Published in final edited form as:

J Biopharm Stat. 2020 ; 30(1): 3–17. doi:10.1080/10543406.2019.1657439.

Response adaptive randomization procedures in seamless phase II/III clinical trials

HONGJIAN ZHU*,

Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston,

Jin Piao,

Keck School of Medicine of the University of Southern California

J. Jack Lee,

Department of Biostatistics, University of Texas MD Anderson Cancer Center

Feifang Hu,

Department of Statistics, George Washington University

Lixin Zhang

School of Mathematical Sciences, Zhejiang University

Abstract

It is desirable to work efficiently and cost effectively to evaluate new therapies in a time-sensitive and ethical manner without compromising the integrity and validity of the development process. The seamless phase II/III clinical trial has been proposed to meet this need, and its efficient, ethical and economic advantages can be strengthened by its combination with innovative response adaptive randomization (RAR) procedures. In particular, well-designed frequentist RAR procedures can target theoretically optimal allocation proportions, and there are explicit asymptotic results. However, there has been little research into seamless phase II/III clinical trials with frequentist RAR because of the difficulty in performing valid statistical inference and controlling the type I error rate. In this paper, we propose the framework for a family of frequentist RAR designs for seamless phase II/III trials, derive the asymptotic distribution of the parameter estimators using martingale processes and offer solutions to control the type I error rate. The numerical studies demonstrate our theoretical findings and the advantages of the proposed methods.

Keywords

Response adaptive randomization; Seamless clinical trials; Type I error

*Corresponding Author: ZHU, HONGJIAN, Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, 1200 Pressler Street, RAS W922, Houston, TX 77030, Hongjian.zhu@uth.tmc.edu, Tel: 713-500-9580.

1. Introduction

It is desirable to work efficiently and cost effectively to evaluate new therapies in a time-sensitive and ethical manner without compromising the integrity and validity of the development process. Adaptive seamless phase II/III designs (ASD) and response-adaptive randomization (RAR) have been proposed to meet this need (Bhatt and Mehta, 2016; Chow and Chang, 2012; Bretz et al., 2009; Hu and Rosenberger, 2006). The seamless phase II/III clinical trial can reduce the lead time between phase II and phase III trials, efficiently combine the data from the two phases for analysis, reduce the number of trials conducted in drug evaluation programs, prolong the monitoring time for patients from the phase II trial for safety reasons, and decrease the sample size. These advantages can be strengthened by its combination with innovative RAR designs that can achieve multiple competing objectives such as maximizing the power of detecting treatment effects and assigning more patients to the better treatment. However, there has been limited theoretical study of the combination of ASD and frequentist RAR designs. This hinders the development and application of this sequential procedure. In this paper, we develop approaches to adaptive seamless phase II/III clinical trials with RAR and study statistical inference for this procedure.

We briefly review the literature on ASD here. The drive to reduce development costs and shorten the time-to-market of new therapies has led to the development of the methodology for ASD. Typically, such trials have two stages separated by an interim analysis (Thall et al., 1988, 1989; Schaid et al., 1990). In the first stage, which resembles a traditional phase II trial, multiple experimental treatments or drug doses are simultaneously compared against a control. The empirically best candidates are then selected for the second stage, which corresponds to a phase III trial. The final analysis includes the patients from both stages and is performed such that the overall type I error rate is controlled. Stallard and Todd (2003) extended the approach to allow sequential analyses and early stopping in the second stage. In addition, Bretz et al. (2006) and Schmidli et al. (2006) followed the approach of Bauer and Kieser (1999) and used the closure principle (Marcus et al., 1976) and combination tests (Bauer and Köhne, 1994) to combine data from the two stages in the final hypothesis test and to control the familywise type I error rate. Liu et al. (2002) offered a solid theoretical foundation for general two-stage adaptive designs. Koenig et al. (2008) proposed the adaptive Dunnett test based on the conditional error rate (Müller and Schäfer, 2001). Most of these approaches can be classified into three categories: the group sequential approach, the combination test approach, and the adaptive Dunnett method. A detailed discussion of ASD can be found in Stallard and Todd (2011) and Korn et al. (2012). Real seamless phase II/III clinical trials include Barnes et al. (2010), Zeymer et al. (2001), Schmoll et al. (2010), and Chen et al. (2015).

Next, we briefly review the literature on RAR. RAR sequentially assigns the next patient based on the previous treatment assignments and responses to achieve greater efficiency and ethical advantages such as assigning more patients to better treatments. Eisele and Woodroffe (1995) and Hu and Zhang (2004) obtained its asymptotic properties. Zhu and Hu (2010, 2012) studied the sequential monitoring of RAR. Other important developments and discussions of the advantages of RAR can be found in Wei and Durham (1978), Rosenberger et al. (2001), Hu and Rosenberger (2003), Ivanova and Rosenberger (2000), Hu et al. (2009),

Zhu and Hu (2009), Biswas et al. (2011), Tymofyeyev et al. (2007), Hu et al. (2015), Wang et al. (2017) and Antognini and Giovagnoli (2010). Real trials using RAR designs include Rout et al. (1993), Tamura et al. (1994), and Andersen (1996).

It is desirable to propose the frequentist RAR designs for seamless trials to combine the advantages of these two strategies since the well-designed frequentist RAR can target the theoretically optimal allocation proportions to achieve efficient and ethical objectives with explicit asymptotic results. However, there have been few theoretical investigations of statistical inference for this procedure. The control of the type I error rate that can be inflated by ASD because of the dual influence of multiplicity and selection (Bauer et al., 2010) is of key importance for seamless phase II/III clinical trials. In addition, RAR designs pose new challenges due to the complicated relationships among the treatment assignments and responses. Besides, the RAR design has to target different allocation proportions in the two stages. Therefore, it is a must and challenging to investigate the statistical inference for this procedure and study how to control the type I error rate and whether the advantages of both RAR and ASD can be kept under this scenario.

Other works provide insight into the topic of this paper. Work on adaptive randomization design and ASD under the Bayesian framework includes but is not limited to Thall and Wathen (2007), Huang et al. (2009), Atkinson and Biswas (2005), Yuan et al. (2011), Brannath et al. (2009), Inoue et al. (2002), Berry (2004, 2012), and Zang and Lee (2014). In addition, Zhu (2017) derived the asymptotic normality for RAR designs with fixed adjusted allocation proportion where the target allocation proportions in the second stage are known and fixed at the beginning of the trial. The results are inspiring for the theoretical investigation of RAR designs with two stages. However, the design with fixed known adjusted allocation proportion is not practical in the real seamless Phase II/III clinical trials where the allocation proportions and the treatments under study in the second stage are usually determined at the interim look.

In this paper, we conducted comprehensive research to address the following questions. First, is there a general mathematical framework for combining ASD and RAR? Second, are there fundamental properties of this sequential procedure that can provide a theoretical foundation for further investigations? Third, is it possible to protect the type I error rate? Fourth, what are the advantages of the combination? We first described the framework for the proposed seamless phase II/III trials with RAR and derived the asymptotic distribution of the parameter estimators by approximating these estimators by martingale processes. This general framework and theoretical results can be used to investigate a variety of research questions for this sequential procedure. Then we used normal responses to demonstrate how to make use of the idea of Cohen and Sackrowitz (1989) and Sampson and Sill (2005) to construct the test statistics and to find the critical value to control the type I error rate for the proposed framework. We employed the large sample theory to overcome the difficulty that the number of patients in each treatment is random and the variances of responses are unknown and need to be estimated, which is nontrivial under the settings of RAR. The numerical studies also showed that our proposed method can control the type I error rate very well, improve the power and assign more people to the better treatment. This paper

opens a door and provides fundamental theoretical foundations for other solutions to controlling the type I error rate for this procedure.

In Section 2, we introduce the notation, describe the framework and state the main theorem. In Sections 3, we perform numerical studies to compare the proposed procedure with other randomization procedures. Conclusions are in Section 4, and technical proofs are given in the Appendix.

2. Framework, theory and implementation of the new design

2.1 Framework

We first describe the framework for the proposed family of RAR designs for seamless phase II/III trials. Suppose there are $(K + 1)$ treatments under study in the clinical trial, the planned total sample size is n , and the planned sample size for the first stage is n_1 . Then the sample size for the second stage is $n_2 = n - n_1$. Let $\mathbf{T}_i = (T_{i0}, T_{i1}, \dots, T_{iK})$ indicate the treatment assignment of the i th subject, where treatment 0 represents the control arm, $T_{ik} = 1, k = 0, 1, \dots, K$ if the i th patient is in treatment k , and $T_{ik} = 0$ otherwise. Let $N_k(m) = \sum_{i=1}^m T_{ik}, k = 0, 1, \dots, K$ be the number of patients in treatment k after m patients have been assigned, and we write $\mathbf{N}(m) = (N_0(m), N_1(m), \dots, N_K(m))$. Let $\mathbf{X}_i = (X_{i0}, X_{i1}, \dots, X_{iK}), i = 1, \dots, n$ be a random matrix of response variables, where $X_{ik}, k = 0, 1, \dots, K$ are d -dimensional random vectors, and only one element of $\mathbf{X}_i$, say X_{ik} , can be observed if $T_{ik} = 1$. Assume that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are independent and identically distributed with unknown parameters $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$. Without loss of generality, we also assume that $\boldsymbol{\theta}_k = E(\mathbf{X}_{ik})$, where $\boldsymbol{\theta}_k$ is the d -dimensional parameter vector $(\theta_{k1}, \dots, \theta_{kd})$ corresponding to treatment k ($k = 0, 1, \dots, K$), since we can transform the response and treat the transformation as responses to make the assumption hold (Gwise, et al. 2008; Hu and Zhang, 2004). After responses from the first m patients have been observed, the following modified sample means are used to estimate the parameter $\boldsymbol{\theta}_k, k = 0, 1, \dots, K$:

$$\hat{\boldsymbol{\theta}}_k(m) = \frac{\sum_{i=1}^m T_{ik} \mathbf{X}_{ik}}{N_k(m) + 1}. \quad (1)$$

We write $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$ and $\hat{\boldsymbol{\theta}}(m) = (\hat{\boldsymbol{\theta}}_0(m), \hat{\boldsymbol{\theta}}_1(m), \dots, \hat{\boldsymbol{\theta}}_K(m))$.

At the first stage, patients are sequentially assigned to all the $(K + 1)$ treatments with RAR, and one treatment, say treatment M, is chosen at the end of this stage based on certain criteria. At the second stage, the planned number of remaining patients will be sequentially assigned to treatment 0 (control arm) and treatment M with RAR. RAR designs achieve various objectives in terms of ethics and efficiency by targeting optimal allocation proportions that are usually obtained based on certain optimal criteria and depend on the unknown parameter $\boldsymbol{\theta}$ (Tymofyeyev et al. 2007; Rosenberger et al. 2001). For example, to maximize the power when comparing two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, the target optimal allocation proportions for the two treatments are $\sigma_1/(\sigma_1 + \sigma_2)$ and $\sigma_2/(\sigma_1 + \sigma_2)$, respectively, which is the so-called Neyman allocation. Let $\mathbf{v}_1 = (v_{10}, v_{11}, \dots, v_{1K}) =$

$\rho_1(\theta) = (\rho_{10}(\theta), \rho_{11}(\theta), \dots, \rho_{1K}(\theta))$ be the target allocation proportions for stage 1, and $v_2 = (v_{20}, v_{2M}) = \rho_2(\theta) = (\rho_{20}(\theta), \rho_{2M}(\theta))$ be the target allocation proportions for stage 2, where $\rho_1(z) = (\rho_{10}(z), \rho_{11}(z), \dots, \rho_{1K}(z)): \mathfrak{R}^{d \times (K+1)} \rightarrow (0,1)^{(K+1)}$ and $\rho_2(z) = (\rho_{20}(z), \rho_{2M}(z)): \mathfrak{R}^{d \times 2} \rightarrow (0,1)^2$ are vector-valued functions satisfying $\rho_j(z)\mathbf{1}' = 1, j = 1, 2$. In practice, we need to sequentially estimate the unknown parameters and update the allocation probability according to previous treatment assignments and responses so that we can target the optimal allocation proportions. Therefore, the allocation probabilities for different patients are different, but the allocation probability will converge for a variety of RAR designs (Hu and Zhang, 2004). Hu and Rosenberger (2003) studied the bias of the randomization procedure from the target allocation proportion and the variability induced by the randomization.

Specifically, we use the following family of DBCD to sequentially allocate patients in seamless phase II/III clinical trials:

- i. Assign $(K+1)m_0$ patients to all the treatments by restricted randomization procedures such as permuted block randomization (Rosenberger and Lachin, 2015) to obtain initial parameter estimates.
- ii. When the m th ($m > (K+1)m_0$) patient who is in the j th ($j = 1, 2$) stage enters the trial, compute $\hat{\theta}(m-1)$ and $\hat{\rho}_j = \rho_j(\hat{\theta}(m-1))$ based on all the previous responses and treatment assignments.
- iii. At the first stage, assign the m th patient to treatment k with probability

$$g_{1k}(N(m-1)/(m-1), \rho_1(\hat{\theta}(m-1))),$$

where $g_{1k}(s, r) = g_{1k}((s_0, s_1, \dots, s_K), (r_0, r_1, \dots, r_K)): (0,1)^{(K+1)} \times (0,1)^{(K+1)} \rightarrow (0,1)$ is the allocation probability function for treatment k and $\sum_{k=0}^K g_{1k}(s, r) = 1$. We write $\mathbf{g}_1 = (g_{10}, g_{11}, \dots, g_{1K})$. This allocation probability function takes into account both the estimated targeted allocation proportions and the current allocation proportions, and it has been well studied in the literature (e.g., Hu and Zhang, 2004; Zhu and Hu, 2010). In the numerical study, we use the following allocation probability function

$$g_{1k}(s, r) = \frac{r_k(r_k/s_k)^2}{\sum_{j=0}^K \left\{ r_j (r_j/s_j)^2 \right\}}.$$

- iv. At the end of the first stage, compare all the treatments and choose one (say treatment M) for continuation into the second stage based on certain criteria, along with the control arm.
- v. At the second stage, assign the m th patient to treatment k , $k = 0, M$ with probability

$$g_{2k}((N_0(m-1)/(m-1), N_M(m-1)/(m-1)), \rho_2(\hat{\theta}(m-1))),$$

where $g_{2k}(s, r): (0,1)^2 \times (0,1)^2 \rightarrow (0,1)$ is the allocation probability function for treatment k and $g_{20}(s, r) + g_{2M}(s, r) = 1$. We write $g_2 = (g_{20}, g_{2M})$. In the numerical study, we use the following allocation probability function

$$g_{2k}(s, r) = \frac{r_k(r_k/s_k)^2}{\{r_0(r_0/s_0)^2 + r_M(r_M/s_M)^2\}}, k = 0, M.$$

Our results are valid under the following conditions for the response, the target allocation proportions, and the allocation probability function:

- (A1) For some $\varepsilon > 0$, $E\|X_{1k}\|^{2+\varepsilon} < \infty$, $k=0, \dots, K$.
- (A2) $g_j(s, r)$, $j=1, 2$, is continuous on $\{(s, v): s\mathbf{1}' = 1\}$ and twice differentiable at (v, v) .
- (A3) $g_j(v, v) = v$, $j=1, 2$, for all $v \in (0,1)^K$.
- (A4) $\rho_j(z)$ is a continuous function and twice continuously differentiable in a small neighborhood of θ .
- (A5) $N_k(n) \rightarrow \infty$, a.s., $k=0, 1, \dots, K$.
- (A6) There is a relatively large number of patients in the second stage compared to the first stage, and $\frac{N(n)}{n} \rightarrow v_2$ a.s. and as the sample size goes to infinity. In addition, by Hu and Zhang (2004), $\frac{N(n)}{n} - v_2 = O\left(\sqrt{\frac{\log \log n}{n}}\right)$ a.s.

Theorem 1 Let $\hat{\theta}'_M = \frac{\sum_{i=n_1+1}^n T_{iM} X_{iM}}{n_{M'}+1}$, where $n'_M = \sum_{i=n_1+1}^n T_{iM}$. Under Conditions

(A1)–(A6), we have

$$\sqrt{N_k(n_1)}(\hat{\theta}_k(n_1) - \theta_k) \rightarrow N(0, \text{Var}(X_{1k})) \text{ in distribution, } k = 0, 1, \dots, K;$$

$$\sqrt{N_0(n)}(\hat{\theta}_0(n) - \theta_0) \rightarrow N(0, \text{Var}(X_{10})) \text{ in distribution,};$$

and

$$\sqrt{n_{M'}}(\hat{\theta}_{M'} - \theta_M) \rightarrow N(0, \text{Var}(X_{1M})) \text{ in distribution.}$$

This theorem offers fundamental theoretical foundations for the proposed general framework and different solutions for controlling the type I error rate might be proposed based on this result. Next, we provide a specific method.

2.2 Performing hypothesis testing for responses from normal distribution.

We have described a general framework for the proposed design and obtained the theoretical support. Next, we apply this framework when the response Y_{ik} of the k th treatment is from the normal distribution $N(\mu_k, \sigma_k^2), k = 0, 1, \dots, K$, and introduce how to construct the test statistics and find critical values to control the type I error rate. In this case, $\theta_k = (\mu_k, \sigma_k^2 + \mu_k^2)$ and $X_{ik} = (Y_{ik}, Y_{ik}^2)$.

Different methods can be used to choose treatment arms for the second stage under the proposed framework. We apply the design setting in Sampson and Sill (2005), i.e., the empirically best treatment and the control arm are continued into the second stage, and stopping for futility is not allowed. Without loss of generality, we assume the ordered outcome $Q = \{\hat{\mu}_1(n_1) > \hat{\mu}_2(n_1) > \dots > \hat{\mu}_K(n_1)\}$ has been observed after the first stage. Then treatment 1 and the control arm enter the phase III trial. This setting is useful in many practical contexts. It can apply when some regimens or formulations of a fixed dose of a compound are under study, and only one best regimen or formulation is needed. For example, when developing topical ointments, we may have several formulations of equal potency that are expected to have minimal adverse effects. We want to choose the empirically best formulation for the phase III trial. At the end of the second stage, we would like to test the hypothesis

$$H_0: \mu_0 = \mu_1 \text{ versus } H_1: \mu_1 > \mu_0. \quad (2)$$

RAR presents tremendous challenges for studying the statistical inference for this procedure. In addition to the complicated relationship between the responses and the treatment assignments, major challenges are the randomness of the patient numbers in each treatment and the variance estimators. We overcame these difficulties by martingales and large sample theory, obtaining the following conclusion under certain conditions.

Theorem 2 Consider the scenario in this section where the specific event Q has been observed and let $\hat{\mu}^* = (\hat{\mu}_2(n_1), \hat{\mu}_3(n_1), \dots, \hat{\mu}_K(n_1))$. Let

$$\bar{Z}' = \frac{N_1(n_1)\hat{\mu}_1(n_1) + n_1\hat{\mu}_1}{N_1(n)},$$

$$W' = \frac{N_1(n)N_0(n)(\bar{Z}' - \hat{\mu}_0(n))}{N_1(n)\hat{\sigma}_0^2(n) + N_0(n)\hat{\sigma}_1^2(n)},$$

$$T' = \frac{N_1(n)\hat{\sigma}_0^2(n)\bar{Z}' + N_0(n)\hat{\sigma}_1^2(n)\hat{\mu}_0(n)}{N_1(n)\hat{\sigma}_0^2(n) + N_0(n)\hat{\sigma}_1^2(n)},$$

$$\bar{Z} = \frac{n_1 v_{11} \hat{\mu}_1(n_1) + (nv_{21} - n_1 v_{11}) \hat{\mu}_1'}{nv_{21}},$$

$$W = \frac{nv_{21} nv_{20} (\bar{Z} - \hat{\mu}_0(n))}{nv_{21} \sigma_0^2 + nv_{20} \sigma_1^2},$$

$$T = \frac{nv_{21} \sigma_0^2 \bar{Z} + nv_{20} \sigma_1^2 \hat{\mu}_0(n)}{nv_{21} \sigma_0^2 + nv_{20} \sigma_1^2}.$$

Then, conditioning on the event Q , the conditional cumulative distribution function of W' given T' and $\hat{\mu}^*$ is asymptotically the conditional cumulative distribution function of W given T and $\hat{\mu}^*$ whose probability density function under H_0 is

$$f(W|T, \hat{\mu}^*) = C_N \exp \left[-\frac{1}{2} W^2 / \left(\frac{nv_{21} nv_{20}}{nv_{21} \sigma_0^2 + nv_{20} \sigma_1^2} \right) \right] \Phi \left[\frac{\sqrt{n_1 v_{11} nv_{21}} \left(T + \frac{\sigma_1^2}{nv_{21}} W - \hat{\mu}_2(n_1) \right)}{\sqrt{(n - n_1) v_{21} \sigma_1}} \right]$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and C is the normalization constant, which involves integrating W over the real line. Therefore, the following two steps based on the observed data can be used to test the hypothesis (2):

- Step 1: We use the observed data to obtain C by integrating the kernel of the above density function over the real line.
- Step 2: We calculate the p-value using the observed data and the derived probability density function from Step 1, and we perform the hypothesis testing and control the type I error rate by comparing the p-value and the significance level.

Theorem 2 offers specific solutions and procedures to control the type I error rate for the proposed design.

3. Numerical Studies

We have derived the large-sample properties and proposed the solution to control the type I error rate based on these properties for this procedure. Next, we study the procedure's finite-sample properties. We first compared our design targeting Neyman allocation (DBCD), the seamless design proposed by Sampson and Sill (2005) (SS), and the seamless designs discussed in Bretz et al. (2006) that used the closure principle (Marcus et al., 1976), combination tests (Bauer and Köhne, 1994) with either the Simes test (Simes, 1986) or the

Dunnett test (Dunnett, 1955) when two experimental treatments and one control arm are under study. Then we studied the performance of our design when more treatments are involved. Take the Indacaterol to Help Achieve New COPD Treatment Excellence (INHANCE) trial (Barnes et al. 2010) as motivation example where there were four experimental treatments and several control arms. To follow the proposed framework of this paper, we revised the scenario in Barnes et al. (2010) and compared the above different designs when four experimental treatments and one control arm were under study and the best performing experimental treatment and the control arm were selected to enter the second stage. Finally, we considered the binary responses and performed the exploratory study using the normal approximation to the Binomial distribution to expand the application of our design and investigate its ability to target different allocation proportions and achieve different objectives.

In Tables 1–4, we consider responses from normal distributions. Suppose 500 patients sequentially enter the trial. There are $(K + 1)$ treatments with responses following normal distributions $N(\mu_k, \sigma_k^2), k = 0, \dots, K, K = 2, 4$, under study at the first stage, and treatment 0 is the control arm. Our objective is to maximize the power. At the first stage, the target allocation proportion is $\rho_1 = (\rho_{10}, \dots, \rho_{1K})$, where $\rho_{1k} = \sigma_k / \sum_{j=0}^K \sigma_j, k = 0, \dots, K$. At the end of the first stage, the treatment arm with the larger mean response is selected to enter the second stage (assuming that a larger value is better). Then we use the following Neyman allocation to maximize the power if the experimental treatment arm M is retained for the second stage:

$$\rho_2 = (\rho_{20}, \rho_{2M}) = \left(\frac{\sigma_0}{\sigma_0 + \sigma_M}, \frac{\sigma_M}{\sigma_0 + \sigma_M} \right).$$

At the end of the trial, the chosen treatment M is compared with the control arm, and the following hypothesis is tested:

$$H_0: \mu_0 = \mu_M \text{ versus } H_1: \mu_M > \mu_0.$$

All the results are based on 10,000 replications, and the nominal type I error rate for the above one-sided test is 0.05.

In Tables 1 and 2, we focus on the type I error rate and the parameter estimators for studies with three treatments and five treatments, respectively. We give results for different proportions ($t = n_1/n$) of the sample size of the first stage when the responses of all the treatments follow $N(0,1)$. The type I error rate can be well controlled by our methods and the designs using the Dunnett test tend to return a slightly higher type I error rate. In addition, our design can estimate the parameter μ_0 very accurately, and the parameter estimates for the selected treatment arm are consistently larger than 0 as expected due to the treatment selection at the interim look. As the proportion of the sample size of the first stage increases, the bias of the parameter estimates for the selected treatment arm becomes greater. This paper focused on hypothesis testing instead of estimation.

In Tables 3 and 4, we focus on the power and report how many times (“Betterwin” in the Tables) out of 10,000 replications the best experimental treatment is selected to be continued at the end of the first stage. Our design resulted in a higher power than other designs including those using the Dunnett test that have relatively higher type I error rate. Our design also performed better than other designs regarding detecting the best treatment, which is of interest to the clinical trial practitioners and represents the convergence rate of our design.

In Tables 5 and 6, we used the normal approximation to the Binomial distribution to expand the application of our design in clinical trials with binary responses. Suppose 500 patients sequentially enter the trial. There are three treatments with responses following Bernoulli distribution with success rate $p_k = 0, 1, 2$, under study at the first stage, and treatment 0 is the control arm. Our objective is to assign more people to the better treatment. At the first stage, the target allocation proportion is $\rho_1 = (p_0, p_1, p_2)$. At the end of the first stage, the treatment arm with the observed larger success rate is selected to enter the second stage. Then we use the following Urn allocation to implement DBCD if treatment $M, M = 1, 2$, is retained for the second stage:

$$\rho_2 = (\rho_{20}, \rho_{2M}) = \left(\frac{1 - p_M}{(1 - p_0) + (1 - p_M)}, \frac{1 - p_0}{(1 - p_0) + (1 - p_M)} \right).$$

At the end of the trial the chosen treatment M is compared with the control arm, and the following hypothesis is tested:

$$H_0: p_0 = p_M \text{ versus } H_1: p_M > p_0.$$

All the results are based on 10,000 replications, and the nominal type I error rate for the above one-sided test is 0.05.

In Table 5, we focus on the type I error rate and the parameter estimators for different proportions (t) of the sample size of the first stage and different success rates ($p_0 = p_1 = p_2 = p$). We found similar conclusions as in Tables 1 and 2. The type I error rate can be well controlled by our methods; the designs using the Dunnett test returned a higher type I error rate; our design can estimate the parameter p_0 very accurately; and the parameter estimates for the selected treatment arm were consistently larger than the estimands as expected due to the treatment selection at the interim look.

In Table 6, we focus on the actual allocation proportions as well as the power and the number of replications out of 10,000 the best experimental treatment is selected to be continued at the end of the first stage. Because a treatment arm has been dropped, the allocation proportions we reported $(\hat{\rho}_0, \hat{\rho}_M)$ were calculated based on final total sample sizes of the control arm and the selected treatment arm. Our design assigned considerably more patients to the better treatment without losing much power, which is not only an ethical advantage but also an attractive feature for recruiting clinical trial participants. In addition, our design also performed better than other designs regarding detecting the best treatment.

4. Conclusion

ASD uses the responses to determine which treatment will be continued to the phase III trial, and even to determine whether to stop or continue the trial. RAR uses the previous responses and treatment assignments to sequentially estimate the parameters and update the allocation probability for the next patient. It is conceptually difficult to combine these two types of adaptive designs because (1) all the responses and treatment assignments are correlated with each other in a complicated manner; (2) the number of patients assigned to each treatment is a random variable; and (3) the data used in the treatment selection are also used for inference at the end of the trial. The main contribution of this paper is to propose a family of frequentist RAR designs which can be practically applied in seamless trials with specific methods to control the type I error rate and explicit asymptotic properties.

There are different directions to continue the research for applying RAR designs in seamless trials.

First, hypothesis testing with type I error control is the primary focus of a seamless phase II/III trial in practice, with estimation being an essential but secondary target (Cohen and Sackrowitz, 1989; Troendle and Yu, 1999; Posch et al., 2005; Stallard et al., 2008; Bowden and Glimm, 2008, 2014; Shen, 2001; Stallard and Todd, 2005; Koopmeiners et al., 2012; Kimani et al., 2013; Robertson et al., 2016). This paper focused on hypothesis testing and investigated the estimation in the numerical studies. The theoretical investigation of the bias and variance of the parameter estimators following the proposed design is worth studying in the future.

Second, RAR achieves a variety of objectives by targeting different allocation proportions. This paper studies Neyman allocation and Urn allocation. Other optimal allocations can be used to implement our design, especially when the responses are binary. For example, the optimal allocation proposed by Rosenberger et al. (2001) can be used to minimize the total failure number while fixing the power. Theoretically speaking, our design can perform better regarding different criteria by targeting different optimal allocations. However, the amount of gain may vary. Based on the numerical studies of this paper, when our design was targeting the Neyman allocation for normal responses, we improved the power, but the improvement was not significant; when our design was targeting the urn allocation for binary responses, we assigned significantly more patients to the better treatment. In practice, the choice of the allocation proportion will depend on the practical need.

Third, we can allow stopping for futility after the interim look and require the response of the best performing treatment at the end of the first stage exceed a threshold value to claim efficacy and to be continued for the second stage. In this case,

$$Q' = \{\hat{\mu}_1(n_1) > \hat{\mu}_2(n_1) > \dots > \hat{\mu}_K(n_1), \hat{\mu}_1(n_1) > a\},$$

where a is the threshold value. Formally, we need to calculate the probability of stopping for futility based on the joint distribution of the parameter estimators at the end of the first stage, and simply adjust the above method by this probability, though no type I error will be spent

due to stopping for futility. It is worth noting that these criteria for choosing the empirically best treatment for the second stage may fail to choose the truly best one. However, one major point of seamless trials is to expedite drug development. Therefore, the “near best” treatment that is safe and effective may be a better choice to benefit the population. An alternative is to choose multiple treatments such as several adjacent doses at the interim look to enter the second stage, especially when several treatments have similar efficacy. For example, in the INHANCE trial (Barnes et al. 2010), the minimum clinically important difference and the difference between the active control and the placebo are used to calculate the threshold value for choosing two adjacent doses.

Fourth, we could study other adaptive randomization designs in seamless trials. In this paper, we employed the doubly adaptive biased coin design proposed by Eisele and Woodroffe (1995) and Hu and Zhang (2004). Another important family of RAR is based on urn models, which includes the randomized play-the-winner rule (Wei and Durham, 1978), sequential estimation-adjusted urn models (Zhang et al., 2006), etc. In addition, as the development of personalized medicine, it is worth studying incorporating covariates into clinical trials. So seamless trials with the covariate-adjusted response-adaptive randomization design (Zhang et al., 2007) is also an important direction.

Fifth, we derived the asymptotic results for the proposed design in seamless phase II/III trials, based on which other specific methods could be explored to control the type I error rate. We leave these for future research.

Acknowledgements:

Special thanks go to anonymous referees, the associate editor and the editor for the constructive comments, which led to a much-improved version of the paper. J. Jack Lee was supported in part by grant CA016672 from the National Cancer Institute. Feifang Hu was supported in part by grant DMS-1612970 from the National Science Foundation and grant 11731011 from the National Natural Science Foundation of China. Lixin Zhang was supported in part by grant 11731012 from the National Natural Science Foundation of China, the 973 Program (No. 2015CB352302), and the Fundamental Research Funds for the Central Universities and Zhejiang Provincial Natural Science Foundation (No: LY17A010016).

APPENDIX

Proof of Theorem 1

Following the proof of Hu and Zhang (2004), we can obtain that

$$\sqrt{N_k(n_1)}(\hat{\theta}_k(n_1) - \theta_k) \rightarrow N(0, \text{Var}(X_{1k})) \text{ in distribution, } k = 0, 1, \dots, K.$$

Next we prove the asymptotic distribution for $\hat{\theta}_0(n)$ that makes use of data from both stages,

and the asymptotic distribution for $\hat{\theta}_{M'} = \frac{\sum_{i=n_1+1}^n T_{iM} X_{iM}}{n_{M'} + 1}$ that makes use of data only

from the second stage.

Let $F_m = \sigma(T_1, \dots, T_m, X_1, \dots, X_m)$ be the σ -algebra generated by the previous m patients. Then if the n th patient is in stage $j, j = 1, 2$, we have

$$E[T_{mk}|F_{m-1}] = g_{jk}(N(m-1)/(m-1), \rho_j(\hat{\theta}(m-1))),$$

and T_m and X_m are independent under F_{m-1} .

Let $\Delta Q_{i,kl} = T_{ik}(X_{ikl} - \theta_{kl})/v_{2k}$, $i = 1, \dots, n$, $k = 1, \dots, K$, $l = 1, \dots, d$. Write

$\Delta Q_i = (\Delta Q_{i,1}, \dots, \Delta Q_{i,K}) = (\Delta Q_{i,kl}; k = 1, \dots, K, l = 1, \dots, d)$ and $Q(n) = \sum_{i=1}^n \Delta Q_i$. Then $Q(n)$ is a sequence of martingales in $\Re^{K \times d}$ and $Q(n) = O(\sqrt{n \log \log n})$ a.s. Then, we have

$$\begin{aligned} \hat{\theta}_{kl}(n) - \theta_{kl} &= \frac{1}{N_k(n) + 1} \sum_{i=1}^n T_{ik} X_{ikl} - \theta_{kl} \\ &= \frac{v_{2k}}{N_k(n) + 1} \left(\sum_{i=1}^{n_1} T_{ik} \frac{X_{ikl} - \theta_{kl}}{v_{2k}} + \sum_{i=n_1+1}^n T_{ik} \frac{X_{ikl} - \theta_{kl}}{v_{2k}} \right) \\ &= \frac{Q_{kl}(n_1) + (Q_{kl}(n) - Q_{kl}(n_1))}{n} + \left(\frac{v_{2k}}{N_k(n) + 1} - \frac{1}{n} \right) (Q_{kl}(n_1) + (Q_{kl}(n) - Q_{kl}(n_1))) \\ &= \frac{Q_{kl}(n_1) + (Q_{kl}(n) - Q_{kl}(n_1))}{n} \text{ a.s.} \end{aligned}$$

Therefore, $\hat{\theta}(n) - \theta = \frac{Q(n)}{n}$ a.s.

It is worth noting that the RAR design in this paper satisfies all the conditions of Hu and Zhang (2004) except that the target allocation proportion is adjusted once after the first stage. Hu and Zhang (2004) actually proved that $Q(n)$ satisfied the Lindeberg's condition; we merely need to derive the variances and covariances of the Martingale differences. When $1 \leq i \leq n_1$, we have

$$\text{Cov}(\Delta Q_{i,kl}, \Delta Q_{i,km} | F_{i-1}) = \frac{1}{v_{2k}^2} g_{1k} \left(\frac{N_k(i-1)}{i-1}, \hat{\rho}_{1k} \right) \text{Cov}(X_{ikl}, X_{ikm})$$

When $n_1 < i \leq n$, we have

$$\text{Cov}(\Delta Q_{i,kl}, \Delta Q_{i,km} | F_{i-1}) = \frac{1}{v_{2k}^2} g_{2k} \left(\frac{N_k(i-1)}{i-1}, \hat{\rho}_{2k} \right) \text{Cov}(X_{ikl}, X_{ikm})$$

Also,

$$\text{Cov}(\Delta Q_{i,kl}, \Delta Q_{i,sm} | F_{i-1}) = 0, k \neq s.$$

Because of Condition (A6), we have

$$\frac{1}{n} \left(\sum_{i=1}^{n_1} g_{1k} \left(\frac{N_k(i-1)}{i-1}, \hat{\rho}_{1k} \right) + \sum_{i=n_1+1}^n g_{2k} \left(\frac{N_k(i-1)}{i-1}, \hat{\rho}_{2k} \right) \right) \rightarrow v_{2k}.$$

Therefore, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}) \rightarrow N(0, \Sigma) \text{ in distribution,}$$

where

$$\Sigma = \text{diag} \left(\frac{1}{v_{20}} \text{Var}(X_{20}), \dots, \frac{1}{v_{2K}} \text{Var}(X_{1K}) \right).$$

Then we can derive

$$\sqrt{N_k(n)}(\hat{\boldsymbol{\theta}}_k(n) - \boldsymbol{\theta}_k) \rightarrow N(0, \text{Var}(X_{1k})) \text{ in distribution, } k = 0$$

Similarly, we can obtain

$$\sqrt{n_M}(\hat{\boldsymbol{\theta}}_{M'} - \boldsymbol{\theta}_M) \rightarrow N(0, \text{Var}(X_{1M})) \text{ in distribution.}$$

Proof of Theorem 2

From Theorem 1, we have

$$\sqrt{N_k(n_1)}(\hat{\mu}_k(n_1) - \mu_k) \sim N(0, \sigma_k^2), \quad k = 0, 1, \dots, K,$$

$$\sqrt{n_1}(\hat{\mu}_1 - \mu_1) \sim N(0, \sigma_1^2),$$

and

$$\sqrt{N_0(n)}(\hat{\mu}_0(n_0) - \mu_0) \sim N(0, \sigma_0^2).$$

Next, we first assume the standard deviations $\sigma_0, \sigma_1, \dots, \sigma_K$ are known and the number of patients in each treatment at both stages are fixed, i.e., $N(n_1) = n_1 v_1$ and $N(n) = n v_2$. Following the idea of Sampson and Sill (2005), we can derive that the conditional pdf for W given T and $\hat{\mu}^*$ under H_0 is

$$f(W|T, \hat{\mu}^*) = C_N \exp \left[-\frac{1}{2} W^2 / \left(\frac{nv_{21}nv_{20}}{nv_{21}\sigma_0^2 + nv_{20}\sigma_1^2} \right) \right] \Phi \left[\frac{\sqrt{n_1v_{11}nv_{21}} \left(T + \frac{\sigma_1^2}{nv_{21}} W - \hat{\mu}_2(n_1) \right)}{\sqrt{(n-n_1)v_{21}\sigma_1}} \right]$$

where C_N is the normalization constant which involves integrating W over the real line. We skip the lengthy calculation to save space.

From previous proof and conditions, we have $\hat{\sigma}_k(n), k = 0, 1, \dots, K$ are consistent estimator of σ_k . We also have $N(n_1) \rightarrow n_1v_1$ in Probability, and $N(n_1) \rightarrow n_1v_2$ in Probability. Let

$\mathbf{A} = (\hat{\mu}_1(n_1), \hat{\mu}_1, \hat{\mu}_0(n))$, $\mathbf{B} = (N_1(n_1), n_1, N_1(n), N_0(n))$, and $\mathbf{c} = (n_1v_{11}, nv_{21} - n_1v_{11}, nv_{21}, nv_{20})$.

Then it is easy to see that W' and T' can be written as $h_1(\mathbf{A}, \mathbf{B})$ and $h_2(\mathbf{A}, \mathbf{B})$ respectively, where h_1 and h_2 are two continuous functions. We also have $W = h_1(\mathbf{A}, \mathbf{c})$ and $T = h_2(\mathbf{A}, \mathbf{c})$. We can prove that for all continuous bounded functions l_1 and l_2 ,

$$E[l_1(h_1(\mathbf{A}, \mathbf{B}))l_2(h_2(\mathbf{A}, \mathbf{B}))] - E[l_1(h_1(\mathbf{A}, \mathbf{c}))l_2(h_2(\mathbf{A}, \mathbf{c}))] \rightarrow 0.$$

So the conditional CDF of $W' = h_1(\mathbf{A}, \mathbf{B})$ given $T' = h_2(\mathbf{A}, \mathbf{B})$ and $\hat{\mu}^*$ is asymptotically the conditional CDF of $W = h_1(\mathbf{A}, \mathbf{c})$ given $T = h_2(\mathbf{A}, \mathbf{c})$ and $\hat{\mu}^*$

A Reference

- Andersen J (1996). Clinical Trials Designs—Made to Order. *J. Biopharm. Stat*, 6, 515–522. [PubMed: 8969983]
- Antognini AB, and Giovagnoli A (2010). Compound optimal allocation for individual and collective ethics in binary clinical trials. *Biometrika*, 97(4), 935–946.
- Atkinson AC, and Biswas A (2005). Bayesian Adaptive Biased-Coin Designs for Clinical Trials with Normal Responses. *Biometrics*, 61(1), 118–125. [PubMed: 15737084]
- Barnes PJ, Pocock SJ, Magnussen H, Iqbal A, Kramer B, Higgins M, and Lawrence D (2010). Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design. *Pulmonary pharmacology and therapeutics*, 23(3), 165–171. [PubMed: 20080201]
- Bauer P, and Kieser M (1999). Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine*, 18(14), 1833–1848. [PubMed: 10407255]
- Bauer P, Koenig F, Brannath W, and Posch M (2010). Selection and bias—two hostile brothers. *Statistics in Medicine*, 29(1), 1–13. [PubMed: 19844944]
- Bauer P, and Köhne K (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics*, 1029–1041. [PubMed: 7786985]
- Berry DA (2004). Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science*, 19 (1), 175–187.
- Berry DA (2012). Adaptive clinical trials in oncology. *Nature reviews Clinical oncology*, 9(4), 199–207.
- Bhatt DL and Mehta C (2016). Adaptive designs for clinical trials. *New England Journal of Medicine*. 375(1), 65–74. [PubMed: 27406349]
- Biswas A, Mandal S, and Bhattacharya R (2011). Multi-treatment optimal response-adaptive designs for phase III clinical trials. *Journal of the Korean Statistical Society*, 40(1), 33–44.

12. Bowden J, and Glimm E (2008). Unbiased estimation of selected treatment means in two-stage trials. *Biometrical Journal*, 50(4), 515–527. [PubMed: 18663760]
13. Bowden J, and Glimm E (2014). Conditionally unbiased and near unbiased estimation of the selected treatment mean for multistage drop-the-losers trials. *Biometrical Journal*, 56(2), 332–349. [PubMed: 24353149]
14. Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, and Racine-Poon A (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine*, 28(10), 1445–1463. [PubMed: 19266565]
15. Bretz F, Koenig F, Brannath W, Glimm E, and Posch M (2009). Adaptive designs for confirmatory clinical trials. *Statistics in medicine*, 28(8), 1181–1217. [PubMed: 19206095]
16. Bretz F, Schmidli H, König F, Racine A, and Maurer W (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. *Biometrical Journal*, 48(4), 623–634. [PubMed: 16972714]
17. Chen YJ, Gesser R and Luxembourg A (2015). A seamless Phase IIB/III adaptive outcome trial: design rationale and implementation challenges. *Clinical Trials*, 12(1), 84–90. [PubMed: 25278227]
18. Chow SC and Chang M (2012). *Adaptive Design Methods in Clinical Trials*, Second Edition Chapman and Hall/CRC.
19. Cohen A, and Sackrowitz HB (1989). Two stage conditionally unbiased estimators of the selected mean. *Statistics and Probability Letters*, 8(3), 273–278.
20. Dunnett CW (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272), 1096–1121.
21. Eisele JR, and Woodroffe MB (1995). Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics*, 23, 234–254.
22. Gwise T, Hu F, and Hu J (2008). Optimal biased coins for two-arm clinical trials. *Statistics and its Interface*, 1(1), 125–135.
23. Hu F, Hu Y, Ma W, Zhang L, and Zhu H (2015). Statistical inference of adaptive randomized clinical trials for personalized medicine. *Clinical Investigation*, 5(4), 415–425.
24. Hu F, and Rosenberger WF (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association*, 98(463), 671–678.
25. Hu F and Rosenberger WF (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley and Sons, Inc., New York.
26. Hu F and Zhang L-X (2004). Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *The Annals of Statistics* 32, 268–301.
27. Hu F, Zhang LX, and He X (2009). Efficient randomized-adaptive designs. *The Annals of Statistics*, 37, 2543–2560.
28. Huang X, Ning J, Li Y, Estey E, Issa JP, and Berry DA (2009). Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Statistics in Medicine*, 28(12), 1680–1689. [PubMed: 19326367]
29. Inoue LY, Thall PF, and Berry DA (2002). Seamlessly expanding a randomized phase II trial to phase III. *Biometrics*, 58(4), 823–831. [PubMed: 12495136]
30. Ivanova A, and Rosenberger WF (2000). A comparison of urn designs for randomized clinical trials of $K > 2$ treatments. *Journal of Biopharmaceutical Statistics*, 10(1), 93–107. [PubMed: 10709804]
31. Jennison C and Turnbull BW (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC.
32. Kimani PK, Todd S, and Stallard N (2013). Conditionally unbiased estimation in phase II/III clinical trials with early stopping for futility. *Statistics in Medicine*, 32(17), 2893–2910. [PubMed: 23413228]
33. Koenig F, Brannath W, Bretz F, and Posch M (2008). Adaptive Dunnett tests for treatment selection. *Statistics in Medicine*, 27(10), 1612–1625. [PubMed: 17876763]

34. Koopmeiners JS, Feng Z, and Pepe MS (2012). Conditional estimation after a twostage diagnostic biomarker study that allows early termination for futility. *Statistics in Medicine*, 31(5), 420–435. [PubMed: 22238117]
35. Korn EL, Freidlin B, Abrams JS and Halabi S (2012). Design issues in randomized phase II/III trials. *Journal of Clinical Oncology*, 30(6), 667–671. [PubMed: 22271475]
36. Liu Q, Proschan MA, and Pledger GW (2002). A unified theory of two-stage adaptive designs. *Journal of the American Statistical Association*, 97(460), 1034–1041.
37. Marcus R, Eric P, and Gabriel KR (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3), 655–660.
38. Müller HH, and Schäfer H (2001). Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*, 57(3), 886–891. [PubMed: 11550941]
39. Posch M, Koenig F, Branson M, Brannath W, Dunger-Baldauf C, and Bauer P (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine*, 24(24), 3697–3714. [PubMed: 16320264]
40. Robertson DS, Prevost AT and Bowden J (2016). Unbiased estimation in seamless phase II/III trials with unequal treatment effect variances and hypothesisdriven selection rules. *Statistics in Medicine*, 35(22), 3907–3922. [PubMed: 27103068]
41. Rosenberger WF, and Lachin JM (2015). *Randomization in clinical trials: theory and practice*. John Wiley & Sons.
42. Rosenberger WF, Stallard N, Ivanova A, Harper CN, and Ricks ML (2001). Optimal adaptive designs for binary response trials. *Biometrics*, 57(3), 909–913. [PubMed: 11550944]
43. Rout CC, Rocke DA, Levin L, Gouws E and Reddy D (1993). A Reevaluation of the Role of Crystalloid Preload in the Prevention of Hypotension Associated with Apinal Anesthesia for Elective Cesarean Section. *Anesthesiology*, 79, 262–269. [PubMed: 8192733]
44. Sampson AR, and Sill MW (2005). Drop-the-losers design: normal case. *Biometrical Journal*, 47(3), 257–268. [PubMed: 16053251]
45. Schaid DJ, Wieand SAM, and Therneau TM (1990). Optimal two-stage screening designs for survival comparisons. *Biometrika*, 77(3), 507–513.
46. Schmidli H, Bretz F, Racine A, and Maurer W (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: applications and practical considerations. *Biometrical Journal*, 48(4), 635–643. [PubMed: 16972715]
47. Schmoll H, Cunningham D, Sobrero A, Karapetis C, Rougier P, Koski SL, Barker P, Mookerjee B, Robertson J and van Cutsem E, (2010) mFOLFOX6 + cediranib vs mFOLFOX6 + bevacizumab in previously untreated metastatic colorectal cancer (MCRC): a randomised, doubleblind, phase II/III study (HORIZON III). *Annals of Oncology*. 21,189–189 [PubMed: 19889615]
48. Shen L (2001). An improved method of evaluating drug effect in a multiple dose clinical trial. *Statistics in Medicine*, 20(13), 1913–1929. [PubMed: 11427949]
49. Simes RJ (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751–754.
50. Stallard N, and Todd S (2003). Sequential designs for phase III clinical trials incorporating treatment selection. *Statistics in medicine*, 22(5), 689–703. [PubMed: 12587100]
51. Stallard N, and Todd S (2005). Point estimates and confidence regions for sequential trials involving selection. *Journal of Statistical Planning and Inference*, 135(2), 402–419.
52. Stallard N, and Todd S (2011). Seamless phase II/III designs. *Statistical Methods in Medical Research*, 20(6), 623–634. [PubMed: 20724313]
53. Stallard N, Todd S, and Whitehead J (2008). Estimation following selection of the largest of two normal means. *Journal of Statistical Planning and Inference*, 138(6), 1629–1638.
54. Tamura RN, Faries DE, Andersen JS and Heiligenstein JH (1994). A Case Study of an Adaptive Clinical Trial in the Treatment of Out-patients with Depressive Disorder. *J. Amer. Statist. Assoc*, 89, 768–776.
55. Thall PF, Simon R, and Ellenberg SS (1988). Two-stage selection and testing designs for comparative clinical trials. *Biometrika*, 75(2), 303–310.

56. Thall PF, Simon R, and Ellenberg SS (1989). A two-stage design for choosing among several experimental treatments and a control in clinical trials. *Biometrics*, 45, 537–547. [PubMed: 2765637]
57. Thall PF, and Wathen JK (2007). Practical Bayesian adaptive randomisation in clinical trials. *European Journal of Cancer*, 43(5), 859–866. [PubMed: 17306975]
58. Troendle JF, and Yu KF (1999). Conditional estimation following a group sequential clinical trial. *Communications in Statistics-Theory and Methods*, 28(7), 1617–1634.
59. Tymofyeyev Y, Rosenberger WF, and Hu F (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*, 102(477), 224–234.
60. Wang L, Chen Y, and Zhu H (2017). Implementing optimal allocation in clinical trials with multiple endpoints. *Journal of Statistical Planning and Inference*, 182, 88–99. [PubMed: 28529406]
61. Wei LJ and Durham S (1978). The Randomized Play-the-winner Rule in Medical Trials. *J. Amer. Statist. Assoc* 73, 840–843.
62. Yuan Y, Huang X, and Liu S (2011). A Bayesian responseadaptive covariatebalanced randomization design with application to a leukemia clinical trial. *Statistics in Medicine*, 30(11), 1218–1229. [PubMed: 21432894]
63. Zang Y, and Lee JJ (2014). Adaptive clinical trial designs in oncology. *Chinese clinical oncology*, 3(4), 49.
64. Zeymer U, Suryapranata H, Monassier JP, Opolski G, Davies J, Rasmanis G, Linssen G, Tebbe U, Schröder R, Tiemann R and Machnig T (2001). The Na⁺/H⁺ exchange inhibitor eniporide as an adjunct to early reperfusion therapy for acute myocardial infarction: Results of the evaluation of the safety and cardioprotective effects of eniporide in acute myocardial infarction (ESCAMI) trial. *Journal of the American College of Cardiology*, 38(6), E1644–E1650.
65. Zhang LX, Hu F, and Cheung SH (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *The Annals of Applied Probability*, 16, 340–369.
66. Zhang L-X, Hu F, Cheung SH and Chan WS (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics*. 35, 1166–1182.
67. Zhu H, and Hu F (2009). Implementing optimal allocation for sequential continuous responses with multiple treatments. *Journal of Statistical Planning and Inference*, 139(7), 2420–2430.
68. Zhu H and Hu F (2010). Sequential monitoring of response-adaptive randomized clinical trials. *The Annals of Statistics*. 38, 2218–2241.
69. Zhu H and Hu F (2012). Interim analysis of clinical trials based on urn models. *Canadian Journal of Statistics*. 40, 550–568.
70. Zhu H (2017). Statistical inference for response adaptive randomization procedures with adjusted optimal allocation proportions. *Journal of Biopharmaceutical Statistics*, 27(5), 732–740. [PubMed: 27937121]

Table 1:

Type I error rate and parameter estimation for seamless trials with three treatments, normal responses, and sample size $n = 500$

Allocation	t	Type I error	$\hat{\mu}_0$	$\hat{\mu}_M$
DBCD	0.1	0.049	$-3.1\text{e-}4(0.06)$	0.008(0.06)
SS	0.1	0.049	$8.9\text{e-}4(0.06)$	0.009(0.06)
Dunnett	0.1	0.054	$-8.9\text{e-}4(0.06)$	0.010(0.06)
Simes	0.1	0.048	$6.7\text{e-}4(0.06)$	0.010(0.06)
DBCD	0.2	0.051	$-1.9\text{e-}4(0.07)$	0.014(0.06)
SS	0.2	0.052	$3.6\text{e-}6(0.07)$	0.013(0.06)
Dunnett	0.2	0.049	$1.3\text{e-}3(0.07)$	0.013(0.06)
Simes	0.2	0.048	$6.0\text{e-}4(0.07)$	0.014(0.06)
DBCD	0.3	0.053	$-6.9\text{e-}4(0.07)$	0.017(0.06)
SS	0.3	0.050	$1.4\text{e-}4(0.07)$	0.019(0.06)
Dunnett	0.3	0.055	$-4.9\text{e-}4(0.07)$	0.018(0.06)
Simes	0.3	0.046	$3.2\text{e-}4(0.07)$	0.017(0.06)

Table 2:

Type I error rate and parameter estimation for seamless trials with five treatments, normal responses, and sample size $n = 500$

Allocation	t	Type I error	$\hat{\mu}_0$	$\hat{\mu}_M$
DBCD	0.2	0.050	1.0e-3(0.07)	0.020(0.07)
SS	0.2	0.051	-2.3e-4(0.07)	0.020(0.07)
Dunnett	0.2	0.056	-4.8e-4(0.07)	0.021(0.07)
Simes	0.2	0.043	3.7e-4(0.07)	0.021(0.07)
DBCD	0.3	0.047	1.5e-3(0.07)	0.027(0.06)
SS	0.3	0.051	2.7e-4(0.07)	0.027(0.07)
Dunnett	0.3	0.058	-8.0e-5(0.07)	0.028(0.07)
Simes	0.3	0.047	2.8e-4(0.07)	0.027(0.07)
DBCD	0.4	0.051	-3.9e-4(0.07)	0.034(0.07)
SS	0.4	0.052	4.4e-4(0.07)	0.035(0.07)
Dunnett	0.4	0.054	-7.4e-4(0.07)	0.033(0.06)
Simes	0.4	0.042	5.2e-4(0.07)	0.034(0.07)

Table 3:

Power of different designs for seamless trials with three treatments, normal responses, and sample size $n = 500$

Allocation	t	$(\mu_0, \sigma_0, \mu_1, \sigma_1, \mu_2, \sigma_2)$	Power	$\hat{\mu}_0$	Betterwin
DBCD	0.1	(1,1,1.4,2.5,1.2,2.5)	0.615	1.00(0.09)	6007
SS	0.1	(1,1,1.4,2.5,1.2,2.5)	0.554	1.00(0.06)	5889
Dunnett	0.1	(1,1,1.4,2.5,1.2,2.5)	0.549	1.00(0.06)	5974
Simes	0.1	(1,1,1.4,2.5,1.2,2.5)	0.521	1.00(0.06)	5957
DBCD	0.2	(1,1,1.5,2.5,1.3,2.5)	0.784	1.00(0.09)	6400
SS	0.2	(1,1,1.5,2.5,1.3,2.5)	0.727	1.00(0.07)	6188
Dunnett	0.2	(1,1,1.5,2.5,1.3,2.5)	0.745	1.00(0.07)	6191
Simes	0.2	(1,1,1.5,2.5,1.3,2.5)	0.741	1.00(0.07)	6350

Table 4:

Power of different designs for seamless trials with five treatments, normal responses, and sample size $n = 500$

Allocation	t	$(\mu_0, \sigma_0, \mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, \mu_4, \sigma_4)$	Power	$\hat{\mu}_0$	Betterwin
DBCD	0.15	(1,1,1.5,2.5,1.4,2.5,1.3,2.5,1.2,2.5)	0.692	1.00(0.09)	3324
SS	0.15	(1,1,1.5,2.5,1.4,2.5,1.3,2.5,1.2,2.5)	0.645	1.00(0.07)	3332
Dunnett	0.15	(1,1,1.5,2.5,1.4,2.5,1.3,2.5,1.2,2.5)	0.657	1.00(0.07)	3283
Simes	0.15	(1,1,1.5,2.5,1.4,2.5,1.3,2.5,1.2,2.5)	0.622	1.00(0.07)	3205
DBCD	0.2	(1,1,1.6,2.5,1.5,2.5,1.4,2.5,1.3,2.5)	0.841	1.00(0.09)	3604
SS	0.2	(1,1,1.6,2.5,1.5,2.5,1.4,2.5,1.3,2.5)	0.783	1.00(0.07)	3454
Dunnett	0.2	(1,1,1.6,2.5,1.5,2.5,1.4,2.5,1.3,2.5)	0.817	1.00(0.07)	3473
Simes	0.2	(1,1,1.6,2.5,1.5,2.5,1.4,2.5,1.3,2.5)	0.772	1.00(0.07)	3490

Table 5:

Type I error rate and parameter estimation for seamless trials with three treatments, binary responses, and sample size $n = 500$

Allocation	p	t	Type I error	$\hat{p}_0$	$\hat{p}_M$
DBCD	0.5	0.1	0.051	0.499(0.03)	0.504(0.03)
SS	0.5	0.1	0.048	0.500(0.03)	0.505(0.03)
Dunnett	0.5	0.1	0.059	0.500(0.03)	0.505(0.03)
Simes	0.5	0.1	0.050	0.500(0.03)	0.505(0.03)
DBCD	0.4	0.2	0.049	0.399(0.03)	0.405(0.03)
SS	0.4	0.2	0.051	0.400(0.03)	0.407(0.03)
Dunnett	0.4	0.2	0.052	0.400(0.03)	0.407(0.03)
Simes	0.4	0.2	0.047	0.400(0.03)	0.407(0.03)
DBCD	0.6	0.3	0.050	0.599(0.03)	0.607(0.03)
SS	0.6	0.3	0.048	0.600(0.03)	0.609(0.03)
Dunnett	0.6	0.3	0.053	0.600(0.03)	0.609(0.03)
Simes	0.6	0.3	0.046	0.600(0.03)	0.609(0.03)

Table 6:Power for seamless trials with three treatments, binary responses, and sample size $n = 500$

Allocation	(p_0, p_1, p_2)	t	Power	$\hat{p}_0$	$\hat{p}_0$	$\hat{p}_M$	Betterwin
DBCD	(0.7, 0.9, 0.8)	0.15	0.972	0.695(0.04)	0.268(0.06)	0.732(0.06)	8579
SS	(0.7, 0.9, 0.8)	0.15	0.970	0.700(0.03)	0.500(0.02)	0.500(0.02)	8370
Dunnett	(0.7, 0.9, 0.8)	0.15	0.975	0.700(0.03)	0.500(0.02)	0.500(0.02)	8416
Simes	(0.7, 0.9, 0.8)	0.15	0.973	0.700(0.03)	0.500(0.02)	0.500(0.02)	8382
DBCD	(0.6, 0.8, 0.7)	0.2	0.955	0.597(0.04)	0.344(0.05)	0.656(0.05)	8525
SS	(0.6, 0.8, 0.7)	0.2	0.949	0.600(0.03)	0.500(0.02)	0.500(0.02)	8333
Dunnett	(0.6, 0.8, 0.7)	0.2	0.958	0.600(0.03)	0.500(0.02)	0.500(0.02)	8310
Simes	(0.6, 0.8, 0.7)	0.2	0.959	0.600(0.03)	0.500(0.02)	0.500(0.02)	8274