Published in final edited form as:

Stat Med. 2008 December 20; 27(29): 6190–6208. doi:10.1002/sim.3426.

Approaches to Handling Data When a Phase II Trial Deviates from the Pre-specified Simon's Two-Stage Design

Yujun Wu¹ and Weichung J. Shih^{2,*}

¹Biostatistics and Programming, Sanofi-Aventis, Bridgewater, New Jersey

²Department of Biostatistics, School of Public Health and Division of Biometrics, The Cancer Institute of New Jersey, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey

Abstract

Simon's 'optimal' and 'minimax' two-stage designs are common methods for conducting phase IIA studies investigating new cancer therapies. However, these designs are rather rigid in their settings because of the pre-specified rejection rules and fixed sample sizes at each stage. In practice, we often encounter the problem that a study is unable to adhere to the event number and sample sizes of the original two-stage design. In this paper, we consider some approaches in handling situations where deviations or interruptions from the original Simon's two-stage design occur because recruitment of patients is slower than expected. We consider four scenarios and use conditional probabilities to address the issues commonly inquired by the scientific review board. We also discuss how to report p-values in these situations.

Keywords

Phase IIA cancer trial; Simon's two-stage design; conditional probability; p-value

I. Introduction

The purpose of Phase IIA cancer clinical trials is to determine whether a new anticancer therapy has sufficient activity against a specific type of tumor to warrant further (phases IIB and III) investigation. These trials are usually conducted as single-arm studies with the primary endpoint being the tumor response rate defined clinically [1]. Denote the true response rate by p . The null hypothesis is $H_0: p \leq p_0$, where p_0 indicates an uninteresting response rate. If the null hypothesis is true, we require the type-I error rate be controlled at a pre-specified level. The study is also designed to have powered at a specified alternative hypothesis $H_A: p \geq p_1$, where $p_1 > p_0$ and is chosen to indicate a response rate which the new therapy may achieve and is worthy of further investigation. For ethical and logistic considerations a two or three-stage design is typically used for these phase II trials [2].

Among many two-stage designs, Simon's [3] 'optimal' and 'minimax' designs are most common in practice. Under the null hypothesis and subject to the constraints of the type-I and II error rates, the optimal two-stage design minimizes the expected sample size while the minimax design minimizes the maximum sample size. Other different extensions and variants of Simon's designs have also been proposed in the literature. For example, Sargent et al. [4] proposed a three-outcome design to allow for a decision of "uncertain" at the

*Corresponding author: W. J. Shih, 195 Little Albany St, New Brunswick, NJ 08901. E-mail address: shihwj@umdnj.edu

second (and final) stage. Jung et al. [5] investigated the class of ‘admissible’ designs for which Simon’s optimal and minimax designs are special cases. Lin and Shih [6] proposed adaptive two-stage designs, which allow the hypothesized p_1 to be dependent on the result of the first stage. Ayanlowo and Redden [7] recently considered incorporation of stochastic curtailment rules during the second stage of the Simon’s optimal and minimax designs.

In practice, we often encounter the following inquiries from the scientific review board or data monitoring committee (or equivalent) that exists at all US-NCI (US National Cancer Institute)-designated cancer centers for oversight of the scientific and operational aspects of clinical trials. A trial has been planned using a Simon’s two-stage design in the protocol. After the trial has begun for a period of time, the investigator reports unexpected slow recruitment of the patients. The review/monitoring board would like to evaluate the study result and ask whether statisticians can help to make recommendations regarding either stopping or continuing the trial on scientific grounds, and perhaps with some modification if needed. The problem of slower than expected accrual rate than expected is not uncommon at cancer centers, as elaborated in Green and Dahlberg [8] and Korn and Simon [9].

In this paper we are concerned with the classical statistical inferences of a trial in such a situation when the original two-stage design is interrupted or deviated unexpectedly. The situation can be classified into the following four scenarios. The first scenario is when a study has not reached the end of the first stage of the original design. The second scenario is when the study has passed the first stage (by the original design’s criterion), but the second stage is yet to complete as originally planned. In these two scenarios, there is only one unplanned evaluation requested by the scientific review board, either before or after the (planned) first stage analysis of the original design. The third scenario is that the study has been evaluated before it reaches the end of the first stage of the original design (as in the first scenario), then a second unplanned evaluation is asked again after the study has passed the first stage but has not yet completed the final stage, thus involving two unplanned evaluations. In all these three scenarios, the (unplanned) request for evaluation of the data, in effect, constitutes additional, unscheduled interim analysis. The fourth scenario also follows the first one where an unplanned evaluation is requested before the completion of the original first stage. But, different from the previous three scenarios, the original design is now totally abandoned and a new decision rule is to be set up to stop or continue the study, and a new second stage design has to be formed if continuing. Our goal in each of the first three scenarios is to form a statistical decision making strategy to either stop or continue the study, determine how the overall type I and II error rates are affected, and consider how to report the observed significance level (p-value) with the results. For the fourth scenario, our goal is to re-design the two stages and come up with a rationale for the new decision rules and to report p-values. We will discuss the first three scenarios as one category of deviation (i.e., the original Simon’s design is interrupted) with the illustration of numerical examples in Sections II to IV. We then discuss the fourth scenario as another category (i.e., the original Simon’s design is abandoned) in Section V. Discussions are given in Section VI.

II. Methods

II.1. Two Stage Designs

For testing the study hypotheses: $H_0: p \leq p_0$ versus $H_a: p \geq p_1$, a Simon’s two-stage design may be represented by the design parameters (r_1, n_1, r, n) , indicating that at the first stage, enter n_1 patients; if there are r_1 or fewer responses, terminate the study and accept H_0 (reject the therapy being promising); otherwise, enter $n - n_1$ additional patients at the second stage; if there are r or fewer responses out of the total n patients, accept H_0 . The design parameters (r_1, n_1, r, n) are set to satisfy the pre-specified type I (α) and type II (β) error rates and other efficiency considerations (e.g., Simon’s optimality and minimax criteria).

We formulate our proposals to handle interruptions around Simon's optimal or minimax designs, but note that the problem and our discussion may apply to other Simon-like two stage designs as well.

II. 2. Deviation Scenario 1

In this scenario, we are asked to evaluate the study when it has not reached the end of the first stage of the original design (i.e., when sample size is less than the pre-specified n_1). We can assist the scientific review committee by posing and answering the following question: Given the current observations, what is the conditional probability of passing the original first stage decision criterion to continue the trial? If this conditional passing probability (CPP) is low, the trial is stopped and H_0 is accepted (i.e., therapy is rejected). If the CPP is reasonably high, the institution may justifiably grant the trial with additional time and resource to continue. We then need to find out how the type I and II error rates are affected because of this additional CPP analysis.

Specifically, let $(x_{1/2}, n_{1/2})$ denote the current data that $x_{1/2}$ responses are observed out of the $n_{1/2}$ patients at the interim stage. Here we use the lower $1/2$ index to mean generally that this unplanned interim analysis happens prior to end of the pre-defined first stage. Given $(x_{1/2}, n_{1/2})$, we calculate CPP , the conditional probability of passing the original first stage decision criterion for continuation (i.e., $> r_1$ responses out of n_1 patients) by assuming $p = p_1$.

$$\begin{aligned} CPP(x_{1/2}, n_{1/2}) &= \Pr(\text{pass the first stage} | x_{1/2}, n_{1/2}, n_1, p=p_1) \\ &= \Pr(X_1 > r_1 | x_{1/2}, n_{1/2}, n_1, p=p_1) \\ &= 1 - B(r_1 - x_{1/2}; n_1 - n_{1/2}, p_1), \end{aligned} \quad (1)$$

where X_1 denotes the responses to be observed out of the n_1 patients at the first stage, and $B(y; m, p)$ is the cumulative probability of the Binomial distribution of observing at most y successes out of total sample size m with success probability p and is equal to 0 if $y < 0$. Let $c_{1/2}$ be a chosen cut-off value for the conditional passing probability, that is, continue the study to the first stage as originally planned if $CPP \geq c_{1/2}$; otherwise, stop (accept H_0).

In addition, since we are using the current observations $(x_{1/2}, n_{1/2})$ to perform a statistical analysis, we need to be mindful about its impact on the type I and type II error rates. Toward this end, first notice that $CPP(x_{1/2}, n_{1/2})$ in (1) is a strictly increasing function of the observed number of responses $x_{1/2}$ when fixing the other parameters. Let

$r_{1/2} = \arg \max_{0 \leq x_{1/2} \leq \min(r_1, n_{1/2})} \{CPP(x_{1/2}, n_{1/2}) < c_{1/2}\}$ the largest value of $x_{1/2} (\leq \min(r_1, n_{1/2}))$ that satisfies the stopping rule of $CPP < c_{1/2}$. Thus, determining if $CPP(x_{1/2}, n_{1/2}) \geq c_{1/2}$ is equivalent to examining if $x_{1/2} \geq r_{1/2}$. It follows that the probability of early terminating the study at the interim stage is

$$PET_{1/2}(p) = B(r_{1/2}; n_{1/2}, p). \quad (2)$$

The probability of early terminating the study at the first stage is

$$\begin{aligned}
\text{PET}_1(p) &= \Pr(\text{pass the interim stage, fail the first stage} | n_1, n_{1/2}) \\
&= \Pr(X_{1/2} > r_{1/2}, X_1 \leq r_1 | n_1, n_{1/2}) \\
&= \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, r_1)} \Pr(X_{1/2}=x_{1/2} | n_{1/2}) \Pr(X_1 \leq r_1 | X_{1/2}=x_{1/2}, n_{1/2}, n_1) \\
&= \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, r_1)} b(x_{1/2}; n_{1/2}, p) B(r_1 - x_{1/2}; n_1 - n_{1/2}, p).
\end{aligned} \tag{3}$$

where $b(y; m, p)$ denotes the probability of observing y responses for a binomial distribution $\text{Bin}(m, p)$. The probability of terminating the study without rejecting H_0 at the final stage is

$$\begin{aligned}
\text{PT}(p) &= \Pr(\text{pass the interim and the first stages, fail the final stage} | n_{1/2}, n_1, n) \\
&= \Pr(X_{1/2} > r_{1/2}, X_1 > r_1, X \leq r | n_{1/2}, n_1, n) \\
&= \sum_{x_1=r_1+1}^{\min(n_1, r)} \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1)} \Pr(X_{1/2}=x_{1/2} | n_{1/2}) \Pr(X_1=x_1 | X_{1/2}=x_{1/2}, n_{1/2}, n_1) \Pr(X \leq r | X_{1/2}=x_{1/2}, X_1=x_1, n_{1/2}, n_1, n) \\
&= \sum_{x_1=r_1+1}^{\min(n_1, r)} \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1)} b(x_{1/2}; n_{1/2}, p) b(x_1 - x_{1/2}; n_1 - n_{1/2}, p) B(r - x_1; n - n_1, p),
\end{aligned} \tag{4}$$

where X denotes the total number of responses observed at the final stage. Hence, with the additional interim analysis, the new type-I error rate is

$$\alpha^* = 1 - [\text{PET}_{1/2}(p_0) + \text{PET}_1(p_0) + \text{PT}(p_0)], \tag{5}$$

and the new power is

$$1 - \beta^* = 1 - [\text{PET}_{1/2}(p_1) + \text{PET}_1(p_1) + \text{PT}(p_1)]. \tag{6}$$

II. 3. Deviation Scenario 2

In this scenario, the study has passed the first stage and continued by the design's criterion (i.e., $x_1 > r_1$), but the second stage is not yet completed as originally planned. For an interim examination of the data, we take the following approach. Given the current data, calculate the conditional power (CP) of rejecting the null hypothesis by assuming that we continue the study to the final stage as originally designed. If the conditional power (CP) is low, the trial is stopped. If the CP is reasonably high, additional time and resources is justifiable and may be granted to continue the trial to the final stage. With a cut-off value for the conditional power to decide whether to stop or continue, the type-I and II error rates will be modified with this additional interim analysis.

Specifically, let $(x_{3/2}, n_{3/2})$ denote that $x_{3/2}$ responses are observed out of the $n_{3/2} (> n_1)$ patients. Here we use the lower index of 3/2 to mean generally that this unplanned interim analysis occurs after the pre-defined first stage but before the completion of the second (final) stages. Given the current observations $(x_{3/2}, n_{3/2})$ and assuming $p = p_1$, the conditional power is

$$\begin{aligned}
CP &= \Pr(\text{reject } H_0 \text{ at the final stage} | x_{3/2}, n_{3/2}, p = p_1) \\
&= \Pr(X > r | x_{3/2}, n_{3/2}, n, p = p_1) \\
&= 1 - B(r - x_{3/2}; n - n_{3/2}, p_1)
\end{aligned} \tag{7}$$

Let $c_{3/2}$ be a chosen cut-off value for the conditional power, that is, continue if $CP \geq c_{3/2}$; otherwise, stop and accept H_0 (i.e., reject the therapy).

Similar to CPP in Scenario 1, CP in (7) is a strictly increasing function of $x_{3/2}$ when fixing other parameters. Let $r_{3/2}$ denote the largest value of $x_{3/2}$ ($\leq \min(r, n_{3/2})$) that satisfies the stopping rule, $CP < c_{3/2}$. The probability of early terminating the study at the first stage is

$$PET_1(p) = B(r_1; n_1, p). \quad (8)$$

It follows that the probability of early terminating the study at the interim stage is

$$\begin{aligned} PET_{3/2}(p) &= \Pr(\text{pass the first stage, fail the interim stage} | n_1, n_{3/2}) \\ &= \Pr(X_1 > r_1, X_{3/2} \leq r_{3/2} | n_1, n_{3/2}) \\ &= \sum_{x_1=r_1+1}^{\min(n_1, r_{3/2})} b(x_1; n_1, p) B(r_{3/2} - x_1; n_{3/2} - n_1, p). \end{aligned} \quad (9)$$

The probability of terminating the study without rejecting H_0 at the final stage is

$$\begin{aligned} PT(p) &= \Pr(\text{pass the first and interim stages, fail the final stage} | n_1, n_{3/2}, n) \\ &= \Pr(X_1 > r_1, X_{3/2} > r_{3/2}, X \leq r | n_1, n_{3/2}, n) \\ &= \sum_{x_{3/2}=r_{3/2}+1}^{\min(n_{3/2}, r)} \sum_{x_1=r_1+1}^{\min(n_1, x_{3/2})} b(x_1; n_1, p) b(x_{3/2} - x_1; n_{3/2} - n_1, p) B(r - x_{3/2}; n - n_{3/2}, p). \end{aligned} \quad (10)$$

Therefore, with the additional interim analysis, the new type-I error rate is

$$\alpha^* = 1 - [PET_1(p_0) + PET_{3/2}(p_0) + PT(p_0)], \quad (11)$$

and the new power is

$$1 - \beta^* = 1 - [PET_1(p_1) + PET_{3/2}(p_1) + PT(p_1)]. \quad (12)$$

II. 4. Deviation Scenario 3

In this scenario, the study has been evaluated once before it reaches the end of the first stage of the original design, and a second interim evaluation is asked again after the study has passed the first stage. Therefore, there are two additional unplanned interim analyses, with outcomes and thresholds $(x_{1/2}, n_{1/2}, r_{1/2})$ for the first and $(x_{3/2}, n_{3/2}, r_{3/2})$ for the second, before and after the first stage of the original design. This situation is less common than the previous two scenarios where only one unplanned interim analysis is requested. Nevertheless, we include it here for completeness.

By the same arguments as before, the probability of early terminating the study at the first interim stage is

$$PET_{1/2}(p) = B(r_{1/2}; n_{1/2}, p).$$

The probability of early terminating the study at the original first stage is

$$PET_1(p) = \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, r_1)} b(x_{1/2}; n_{1/2}, p) B(r_1 - x_{1/2}; n_1 - n_{1/2}, p). \quad (13)$$

The probability of early terminating the study at the second interim stage is

$$PET_{3/2}(p) = \sum_{x_1=r_1+1}^{\min(n_1, r_{3/2})} \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1)} b(x_{1/2}; n_{1/2}, p) b(x_1 - x_{1/2}; n_1 - n_{1/2}, p) B(r_{3/2} - x_1; n_{3/2} - n_1, p). \quad (14)$$

The probability of terminating the study without rejecting H_0 at the final stage is

$$PT(p) = \sum_{x_{3/2}=r_{3/2}+1}^{\min(n_{3/2}, r)} \sum_{x_1=r_1+1}^{\min(n_1, x_{3/2})} \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1)} b(x_{1/2}; n_{1/2}, p) b(x_1 - x_{1/2}; n_1 - n_{1/2}, p) b(x_{3/2} - x_1; n_{3/2} - n_1, p) B(r - x_{3/2}; n - n_{3/2}, p). \quad (15)$$

Hence, with the additional two interim analyses, the new type-I error rate is

$$\alpha^* = 1 - [PET_{1/2}(p_0) + PET_1(p_0) + PET_{3/2}(p_0) + PT(p_0)], \quad (16)$$

and the new power is

$$1 - \beta^* = 1 - [PET_{1/2}(p_1) + PET_1(p_1) + PET_{3/2}(p_1) + PT(p_1)]. \quad (17)$$

III. Calculation of p-values

The above discussion is to help the scientific review board and the principle investigator to decide the future course of the study, and to investigate the effect of the additional interim analyses on the overall type I error rate and power. When the study is finished, either early stopped due to the low CP or CPP or other resources related reasons, or has completed the final stage, we need to analyze the data as it is. Notice that, unlike Fleming's [10] design, Simon's two-stage designs only stop for early futility, not for early efficacy, there is no concern of inflating the type-I error rate even with the occurrence of design deviation. Nevertheless, investigators are often interested in reporting a p-value under these circumstances. When the p-value is reported, we also should accompany it by the new type-I error rate α^* given in the previous sections to judge its significance.

We consider the following common definition of a p-value: the null probability of obtaining a test statistic value at least as extreme as that observed. Therefore, we must specify how to order different outcomes with early stopping taken into account. As in the case of group sequential designs, the sufficient statistic is a pair $(T, X_T/n_T) = (t, x_t/n_t)$ of the observed stopping time and the observed proportion of responders (successes) at $T=t$. For our situation with original 2 stages ($t=1, 2$) and 2 possible additional interim analyses ($t=1/2, 3/2$), t

obviously can only be $1/2$, 1 , $3/2$, or 2 , using the notation in the previous sections and letting $x_2 = x$ and $n_2 = n$ for $t = 2$. The ‘stage-wise ordering’, often used for group sequential trials with continuous observations [11], may be adopted here with some modifications., which are needed since in our situation the early stop is for futility (acceptance of H_0), not for efficacy (rejection of H_0). We first order the outcomes according to their evidence of futility and calculate the probability under H_0 of more extreme cases supporting H_0 than observed, and then the exact one-sided p-value would be one minus this probability. Jung et al. [12] has also considered this stage-wise ordering as well as other orderings for multistage phase II trials. Our consideration of interruption or deviation will actually result in an (unplanned) multistage situation and the derivations below are more specific to our situations.

In the two-dimensional region, we order the outcome based on the observed stopping time t first, with earlier stopping time indicating stronger evidence of futility. When the observed stopping times are the same, the smaller test statistic x_t/n_t , the stronger evidence of futility. If the test statistic x_t/n_t is also the same, we consider the outcomes equivalent in supporting H_0 or H_a , since the identical sufficient statistics provide the same information about the parameter p . This stage-wise ordering has the advantage that it only involves outcomes occurring up to the stopping time, not outcomes yet to occur after the stopping time. In particular, suppose we have observed $(t, x_t/n_t)$, by definition,

$$\begin{aligned} \text{p-value} &= 1 - [\Pr(T, X_T/n_T \text{ more extreme than } (t, x_t/n_t) \text{ to support } H_0 | p=p_0)] \\ &= 1 - \Pr[(T < t) \text{ or } (T = t \text{ and } X_T < x_t) | n_t, p=p_0] \end{aligned}$$

In the following we derive specific formulas for the original design and different deviation scenarios.

III.1 Original Design

When the trial follows the original two-stage design with no additional interim analyses, the possible observed stopping times are $t = 1$, or 2 . For $t = 1$, with $x_1 (\leq r_1)$ responses out of n_1 patients,

$$\text{p-value} = 1 - \Pr[X_1 < x_1 | n_1, p=p_0] = 1 - B(x_1 - 1; n_1, p_0). \quad (18a)$$

For $t = 2$, with $X_2 = x$ responses out of $n_2 = n$ patients,

$$\begin{aligned} \text{p-value} &= 1 - \Pr[(T=1) \text{ or } (T=2, X_2 < x) | n_1, n, p=p_0] \\ &= 1 - [\Pr(X_1 \leq r_1 | n_1, p=p_0) + \Pr(X_1 > r_1, X_2 < x | n_1, n, p=p_0)] \\ &= 1 - \left[B(r_1; n_1, p_0) + \sum_{x_1=r_1+1}^{\min(n_1, x-1)} b(x_1; n_1, p_0) B(x - x_1 - 1; n - n_1, p_0) \right]. \end{aligned} \quad (18b)$$

III.2 Deviation Scenario 1

In this scenario, the possible observed stopping times are $t = 1/2$, 1 , or 2 , when there is an interruption at $t = 1/2$ with stopping boundary $r_{1/2}$. For $t = 1/2$, with $x_{1/2} (\leq r_{1/2})$ responses out of $n_{1/2}$ patients,

$$\text{p-value} = 1 - \Pr[X_{1/2} < x_{1/2} | n_{1/2}, p=p_0] = 1 - B(x_{1/2} - 1; n_{1/2}, p_0). \quad (19a)$$

For $t = 1$, with $x_1 (\leq r_1)$ responses out of n_1 patients

$$\begin{aligned}
\text{p-value} &= 1 - \Pr[(T=1/2) \text{ or } (T=1, X_1 < x_1) | n_{1/2}, n_1, p=p_0] \\
&= 1 - \left[B(r_{1/2}; n_{1/2}, p_0) + \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1-1)} b(x_{1/2}; n_{1/2}, p_0) B(x_1 - x_{1/2} - 1; n_1 - n_{1/2}, p_0) \right].
\end{aligned} \tag{19b}$$

For $t=2$, with $X_2=x$ responses out of $n_2=n$ patients,

$$\begin{aligned}
\text{p-value} &= 1 - \Pr[(T=1/2) \text{ or } (T=1) \text{ or } (T=2, X_2 < x) | n_{1/2}, n_1, n, p=p_0] \\
&= 1 - [\Pr(X_{1/2} \leq r_{1/2} | n_{1/2}, p=p_0) + \Pr(X_{1/2} > r_{1/2}, X_1 \leq r_1 | n_{1/2}, n_1, p=p_0) + \Pr(X_{1/2} > r_{1/2}, X_1 > r_1, X < x | n_{1/2}, n_1, n, p=p_0)] \\
&= 1 - \left[B(r_{1/2}; n_{1/2}, p_0) + \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, r_1)} b(x_{1/2}; n_{1/2}, p_0) B(r_1 - x_{1/2}; n_1 - n_{1/2}, p_0) + \sum_{x_1=r_1+1}^{\min(n_1, x-1)} \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1)} b(x_{1/2}; n_{1/2}, p_0) b(x_1 - x_{1/2}; n_1 - n_{1/2}, p_0) \right].
\end{aligned} \tag{19c}$$

III.3 Deviation Scenario 2

In this scenario, the possible observed stopping times are $t=3/2$ or 2, when there is an interruption at $t=3/2$ with stopping boundary $r_{3/2}$. Similarly as in the previous section, for $t=3/2$, with $x_{3/2}(\leq r_{3/2})$ responses out of $n_{3/2}$ patients,

$$\begin{aligned}
\text{p-value} &= 1 - \Pr[(T=1) \text{ or } (T=3/2, X_{3/2} < x_{3/2}) | n_1, n_{3/2}, p=p_0] \\
&= 1 - \left[B(r_1; n_1, p_0) + \sum_{x_1=r_1+1}^{\min(n_1, x_{3/2}-1)} b(x_1; n_1, p_0) B(x_{3/2} - x_1 - 1; n_{3/2} - n_1, p_0) \right].
\end{aligned} \tag{20a}$$

For $t=2$, with $X_2=x$ responses out of $n_2=n$ patients,

$$\begin{aligned}
\text{p-value} &= 1 - \Pr[(T=1) \text{ or } (T=3/2) \text{ or } (T=2, X_2 < x) | n_1, n_{3/2}, n, p=p_0] \\
&= 1 - \left[B(r_1; n_1, p_0) + \sum_{x_1=r_1+1}^{\min(n_1, r_{3/2})} b(x_1; n_1, p_0) B(r_{3/2} - x_1; n_{3/2} - n_1, p_0) + \sum_{x_{3/2}=r_{3/2}+1}^{\min(n_{3/2}, x-1)} \sum_{x_1=r_1+1}^{\min(n_1, x_{3/2})} b(x_1; n_1, p_0) b(x_{3/2} - x_1; n_{3/2} - n_1, p_0) \right].
\end{aligned} \tag{20b}$$

III.4 Deviation Scenario 3

In this scenario, there are two interruptions at $t=1/2$ and $3/2$ with stopping boundaries $r_{1/2}$ and $r_{3/2}$, respectively, and the possible observed stopping times are $t=1/2, 1, 3/2$ or 2. The p-values for $t=1/2$ and 1 are the same as defined in (19a) and (19b), respectively. For $t=3/2$, similar as (19c),

$$\begin{aligned}
\text{p-value} &= 1 - \Pr[(T=1/2) \text{ or } (T=1) \text{ or } (T=3/2, X_{3/2} < x_{3/2}) | n_{1/2}, n_1, n_{3/2}, p=p_0] \\
&= 1 - \left[B(r_{1/2}; n_{1/2}, p_0) + \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, r_1)} b(x_{1/2}; n_{1/2}, p_0) B(r_1 - x_{1/2}; n_1 - n_{1/2}, p_0) + \sum_{x_1=r_1+1}^{\min(n_1, x_{3/2}-1)} \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1)} b(x_{1/2}; n_{1/2}, p_0) b(x_1 - x_{1/2}; n_1 - n_{1/2}, p_0) \right]
\end{aligned}
\tag{21a}$$

For $t=2$, with $X_2=x$ responses out of $n_2(=n)$ patients,

$$\begin{aligned}
\text{p-value} &= 1 - \Pr[(T=1/2) \text{ or } (T=1) \text{ or } (T=3/2) \text{ or } (T=2, X_2 < x) | n_{1/2}, n_1, n_{3/2}, n, p=p_0] \\
&= 1 - \left[B(r_{1/2}; n_{1/2}, p_0) + \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, r_1)} b(x_{1/2}; n_{1/2}, p_0) B(r_1 - x_{1/2}; n_1 - n_{1/2}, p_0) + \sum_{x_1=r_1+1}^{\min(n_1, r_{3/2})} \sum_{x_{1/2}=r_{1/2}+1}^{\min(n_{1/2}, x_1)} b(x_{1/2}; n_{1/2}, p_0) b(x_1 - x_{1/2}; n_1 - n_{1/2}, p_0) \right]
\end{aligned}
\tag{21b}$$

IV. Numerical examples

IV.1. Effect of deviations on type-I error rate and power

In this section we provide numerical examples to illustrate the use of the conditional passing probability (*CPP*) and conditional probability (*CP*) and for scenarios 1, 2 and 3, as appropriate, with some commonly seen parameters of p_0 and p_1 for phase IIA studies in cancer. We compare the original type-I error rate ($\alpha = 0.05$) and power ($1-\beta = 0.80$) with the new type-I error rate (α^*) and power ($1-\beta^*$) when the original designs are interrupted at $n_{1/2}$ ($< n_1$) or $n_{3/2}$ ($< n$) patients. For the first scenario, we illustrate with $n_{1/2} = 2n_1/3$. For the second scenario, we choose $n_{3/2}$ to be the middle of n_1 and n . We investigate cut-off values $c_{1/2}$ for *CPP* and $c_{3/2}$ for *CP* so that they may be specified properly in practice.

Table 1 and 2 give the results for the three deviation scenarios with Simon's optimal and minimax designs respectively. Recall that Scenario 1 is when the study is paused for examination before completing the pre-planned stage 1, and the conditional passing probability (*CPP*) is used to decide continuing or stopping the study. We have calculated the new type-I error rate and power for different *CPP* cut-off values and posted in the tables for $c_{1/2} = 0.6$ and 0.8 . We can see that when the study is paused (due to slow enrollment) at about $2/3$ of n_1 , the type-I error rate will be somewhat conservative and the power will suffer much if we use $c_{1/2} = 0.8$ as opposed to using $c_{1/2} = 0.6$, compared to the original type-I error and power. Of course, lower *CPP* cut-off implies easier to continue the study, hence not altering the design features too much. Higher *CPP* cut-off, on the other hand, gives more chance to stop the study, hence leading to greater difference between the original and the new type-I and II error rates.

The middle panels of Table 1 and 2 give the results for the second scenario with Simon's optimal and minimax designs respectively, when the study is interrupted in the middle of the pre-planned stage 2, and the conditional probability (*CP*) is the criterion to continue or stop the study. Likewise, we have calculated the new type-I error rate and power for different *CP*

cut-off values and posted in the tables for $c_{3/2} = 0.6$ and 0.8 . We can see that when the study is paused at about the middle of n_1 and n , the type- I error rate will not be too conservative and the power will not suffer much in most of the (p_0, p_1) configurations when we use the *CP* cut-off $c_{3/2} = 0.6$, but they could be affected somewhat when choosing $c_{3/2} = 0.8$.

Other *CP* cut-offs such as $c_{3/2} = 0.5$ and 0.7 have also been examined (not shown here). We highlight $c_{3/2} = 0.6$ and 0.8 here from the following perspective. It is interesting to examine the conditional power of the original Simon's designs. The optimal design is with the consideration of stopping the trial early for futility when the null hypothesis is true. Hence the n_1 is small. The minimax design is with the consideration of saving the maximum sample size. Hence the n is small. It is enlightening to see in Table 3 that, when observing r_1+1 responses out of n_1 subjects, the conditional power of the minmax design is around 0.35 to 0.73 , much lower than that of the optimal design, which ranges from 0.72 to 0.91 . Since the deviation scenario 2 is when interruption occurs after the stage 1, we would naturally use a conditional power at least higher than that at Stage 1. This argues in favor of considering $c_{3/2}$ about 0.8 for the optimal design and $c_{3/2}$ about 0.6 for the minimax design, as we have used in Table 1 and 2, when deviation scenario 2 occurs.

The right panels of Table 1 and 2 give the results for the third deviation scenario with Simon's optimal and minimax designs respectively, where the study is interrupted first during stage 1, then during stages 2 again. The conditional passing probability (*CPP*) with its cut-off value $c_{1/2}$ and the conditional probability (*CP*) with its cut-off value $c_{3/2}$ are used as the criteria to continue or stop the study at each interruption, respectively. Same as above can be said with this Scenario, only to a more degree of conservativeness of the type-I error and power loss, since additional interim analysis is performed here. These results should give the investigator some idea of the consequences from the deviation of the original design and choose the cut-off values appropriately.

IV.2. Calculation of CPP, CP and p-values

We illustrate the first deviation scenario with Simon's minimax design for testing $H_0: p \leq 0.6$ versus $H_a: p \geq 0.75$. Table 2 shows that with $\alpha = 0.05$ and $1-\beta = 0.80$, the design parameters $(r_1, n_1, r, n) = (18, 30, 43, 62)$. The study is interrupted at $n_{1/2} = 20$ patients due to slow enrollment. Using *CPP* cut-off $c_{1/2} = 0.6$, we obtain $r_{1/2} = 11$. The type-I error rate now becomes $\alpha^* = 0.046$. Suppose the interim data shows $x_{1/2} = 10$ responses, not exceeding $r_{1/2} = 11$. Hence we should stop the study. We also calculate *CPP* from Eq. (1) to confirm that we do not have enough conditional power to pass the original design's first stage:

$$CPP(x_{1/2}, n_{1/2}) = 1 - B(r_1 - x_{1/2}; n_1 - n_{1/2}, p_1) = 1 - B(8, 10; 0.75) = 0.24 < 0.6.$$

From (19a), the p-value for this outcome is:

$$p\text{-value} = 1 - \Pr[X_{1/2} < x_{1/2} | n_{1/2}, p = p_0] = 1 - B(x_{1/2} - 1; n_{1/2}, p_0) = 0.872.$$

However, suppose that the interim data shows $x_{1/2} = 13$ responses, exceeding $r_{1/2} = 11$. We then continue the study. We also calculate *CPP* from Eq. (1) to confirm that we have enough conditional power to pass the original design's first stage:

$$CPP(x_{1/2}, n_{1/2}) = 1 - B(r_1 - x_{1/2}; n_1 - n_{1/2}, p_1) = 1 - B(5, 10; 0.75) = 0.92.$$

If we then observe $(x_1, n_1) = (17, 30)$ and stop the study there, the p-value for this outcome follows (19b) as:

$$p\text{-value} = 1 - \Pr[(T=1/2) \text{ or } (T=1, X_1 < x_1) | n_{1/2}, n_1, p=p_0] = 0.555.$$

Notice that for the same $(x_1, n_1) = (17, 30)$ without the interruption the p-value would be 0.715.

We next illustrate the deviation scenario 2 with Simon's optimal design for testing $H_0: p \leq 0.4$ versus $H_a: p \geq 0.6$. Table 1 shows that with $\alpha = 0.05$ and $1-\beta = 0.80$, the design parameters $(r_1, n_1, r, n) = (7, 16, 23, 46)$. Suppose that the study passes stage 1, then is interrupted at $n_{3/2} = 31$ patients. With CP cut-off $c_{3/2} = 0.7$, we obtain $r_{3/2} = 15$. The new type-I error rate is $\alpha^* = 0.0415$. Suppose the interim data shows $x_{3/2} = 14$, we calculate CP from Eq. (7)

$$CP = 1 - B(r - x_{3/2}; n - n_{3/2}, p_1) = 1 - B(9; 15, 0.6) = 0.40$$

Either fact that $x_{3/2} < r_{3/2}$ or that $CP < c_{3/2}$ will recommend the study be stopped. We may report a p-value for this outcome as from (20a):

$$p\text{-value} = 1 - \Pr[(T=1) \text{ or } (T=3/2, X_{3/2} < x_{3/2}) | n_1, n_{3/2}, p=p_0] = 0.205.$$

V. Deviation Scenario 4 and Design Modification with Stochastic Curtailment

In the above approaches to deviation scenarios 1-3, we have incorporated the original 2-stage design with the additional unscheduled look(s) by checking the CPP or CP , hence increased the total number of analyses. In practice, however, the investigator may want to stay with two-stages, not to increase the number of analyses. This is also the most common situation we encounter with slow recruitment when the study is interrupted before completing the first stage of the original design. In this scenario, different from Scenario 1, we consider the unscheduled look as the new first stage and form a new second (final) stage, if continuing, and abandon the original (Simon's) two-stage design's specifications entirely. Our strategy to accomplish this is to consider the conditional power again, i.e., using the stochastic curtailment method. Our situation is also different from that of Ayanlowo and Redden [7] in that they considered stochastic curtailment during the second stage, while we consider it during the first stage; hence we limit the maximum looks in two.

Following the previous notation, suppose that there is an interruption at $(x_{1/2}, n_{1/2})$ prior to the pre-defined first stage. For the new (rescue) two-stage design, we take this unscheduled look as the new first stage $(x_1^*, n_1^* = (x_{1/2}, n_{1/2}))$, and let r_1^* be the threshold for continuing. If $x_1^* > r_1^*$, additional $n^* - n_1^*$ patients will be enrolled in the study at the new second stage or final stage with r^* to be the new final stage decision rule. The main question is how to construct r_1^* , r^* and n^* . We discuss their construction as follows.

V.1. Construction of new design specifications r_1^* , r^* and n^*

In reviewing Simon's two-stage optimal and minimax designs, we may notice that the decision rule r_1 at the first stage (given n_1) often is the integer value resulting in numerically the proportion r_1/n_1 closest to the null hypothesis value p_0 among all integers between 1 and n_1 . For example, in Table 1, the first row, $p_0 = 0.2 < 3/13 = 0.23$; in Table 2, the last row, $p_0 =$

0.6 = 18/30. Therefore, when the study is continued, i.e., when observing at least r_1+1 responses out of the n_1 patients, the point estimate at the first stage is often greater than p_0 . For the new first stage when the original design is deviated, we would apply a similar criterion by choosing the decision rule r_1^* such that $r_1^* = \lceil [n_1^*(p_0 + p_1)/2] \rceil$, where the notation $\lceil f \rceil$ means the smallest integer larger than the fraction f . This rule is stringent than the other possible choice of $r_1^* = \lceil [r_1^*p_0] \rceil$ we may also consider. We prefer the more stringent rule for continuing the study, since, under the situation of slow patient recruitment, the study really needs to show a stronger result to be justifiably granted with additional time and resource to continue. On the other hand, it is neither desirable to set r_1^* too high (such as $r_1^* = \lceil [n_1^*p_1] \rceil$) in order to maintain a certain overall power for the study. Furthermore, because r_1^* should be less than n_1^* , we modify r_1^* as

$$r_1^* = \min \{ \lceil [n_1^*(p_0 + p_1)/2] \rceil, n_1^* - 1 \}. \quad (22)$$

We emphasize that it is important to fix the r_1^* as prescribed here before we encounter this scenario of deviation when (x_1^*, n_1^*) becomes known.

Given the current observations (x_1^*, n_1^*) with $x_1^* > r_1^*$ and assuming $p = p_1$, the conditional power is

$$\begin{aligned} CP &= \Pr(\text{reject } H_0 \text{ at the final stage} | x_1^*, n_1^*, n^*, p = p_1) \\ &= \Pr(X > r^* | x_1^*, n_1^*, n^*, p = p_1) \\ &= 1 - B(r^* - x_1^*; n^* - n_1^*, p_1). \end{aligned} \quad (23)$$

The overall type I error is

$$\alpha = 1 - \left[B(r_1^*; n_1^*, p_0) + \sum_{x_1^*=r_1^*+1}^{\min(n_1^*, r^*)} b(x_1^*; n_1^*, p_0) B(r^* - x_1^*; n^* - n_1^*, p_0) \right]. \quad (24)$$

The values of r^* and n^* in the rescue design are chosen to control the overall type I error at the pre-specified level, e.g., 5%, and also to achieve the conditional power CP at a desired level, e.g., $\geq 80\%$, when assuming a conservative observation $x_1^* = r_1^* + 1$. However, notice that the pair (r^*, n^*) satisfying the above conditions are often not unique. Of all the solutions, we would choose the smallest n^* and pick the r^* that maximizes the conditional power.

V.2. Advantages of the rescue design

One obvious advantage of the new (rescue) design is its flexibility in n_1^* , the patients at the first stage. As a rescue, n_1^* in the new design is not a fixed design parameter prior to the study start, but is realized as the study comes for an evaluation, and the second stage is designed based upon the value of n_1^* , whereas for Simon's optimal and minimax designs, n_1 is a design parameter that is pre-specified before study starts.

Another advantage of the new design is the possibility of requiring fewer patients, as desired by the rescue situation. In Table 4 and Table 5, we display, as examples, some common Simon's optimal designs and minimax designs in contrast with their corresponding rescue two-stage designs for different hypotheses. For the new designs, we illustrate that the deviation occurs when $2n_1/3$ patients are enrolled. The designs are based upon controlling 5% overall type I error rate, maintaining 80% overall power for Simon's optimal and minimax designs, and achieving at least 80% conditional power for the rescue designs, and

for the rescue design, the rule of choosing r^*_1 is based on (22). (Similar results are observed for other settings of type I errors and powers and are not shown here.) Compared to the optimal and minimax designs, the corresponding new design requires smaller sample size (i.e., $n^* < n$) in all cases. This advantage is acquired by having a more stringent rule to continue the study than the (original) optimal or minimax design.

V.3. Reporting p-value or maximum type-II error for the rescue design

When the trial follows the new (rescue) two-stage design, the possible observed stopping times are still $t = 1$, or 2. The calculation of p-values is the same as that for the Simon's two-stage design, only now with the observation of x^*_1 and the new design parameters of r^*_1 , r^* and n^* .

Furthermore, in the new design, the first stage passing rule of r^*_1 is chosen (according to Eq. (22)) mainly based on the statistical consideration we discussed above. In practice, however, medical investigators may not necessarily follow this statistical rule. Instead, they may decide not to continue the study (and accept H_0) even though x^*_1 is greater than r^*_1 , perhaps in the consideration of some other factors, e.g., safety or resource issues. In this case, we may reasonably assume that the decision boundary r^*_1 (by (22)) is over-ruled by the implicit new boundary $r^*_1 = x^*_1$. In this case, we provide the minimum type II error rate, which is simply equal to $B(x^*_1, n^*_1, p_1)$, as a useful statistic in making such an over-ruling decision.

V.4. Example

A phase II study was conducted at the Cancer Institute of New Jersey to test the hypothesis that the sensitivity of cervical cancer to paclitaxel can be increased by interferon alpha-2b (Intron-A) and 13-cis retinoic acid (CRA) [13]. The protocol stated that the combination therapy of paclitaxel, CRA, and Intron-A will be of interest only if the true response rate is greater than 18% (H_0), and that the study targets the true response rate at $\geq 30\%$ (H_A) for further phase III investigation. The Simon's optimal design $(r_1, n_1, r, n) = (5, 27, 15, 66)$ was employed in the protocol, which controls type-I error rate = 10% and satisfies power = 80%. The study accrued 23 patients during 3/12/2001 to 2/29/2008 and observed 5 responders. On March 4, 2008 the principal investigator expressed to the Scientific Review Board that the accrual has been slow and requested an early evaluation of the study. We may view this situation as Scenario 1 deviation from the optimal design and calculate the conditional probability of passing the first stage (CPP). From Eq. (1) we obtain $CPP = 0.76$. Given that the original design's total $n = 66$, there is a concern that even if the study passes the first stage, it is difficult to follow the original plan with the current slow accrual rate. Hence we also propose to handle this situation as a Scenario 4 deviation. As such, the new rescue design has $n^*_1 = 23$ and $r^*_1 = \min \{ \lceil [n_1^*(p_0 + p_1)/2] \rceil, n_1^* - 1 \} = 6$ (from Eq. (22)) for its first stage. Since $x (=5)$ does not exceed 6, we then stop the study under the new rescue design and report p-value = 0.40. There is no need to construct the specification for the new second stage.

VI. Discussion

In this paper we have attempted to address the issue of deviations from Simon's two-stage designs in phase II cancer clinical trials. We have dealt with four scenarios of the deviations and provided calculations of the new type-I and II errors when deviations occur and handled by our proposed approaches. Some additional comments are as follows. First, the conditional probability of passing the first stage (CPP) for Scenario 1 (see Eq. (1)) and the conditional power (CP) of rejecting the null hypothesis assuming that we continue the study to the final stage for Scenarios 2 and 3 (see Eq. (7)) were calculated by assuming $p = p_1$. In fact CPP and CP can be calculated with any value of p , e.g., under the null hypothesis ($p =$

p_0) to give a pessimistic value or under the current trend, $p = x_{1/2}/n_{1/2}$, which might be more realistic. Second, we have noted that with interruptions and deviations, the new type-I error rate α^* would be somewhat less than the original type-I error rate α . We have taken the approach of reporting the new type-I error rate α^* with the p-value to judge the significance since it is possible that a p-value less than 0.05 would be greater than the new α^* . Of course the other more involved approach would be to alter the decision rule (r, n) of the original design to control the type-I error at the original α level. This latter approach is equivalent to abandoning the original design, similar to the approach of Scenario 4, only adding more steps. Third, for Scenario 4, in addition to its CP being also able to be calculated with any value of p as above (see Eq. (23)), this rescue design is most interesting in the sense that it actually fits in the general sample size re-estimation (SSR) framework of Shih [14] for any type of response. In SSR, the original sample size only serves as guidance, not cast-stoned, for the trial to be flexible after an interim analysis. When the patient accrual is slow, we need to re-design the study. The rescue design as proposed maintains the two-stage structure with revised decision rules and sample size, and has attractive advantages as we have demonstrated.

Acknowledgments

This work was partially supported by the NCI CCSG grant CA-72720-11 (WJS) and an UMDNJ Foundation Grant (YW during 2007 when he was at UMDNJ).

REFERENCES

1. Leventhal, BG.; Wittes, RE. Research Methods in Clinical Oncology. Raven Press; New York: 1988.
2. Herson, J. "Statistical aspects in the design and analysis of phase II clinical trials". In: Buyse, ME.; Staguet, MJ.; Sylvester, RJ., editors. Cancer Clinical Trials: Methods and Practice. Oxford University Press; Oxford: 1984.
3. Simon R. "Optimal two-stage designs for Phase II clinical trials". Controlled Clinical Trials. 1989; 10:1–10. [PubMed: 2702835]
4. Sargent DJ, Chan V, Goldberg RM. "A three-outcome design for phase II clinical trials". Controlled Clinical Trials. 2001; 22:117–125. [PubMed: 11306150]
5. Jung SH, Lee T, Kim KM, George SL. "Admissible two-stage designs for phase II cancer clinical trials". Statistics in Medicine. 2004; 23:561–5609. [PubMed: 14755389]
6. Lin Y, Shih WJ. "Adaptive two-stage designs for single-arm phase IIA cancer clinical trials". Biometrics. 2004; 60:482–490. [PubMed: 15180674]
7. Ayanlowo AO, Redden DT. "Stochastically curtailed phase II clinical trials". Statistics in Medicine. 2007; 26:1462–1472. [PubMed: 16900560]
8. Green SJ, Dahlberg S. "Planned versus attained design in phase II clinical trials". Statistics in Medicine. 1992; 11:853–862. [PubMed: 1604065]
9. Korn EL, Simon R. "Data monitoring committees and problems of lower-than-expected accrual or event rates". Controlled Clinical Trials. 1996; 17:526–535. [PubMed: 8974211]
10. Fleming TR. "One sample multiple testing procedures for phase II clinical trials". Biometrics. 1982; 43:143–151. [PubMed: 7082756]
11. Tsiatis AA, Rosner GL, Mehta CR. "Exact confidence intervals following a group sequential test". Biometrics. 1989; 40:797–803. [PubMed: 6518248]
12. Jung SH, Owzar K, George SL. "P-value calculation for multistage phase II cancer clinical trials". Journal of Biopharmaceutical Statistics. 2006; 16:765–775. [PubMed: 17146978]
13. Rodriguez-Rodriguez, L. "A Phase II Trial of Chemosensitization with Paclitaxel, 13-cis Retinoic Acid and Interferon Alpha-2b in Advanced Uterine Cervical Carcinoma". The Cancer Institute of New Jersey Oncology Group. 2001. protocol #100101 (CINJOG #NJ 1703)

14. Shih WJ. "Commentary: Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: a comparison". *Statistics in Medicine*. 2006; 25:933–941. [PubMed: 16220505]

Table 1

Simon's optimal design and deviations

Design ($\alpha = 0.05, 1-\beta = 0.80$)				Deviation Scenario 1				Deviation Scenario 2				Deviation Scenario 3			
				$c_{1/2} = 0.6$		$c_{1/2} = 0.8$		$c_{3/2} = 0.6$		$c_{3/2} = 0.8$		$c_{1/2} = c_{3/2} = 0.6$		$c_{1/2} = c_{3/2} = 0.8$	
p_0	p_1	r_1/n_1	r/n	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$
0.2	0.4	3/13	12/43	0.040 ⁺	0.712 ⁺	0.040 ⁺	0.712 ⁺	0.045	0.788	0.037	0.754	0.037	0.703	0.031	0.677
0.4	0.6	7/16	23/46	0.046	0.779	0.038	0.692	0.046	0.793	0.032	0.722	0.044	0.773	0.026	0.637
0.6	0.8	7/11	30/43	0.045 ⁺	0.760 ⁺	0.045 ⁺	0.760 ⁺	0.044	0.788	0.044	0.750	0.041	0.747	0.033	0.714
0.2	0.35	5/22	19/72	0.043	0.744	0.032	0.617	0.046	0.792	0.041	0.773	0.041	0.737	0.027	0.603
0.4	0.55	11/26	40/84	0.047	0.786	0.042	0.732	0.044	0.785	0.038	0.759	0.042	0.768	0.033	0.696
0.6	0.75	17/27	46/67	0.046	0.791	0.042	0.753	0.044	0.786	0.029	0.693	0.043	0.777	0.027	0.662

⁺The α^* and $1-\beta^*$ are the same for $c_{1/2} = 0.6$ and $c_{1/2} = 0.8$ because both lead to the same $r_{1/2}$ for this case.

Table 2

Simon's minimax design and deviations

Design ($\alpha = 0.05$, $1-\beta = 0.80$)				Deviation Scenario 1				Deviation Scenario 2				Deviation Scenario 3			
				$c_{1/2} = 0.6$		$c_{1/2} = 0.8$		$c_{3/2} = 0.6$		$c_{3/2} = 0.8$		$c_{1/2} = c_{3/2} = 0.6$		$c_{1/2} = c_{3/2} = 0.8$	
p_0	p_1	r_1/n_1	r/n	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$	α^*	$1-\beta^*$
0.2	0.4	4/18	10/33	0.043	0.779	0.035	0.699	0.034	0.738	0.034	0.738	0.033	0.722	0.028	0.659
0.4	0.6	17/34	20/39	0.041	0.753	0.032	0.672	0.048	0.799	0.038	0.757	0.041	0.751	0.026	0.651
0.6	0.8	8/13	25/35	0.048	0.788	0.048	0.788	0.047	0.798	0.039	0.753	0.045	0.779	0.038	0.738
0.2	0.35	6/31	15/53	0.049	0.795	0.046	0.771	0.036	0.726	0.024	0.636	0.036	0.723	0.023	0.625
0.4	0.55	28/59	34/70	0.041	0.746	0.034	0.685	0.041	0.771	0.041	0.771	0.035	0.724	0.030	0.669
0.6	0.75	18/30	43/62	0.046	0.793	0.043	0.765	0.042	0.772	0.034	0.718	0.041	0.765	0.032	0.696

Table 3Conditional power when observing $r_1 + 1$ responses at the first stage of the Simon's two-stage design

Hypothesis		Optimal design			Minimax design		
P_0	p_1	r_1/n_1	r/n	CP	r_1/n_1	r/n	CP
0.2	0.4	3/13	12/43	0.906	4/18	10/33	0.597
0.4	0.6	7/16	23/46	0.825	17/34	20/39	0.683
0.6	0.8	7/11	30/43	0.910	8/13	25/35	0.733
0.2	0.35	5/22	19/72	0.884	6/31	15/53	0.353
0.4	0.55	11/26	40/84	0.815	28/59	34/70	0.633
0.6	0.75	17/27	46/67	0.715	18/30	43/62	0.432

Table 4

Simon's optimal and minimax 2-stage designs and the corresponding rescue new 2-stage designs in Scenario 1 of deviation at $n_1^* = 2n_1/3$ $p_1 - p_0 = 0.2$ $r_1^* = \min\{[r_1^*(p_0 + p_1)/2], n_1^* - 1\}$

Hypothesis		Optimal Design				Minimax			
		Original $\alpha = 0.05, 1-\beta = 0.80$		New (rescue) $n_1^* = 2n_1/3$ $\alpha = 0.05, CP = 0.80$		Original $\alpha = 0.05, 1-\beta = 0.80$		New (rescue) $n_1^* = 2n_1/3$ $\alpha = 0.05, CP = 0.80$	
p_0	p_1	r_1/n_1	r/n	r_1^*/n_1^*	r^*/n^*	r_1/n_1	r/n	r_1^*/n_1^*	R^*/n^*
0.05	0.25	0/9	2/17	1/6	2/12	0/12	2/16	2/8	3/14
0.1	0.3	1/10	5/29	2/7	3/12	1/15	5/25	3/10	4/15
0.2	0.4	3/13	12/43	3/9	5/16	4/18	10/33	4/12	6/19
0.3	0.5	5/15	18/46	5/10	6/13	6/19	16/39	6/13	7/16
0.4	0.6	7/16	23/46	6/11	10/19	17/34	20/39	12/23	16/31
0.5	0.7	8/15	26/43	7/10	8/12	12/23	23/37	10/16	18/29
0.6	0.8	7/11	30/43	6/8	11/15	8/13	25/35	7/9	8/10
0.7	0.9	4/6	22/27	3/4	14/17	19/23	21/26	13/16	19/23

Table 5

Simon's optimal and minimax 2-stage designs and the corresponding rescue new 2-stage designs in Scenario 1 of deviation at $n_1^* = 2n_1/3$ ($p_1 - p_0 = 0.15$) $r_1^* = \min\{[n_1^*(p_0 + p^1)/2], n_1^* - 1\}$

Hypothesis		Optimal Design				Minimax Design			
		Original $\alpha = 0.05, 1-\beta = 0.80$		New (rescue) $n_1^* = 2n_1/3, \alpha = 0.05, CP = 0.80$		Original $\alpha = 0.05, 1-\beta = 0.80$		New (rescue) $n_1^* = 2n_1/3, \alpha = 0.05, CP = 0.80$	
P_0	P_1	r_1/n_1	r/n	r_1^*/n_1^*	r^*/n^*	r_1/n_1	r/n	r_1^*/n_1^*	R^*/n^*
0.05	0.2	0/10	3/29	1/7	2/15	0/13	3/27	2/9	3/17
0.1	0.25	2/18	7/43	3/12	4/18	2/22	7/40	3/15	4/21
0.2	0.35	5/22	19/72	5/15	6/19	6/31	15/53	6/21	11/39
0.3	0.45	9/27	30/81	7/18	17/44	16/46	25/65	12/31	20/52
0.4	0.55	11/26	40/84	9/18	21/43	28/59	34/70	20/40	25/51
0.5	0.65	15/28	48/83	11/19	30/51	39/66	40/68	26/44	37/63
0.6	0.75	17/27	47/67	13/18	23/33	18/30	43/62	14/20	29/42
0.7	0.85	14/19	46/59	11/13	12/14	16/23	39/49	13/16	21/26
0.8	0.95	7/9	26/29	5/6	19/21	7/9	26/29	5/6	19/21