Statistics in Biopharmaceutical Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usbr20

Modified Simon's Two-Stage Design for Phase IIA Clinical Trials in Oncology—Dynamic Monitoring and More Flexibility

Weichung Joe Shih, Yunqi Zhao & Tai Xie

To cite this article: Weichung Joe Shih, Yunqi Zhao & Tai Xie (2023): Modified Simon's Two-Stage Design for Phase IIA Clinical Trials in Oncology—Dynamic Monitoring and More Flexibility, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2023.2177332

To link to this article: https://doi.org/10.1080/19466315.2023.2177332

Published online: 10 Mar 2023.

STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2023, VOL. 00, NO. 0, 1–7
https://doi.org/10.1080/19466315.2023.2177332

Modified Simon's Two-Stage Design for Phase IIA Clinical Trials in Oncology—Dynamic Monitoring and More Flexibility

Weichung Joe Shih [a], Yunqi Zhao [b,c], and Tai Xie [c]

[a]Department of Biostatistics and Epidemiology and Cancer Institute of New Jersey, Rutgers Biomedical Health Sciences, New Brunswick, NJ; [b]Statistical and Quantitative Sciences, Takeda Pharmaceuticals, Cambridge, MA; [c]CIMS Global, Somerset, NJ

ABSTRACT

The traditional Simon's two-stage design for phase IIA clinical trials is modified to enhance the flexibility in conducting the interim analysis and sample size adjustment. The modification is based on the well-established methodology in adaptive designs using the conditional probability and allows for early termination as well as extension with sample size adjustment. The dynamic data monitoring system is naturally suitable for basket trials where several tumor types are monitored simultaneously with different enrollment rates.

ARTICLE HISTORY
Received May 2022
Accepted January 2023

KEYWORDS
Adaptive design; Conditional probability; DDM; Dynamic data monitoring; Sample size re-estimation; SSR

1. Introduction

Phase IIA oncology clinical trials in drug development are conducted 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 set up by single arm of the test compound with the primary endpoint being the tumor response rate defined clinically (Leventhal and Wittes 1988). Simon's "optimal" and "minimax" two-stage designs (Simon 1989) have been by far the most commonly used methods for these Phase IIA studies (Koyama and Chen 2008). However, these designs are rather rigid in their settings because of the assumed response rate, pre-specified rejection rules and fixed sample sizes at each stage. Several authors have proposed methods to amend or modify Simon's designs when the original expectation and/or configuration is deviated in practice. For example, Lin and Shih (2004) considered an adaptive design that used observed response rate at the first stage to redesign optimally the second stage. Their method is very practical since the target response rate in an early phase trial can be quite uncertain at the design stage and should be monitored closely. Wu and Shih (2008) considered four scenarios when unplanned evaluations were called for either before or after the first stage, while before the pre-specified final (second) stage ends, due to either administrative reasons (such as slow patient enrollment or regulatory/scientific review board's inquiry) or change of certain design assumptions from other trials. Other related discussions on phase II oncology trials include Fleming (1982), Green and Dahlberg (1992), Banerjee and Tsiatis (2006), Koyama and Chen (2008), and Englert and Kieser (2012a, 2012b), among others.

Adaptive design methodology has been developed in the last decades to allow performing trial modifications in the mid-stream while controlling the Type I error rate. Most of the methods developed have been for situations where a large sample theory (Scharfstein, Tsiatis, and Robins 1997) may be applied to form an approximate z-test based on either the arcsine transformation or the class of test statistics that are Regular and Asymptotically Linear (RAL); see Xie et al. (2021) for a recent review and further development. For early phase trials such as the phase IIA in oncology, however, we need to consider the very limited sample size involved, compared to larger sample size in late phase trials. For binary response rate with limited sample size, we use the exact binomial distribution, as it was done in the original Simon's designs, rather than the approximation by a z-test. Nevertheless, the ideas of adaptive designs developed previously, including flexible two-stage sequential alpha-spending with possible early stop and/or expansion with Sample Size Re-estimation (SSR) methods, are still applicable. In fact, Englert and Kieser (2012b) applied the conditional error rate principle by Muller and Schafer (2004) and Proschan and Hunsberger (1995) and proposed a general framework of discrete conditional error functions to construct more efficient phase II designs that allow flexible design modifications. Li et al. (2002) commented on the use of conditional error functions in Proschan and Hunsberger (1995) in terms of their mathematical form that lacks clear interpretation. Li et al. (2002) instead advocated the use of conditional probability directly. The conditional probability approach is easily interpretable since it is the conditional Type-I error rate under the null hypothesis and is the conditional power under the alternative hypothesis; the former is used for considering alpha preservation and the latter

CONTACT Weichung Joe Shih w.joe.shih@rutgers.edu Department of Biostatistics and Epidemiology, Rutgers University, 16 Ambrose Valley Lane, Piscataway, NJ 08854.
Weichung Joe Shih is Professor Emeritus of the Rutgers University.

is used for enhancing power by SSR. Furthermore, the approach of Li et al. (2002) considered early stop for both overwhelming efficacy and binding or nonbinding futility, as put forward by Shih, Li, and Wang (2016). The Dynamic Data Monitoring (DDM) system with a "radar screen" display developed by Xie et al. (2021) that allows for "anytime" accessibility to data can also be applied to monitoring the Phase IIA trial especially for the increasingly popular basket trials. In this article, we delineate the conditional probability approach to the discrete binary response rate for single-armed phase IIA trials in oncology. In Section 2, we set up the design and notation. In Sections 3–5, we describe the modification of the traditional Simon's two-stage design including the flexible monitoring schedule for interim analysis, monitoring regions, and extension of sample size. In Section 6, we illustrate with a numerical example and summarize the proposed design. In Section 7, the DDM system is used to monitor for phase IIA trials simultaneously in the basket trial setting. Finally, we conclude with discussion in Section 8.

2. Notation and Design Setup

Let $X_i = 1$ with probability $p, X_i = 0$ with probability $1 - p$ for $i = 1, 2, \ldots, N$, where N is the number of patients, $X_i = 1$ while observing response for patient i, and $X_i = 0$ if patient i is a nonresponder. Consider the following hypothesis test with binary data:

$$H_0 : p = p_0$$
$$H_a : p = p_1$$

where p_0 is the response rate for the standard therapy and p_1 is the targeted response rate for the experimental therapy under investigation. Without loss of generality, it is assumed that $p_1 > p_0$. We first consider the fixed sample size design without any interim analysis to obtain the maximal sample size N. Suppose we would like to have a power $1 - \beta$, and a Type I error rate α. Let $X_n = \sum_{i=1}^{n} X_i$ denote the total number of responses out of n patients. X_n has a Binomial distribution with probability density function $b(x; n, p)$ and cumulative probability function $B(r, n, p) = P(X \le r; n, p) = \sum_{x=0}^{r} b(x, n, p)$. Then, to determine a sample size N and a critical boundary R, such that, if $X_N \ge R$, we reject H_0 and claim the drug is worthy of further study; otherwise, we do not reject H_0 and dismiss the therapy. Thus, N and R will be the minimum values that satisfy

- $P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(X_N \ge R \mid p = p_0)$
  $= 1 - B(R - 1, N, p_0) \le \alpha$
- $1 - B(R - 1, N, p_1) \ge 1 - \beta.$ (1)

Next, we consider two-stage designs. Recall Simon's designs are represented by configuration (r_1, n_1, r, n), where the trial will stop for futility (accepting H_0) at the first stage of n_1 patients if r_1 or fewer responses are observed; otherwise, the trial continues with another n_2 patients. The H_0 is accepted at the second stage if r or fewer responses are observed out of the total $n = n_1 + n_2$ patients. The optimal design's configuration is determined such that the expected sample size under H_0 is the smallest among all feasible designs. (A design is feasible if its configuration satisfies the constraints of the Type I and Type II error rates). The minimax design's configuration is determined such that the maximum sample size is the smallest one among all feasible designs. The algorithms use the fixed design sample size as the basis to do the grid search (see Simon 1989). As commented earlier, in practice, we often encounter deviations from the configuration. We seek more flexibility by following the general framework given in Li et al. (2002) and Shih, Li, and Wang (2016) as follows.

Still, N is the maximal sample size from the fixed size design. Consider an interim analysis to be conducted when n_1 patients completed the study. Let $t = \frac{n_1}{N}$, $X_{n_1} = \sum_{i=1}^{n_1} X_i$ and $X_N = \sum_{i=1}^{N} X_i$ be the number of responses out of n_1 and N patients, respectively. Under the alternative hypothesis, the conditional power (CP_a) can be expressed as

$$\begin{aligned} CP_a &= P(X_N \ge R \mid X_{n_1} = x, p = p_1) \\ &= P(X_N - X_{n_1} \ge R - x \mid p = p_1) \\ &= 1 - P(X_N - X_{n_1} \le R - x - 1 \mid p = p_1) \\ &= 1 - B(R - x - 1, N - n_1, p_1). \end{aligned} \quad (2)$$

The conditional Type-I error rate is $CP_0 = 1 - B(R - x - 1, N - n_1, p_0)$. Under the current trend, $\hat{p} = \frac{x_{n_1}}{n_1}$, the conditional power is simply

$$\begin{aligned} CP_c &= P(X_N \ge R \mid X_{n_1} = x, p = \hat{p}) \\ &= 1 - B(R - x - 1, N - n_1, \hat{p}). \end{aligned} \quad (3)$$

Other choices of p such as $\overline{p} = \frac{p_0 + p_1}{2}$ are also possible.

Notice that different from Simon's designs, we do not consider when it is an "optimal" time $t = \frac{n_1}{N}$ for the interim analysis. In fact, we may plan but allow changing the time, or even consider "anytime" to perform a DDM with a radar-like display of different regions describing the status of the trial in terms of conditional powers, as laid out in Xie et al. (2021); see more discussion in the following sections.

3. Trial Monitoring

Simon's two-stage designs only consider early stop for futility, when the number of responses is less than a critical value at the first stage. A possible extension is to allow for early stop for overwhelming efficacy (Fleming 1982) as well, for which an alpha spending strategy needs to be factored in. If a total of R responses is necessary to reject the null hypothesis at the final analysis, as obtained from the system of equations (1), then to control the overall Type I error we have:

$$P(X_{n_1} \ge m \mid p_0) + P(r_1 \le X_{n_1} < m, X_N \ge R \mid p_0) \le \alpha \quad (4)$$

where m is the threshold number of responses needed for early termination of the trial for overwhelming efficacy at the interim analysis. Let $\alpha_1 = P(X_{n_1} \ge m \mid p_0) = 1 - B(m - 1, n_1, p_0)$ be the alpha spent at the first/interim stage. Then (4) is

$$\begin{aligned} &\alpha \ge \alpha_1 + P(X_{n_1} + X_{N-n_1} \ge R \mid r_1 \le X_{n_1} < m)\, P(r_1 \le X_{n_1} < m) \\ &= \alpha_1 + \sum_{y=r_1}^{m-1} P(X_{N-n_1} \ge R - y \mid X_{n_1} = y)\, P(X_{n_1} = y) \\ &= \alpha_1 + \sum_{y=r_1}^{m-1} \left(1 - B(R - y - 1, N - n_1, p_0)\right) b(y, n_1, p_0). \end{aligned}$$

$$\text{Thus, } \sum_{y=r_1}^{m-1} \left(1 - B(R - y - 1, N - n_1, p_0)\right) b(y, n_1, p_0) \le \alpha - \alpha_1. \quad (5)$$

If α_1 and m are given, then the timing (n_1) of the interim analysis is also fixed. If the timing (n_1) and α_1 are pre-specified, then the critical value of m can be obtained. Notice that, due to the discreteness of the binomial distribution, we often need to find an appropriate level of $\alpha_1 = 1-B\left(m-1, n_1, p_0\right)$ when planning the interim analysis with given (m, n_1). For Simon's design, there is no early stop for efficacy, thus, $\alpha_1 = 0, m > n_1$. Furthermore, since (N, R) is calculated for the fixed sample size design from (1), we need to adjust R to $R'(\geq R)$ by (5) once α_1 is settled. An example is given in Section 6 to illustrate these points.

In (4) and (5) we also include r_1 as the futility bound for X_{n_1}. Setting $r_1 = 0$ means that the futility is not used in "buying" the alpha level, since futility, unlike efficacy, is usually nonbinding, that is, it could be over-ruled by the sponsor. In contrast, the futility rule at the interim stage in Simon's designs is a design configuration and must be reinforced.

4. Monitoring Regions

Denote CP_a or CP_c as CP_*, where $* = a$ or c. Shih, Li, and Wang (2016) followed Li et al. (2002) and defined three regions as follows:

- $CP_* \geq q_1$: "favorable region"
- $q_2 \leq CP_* < q_1$: "hopeful region" (6)
- $CP_* < q_2$: "unfavorable region."

Alternatively, via formulas (2) and (3) we can easily convert the above regions in term of CP_* to equivalent regions in term of number of responses X_{n_1}:

- If $X_{n_1} \geq x_u$, then $CP_* \geq q_1$: "favorable region"
- If $x_l \leq X_{n_1} < x_u$, then $q_2 \leq CP_* < q_1$: "hopeful region" (7)
- If $X_{n_1} < x_l$, then $CP_* < q_2$: "unfavorable region.

Specification (6) is more interpretable, while (7) is easier for execution. When the interim result falls in the favorable region, the trial is overwhelmingly positive and may be terminated early. This is linked to $\alpha_1 = 1 - B\left(m-1, n_1, p_0\right)$, considered previously under the null hypothesis. Thus, $x_u = m$ for given n_1. When the interim result falls in the unfavorable region, the trial is hopeless and may be terminated early due to futility. Thus, $x_l = r_1$ for given n_1. When the interim result is in-between the two regions, the trial is hopeful and should continue. When continuing the trial, we may either keep the original sample size N or increase the sample size beyond N to enhance the power, depending on the observed response rate at interim; see more discussions in the following section.

We note two additional comments. Lee and Liu (2008) considered the use of Bayesian predictive probability, which is computationally more intensive than the more popular non-Bayesian conditional probability approach we considered here. Chen, DeMets, and Lan (2004) and Mehta and Pocock (2011) found a sub-region called "promising zone" in the "hopeful region," where increasing sample size would not alter the original critical value of the final z-test. In the next section, we shall see that our new critical value R^* here would be different from R' when sample size changes from N to N^*.

5. Sample Size Re-estimation

When the interim result falls into the "hopeful region," SSR may be conducted to increase the sample size to enhance the probability of rejecting H_0 at the final analysis. In the z-test case, Chen, DeMets, and Lan (2004) and Mehta and Pocock (2011) termed "promising zone" of the sub-region of "hopeful region" (Chen et al.'s proposal being conditional power under the current trend > 0.5) where increasing sample size is performed without changing the final critical value of the Wald z-test. However, their rule is not applicable for the binomial case where the variance and mean are not distinct parameters. A more general SSR strategy in the hopeful region that allows conditional power lower than 0.5 and uses a revised critical value for the final test is as follows.

To preserve the Type-I error rate, a sufficient condition for the new rejection boundary adjusted to R^* is that,

$$\begin{aligned} &CP_0\left(N^*, R^*\right) \\ &= P\left(X_{N^*} \geq R^* \text{ out of } N^* \text{ subjects} | X_{n_1} = x \text{ out of } n_1 \text{ subjects}, p_0\right) \\ &\leq CP_0 = P\left(X_N \geq R' \text{ out of } N \text{ subjects} | X_{n_1} = x \text{ out of } n_1 \text{ subjects}, p_0\right), \end{aligned} \tag{8}$$

which follows the conditional error rate principle by Muller and Schafer (2004).

Moreover, the conditional power based on X_{n_1}, N^*, and R^* satisfies a desired level. That is,

$$CP_*(N^*, R^*) \geq 1 - \beta' \tag{9}$$

where $CP_*(N^*, R^*) = P\big(X_{N^*} \geq R^* \text{ out of } N^* \text{ subjects} | X_{n_1} = x \text{ out of } n_1 \text{ subjects}, p\big)$

$$\begin{aligned} &= P\left(X_{N^*} - X_{n_1} \geq R^* - x \middle| N^* - n_1, p\right) \\ &= 1 - P\left(X_{N^*} - X_{n_1} < R^* - x \middle| N^* - n_1, p\right) \\ &= 1 - B(R^* - x - 1, N^* - n_1, p) \end{aligned}$$

with $p = p_1$ (for $CP_a(N^*, R^*)$) or $p = \hat{p}$ (for $CP_c(N^*, R^*)$). By (8) and (9) we solve for (N^*, R^*).

The general search algorithm is in the Appendix. Note that all solutions that satisfy (8) and (9) are "feasible solutions." The algorithm basically provides an "optimal" (N^*, R^*) in the sense that N^* is the smallest among all feasible solutions. In practice though, there may be other considerations the trialist needs to incorporate such as a minimum sample size for safety or pharmacokinetics evaluations. Sometimes a larger N^* in the set of feasible solutions may be necessary. On the other hand, there is always a budgetary limit to cap the largest N^* the trialist can practically afford. In our experiences, for adaptive designs, flexibility is a more desirable property than some optimality, especially for phase 2 trials' design and conduct.

6. Numerical Example

Consider the following hypothesis test with binary data:

$$\begin{aligned} H_0 &: p = 0.5 \\ H_a &: p = 0.8. \end{aligned}$$

To control $\alpha = 0.025$ and achieve a power of $1 - \beta = 80\%$, using (1) for a fixed size design, $N = 23$ patients are needed. We would

claim the response rate of this new therapy greater than 0.5 if at least 17 responses (R = 17) are observed out of the 23 patients.

The configuration of Simon's optimal design is $(r_1, n_1, r, n) = (4, 7, 16, 24)$ and $(r_1, n_1, r, n) = (5, 10, 14, 20)$ for the minimax design. Suppose, however, despite the plan, an interim analysis is conducted when 11 patients completed the study. What do we do then? Suppose the investigator has an interest to stop the trial early if the response rate is overwhelmingly high among the 11 patients, how can we include this option in the design? If the investigator is willing to expand the study beyond 23 patients when the interim response rate looks "hopeful," how should this consideration be flexibly written in the protocol or the data analysis plan?

Let us start from the setting obtained from the fixed-size design ($N = 23, R = 17; \alpha = 0.025, 1 - \beta = 80\%$). Then, first to illustrate an alpha-spending when data from $n_1 = 11$ patients are available to perform an interim analysis, suppose we want to spend a small amount of α_1 for early stop for overwhelmingly strong efficacy and $r_1 = 0$ for not to "buy back" any nonbinding futility. Because $\alpha_1 = 1 - B\left(m - 1, n_1, p_0\right) = 1 - B\,(9, 11, 0.5) = 0.00586$ (about right), and $1 - B\,(8, 11, 0.5) = 0.03271$ (too high), we would reasonably choose m = 10 for $\alpha_1 = 0.00586$ as the criterion for "overwhelming" early efficacy. With this planned interim analysis and α_1 settled, we adjust the critical value R to $R'(\geq R)$. By (5), solving for R' in $\sum_{y=0}^{9}(1 - B\left(R' - y - 1, 23 - 11, 0.5\right))b(y, 11, 0.5) \leq 0.025 - 0.00586 = 0.01914$ leads to $R' = R = 17$ in this case; the resulting left-hand side value of the above inequality is 0.014966. Note that in this numerical example, the exact Type I error rate is $1 - B\left(R - 1, N, p_0\right) = 1 - B\,(17 - 1, 23, 0.5) = 0.01734$, which gives us some room to allow for nonzero but small $\alpha_1 = 0.00586$ with the same boundary $R' = R = 17$. In general, $R' \geq R$, which is used on the right-hand side of (8).

Next, suppose out of the 11 patients, 8 responses are observed. Being in the "hopeful" region (8 < m, where m=10), we calculate the conditional power under the current trend,

$$\begin{aligned} CP_c &= P\left(X_N \geq 17 \middle| X_{n_1} = 8, p = \frac{8}{11} = 0.727\right) \\ &= 1 - P\left(X_N - X_n \leq 8 \middle| N - n_1 = 12, p = \frac{8}{11}\right) \\ &= 0.58. \end{aligned}$$

To increase the sample size to achieve at least 90% conditional power under the current trend for the new design, we apply (8) and (9), respectively:

- $1 - B\left(R^* - 8 - 1, N^* - 11, p_0 = 0.5\right) \leq 1 - B(17 - 8 - 1, 23 - 11, p_0 = 0.5)$;
- $1 - B\left(R^* - 8 - 1, N^* - 11, \frac{x}{n_1} = \frac{8}{11} = 0.727\right) \geq 90\%$.

Solving for N^* and R^*, we get $N^* = 47$ and $R^* = 31$. As a reference, we may optimistically assume $p = 0.8$ under the alterative hypothesis to do the calculations and obtain $N^* = 31$ and $R^* = 22$. The 90% *CP* is based on the guidance given by Shih, Li, and Wang (2016) for the interim analysis information fraction here $t = 11/23 = 0.478$. However, if for a budgetary reason, a lower $CP = 80\%$ is desired, then we get $(N^*, R^*) = (38, 26)$ under the current trend (response rate of 0.727), or (26, 19) using the response rate 0.8 of the alternative hypothesis.

From the illustration, we summarize the proposed modified Simon's design (as flexible two-stage design) as follows:

1. Consider the fixed-size design for the maximal sample size and rejection boundary (N,R) first. This maximal sample size helps planning the study budget as well as statistical power. Keep in mind that a two-stage flexible design will be developed next, with an interim analysis to provide opportunity to revise the sample size and to adjust the final rejection boundary.
2. Sketch tentatively, a Simon's two-stage *type* of design as a starting point in the protocol, perhaps with a very small α_1 spending if willing, but not allow the early futility to "buy" (gain) the final stage alpha-level, that is, nonbinding futility, meaning no need to follow (n_1, r_1) exactly and not to make r_1 affect the final test rule. With planned (n_1, m) the α_1 level needs to be discussed in the protocol about its approximate magnitude according to what (n_1, m) may turn out to be; the magnitude of α_1 is approximate because of the discreteness nature of the binomial distribution as well as the yet-realized and subject-to-change n_1 (see the next point). In other words, the three regions are approximately pre-specified, with solid boundary between the "favorable" and "hopeful" regions and soft boundary between the "hopeful" and "unfavorable" regions. Once α_1 is ensured, adjust the rejection boundary R to $R'(\geq R)$. R' is used with the original N if no SSR is done, or is updated to R^* with N^* via (8) and (9) when considering SSR.
3. Plan n_1, but be flexible; only when n_1 is realized at the interim analysis, we observe the interim response rate to see where it is in terms of the pre-specified three regions.
4. We may choose a desirable future $CP_*(N^*, R^*) \geq 1 - \beta'$ level ((8) and (9)) for the new sample size and critical value to proceed. This *CP* level does not need to be specified in the protocol, only to be decided prior to the interim analysis database lock.
5. The new (N^*, R^*) will be documented in the update of the data analysis plan for continuing the trial. At the final analysis, hopefully there is no major deviation from (N^*, R^*); otherwise, (8) and (9) need to be re-evaluated for verification or adjustment.

7. Extension to Basket Trials with Dynamic Data Monitoring (DDM)

The advantage in the flexibility of the proposed two-stage adaptive design for phase IIA single arm trials beyond the rigid original Simon's two-stgae designs is obvious, as illustrated above. The timing of the interim analysis (n_1) is not as rigid as in the Simon's design. It usually occurs near half the initially planned sample size ($= N/2$). Recently, the idea of basket trials has been practiced widely to include several phase IIA studies as the expansion part of the so-called seamless Phase I/IIA basket trials. See, for example, NCT04913220 and NCT04761198 in the reference. A characteristic of the basket trial design is that the several phase IIA studies for different tumor types are on naturally different enrollment schedules but are monitored simultaneously. Hence, the DDM system, advocated in Xie et al.

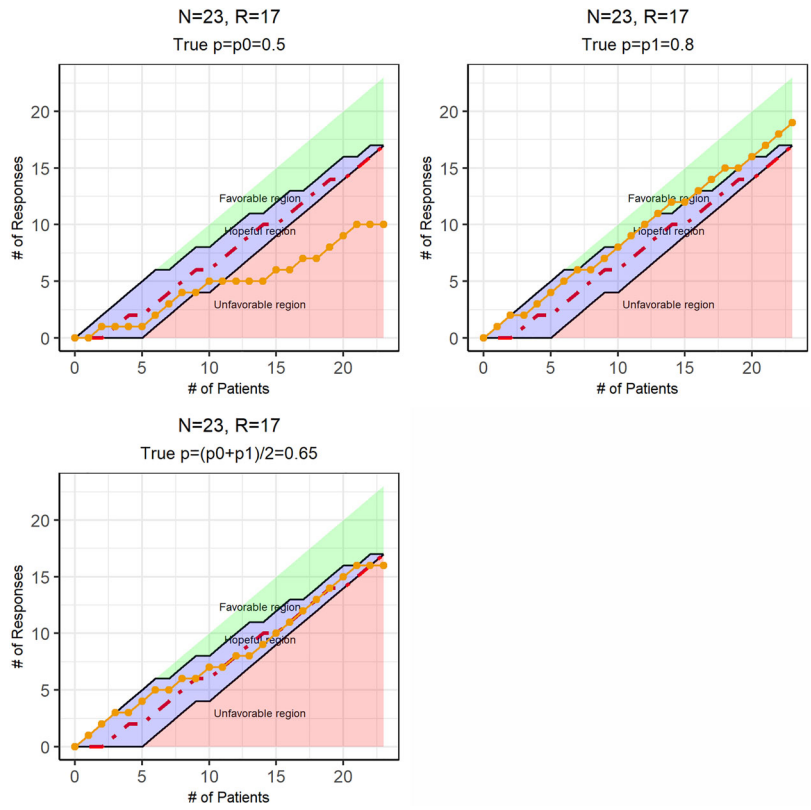

Figure 1. Dynamic Data Monitoring (DDM) system radar screen displays. Use the conditional power under the alternative hypothesis $0.05 \leq CP_a < 0.90$ to define the hopeful (and other two) regions. The red dotted line is for $CP_a = 0.5$. Illustrations with data generated (the orange dotted line) when the true response rates are p_0, p_1, and $\bar{p} = \frac{p_0+p_1}{2}$, respectively, as indicated in the figure title.

(2021), is very useful for this situation. DDM uses modern data management technologies so that "anytime" monitoring may be carried out with a "radar screen" system for each possible n_1 (for different tumor types) as patients enroll into the trial and response rate is updated anytime when data is ready. Usually, the main monitoring goal of a basket trial is to prune the branches of ineffective tumor types and grow the branches of the effective tumor types, which correspond to futility analyses and sample size re-estimations, respectively, as we alluded to in describing features of our proposed modified Simon's design. Of course, the "radar screen" can be the same or be displayed in different panels, depending on whether those tumor types are tested with the same or different hypotheses regarding p_0 versus p_1, or monitored regarding the same or different criteria of conditional powers.

Using the conditional power formulas (2) and (3), the DDM system radar screen display is shown in the following illustrations for the numerical example in Section 6.

In the first set of illustration, we use the conditional power under the alternative hypothesis $0.05 \leq CP_a < 0.90$ to define the hopeful (and other two) regions. The red line in the middle of the hopeful region corresponds to $CP_a = 0.50$. Figure 1 displays a simulated trial as discussed in Section 6, where $N = 23$ and $R = 17$, with data (orange dotted line) generated when the true response rates are p_0, p_1, and $\bar{p} = \frac{p_0+p_1}{2}$, respectively.

In the second set of illustration, we use the conditional power under the current trend $0.05 \leq CP_c < 0.90$ to define the hopeful (and other two) regions. The red line in the middle of the hopeful region corresponds to $CP_c = 0.50$. Else are the same as described above (Figure 2).

Note that we only illustrate with one hypothesis of response rate corresponding to monitoring one tumor type in these figures. DDM can display multiple radar screens when monitoring different tumor types simultaneously in a basket trial.

8. Discussion

Phase IIA trials need to be designed with much flexibility than the current practice with the traditional Simon's designs. Our intention in this article is to propose an alternative, more flexible design to the popular Simon's two-stage design for phase IIA single-arm trials in oncology rather than to do a comprehensive review of the literature. We propose a modified Simon's design to enhance flexibility and suitability for dynamic data monitoring. In the framework of basket trial designs, it is often occurring that different tumor types exhibit different enrollment rates but are monitored simultaneously. Dynamic data monitoring is a natural and helpful setup. Methodologically, the exact calculation of conditional probability based on the binomial distribution is the basic feature, different from the well-established adaptive designs for late phase trials where the normal approximation is commonly applied. We limit our discussion on the design and monitoring parts in this article, in which the hypothesis testing plays the major role. Finally, the estimation inference with confidence intervals after adaptation has been discussed in the literature, including, for example, Koyama and Chen (2008),

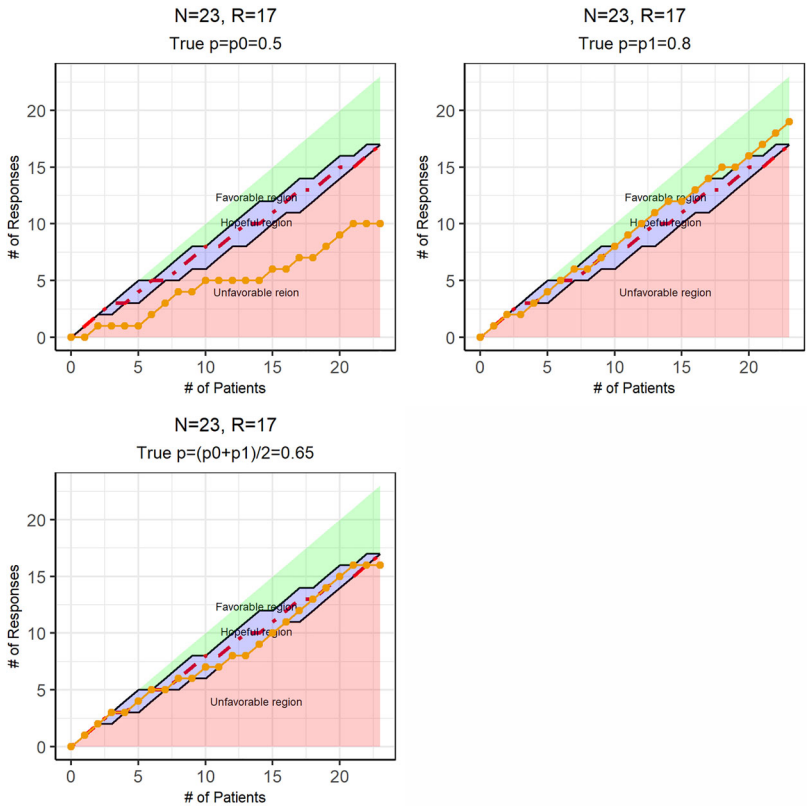

Figure 2. Dynamic Data Monitoring (DDM) system radar screen displays. Use the conditional power under the current trend $0.05 \leq CP_c < 0.90$ to define the hopeful (and other two) regions. The red dotted line is for $CP_c = 0.5$. Illustrations with data generated (the orange dotted line) when the true response rates are p_0, p_1, and $\bar{p} = \frac{p_0+p_1}{2}$, respectively, as indicated in the figure title.

Wang, Li, and Shih (2010), Luo et al. (2012), and Bowden and Trippa (2017), among others. We leave the estimation inference of this proposal to future research.

Appendix

When the interim result falls into the "hopeful region," one may consider increasing the sample size from N to N^*, the boundary should be increased from R' to R^*. Thus, $N < N^*$ and $R' \leq R^*$. Algorithm for finding (N^*, R^*): Given an upper limit (denoted as 'Nup') for the new sample size, the N^* is searched from (N+1) to *Nup*. Under each candidate value of N^*, R^* is searched from R' to N^*. For each combination of (N^*, R^*), we evaluated whether the (8) and (9) are satisfied. If satisfied, the pair belongs to the "feasible solution" set. We identify the first pair that satisfies both inequalities and output as (N^*, R^*). To request access to the software please address to Tai Xie at tomx@cims-global.com.

Acknowledgments

The work was conducted while Dr. Zhao was with CIMS Global. The authors wish to thank the referees and editors for their careful review and constructive comments and suggestions.

Disclosure Statement

The authors report there are no competing interests to declare.

ORCID

Weichung Joe Shih https://orcid.org/0000-0003-1167-6121
Tai Xie https://orcid.org/0000-0002-8408-0911

References

Banerjee, A., and Tsiatis, A. A. (2006), "Adaptive Two-Stage Designs in Phase II Clinical Trials," *Statistics in Medicine*, 25, 3382–3395. [1]

Bowden, J., and Trippa L. (2017), "Unbiased Estimation for Response Adaptive Clinical Trials," *Statistical Methods in Medical Research*, 26, 2376–2388. [6]

Chen, Y. H., DeMets, D. L., and Lan, K. K. G. (2004), "Increasing the Sample Size When the Unblinded Interim Result Is Promising," *Statistics in Medicine*, 23, 1023–1038. [3]

Englert, S., and Kieser, M. (2012a), "Adaptive Designs for Single-Arm Phase II Trials in Oncology," *Pharmaceutical Statistics*, 11, 241–249. [1]

Englert, S., and Kieser, M. (2012b), "Improving the Flexibility and Efficiency of Phase II Designs for Oncology Trials," *Biometrics*, 68, 886–892. [1]

Fleming, T. R. (1982), "One-Sample Multiple Testing Procedure for Phase II Clinical Trials," *Biometrics*, 38, 143–151. [1,2]

Green, S. J., and Dahlberg, S. (1992), "Planned versus Attained Design in Phase II Clinical Trials," *Statistics in Medicine*, 11, 853–862. [1]

Koyama, T., and Chen, T. (2008), "Proper Inference from Simon's Two-Stage Designs," *Statistics in Medicine*, 27, 3145–3154. [1,5]

Lee, J. J., and Liu, D. D. (2008), "A Predictive Probability Design for Phase II Cancer Clinical Trials," *Clinical Trials*, 5, 93–106. [3]

Leventhal, B. G., and Wittes, R. E. (1988), *Research Methods in Clinical Oncology*, New York: Raven Press. [1]

Li, G., Shih, W. J., Xie, T., and Lu, J. (2002), "A Sample Size Adjustment Procedure for Clinical Trials based on Conditional Power," *Biostatistics*, 3, 277–287. [1,2,3]

Lin, Y., and Shih, W. J. (2004), "Adaptive Two-Stage Designs for Single-Arm Phase IIA Cancer Clinical Trials," *Biometrics*, 60, 482–490. [1]

Luo, X., Li, M., Shih, W. J., and Ouyang, P. (2012), "Estimation of Treatment Effect Following a Clinical Trial with Adaptive Design," *Journal of Biopharmaceutical Statistics*, 22, 700–718. DOI: 10.1080/10543406.2012.676534. [6]

Mehta, C. R., and Pocock, S. J. (2011), "Adaptive Increase in Sample Size When Interim Results are Promising: A Practical Guide with Examples," *Statistics in Medicine*, 30, 3267–3284. [3]

Muller, H. H., and Schafer, H. (2004), "A General Statistical Principle for Changing a Design Anytime during the Course of a Trial," *Statistics in Medicine*, 23, 2497–2508. [1,3]

NCT04761198. "A Phase 1b/2 Open-Label Study of the Efficacy and Safety of Etigilimab (MPH313) Administered in Combination With Nivolumab to Subjects With Locally Advanced or Metastatic Solid Tumors (ACTIVATE)." Available at *https://www.clinicaltrials.gov/ct2/show/NCT04761198?term=NCT04761198&draw=2&rank=1*

NCT04913220. "A Phase 1/2 Non-randomized, Open-label, Multi-cohort, Multi-center Study Assessing the Clinical Benefit of SAR444245 (THOR- 707) Combined With Cemiplimab for the Treatment of Participants With Advanced Unresectable or Metastatic Skin Cancers." Available at *https://www.clinicaltrials.gov/ct2/show/NCT04913220?term=NCT04913220&draw=2&rank=1*

Proschan, M. A., and Hunsberger, S. A. (1995), "Designed Extension of Studies based on Conditional Power," *Biometrics*, 51, 1315–1324. [1]

Scharfstein, D. O., Tsiatis, A. A., and Robins, J. M. (1997), "Semiparametric Efficiency and its Implication on the Design and Analysis of Group-Sequential Studies," *Journal of the American Statistical Association*, 92, 1342–1350. [1]

Shih, W. J., Li, G., and Wang, Y. (2016), "Methods for Flexible Sample-Size Design in Clinical Trials: Likelihood, Weighted, Dual Test, and Promising Zone Approaches," *Contemporary Clinical Trials*, 47, 40–48. [2,3,4]

Simon, R. (1989), "Optimal Two-Stage Designs for Phase II Clinical Trials," *Controlled Clinical Trials*, 10, 1–10. [1,2]

Wang, Y., Li, G., and Shih, W. J. (2010), "Estimation and Confidence Intervals for Two-Stage Sample-Size-Flexible Design with LSW Likelihood Approach," *Statistics in Biosciences*, 2, 180–190. [6]

Wu, Y., and Shih, W. J. (2008), "Approaches to Handling Data When a Phase II Trial Deviates from the Pre-specified Simon's Two-Stage Design," *Statistics in Medicine*, 27, 6190–6208. [1]

Xie, T., Zhang, P., Shih, W. J., Tu, Y., and Lan, K. K. G. (2021), "Dynamic Monitoring of Ongoing Clinical Trials," *Statistics in Biopharmaceutical Research*, 14, 580–591. DOI:10.1080/19466315.2021.1880965 [1,2,5]