0.1566	0.1200	0.0250
0.05	α_1	0.005	0.010	0.015	0.020	0.025
	α_2	0.3333	0.3155	0.2967	0.2767	0.2554

Source: Chang [19]. Statistics in Medicine, 26, 2772-2784

Table 3 Stopping boundaries for two-stage efficacy and futility designs.

Thus, it can be verified that

$$\alpha = \begin{cases} \alpha_1 + \alpha_2(\beta_1 - \alpha_1) - \frac{1}{2}(\beta_1^2 - \alpha_1^2) & \text{for } \beta_1 < \alpha_2 \\ \alpha_1 + \frac{1}{2}(\alpha_2 - \alpha_1)^2 & \text{for } \beta_1 \geq \alpha_2 \end{cases} \tag{15}$$

Similarly, under (15), various boundaries can be obtained with appropriate choices ofα_1 ,α_2 ,   andα_1 (Table 3).The adjusted p-value is given by

$$p(t;k) = \begin{cases} t & \text{if } k = 1 \\ \alpha_1 + t(\beta_1 - \alpha_1) - \frac{1}{2}(\beta_1^2 - \alpha_1^2) & \text{if } k = 2 \text{ and } \beta_1 < \alpha_2 \\ \alpha_1 + \frac{1}{2}(t - \alpha_1)^2 & \text{if } k = 2\ \beta_1 \geq \alpha_2 \end{cases} \tag{16}$$

Where $t = p_1$ if the trial stops at Stage 1 and $t = p_1 + p_2$ if the trial stops at Stage 2.

Early futility stopping – A trial featuring early futility stopping is a special case of the previous design, where $\alpha_1 = 0$ in equation (15). Hence, we have

$$\alpha = \begin{cases} \alpha_2\beta_1 - \frac{1}{2}\beta_1^2 & \text{for } \beta_1 < \alpha_2 \\ \frac{1}{2}\alpha_2^2 & \text{for } \beta_1 \geq \alpha_2 \end{cases} \tag{17}$$

Solving for α_2, it can be obtained that

$$\alpha_2 = \begin{cases} \frac{\alpha}{\beta_1} + \frac{1}{2}\beta_1 & \text{for } \beta_1 < \sqrt{2\alpha} \\ \sqrt{2\alpha} & \text{for } \beta_1 \geq \alpha_2 \end{cases} \tag{18}$$

Examples of the stopping boundaries generated using equation (18) are presented in Table 4. The adjusted p-value can be obtained from equation (16), where $\alpha_1 = 0$, that is,

One-sided α	β_1	0.1	0.2	0.3	≥ 0.4
0.025	α_2	0.3000	0.2250	0.2236	0.2236
0.05	α_2	0.5500	0.3500	0.3167	0.3162

Source: Chang [19].Statistics in Medicine, 26, 2772-2784.

Table 4: Stopping boundaries for two-stage futility design.

$$p(t;k) = \begin{cases} t & \text{if } k = 1 \\ \alpha_1 + t\beta_1 - \frac{1}{2}\beta_1^2 & \text{if } k = 2 \text{ and } \beta_1 < \alpha_2 \\ \alpha_1 + \frac{1}{2}t^2 & \text{if } k = 2\ \beta_1 \geq \alpha_2 \end{cases} \tag{19}$$

Conditional power

Conditional power with or without clinical trial simulation is often considered for sample size re-estimation in adaptive trial designs. As discussed earlier, since the stopping boundaries for the most existing methods are either based on z-scale or p-value, to link a z-scale and a p-value, we will consider $p_k = 1-\Phi(z_k)$ or inversely, $z_k = \Phi^{-1}(1-p_k)$, where z_k and p_k are the normal z -score and the p-value from the sub-sample at the k th stage, respectively. It should be noted that z_2 has asymptotically normal distribution with $N(\delta / se(\hat{\delta}_2), 1)$ under the alternative hypothesis, where $\hat{\delta}_2$ is the estimation of treatment difference in the second stage and

$$se(\hat{\delta}_2) = \sqrt{2\hat{\sigma}^2 / n_2} \approx \sqrt{2\sigma^2 / n_2}.$$

The conditional power can be evaluated under the alternative hypothesis when rejecting the null hypothesis H_0. That is,

$$z_2 \geq B(\alpha_2, p_1). \tag{20}$$

Thus, the conditional probability given the first stage naïve p-value, p_1 at the second stage is given by

$$P_C(p_1, \delta) = 1 - \Phi\left(B(\alpha_2, p_1) - \frac{\delta}{\sigma}\sqrt{\frac{n_2}{2}} \right), \alpha_1 < p_1 \leq \beta_1. \tag{21}$$

As an example, for the method based on the product of stage-wise p-values (MPP), the rejection criterion for the second stage is

$$p_1 p_2 \leq \alpha_2, i.e., z_2 \geq \Phi^{-1}(1 - \alpha_2 / p_1).$$

Therefore, $B(\alpha_2, p_1) = \Phi^{-1}(1-\alpha_2 / p_1)$.

Similarly, for the method based on the sum of stage-wise p-values (MSP), the rejection criterion for the second stage is

$$p_1 + p_2 \leq \alpha_2, i.e., z_2 = B(\alpha_2, p_1) = \Phi^{-1}(1 - \max(0, \alpha_2 - p_1)).$$

On the other hand, for the inversenormal method [21] the rejection criterion for the second stage is

$$w_1 z_1 + w_2 z_2 \geq \Phi^{-1}(1-\alpha_2),$$

$$\text{i.e., } z_2 \geq (\Phi^{-1}(1-\alpha_2) - w_1\Phi^{-1}(1-p_1)) / w_2,$$

where w_1 and w_2 are prefixed weights satisfying the condition of $w_1^2 + w_2^2 = 1$. Note that the group sequential design and CHW method [9] are special cases of the inverse-normal method. Since the inverse normal method requires two additional parameters (w_1 and w_2), for simplicity, we will only compare the conditional powers of MPP and MSP. For a valid comparison, the same α_1 is used for both methods. As it can be seen from equation (21), the comparison of the conditional power is equivalent to the comparison of function $B(\alpha_2, p_1)$. Equating the two $B(\alpha_2, p_1)$, we have

$$\frac{\hat{\alpha}_2}{p_1} = \tilde{\alpha}_2 - p_1, \tag{22}$$

where $\hat{\alpha}_2$ and $\tilde{\alpha}_2$ are the final rejection boundaries for MPP and MSP, respectively. Solving (22) for p_1, we obtain the critical point for p_1

$$\eta = \frac{\tilde{\alpha}_2 \mp \sqrt{\tilde{\alpha}_2^2 - 4\tilde{\alpha}_2}}{2}. \tag{23}$$

Equation (23) indicates that when $p_1 < \eta_1$ or $p_2 > \eta_2$, MPP has a higher conditional power than that of MSP. When $\eta_1 < p_1 < \eta_2$, MSP has a higher conditional power than MPP.As an example, for one-sided teat at $\alpha = 0.025$, ifwe choose $\alpha_1 = 0.01$ and $\beta_1 = 0.3$, then $\hat{\alpha}_2 = 0.0044$, and $\tilde{\alpha}_2 = 0.2236$, which result in $\eta_1 = 0.0218$, $\eta_2 = 0.2018$ by equation (23).

Note that the unconditional power P_w is nothing but the expectation of conditional power, i.e.

$$P_w = E_\delta[P_C(p_1, \delta)]. \tag{24}$$

Therefore, the difference in unconditional power between MSP and MPP is dependent on the distribution of p_1, and consequently, dependent on the true difference δ, and the stopping boundaries at the first stage α_1, β_1.

Note that in Bauer and Kohne's [6] method using Fisher's combination, which leads to the equation $\alpha_1 + \ln(\beta_1 / \alpha_1)e^{-(1/2)\chi^2_{4,1-\alpha}} = \alpha$, it is obvious that determination of β_1 leads to a unique α_1, consequently α_2. This is a non-flexible approach. However, it can be verified that the method can be generalized to $\alpha_1 + \alpha_2 \ln \beta_1 / \alpha_1 = \alpha$, where α_2 does not have to be $e^{-(1/2)\chi^2_{4,1-\alpha}} = \alpha$

Note that Tsiatis and Mehta [18] indicated that for any adaptive design with sample size adjustment, there exists a more powerful group sequential design. It, however, should be noted that the efficacy gain by the classic group sequential design is at the price of a cost. For example, as the number of interim analyses increases (e.g. from 3 to 10), the associated cost may increases substantially. Also, the optimal design is under the condition of a pre-specified error-spending function, but adaptive designs do not require in general a fixed error-spending function.

Analysis for category II adaptive designs

Now, consider a Category II two-stage phase II/III seamless adaptive designs which have same study objectives but different study endpoints (continuous endpoints). Let x_i be the observed value of the study endpoint (e.g., a biomarker) from the i^{th} subject in phase II (Stage 1), $i = 1, \ldots, n$ and y_j be the observed value of the study endpoint (i.e. the primary clinical endpoint) from the j^{th} subject in phase III (Stage 2), $j = 1, \ldots, m$. Suppose that $x_i's$ and $y_j's$ are independently and identically distributed with $E(x_i) = \nu$ and $Var(x_i) = \tau^2$, and $E(y_j) = \mu$ and $Var(y\ \ )$, respectively. Chow, Lu and Tse [22] proposed obtaining predicted values of the clinical endpoint based on data collected from the biomarker (or surrogate endpoint) under an established relationship between the biomarker and the clinical endpoint. These predicted values are then be combined with the data collected at the confirmatory phase (Stage 2) to derive a statistical inference on the treatment effect under investigation. For simplicity, suppose that x and y can be correlated in the following straight-line relationship

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{25}$$

where ε is the random error with zero mean and variance ς^2. ε is assumed to be independent of x. In practice, we assume that this relationship is well-established. In other words, the parameters β_0 and β_1 are assumed known. Based on equation (25), the observations x_i observed in the first stage can then be transformed $\beta_0 + \beta_1 x_i$ (denoted by $\hat{y}_i$). $\hat{y}_i$ is then considered as the observation of the clinical endpoint and combined with those observations y_i collected in the second stage to estimate the treatment mean. Chow, Lu and Tse [22] proposed the following weighted-mean estimator,

$$\hat{\mu} = \omega\overline{y} + (1-\omega)\overline{y} \tag{26}$$

where $\overline{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{y} = \frac{1}{m}\sum_{j=1}^{m} y_j$ and $0 \le \omega \le 1$. It should be noted that $\hat{\mu}$ is the minimum variance unbiased estimator among all weighted-mean estimators when the weight is given by

$$\omega = \frac{n/(\beta_1^2\tau^2)}{n/(\beta_1^2\tau^2) + m/\sigma^2} \tag{27}$$

if β_1, τ^2 and σ^2 are known. In practice, τ^2 and σ^2 are usually unknown and ω is commonly estimated by

$$\hat{\omega} = \frac{n/s_1^2}{n/s_1^2 + m/s_2^2} \tag{28}$$

where s_1^2 and s_2^2 are the sample variances of $\hat{y}_i$'s and y_j's, respectively. The corresponding estimator of μ, which is denoted by

$$\hat{\mu}_{GD} = \hat{\omega}\overline{y} + (1-\omega)\overline{y}, \tag{29}$$

and is referred to as the Graybill-Deal (GD) estimator of μ. Note that Meier [23] proposed an approximate unbiased estimator of the variance of the GD estimator, which has bias of order $O(n^{-2}+m^{-2})$. Khatri and Shah [24] gave an exact expression of the variance of this estimator in the form of an infinite series, which is given as.

$$\overset{\square}{V}ar(\hat{\mu}_{GD}) = \frac{1}{n/S_1^2 + m/S_2^2}\left[1 + 4\hat{\omega}(1-\hat{\omega})\left(\frac{1}{n-1} + \frac{1}{m-1}\right)\right].$$

Based on the GD estimator, the comparison of the two treatments can be made by testing the following hypotheses

$$H_0: \mu_1 = \mu_2 \qquad v.s. \qquad H_1: \mu_1 \ne \mu_2 \tag{30}$$

Let $\hat{y}_{ij}$ be the predicted value (based on $\beta_0 + \beta_1 x_{ij}$), which is used as the prediction of y for the jth subject under the ith treatment in phase II (Stage 1). From equation (29), the GD estimator of μ_i is given by

$$\hat{\mu}_{GDi} = \hat{\omega}_i\overline{y}_i + (1-\omega_i)\overline{y}_i, \tag{31}$$

where $\overline{\hat{y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, $\overline{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$ and $\hat{\omega}_i = \frac{n_i/S_{1i}^2}{n_i/S_{1i}^2 + m_i/S_{2i}^2}$

with S_{1i}^2 and S_{2i}^2 being the sample variances of $(\hat{y}_{i1},\cdots,y_{in_i})$ and $(y_{i1},\cdots,y_{im_i})$, respectively. For hypotheses (30), consider the following test statistic,

$$\tilde{T}_1 = \frac{\hat{\mu}_{GD1} - \mu_{GD2}}{\sqrt{\overset{\square}{V}ar(\hat{\mu}_{GD1}) + \overset{\square}{V}ar(\mu_{GD2})}} \tag{32}$$

where

$$\overset{\square}{V}ar(\hat{\mu}_{GDi}) = \frac{1}{n_i/S_{1i}^2 + m_i/S_{2i}^2}\left[1 + 4\hat{\omega}_i(1-\omega_i)\left(\frac{1}{n_i-1} + \frac{1}{m_i-1}\right)\right]$$

is an estimator of $Var(\hat{\mu}_{GDi})$, i =1, 2. Consequently, an approximate $100(1-\alpha)\%$ confidence interval of $\mu_1 - \mu_2$ is given as

$$\left(\hat{\mu}_{GD1} - \mu_{GD2} - z_{\alpha/2}\sqrt{V_T},\ \ \mu_{GD1} - \mu_{GD2} + z_{\alpha/2}\sqrt{V_T}\right) \tag{33}$$

Where $V_T = Var(\hat{\mu}_{GD1}) + Var(\hat{\mu}_{GD2})$. As a result, the null hypothesis H_0 is rejected if the above confidence interval does not contain 0. Thus, under the local alternative hypothesis that $H_1: \mu_1 - \mu_2 = \delta \ne 0$, the required sample size to achieve a $1-\beta$ power satisfies

$$-z_{\alpha/2} + |\delta| \big/ \sqrt{Var(\hat{\mu}_{GD1}) + Var(\mu_{GD2})} \ = z_\beta .$$

Thus, if we let $m_i=\rho n_i$ and $n_2=\gamma n_1$. Then, denoted by N_T, the total sample size required for achieving a desired power for detecting a clinically meaningful difference between the two treatments is (1+ρ)(1+γ)n_1, which is given by

$$n_1 = \frac{1}{2}AB\left(1 + \sqrt{1 + 8(1+\rho)A^{-1}C}\right) \tag{34}$$

where $A = \frac{(z_{\alpha/2} + z_\beta)^2}{\delta^2}, B = \frac{\sigma_1^2}{\rho + r_1^{-1}} + \frac{\sigma_2^2}{\gamma(\rho + r_2^{-1})}$ and $C = B^{-2}\left[\frac{\sigma_1^2}{r_1(\rho + r_1^{-1})^3} + \frac{\sigma_2^2}{\gamma^2 r_2(\rho + r_2^{-1})^3}\right]$ with $r_i = \beta_1^2\tau_i^2/\sigma_i^2$, i = 1, 2.

If one wishes to test for the following superiority hypotheses

H_1: $\mu_1 - \mu_2 = \delta_1 > \delta$.

The required sample size for achievng $1-\beta$ power satisfies

$$-z_\alpha + (\delta_1 - \delta)\big/\sqrt{Var(\hat{\mu}_{GD1}) + Var(\mu_{GD2})} \ = z_\beta$$

This gives

$$n_1 = \frac{1}{2}DB\left(1 + \sqrt{1 + 8(1+\rho)D^{-1}C}\right) \tag{35}$$

where $D = \frac{(z_\alpha + z_\beta)^2}{(\delta_1 - \delta)^2}$. For the case of testing for equivalence with a significance level α, consider the local alternative hypothesis thatH_1: $\mu_1 - \mu_2 = \delta_1$ with $|\delta_1| < \delta$. The required sample size to achieve $1-\beta$ power satisfies

$$-z_\alpha + (\delta - \delta_1)\big/\sqrt{Var(\hat{\mu}_{GD1}) + Var(\mu_{GD2})} \ = z_\beta$$

Thus, the total sample size for two treatment groups is $(1+\rho)(1+\gamma)n_1$ with n_1 given

$$n \ = -EB\left(1 + \sqrt{1 + 8(1+ \quad)E \quad C}\right) \tag{36}$$

where $E = \frac{(z_\alpha + z_{\beta/2})^2}{(\delta - |\delta_1|)^2}$.

Note that formulas for sample size calculation and allocation for testing equality, non-inferiority, superiority, and equivalence for other data types such as binary response and time-to-event endpoints can be similarly obtained.

Analysis for Category III and IV Adaptive Designs

In this section, statistical inference for Category III and IV phase II/III seamless adaptive designs will be discussed. For a Category III design, the study objectives at different stages are different (e.g., dose selection versus efficacy confirmation) but the study endpoints are same at different stages. For a Category IV design, both study objectives and endpoints at different stages are different (e.g., dose selection versus efficacy confirmation with surrogate endpoint versus clinical study endpoint).

As indicated earlier, how to control the overall type I error rate at a pre-specified level is one of the major regulatory concerns when adaptive design methods are employed in confirmatory clinical trials. Another concern is how to perform power analysis for sample size calculation/allocation for achieving individual study objectives originally set by the two separate studies (different stages). In addition, how to combine data collected from both stages for a combined and valid final analysis. Under a Category III or IV phase II/III seamless adaptive design, in addition, the investigator plans to have an interim analysis at each stage. Thus, if we consider the initiation of the study, first interim analysis, end of Stage 1 analysis, second interim analysis, and final analysis as critical milestones, the two-stage adaptive design becomes a 4-stage transitional seamless trial design. In what follows, we will focus on analysis of a four-stage transitional seamless design without (non-adaptive version) and with (adaptive version) adaptations, respectively.

Non-adaptive version

For a given clinical trial comparing *k* treatments groups, $E_1,\ldots,E_k$ with a control group *C*, suppose a surrogate (biomarker) endpoint and a well-established clinical endpoint are available for assessment of the treatment effect. Denoted by θ_i and$\psi_i, i=1,\ldots,k$ the treatment effect comparing E_i with *C* assessed by the surrogate (biomarker) endpoint and the clinical endpoint, respectively. Under the surrogate and clinical endpoints, the treatment effect can be tested by the following hypotheses:

$$H_{0,2}:\psi_1=\cdots=\psi_k, \tag{37}$$

which is for the clinical endpoint, while the hypothesis

$$H_{0,1}:\theta_1=\cdots\theta_k, \tag{38}$$

is for the surrogate (biomarker) endpoint. Cheng and Chow (2015) assumed that ψ_i is a monotone increasing function of the corresponding θ_i and proposed to test the hypotheses (37) and (38) at 3 stages (i.e., stage 1, stage 2a, stage 2b, and stage 3) based on accrued data at 4 interim analyses. Their proposed tests are briefly described below. For simplicity, he variances of the surrogate (biomarker) endpoint and the cli lenoted by σ^2 and τ^2, which are assumed known.

Stage 1 – At this stage, $(k+1)n_1$ subjects are randomly assigned to receive either one of the k treatments or the control at a 1:1 ratio. In this case, we have n_1 subjects in each group. At the first interim analysis, the most effective treatment will be selected based on the surrogate (biomarker) endpoint and proceed to subsequent stages. For pairwise comparison, consider test statistics $\hat{\theta}_{i,1}, i=1,\ldots,k$ and $S=argmax_{1\le j\le k}\,\overset{\wedge}{\theta}_{i,1}$. Thus, if $\hat{\theta}_{S,1}\le c_1$ for some pre-specified critical value c_1, then the trial is stopped and we are in favor of $H_{0,1}$. On the other hand, if $\hat{\theta}_{S,1}>c_{1,1}$, then we conclude that the treatment E_S is considered the most promising treatment and proceed to subsequent stages. Subjects who receive either the promising treatment or the control will be followed for the clinical endpoint. Treatment assessment for all other subjects will be terminated but will undergo necessary safety monitoring.

Stage 2a – At Stage 2a, $2n_2$ additional subjects will be equally randomized to receive either the treatment E_S or the control C. The second interim analysis is scheduled when the short term surrogate measures from these $2n_2$ Stage 2 subjects and the primary endpoint measures from those $2n_1$ Stage 1 subjects who receive either the treatment E_S or the control C become available. Let $T_{1,1}=\hat{\theta}_{S,1}$ and $T_{1,2}=\hat{\psi}_{S,1}$ be the pair-wise test statistics from Stage 1 based on the surrogate endpoint and the primary endpoint, respectively, and $\hat{\theta}_{S,2}$ be the statistic from Stage 2 based on the surrogate. If

$$T_{2,1}=\sqrt{\frac{n_1}{n_1+n_2}}\hat{\theta}_{S,1}+\sqrt{\frac{n_2}{n_1+n_2}}\theta_{S,2}\le c_{2,1},$$

then stop the trial and accept $H_{0,1}$. If $T_{2,1}>c_{2,1}$ and $T_{1,2}>c_{1,2}$, then stop the trial and reject both $H_{0,1}$ and $H_{0,2}$. Otherwise, if $T_{2,1}>c_{2,1}$ but $T_{1,2}\le c_{1,2}$, then we will move on to Stage 2b.

Stage 2b – At Stage 2b, no additional subjects will be recruited. The third interim analysis will be performed when the subjects in Stage 2a complete their primary endpoints. Let

$$T_{2,2}=\sqrt{\frac{n_1}{n_1+n_2}}\hat{\psi}_{S,1}+\sqrt{\frac{n_2}{n_1+n_2}}\psi_{S,2},$$

where $\hat{\psi}_{S,2}$ is the pair-wise test statistic from stage 2b. If $T_{2,2}>c_{2,2}$, then stop the trial and reject $H_{0,2}$. Otherwise, we move on to Stage 3.

Stage 3 – At Stage 3, the final stage, $2n_3$ additional subjects will be recruited and followed till their primary endpoints. At the fourth interim analysis, define

$$T_3=\sqrt{\frac{n_1}{n_1+n_2+n_3}}\hat{\psi}_{S,1}+\sqrt{\frac{n_2}{n_1+n_2+n_3}}\psi_{S,2}+\sqrt{\frac{n_1}{n_1+n_2+n_3}}\psi_{S,3},$$

where$\hat{\psi}_{S,3}$ is the pair-wise test statistic from stage 3. If $T_3>c_3$, then stop the trial and reject $H_{0,2}$; otherwise, accept $H_{0,2}$. The parameters in the above designs, n_1, n_2, n_3, $c_{1,1}$, $c_{1,2}$, $c_{2,1}$, $c_{2,2}$, and c_3 are determined such that the procedure will have a controlled type I error rate of α and a target power of $1-\beta$.

In the above design, the surrogate data in the first stage are used to select the most promising treatment rather than assessing$H_{0,1}$. This means that upon completion of stage one a dose does not need to be significance in order to be used in subsequent stages. In practice, it is recommended that the selection criterion be based on precision analysis (desired precision or maximum error allowed) rather than power analysis (desired power). This property is attractive to the investigator since it does not suffer from any lack of power because of limited sample sizes.

As discussed above, under the 4-stage transitional seamless design, two sets of hypotheses, namely $H_{0,1}$ and $H_{0,2}$ are to be tested. Since the rejection of $H_{0,2}$ leads to the claim of efficacy, it is considered the hypothesis of primary interest. However, in the interest of controlling the overall type I error rate at a pre-specified level of significance, $H_{0,1}$ need to be tested following the principle of closed testing procedure to avoid any statistical penalties.

In summary, the two-stage phase II/III seamless adaptive design is attractive due to its efficiency, such as potentially reducing the lead time between studies (i.e., a phase II trial and a phase III study) and flexibility, such as making an early decision and taking appropriate actions (e.g. stop the trial early or delete/add dose groups).

Adaptive version

The approach for trial design with non-adaptive version discussed in the previous section is basically a group sequential procedure with treatment selection at interim. There are no additional adaptations involved. With additional adaptations (adaptive version), Tsiatis and Metha [18] and Jennison and Turnbull [25] argue that adaptive designs typically suffer from loss of efficiency and hence are typically not recommended in regular practice. Proschan et al. [26] however, also indicated that in some scenarios, particularly when there is not enough primary outcome information available, it is appealing to use an adaptive procedure as long as it is statistically valid and justified. The transitional feature of the multiple stage design enables us not only to verify whether the surrogate (biomarker) endpoint is predictive of the clinical outcome, but also to modify the design adaptively after the review of interim data. A possible modification is to adjust the treatment effect of the clinical outcome while validating the relationship between the surrogate (e.g. biomarker) endpoint and the clinical outcome. In practice, it is often assumed that there exists a local linear relationship between ψ and θ, which is a reasonable assumption if we focus only on the values at a neighborhood of the most promising treatment E_S. Thus, at the end of Stage 2a, we can re-estimate the treatment effect of the primary endpoint using

$$\hat{\delta}_S = \frac{\hat{\psi}_{S,1}}{\hat{\theta}_{S,1}} T_{2,1}.$$

Consequently, sample size can be re-assessed at Stage 3 based on a modified treatment effect of the primary endpoint $\delta = \max\{\delta_S, \delta_0\}$, where δ_0 is a minimally clinically relevant treatment effect. Suppose m is the re-estimated Stage 3 sample size based on δ. Then, there is no modification for the procedure if $m \le n_3$.On the other hand, if $m > n_3$, then m (instead of n_3 as originally planned) subjects per arm will be recruited at Stage 3. The detailed justification of the above adaptation can be found in Cheng and Chow [5].

A case study – hepatitis C infection

A pharmaceutical company is interested in conducting a clinical trial for evaluation of safety, tolerability and efficacy of a test treatment for patients with hepatitis C virus infection. For this purpose, a two-stage seamless adaptive design is considered. The proposed trial design is to combine two independent studies (one phase IIb study for treatment selection and one phase III study for efficacy confirmation) into a single study. Thus, the study consists of two stages: treatment selection (Stage 1) and efficacy confirmation (Stage 2).The study objective at the first stage is for treatment selection, while the study objective at Stage 2 is to establish the non-inferiority of the treatment selected from the first stage as compared to the standard of care (SOC). Thus, this is a typical Category IV design (a two-stage adaptive design with different study objectives at different stages).

For genotype 1 HCV patients, the treatment duration is usually 48 weeks of treatment followed by a 24 weeks follow-up. The well-established clinical endpoint is the sustained virologic response (SVR) at week 72. The SVR is defined as an undetectable HCV RNA level (< 10 IU/mL) at week 72. Thus, it will take a long time to observe a response. The pharmaceutical company is interested in considering a biomarker or a surrogate endpoint such as a regular clinical endpoint with short duration to make early decision for treatment selection of four active treatments under study at end of Stage 1. As a result, the clinical endpoint of early virologic response (EVR) at week 12 is considered as a surrogate endpoint for treatment selection at Stage 1. At this point, the trial design has become a typical Category IV adaptive trial design (i.e., a two-stage adaptive design with different study endpoints and different study objectives at different stages). The resultant Category IV adaptive design is briefly outline below (Figure 1):

Stage 1 –At this stage, the design begins with five arms (4 active treatment arms and one control arm). Qualified subjects are randomly assigned to receive one of the five treatment arms at a 1:1:1:1:1 ratio. After all Stage 1 subjects have completed Week 12 of the study, an interim analysis will be performed based on EVR at week 12 for treatment selection. Treatment selection will be made under the assumption that the 12 week EVR is predictive of 72 week SVR. Under this assumption, the most promising treatment arm will be selected using precision analysis under some pre-specified selection criteria. In other words, the treatment arm with highest confidence level for achieving statistical significance (i.e., the observed difference as compared to the control is not by chance alone) will be selected. Stage 1 subjects who have not yet completed the study protocol will continue with their assigned therapies for the remainder of the planned 48 weeks, with final follow-up at Week 72. The selected treatment arm will then proceed to Stage 2.

Stage 2 –At Stage 2, the selected treatment arm from Stage 1 will be test for non-inferiority against the control (SOC). A separate cohort of subjects will be randomized to receive either the selected treatment from Stage 1 or the control (SOC)at a 1:1 ratio. A second interim analysis will be performed when all Stage 2 subjects have completed Week 12 and 50% of the subjects (Stage 1 and Stage 2 combined) have completed 48 weeks treatment and follow-up of 24 weeks. The purpose of this interim analysis is two-fold. First, it is to validate the assumption that EVR at week 12 is predictive of SVR at week 72. Second, it is to perform sample size re-estimation to determine whether the trial will achieve study objective (establishing non-inferiority) with the desired power if the observed treatment preserves till the end of the study.

Statistical tests as described in the previous section will be used to test non-inferiority hypotheses at interim analyses and at end of stage analyses. For the two planned interim analyses, the incidence of EVR at week 12 as well as safety data will be reviewed by an independent data safety monitoring board (DSMB). The commonly used O'Brien-Fleming type of conservative boundaries will be applied for controlling the overall Type I error rate at 5% [27]. Adaptations such as stopping the trial early, discontinuing selected treatment arms, and re-estimating the sample size based on the pre-specified criteria may be applied as recommended by the DSMB. Stopping rules for the study will be designated by the DSMB, based on their ongoing analyses of the data and as per their charter.

Figure 1. A diagram of 4-stage transitional seamless trial design

Concluding Remarks

Chow and Chang [2] pointed out that the standard statistical methods for a group sequential trial (with one planned interim

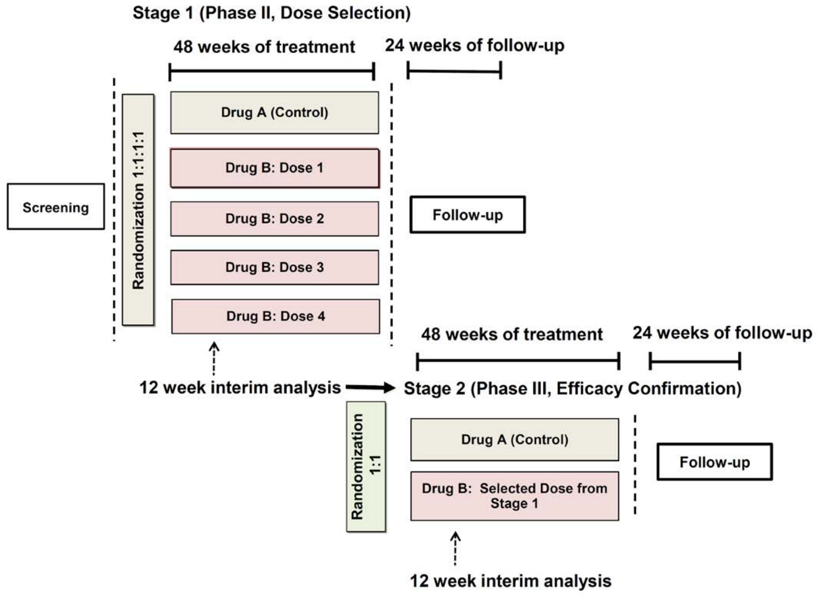

Figure 1: A diagram of 4-stage transitional seamless trial design.

analysis) is often applied for planning and data analysis of a two-stage adaptive design regardless whether the study objectives and/or the study endpoints are the same at different stages. As discussed earlier, two-stage seamless adaptive designs can be classified into four categories depending upon the study objectives and endpoints used at different stages. The direct application of standard statistical methods leads to the concern that the obtained p-value and confidence interval for assessment of the treatment effect may not be correct or reliable. Most importantly, sample size required for achieving a desired power obtained under a standard group sequential trial design may not be sufficient for achieving the study objectives under the two-stage seamless adaptive trial design, especially when the study objectives and/or study endpoints at different stages are different.

As indicated in the 2010 FDA draft guidance on adaptive clinical trial design, adaptive designs were classified as either well understood designs or less well understood designs depending upon the availability of well-established statistical methods of specific designs [1]. In practice, most of the adaptive designs (including the two-stage seamless adaptive designs discussed in this article) are considered less well understood designs. Thus, the major challenge is not only the development of valid statistical methods for those less well understood designs, but also the development of a set of criteria for choosing an appropriate design among these less well understood designs for valid and reliable assessment of test treatment under investigation.

Disclaimer

The views presented in this article have not been formally disseminated by the U.S. Food and Drug Administration and should not be construed to represent any agency determination or policy.

References

1. FDA (2010) Draft Guidance for Industry – Adaptive Design Clinical Trials for Drugs and Biologics.
2. Chow SC, Chang M (2011) Adaptive Design Methods in Clinical Trials (2nd Edn). Chapman and Hall/CRC, Taylor and Francis, New York
3. Chow SC (2011) Controversial Issues in Clinical Trials. Chapman and Hall/CRC, Taylor and Francis, New York, New York.
4. Chow SC, Tu YH (2008) On Two-stage Seamless Adaptive Design in Clinical Trials. J Formos Med Assoc 107: 52-60.
5. Cheng B, Chow SC (2015) Statistical inference for a multiple-stage transitional seamless trials designs with different study objectives and endpoints. Submitted.
6. Bauer P, Kohne K (1994) Evaluation of experiments with adaptive interim analyses. Biometrics 50: 1029-41.
7. Bauer P, Rohmel J (1995) An Adaptive Method for Establishing a Dose-Response Relationship. Stat Med 14: 1595-1607.
8. Posch M, Bauer P (2000) Interim analysis and sample size reassessment. Biometrics 56: 1170-1176.
9. Cui L, Hung HMJ, Wang SJ (1999) Modification of sample size in group sequential clinical trials. Biometrics 55: 853-857.
10. Liu Q, Chi GYH (2001) On sample size and inference for two-stage adaptive designs. Biometrics 57: 172-177.
11. Proschan MA, Hunsberger SA (1995) Designed extension of studies based on conditional power. Biometrics 51: 1315-1324.
12. Li G, Shih WCJ, Wang YN (2005) Two-stage adaptive design for clinical trials with survival data. J Biopharm Stat 15: 707-718.
13. Muller HH, Schafer H (2001) Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. Biometrics 57: 886-891.
14. Rosenberger WF, Lachin JM (2002) Randomization in Clinical Trials. John Wiley & Sons, Inc., New York.
15. Chow SC, Chang M, Pong A (2005) Statistical consideration of adaptive methods in clinical development. J Biopharm Stat 15: 575-591.
16. Hommel G, Lindig V, Faldum A (2005) Two-stage adaptive designs with correlated test statistics. J Biopharm Stat 15: 613-623.
17. Todd S (2003) An adaptive approach to implementing bivariate group sequential clinical trial designs. Journal of biopharm stat 13: 605-19.
18. Tsiatis AA, Mehta C (2003) On the inefficiency of the adaptive design for monitoring clinical trials. Biometrika 90: 367-378.

19. Chang M (2007) Adaptive design method based on sum of p-values. Stat Med 26: 2772-2784.

20. Jennison C, Turnbull BW (2000) Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall/CRC, London/Boca Raton, FL.

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

22. Chow SC, Lu QS, Tse SK (2007) Statistical analysis for two-stage seamless design with different study endpoints. J Biopharm Stat 17: 1163-1176.

23. Meier P (1953) Variance of a Weighted Mean. Biometrics 9: 59-73.

24. Khatri CG, Shah KR (1974) Estimation of location of parameters from two linear models under normality. Communications in Statistics-Theory and Methods 3: 647-663.

25. Jennison C, Turnbull BW (2006) Adaptive and nonadaptive group sequential tests. Biometrika 93: 1-21.

26. Proschan MA, Lan GKK, Wittes JT (2006) Statistical Monitoring of Clinical Trials: A Unified Approach.

27. Obrien PC, Fleming TR (1979) A Multiple Testing Procedure for Clinical-Trials. Biometrics 35: 549-556.