The New England Journal of Statistics in Data Science Volume 0, 1–9 (2023)
DOI: https://doi.org/10.51387/23-NEJSDS25

Seamless Clinical Trials with Doubly Adaptive Biased Coin Designs

Hongjian ZHU, Jun YU*, Dejian LAI, and Li WANG

Abstract

In addition to scientific questions, clinical trialists often explore or require other design features, such as increasing the power while controlling the type I error rate, minimizing unnecessary exposure to inferior treatments, and comparing multiple treatments in one clinical trial. We propose implementing adaptive seamless design (ASD) with response adaptive randomization (RAR) to satisfy various clinical trials' design objectives. However, the combination of ASD and RAR poses a challenge in controlling the type I error rate. In this paper, we investigated how to utilize the advantages of the two adaptive methods and control the type I error rate. We offered the theoretical foundation for this procedure. Numerical studies demonstrated that our methods could achieve efficient and ethical objectives while controlling the type I error rate.

KEYWORDS AND PHRASES: Adaptive design, Ethics, Efficiency, Response adaptive randomizations, Type I error.

1. INTRODUCTION

The significance of streamlining clinical trials has been emphasized in the Critical Path Opportunities Report and List [60]. The FDA [61] revised their guidance on seamless clinical trials and re-iterated the importance of moving towards the broadening acceptance of seamless trials. The FDA [61] outlined the need to evaluate new therapies in a time-sensitive, cost-effective and ethical manner without compromising the integrity and validity of the development process.

The seamless phase II/III clinical trial can reduce the lead time between different phases, reduce the number of trials for comparing multiple treatments, efficiently combine the data from both phases, monitor patients from the phase II trial longer for safety issues, and decrease the sample size while maintaining power. Typically, multiple experimental treatments are compared against a control in the first stage. The empirically best candidates are then selected to enter the second stage together with the control arm. The final analysis based on the patients from both stages is performed such that the overall type I error rate is controlled [44, 55, 56]. Until 2016, there have been more than 40 active, first-in-human cancer trials that are using the seamless strategy [39]. A motivating example is the Indacaterol to Help Achieve New COPD Treatment Excellence (INHANCE) trial [4], an adaptive seamless phase II/III clinical trial of inhaled indacaterol to treat chronic obstructive pulmonary disease (COPD). Other real seamless phase II/III clinical trials include [68] and [16].

In practice, hypothesis testing with type I error control is the primary focus of a seamless phase II/III trial, with estimation being an essential but secondary target [9, 10, 19, 32, 33, 38, 40, 46, 51, 52, 58]. This paper will focus on the control of type I error rate, as well as the investigation of the advantages of implementing DBCD in seamless clinical trials. The closure principle [37] has been proposed to handle the multiple testing problem; certain combination methods such as the inverse χ^2 method [5] and the weighted inverse normal method [34] have been proposed to combine data from the two stages; and different approaches such as the Simes test [47] and the Dunnett test [20] have been proposed to test the intersection of more than two hypotheses constructed for applying the closure principle. [14] and [45] made use of these methods to control the familywise type I error rate (FWER) for ASD. This paper will employ this framework since FDA and the pharmaceutical industry will readily accept it. [31, 49, 50] allowed more than one experimental treatment to continue beyond the first interim analysis and sequential analyses in the second stage. [63] proposed a multi-stage drop-the-losers design and discussed the required sample size. [36] proposed methods for any number of treatment arms, any number of stages and any number of patients per treatment per stage in such trials. [35] provided the theoretical foundation for a general family of two-stage adaptive designs. ASD with different study endpoints in the two stages has been investigated by [18, 48, 57]. We leave all these extensions for future research on our proposed procedure.

Next, we introduce RAR. Clinical trials are complex and usually have multiple objectives such as increasing the power of detecting treatment differences, and assigning more patients to better treatments. Two families of RAR have been proposed to achieve these objectives: DBCD [25, 62, 70, 71] and urn models [64, 65, 69, 72]. RAR can achieve greater efficiency and ethical advantages by skewing the allocation

*Corresponding author.

proportion based on previous treatment assignments and responses. A popular formal RAR framework contains three steps. First, the design objectives are mathematically formulated, and it is usually expressed in an optimization problem. Second, the optimal allocation proportions to achieve this objective as the solution of the optimization problem are derived. Third, a specific RAR design is implemented to target the optimal allocation proportion. [25, 71, 72] studied the asymptotic properties and sequential monitoring of RAR. [24] showed that RAR could increase efficiency in certain clinical trials. [59] explored the derivation of optimal allocation proportion. Other discussions of the advantages of RAR can be found in [2, 8, 21, 26, 27, 30, 41]. Clinical trials using RAR designs include [1, 42, 53]. This paper focuses on using DBCD as randomization in seamless clinical trials.

Therefore, it is desirable to study how to benefit from the advantages of ASD and RAR in one clinical trial. However, both ASD and RAR pose a challenge in controlling the type I error rate, which is critical in confirmatory clinical trials. ASD tends to increase the type I error rate due to multiple testing and treatment selection at the interim look. RAR introduced extra difficulties with correlated responses and treatment assignments. In this paper, we overcame these difficulties and studied its asymptotic and finite-sample properties.

In Section 2, we introduce the notation, our proposed methods, and theoretical findings. In Section 3, we offer results from numerical studies via simulations. Conclusions are in Section 4, and the proof is in the Supplementary materials.

2. SEAMLESS CLINICAL TRIALS WITH DBCD

2.1 Adaptive Seamless Design with DBCD

We first introduce the notation for DBCD with multiple treatments. Suppose $(K+1)$ treatments are under study in a clinical trial with sample size n. Let $\boldsymbol{T}_i = (T_{i0}, T_{i1}, \ldots, T_{iK})$ denote the ith patient's treatment assignment, where treatment 0 indicates the control arm, $T_{ik} = 1$, $k = 0, 1, \ldots, K$ if the ith patient is in treatment k, and $T_{ik} = 0$ otherwise. Let $\boldsymbol{N}(m) = (N_0(m), N_1(m), \ldots, N_K(m))$, where $N_k(m) = \sum_{i=1}^{m} T_{ik}$, $k = 0, 1, \ldots, K$ is the number of patients assigned to treatment k after m patients have entered the trial. Let $\boldsymbol{X}_i = (\boldsymbol{X}_{i0}, \boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iK})$, $i = 1, \ldots, n$ be a random matrix of response variables, where $\boldsymbol{X}_{ik}, k = 0, 1, \ldots, K$, are d-dimensional random vectors. Here, if the ith patient is assigned to treatment k, only $\boldsymbol{X}_{ik}$ can be observed. In other words, $\boldsymbol{X}_{ik}$ is the ith patient's response in the presence of treatment k and only observed if $T_{ik} = 1$. Therefore, the variable T_{ik} does not influence the expectation of $\boldsymbol{X}_{ik}$; it only determines if it is observed. Without loss of generality, we assume $\boldsymbol{\theta}_k = E(\boldsymbol{X}_{ik}) = (\theta_{k1}, \ldots, \theta_{kd})$, $k = 0, 1, \ldots, K$. Then the parameter estimator after responses of m patients have been observed is

$$\hat{\boldsymbol{\theta}}_k(m) = \frac{\sum_{i=1}^{m} T_{ik} \boldsymbol{X}_{ik}}{N_k(m)}. \tag{2.1}$$

Write

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$$

and

$$\hat{\boldsymbol{\theta}}(m) = (\hat{\boldsymbol{\theta}}_0(m), \hat{\boldsymbol{\theta}}_1(m), \ldots, \hat{\boldsymbol{\theta}}_K(m)).$$

RAR can achieve various objectives by targeting different allocation proportions that will be functions of unknown parameters [59]. Let $\boldsymbol{\rho}_l(\boldsymbol{\theta}) = (\rho_{l0}(\boldsymbol{\theta}), \rho_{l1}(\boldsymbol{\theta}), \ldots, \rho_{lK}(\boldsymbol{\theta}))$, $l = 1, 2$, is the target allocation proportions for stage l, where $\boldsymbol{\rho}_l(\boldsymbol{\theta}) : \Re^{d\times(K+1)} \to (0,1)^{(K+1)}$ is the vector-valued functions satisfying $\boldsymbol{\rho}_l(\boldsymbol{\theta})\mathbf{1}' = 1$. Specific examples can be seen in Section 3.

Next, we introduce the procedure to conduct a seamless phase II/III clinical trials with a family of DBCD:

(i) In the first stage, we first assign m_0 patients to each of the $K+1$ treatments by fixed design to obtain initial parameter estimates. When the mth $(m > (K+1)m_0)$ patient enters the first stage of the trial, calculate $\hat{\boldsymbol{\theta}}(m-1)$ and $\hat{\boldsymbol{\rho}}_1 = \boldsymbol{\rho}_1(\hat{\boldsymbol{\theta}}(m-1))$ based on all the previous responses and treatment assignments.

(ii) Assign the mth patient to treatment k with probability

$$g_{1k}\left(\boldsymbol{N}(m-1)/(m-1), \boldsymbol{\rho}_1(\hat{\boldsymbol{\theta}}(m-1))\right),$$

where $g_{1k}(\boldsymbol{s}, \boldsymbol{r}) = g_{1k}((s_0, s_1, \ldots, s_K), (r_0, r_1, \ldots, r_K)) : (0,1)^{(K+1)} \times (0,1)^{(K+1)} \to (0,1)$ satisfies $\sum_{k=0}^{K} g_{1k}(\boldsymbol{s}, \boldsymbol{r}) = 1$ [25]. We write $\boldsymbol{g}_1 = (g_{10}, g_{11}, \ldots, g_{1K})$. [25] proposed the following allocation probability function to the treatment k for the mth patient

$$g_{1k}(\boldsymbol{s}, \boldsymbol{r}) = \frac{r_k (r_k/s_k)^2}{\sum_{j=0}^{K} \{r_j (r_j/s_j)^2\}}, \tag{2.2}$$

where $s_k = N_k(m-1)/(m-1)$ and $r_k = \rho_{1k}(\hat{\boldsymbol{\theta}}(m-1))$.

(iii) At the end of the first stage, choose one (say treatment M) based on certain criteria to enter the second stage, along with the control arm. For example, we can choose the experimental treatment arm with the largest treatment effect to enter the second stage; we can also incorporate safety data into the criteria for choosing a treatment arm for the second stage.

(iv) Because we have only two treatment arms under study in the second stage, let $\boldsymbol{\rho}_2(\boldsymbol{\theta}) = (\rho_{20}(\boldsymbol{\theta}), \rho_{2M}(\boldsymbol{\theta}))$ be the target allocation proportions for the second stage, where $\boldsymbol{\rho}_2(\boldsymbol{\theta}) : \Re^{d\times 2} \to (0,1)^2$ is the vector-valued functions satisfying $\boldsymbol{\rho}_2(\boldsymbol{\theta})\mathbf{1}' = 1$. At the second stage, assign the mth patient to treatment $k, k = 0, M$ with probability

$$g_{2k}\left((N_0(m-1)/(m-1), N_M(m-1)/(m-1)), \boldsymbol{\rho}_2(\hat{\boldsymbol{\theta}}(m-1))\right),$$

where $g_{2k}(\boldsymbol{s},\boldsymbol{r}):(0,1)^2\times(0,1)^2\to(0,1)$ satisfies $g_{20}(\boldsymbol{s},\boldsymbol{r})+g_{2M}(\boldsymbol{s},\boldsymbol{r})=1$. We write $\boldsymbol{g}_2=(g_{20},g_{2M})$.

The above DBCD considers both the estimated targeted allocation proportions and the current allocation proportions in order to achieve different ethical and efficient objectives. A specific family of allocation probability functions will be given in Section 3. Other discussions and properties of DBCD can be seen in [71, 25].

2.2 Data Analysis Procedure

At the end of the clinical trial, one considers a general hypothesis test:

$$H_{0,M}: h(\boldsymbol{\theta}_M)=h(\boldsymbol{\theta}_0) \text{ versus } H_{1,M}: h(\boldsymbol{\theta}_M)>h(\boldsymbol{\theta}_0)$$

where $h(\boldsymbol{\theta}_j)$ is a $\Re^d\to\Re$ continuous and twice differentiable function in a small neighborhood of $\boldsymbol{\theta}_j,\ j=0,M$.

We test the above hypothesis with the combined data from the two stages and follow the closure principle [37] to control the familywise type I error rate. The closure principle rejects $H_{0,M}$ at level α if each intersection hypothesis $H_{0,I}$ with $M\in I,\ I\subseteq\{1,\ldots,K\}$, is rejected at level α, where $H_{0,I}=\cap_{k\in I}H_{0,k}$ with $H_{0,k}: h(\boldsymbol{\theta}_k)=h(\boldsymbol{\theta}_0)$. Each $H_{0,I}$ can be tested with the following inverse χ^2 method. Let $P_{1,I}$ and $P_{2,I}$ denote the p-values for $H_{0,I}$ based on the data from the first stage and the second stage, respectively. Then we reject $H_{0,I}$ if $-\log(P_{1,I}P_{2,I})>\chi^2_4(1-\alpha)/2$, where $\chi^2_4(1-\alpha)$ is the $(1-\alpha)$th quantile of the χ^2 distribution with 4 degrees of freedom. To calculate the adjusted p-values for each stage, $P_{1,I}$ and $P_{2,I}$, we use the Simes test [47] with the following test statistics for the elementary hypotheses $H_{0,k}$ in the intersection hypothesis $H_{0,I}$,

$$Z_k\left(\frac{\boldsymbol{N}(n)}{n},\hat{\boldsymbol{\theta}}(n)\right)=\frac{h\left(\hat{\boldsymbol{\theta}}_k(n)\right)-h\left(\hat{\boldsymbol{\theta}}_0(n)\right)}{\sqrt{\hat{Var}\left(h\left(\hat{\boldsymbol{\theta}}_k(n)\right)\right)+\hat{Var}\left(h\left(\hat{\boldsymbol{\theta}}_0(n)\right)\right)}}. \tag{2.3}$$

Here $\hat{Var}(h(\hat{\boldsymbol{\theta}}_k(n)))$ and $\hat{Var}(h(\hat{\boldsymbol{\theta}}_0(n)))$ are some consistent estimators of the variances of $h(\hat{\boldsymbol{\theta}}_k(n))$ and $h(\hat{\boldsymbol{\theta}}_0(n))$ respectively. We assume that for some functions ν_k and v_0

$$\begin{aligned}&n\hat{Var}\left(h\left(\hat{\boldsymbol{\theta}}_j(n)\right)\right)\\&\quad=\nu_j\left(\frac{\boldsymbol{N}(n)}{n},\hat{\boldsymbol{\theta}}(n)\right)(1+o(1))\qquad \text{a.s. } j=0,k.\end{aligned}$$

Both $\nu_j(\boldsymbol{y},\boldsymbol{z})$ and $Z_k(\boldsymbol{y},\boldsymbol{z})$ are $\Re^{(K+1)(1+d)}\to\Re$ function, where $\boldsymbol{y}$ is a $(K+1)$-dimensional vector and $\boldsymbol{z}$ is a $(K+1)d$-dimensional vector. Examples of using this formulation are given in Section 3.

2.3 Asymptotic Results

Before we give the main theorem, we need the following conditions.

(A1) For some $\varepsilon>0$, $E\|\boldsymbol{X}_1\|^{2+\varepsilon}<\infty$;

(A2) $g_{1k}(\boldsymbol{s},\boldsymbol{r}),k=0,1,\ldots,K$, is jointly continuous and twice differentiable at $(\boldsymbol{\rho}_1,\boldsymbol{\rho}_1)$, and $g_{2k}(\boldsymbol{s},\boldsymbol{r}),k=0,M$, is jointly continuous and twice differentiable at $(\boldsymbol{\rho}_2,\boldsymbol{\rho}_2)$;

(A3) $g_{1k}(\boldsymbol{r},\boldsymbol{r})=r_k,\ k=0,1,\ldots,K$, for all $\boldsymbol{r}\in(0,1)^{(K+1)}$ and $\sum_{k=0}^K r_k=1$, $g_{1k}(\boldsymbol{s},\boldsymbol{r})$ is strictly decreasing in $\boldsymbol{s}\in(0,1)^{(K+1)}$ and strictly increasing in $\boldsymbol{r}\in(0,1)^{(K+1)}$, $g_{2k}(\boldsymbol{r},\boldsymbol{r})=r_k,k=0,M$, for all $\boldsymbol{r}\in(0,1)^2$ and $r_0+r_M=1$, and $g_{2k}(\boldsymbol{s},\boldsymbol{r})$ is strictly decreasing in $\boldsymbol{s}\in(0,1)^2$ and strictly increasing in $\boldsymbol{r}\in(0,1)^2$;

(A4) $\boldsymbol{\rho}_{1k}(\boldsymbol{\theta}),k=0,1,\ldots,K$ and $\boldsymbol{\rho}_{2k}(\boldsymbol{\theta}),k=0,M$ are continuous functions and twice continuously differentiable in a small neighborhood of $\boldsymbol{\theta}$;

(A5) $\nu_j(\boldsymbol{y},\boldsymbol{z})$ is jointly continuous and twice differentiable in a small neighborhood of $(\boldsymbol{\rho},\boldsymbol{\theta})$;

(A6) $Z_k(\boldsymbol{y},\boldsymbol{z})$ is a continuous function and it is twice continuously differentiable in a small neighborhood of vector $(\boldsymbol{\rho},\boldsymbol{\theta})$.

Theorem 2.1. *Under Conditions (A1)–(A6), a valid type I error rate can be asymptotically obtained for the Simes test with the test statistics $Z_k,k=1,\ldots,K$, for the proposed procedure. That is, for a given significance level α, when $H_{0,M}$ holds, the probability that we reject $H_{0,M}$ has a limit that is not larger than α.*

Theorem 2.1 offers the theoretical justification for controlling the type I error rate for our procedure. All these conditions are easily satisfied. The well-known family of DBCD [25] meets all these requirements. Condition (A1) ensures consistency and asymptotic normality. All the examples in Chapter 5 [23] meet Conditions (A4)–(A6). In particular, Condition (A3) has practical meaning in clinical trials: if the current actual allocation proportion is equal to the target allocation proportion, the allocation probability for the next patient will equal to the target allocation proportion ($g_{jk}(\boldsymbol{r},\boldsymbol{r})=r_k$). On top of that, because the allocation probability function is strictly decreasing in the actual allocation proportion and strictly increasing in the estimated target allocation proportion, the proposed RAR design will asymptotically drive the actual allocation proportion to approach the theoretically targeted one for each stage ($\boldsymbol{\rho}_1$ for stage 1 and $\boldsymbol{\rho}_2$ for stage 2), which is proved in [25]. The actual final allocation proportion for the two-stage seamless trial when the sample size is finite will be studied in the next section.

3. NUMERICAL STUDIES

In this section, we study the finite-sample properties of our proposed procedure and offer three specific targeted allocation proportions.

Suppose 300 patients sequentially enter the trial with two experimental treatments and one control in the first stage. Let the responses $X_{ik},i=1,\ldots,300,\ k=0,1,2$, follow the Bernoulli distribution with success rate p_k, respectively.

These patients will be sequentially randomly allocated to the treatment k with the following allocation probability function [25]

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

where $s_k = N_k(m-1)/(m-1)$ and $r_k = \rho_{1k}(\hat{\boldsymbol{\theta}}(m-1))$ when we are calculating the allocation probability for the mth patient. We will discuss three specific targeted allocation proportions later. The experimental treatment arm with a larger treatment effect, say treatment M, is chosen to continue to the second stage. In the second stage, 500 patients are sequentially randomly allocated to the control arm and treatment M with the following allocation probability function

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

At the end of the trial, we test

$$H_{0,M} : p_M = p_0 \ vs. \ H_{1,M} : p_M > p_0.$$

In this case, $d = 1$, $\theta_k = p_k$, and

$$Z_k = (\hat{p}_k - \hat{p}_0)/\sqrt{\hat{p}_k(1-\hat{p}_k)/N_k + \hat{p}_0(1-\hat{p}_0)/N_0}, k = 1, \ldots, K.$$

The significance level is 0.025 for all the tests. All the results are based on 10, 000 replications.

In the first scenario, let

$$\boldsymbol{\rho}_1(\boldsymbol{\theta}) = \left(\frac{p_0}{p_0+p_1+p_2}, \frac{p_1}{p_0+p_1+p_2}, \frac{p_2}{p_0+p_1+p_2}\right)$$

and

$$\boldsymbol{\rho}_2(\boldsymbol{\theta}) = \left(\frac{q_M}{q_0+q_M}, \frac{q_0}{q_0+q_M}\right)$$

that is the urn allocation [64]. Urn allocation is used to assign more patients to the better treatment.

In the second scenario, let

$$\boldsymbol{\rho}_1(\boldsymbol{\theta}) = \left(\frac{\sqrt{p_0}}{\sqrt{p_0}+\sqrt{p_1}+\sqrt{p_2}}, \frac{\sqrt{p_1}}{\sqrt{p_0}+\sqrt{p_1}+\sqrt{p_2}}, \frac{\sqrt{p_2}}{\sqrt{p_0}+\sqrt{p_1}+\sqrt{p_2}}\right)$$

and

$$\boldsymbol{\rho}_2(\boldsymbol{\theta}) = \left(\frac{\sqrt{p_0}}{\sqrt{p_0}+\sqrt{p_M}}, \frac{\sqrt{p_M}}{\sqrt{p_0}+\sqrt{p_M}}\right)$$

that is the optimal allocation [41]. The optimal allocation is used to minimize the total number of failures while fixing the power.

In the third scenario, let

$$\boldsymbol{\rho}_1(\boldsymbol{\theta}) = \left(\frac{p_0}{p_0+p_1+p_2}, \frac{p_1}{p_0+p_1+p_2}, \frac{p_2}{p_0+p_1+p_2}\right)$$

Table 1. Performance of DBCD targeting the urn allocation under H_0 when three treatments are under study.

(p_1, p_2, p_0)	Design	α	$\hat{p}_0$	ρ_0	Failure
(0.5, 0.5, 0.5)	DBCD	0.024	0.500(0.027)	0.438(0.020)	400(14)
(0.5, 0.5, 0.5)	CR	0.024	0.500(0.027)	0.438(0.017)	400(14)
(0.6, 0.6, 0.6)	DBCD	0.022	0.599(0.026)	0.437(0.022)	320(14)
(0.6, 0.6, 0.6)	CR	0.022	0.600(0.026)	0.438(0.017)	320(14)
(0.7, 0.7, 0.7)	DBCD	0.024	0.700(0.025)	0.437(0.025)	240(13)
(0.7, 0.7, 0.7)	CR	0.023	0.700(0.025)	0.438(0.017)	240(13)
(0.8, 0.8, 0.8)	DBCD	0.023	0.799(0.021)	0.437(0.032)	160(11)
(0.8, 0.8, 0.8)	CR	0.025	0.800(0.021)	0.438(0.017)	160(11)

Table 2. Performance of DBCD targeting the optimal allocation under H_0 when three treatments are under study.

(p_1, p_2, p_0)	Design	α	$\hat{p}_0$	ρ_0	Failure
(0.5, 0.5, 0.5)	DBCD	0.023	0.499(0.027)	0.438(0.012)	400(14)
(0.5, 0.5, 0.5)	CR	0.024	0.500(0.027)	0.438(0.017)	400(14)
(0.6, 0.6, 0.6)	DBCD	0.024	0.600(0.026)	0.438(0.011)	320(14)
(0.6, 0.6, 0.6)	CR	0.022	0.600(0.026)	0.438(0.017)	320(14)
(0.7, 0.7, 0.7)	DBCD	0.026	0.700(0.025)	0.438(0.010)	240(13)
(0.7, 0.7, 0.7)	CR	0.023	0.700(0.025)	0.438(0.017)	240(13)
(0.8, 0.8, 0.8)	DBCD	0.024	0.800(0.022)	0.438(0.010)	160(11)
(0.8, 0.8, 0.8)	CR	0.025	0.800(0.021)	0.438(0.017)	160(11)

Table 3. Performance of DBCD targeting the intuitively ethical allocation proportion under H_0 when three treatments are under study.

(p_1, p_2, p_0)	Design	α	$\hat{p}_0$	ρ_0	Failure
(0.5, 0.5, 0.5)	DBCD	0.027	0.500(0.027)	0.438(0.020)	400(14)
(0.5, 0.5, 0.5)	CR	0.024	0.500(0.027)	0.438(0.017)	400(14)
(0.6, 0.6, 0.6)	DBCD	0.024	0.599(0.026)	0.437(0.017)	320(14)
(0.6, 0.6, 0.6)	CR	0.022	0.600(0.026)	0.438(0.017)	320(14)
(0.7, 0.7, 0.7)	DBCD	0.026	0.699(0.025)	0.437(0.014)	240(13)
(0.7, 0.7, 0.7)	CR	0.023	0.700(0.025)	0.438(0.017)	240(13)
(0.8, 0.8, 0.8)	DBCD	0.022	0.800(0.022)	0.438(0.012)	160(11)
(0.8, 0.8, 0.8)	CR	0.025	0.800(0.021)	0.438(0.017)	160(11)

and

$$\boldsymbol{\rho}_2(\boldsymbol{\theta}) = \left(\frac{p_0}{p_0+p_M}, \frac{p_M}{p_0+p_M}\right).$$

This target allocation proportion is intuitively assigning more patients to the better treatment.

In Tables 1–3, we studied and compared the performance of our methods under each of the above scenarios and complete randomization (CR) under $H_{0,M}$. In these tables, we found that, under $H_{0,M}$, our method can control the type I error rate (α) well. We reported $\hat{p}_0$ as a representative of the parameter estimators. We also reported the actual allocation proportion to the control group (ρ_0) and the total number of failures (Failure). The standard deviations are in the parentheses. In all the tables, our methods and CR return almost the same values in terms of the allocation proportion and the total number of failures under $H_{0,M}$, since our designs are also targeting the equal allocation under $H_{0,M}$. Our methods can also estimate the parameter accurately.

Table 4. Performance of DBCD targeting the urn allocation under H_1 when three treatments are under study.

(p_1, p_2, p_0)	Design	Power	$\hat{p}_0$	ρ_0	Failure
$(0.4, 0.45, 0.3)$	DBCD	0.939	0.300(0.027)	0.379(0.020)	494(15)
$(0.4, 0.45, 0.3)$	CR	0.946	0.300(0.025)	0.438(0.017)	501(15)
$(0.5, 0.55, 0.4)$	DBCD	0.926	0.400(0.028)	0.378(0.021)	414(16)
$(0.5, 0.55, 0.4)$	CR	0.930	0.400(0.026)	0.438(0.017)	420(15)
$(0.6, 0.65, 0.5)$	DBCD	0.925	0.500(0.029)	0.371(0.023)	333(16)
$(0.6, 0.65, 0.5)$	CR	0.929	0.500(0.027)	0.438(0.017)	340(15)
$(0.7, 0.75, 0.6)$	DBCD	0.952	0.599(0.029)	0.358(0.026)	251(15)
$(0.7, 0.75, 0.6)$	CR	0.955	0.600(0.026)	0.438(0.017)	260(14)

Table 5. Performance of DBCD targeting the optimal allocation under H_1 when three treatments are under study.

(p_1, p_2, p_0)	Design	Power	$\hat{p}_0$	ρ_0	Failure
$(0.4, 0.45, 0.3)$	DBCD	0.938	0.300(0.026)	0.395(0.015)	495(15)
$(0.4, 0.45, 0.3)$	CR	0.946	0.300(0.025)	0.438(0.017)	501(15)
$(0.5, 0.55, 0.4)$	DBCD	0.928	0.400(0.027)	0.404(0.013)	417(15)
$(0.5, 0.55, 0.4)$	CR	0.930	0.400(0.026)	0.438(0.017)	420(15)
$(0.6, 0.65, 0.5)$	DBCD	0.928	0.500(0.028)	0.410(0.017)	337(15)
$(0.6, 0.65, 0.5)$	CR	0.929	0.500(0.027)	0.438(0.017)	340(15)
$(0.7, 0.75, 0.6)$	DBCD	0.951	0.600(0.027)	0.414(0.010)	258(14)
$(0.7, 0.75, 0.6)$	CR	0.955	0.600(0.026)	0.438(0.017)	260(14)

Table 6. Performance of DBCD targeting the intuitively ethical allocation proportion under H_1 when three treatments are under study.

(p_1, p_2, p_0)	Design	Power	$\hat{p}_0$	ρ_0	Failure
$(0.4, 0.45, 0.3)$	DBCD	0.939	0.299(0.028)	0.354(0.027)	491(15)
$(0.4, 0.45, 0.3)$	CR	0.946	0.300(0.025)	0.438(0.017)	501(15)
$(0.5, 0.55, 0.4)$	DBCD	0.926	0.399(0.029)	0.372(0.023)	413(16)
$(0.5, 0.55, 0.4)$	CR	0.930	0.400(0.026)	0.438(0.017)	420(15)
$(0.6, 0.65, 0.5)$	DBCD	0.927	0.500(0.029)	0.383(0.019)	334(15)
$(0.6, 0.65, 0.5)$	CR	0.929	0.500(0.027)	0.438(0.017)	340(15)
$(0.7, 0.75, 0.6)$	DBCD	0.954	0.599(0.028)	0.391(0.016)	255(14)
$(0.7, 0.75, 0.6)$	CR	0.955	0.600(0.026)	0.438(0.017)	260(14)

In Tables 4–6, we studied and compared the performance of our methods under each of the above scenarios and CR under $H_{1,M}$. In Table 4, we can see that our method can save up to around 10 patients while keeping the power at the same level as CR under $H_{1,M}$ for the first scenario. In Table 5, we can see that our method can assign more patients to the better treatment while keeping the power at the same level as CR under $H_{1,M}$. In Table 6, we found that the DBCD targeting this allocation proportion can also save up to 10 patients under the reported settings without sacrificing the power.

We further performed numerical studies for clinical trials with one control arm and three experimental treatment arms representing the low, medium, and high doses of the experimental drugs. The success rates for the control arm and three experimental treatment arms are p_0, p_1, p_2 and p_3, respectively. We keep the same sample size for Stage 2 as Tables 1–3, but increase the sample size to 400 for Stage 1, considering we have four treatment arms in this stage.

Table 7. Performance of DBCD targeting the urn allocation under H_0 when four treatments are under study.

(p_1, p_2, p_3, p_0)	Design	α	$\hat{p}_0$	ρ_0	Failure
$(0.5, 0.5, 0.5, 0.5)$	DBCD	0.023	0.499(0.027)	0.389(0.018)	450(15)
$(0.5, 0.5, 0.5, 0.5)$	CR	0.025	0.500(0.027)	0.389(0.016)	450(15)
$(0.6, 0.6, 0.6, 0.6)$	DBCD	0.025	0.599(0.026)	0.389(0.020)	360(15)
$(0.6, 0.6, 0.6, 0.6)$	CR	0.023	0.600(0.026)	0.389(0.016)	360(15)
$(0.7, 0.7, 0.7, 0.7)$	DBCD	0.022	0.699(0.025)	0.389(0.022)	270(14)
$(0.7, 0.7, 0.7, 0.7)$	CR	0.025	0.700(0.025)	0.389(0.016)	270(14)
$(0.8, 0.8, 0.8, 0.8)$	DBCD	0.022	0.799(0.022)	0.389(0.028)	180(12)
$(0.8, 0.8, 0.8, 0.8)$	CR	0.025	0.800(0.022)	0.389(0.016)	180(12)

Table 8. Performance of DBCD targeting the optimal allocation under H_0 when four treatments are under study.

(p_1, p_2, p_3, p_0)	Design	α	$\hat{p}_0$	ρ_0	Failure
$(0.5, 0.5, 0.5, 0.5)$	DBCD	0.023	0.500(0.027)	0.389(0.011)	450(15)
$(0.5, 0.5, 0.5, 0.5)$	CR	0.025	0.500(0.027)	0.389(0.016)	450(15)
$(0.6, 0.6, 0.6, 0.6)$	DBCD	0.023	0.600(0.026)	0.389(0.010)	360(15)
$(0.6, 0.6, 0.6, 0.6)$	CR	0.023	0.600(0.026)	0.389(0.016)	360(15)
$(0.7, 0.7, 0.7, 0.7)$	DBCD	0.026	0.700(0.024)	0.389(0.009)	270(14)
$(0.7, 0.7, 0.7, 0.7)$	CR	0.025	0.700(0.025)	0.389(0.016)	270(14)
$(0.8, 0.8, 0.8, 0.8)$	DBCD	0.023	0.800(0.021)	0.389(0.008)	180(12)
$(0.8, 0.8, 0.8, 0.8)$	CR	0.025	0.800(0.022)	0.389(0.016)	180(12)

Table 9. Performance of DBCD targeting the intuitively ethical allocation proportion under H_0 when four treatments are under study.

(p_1, p_2, p_3, p_0)	Design	α	$\hat{p}_0$	ρ_0	Failure
$(0.5, 0.5, 0.5, 0.5)$	DBCD	0.022	0.499(0.027)	0.389(0.018)	450(15)
$(0.5, 0.5, 0.5, 0.5)$	CR	0.025	0.500(0.027)	0.389(0.016)	450(15)
$(0.6, 0.6, 0.6, 0.6)$	DBCD	0.025	0.599(0.026)	0.389(0.016)	360(15)
$(0.6, 0.6, 0.6, 0.6)$	CR	0.023	0.600(0.026)	0.389(0.016)	360(15)
$(0.7, 0.7, 0.7, 0.7)$	DBCD	0.024	0.700(0.025)	0.389(0.013)	270(14)
$(0.7, 0.7, 0.7, 0.7)$	CR	0.025	0.700(0.025)	0.389(0.016)	270(14)
$(0.8, 0.8, 0.8, 0.8)$	DBCD	0.020	0.800(0.021)	0.389(0.011)	180(12)
$(0.8, 0.8, 0.8, 0.8)$	CR	0.025	0.800(0.022)	0.389(0.016)	180(12)

In the first scenario (Tables 7 and 10 for $H_{0,M}$ and $H_{1,M}$), let

$$\boldsymbol{\rho}_1(\boldsymbol{\theta}) = \Big(\frac{p_0}{p_0+p_1+p_2+p_3}, \frac{p_1}{p_0+p_1+p_2+p_3}, \frac{p_2}{p_0+p_1+p_2+p_3}, \frac{p_3}{p_0+p_1+p_2+p_3}\Big)$$

and $\boldsymbol{\rho}_2(\boldsymbol{\theta}) = \Big(\frac{q_M}{q_0+q_M}, \frac{q_0}{q_0+q_M}\Big)$ that is the urn allocation [64]. In the second scenario (Table 8 and 11 for H_0 and H_1), let $\rho_{1k}(\boldsymbol{\theta}) = \frac{\sqrt{p_k}}{\sqrt{p_0}+\sqrt{p_1}+\sqrt{p_2}+\sqrt{p_3}}$, $k = 0, 1, 2, 3$, and $\boldsymbol{\rho}_2(\boldsymbol{\theta}) = \Big(\frac{\sqrt{p_0}}{\sqrt{p_0}+\sqrt{p_M}}, \frac{\sqrt{p_M}}{\sqrt{p_0}+\sqrt{p_M}}\Big)$ that is the optimal allocation [41]. In the third scenario (Table 9 and 12 for H_0 and H_1), we use the intuitively better allocation proportion

$$\boldsymbol{\rho}_1(\boldsymbol{\theta}) = \Big(\frac{p_0}{p_0+p_1+p_2+p_3}, \frac{p_1}{p_0+p_1+p_2+p_3}, \frac{p_2}{p_0+p_1+p_2+p_3}, \frac{p_3}{p_0+p_1+p_2+p_3}\Big)$$

Table 10. Performance of DBCD targeting the urn allocation under H_1 when four treatments are under study.

(p_1, p_2, p_3, p_0)	Design	Power	$\hat{p}_0$	ρ_0	Failure
$(0.35, 0.4, 0.45, 0.3)$	DBCD	0.910	0.299(0.026)	0.339(0.019)	559(17)
$(0.35, 0.4, 0.45, 0.3)$	CR	0.910	0.300(0.025)	0.389(0.016)	566(16)
$(0.45, 0.5, 0.55, 0.4)$	DBCD	0.892	0.400(0.028)	0.338(0.019)	469(17)
$(0.45, 0.5, 0.55, 0.4)$	CR	0.896	0.400(0.026)	0.389(0.016)	476(17)
$(0.55, 0.6, 0.65, 0.5)$	DBCD	0.894	0.499(0.029)	0.332(0.021)	379(17)
$(0.55, 0.6, 0.65, 0.5)$	CR	0.896	0.500(0.027)	0.389(0.016)	386(16)
$(0.65, 0.7, 0.75, 0.6)$	DBCD	0.923	0.598(0.029)	0.320(0.024)	287(17)
$(0.65, 0.7, 0.75, 0.6)$	CR	0.922	0.600(0.026)	0.389(0.016)	296(16)

Table 11. Performance of DBCD targeting the optimal allocation under H_1 when four treatments are under study.

(p_1, p_2, p_3, p_0)	Design	Power	$\hat{p}_0$	ρ_0	Failure
$(0.35, 0.4, 0.45, 0.3)$	DBCD	0.907	0.300(0.026)	0.353(0.014)	561(16)
$(0.35, 0.4, 0.45, 0.3)$	CR	0.910	0.300(0.025)	0.389(0.016)	566(16)
$(0.45, 0.5, 0.55, 0.4)$	DBCD	0.894	0.400(0.027)	0.361(0.012)	473(17)
$(0.45, 0.5, 0.55, 0.4)$	CR	0.896	0.400(0.026)	0.389(0.016)	476(17)
$(0.55, 0.6, 0.65, 0.5)$	DBCD	0.894	0.500(0.028)	0.366(0.011)	383(17)
$(0.55, 0.6, 0.65, 0.5)$	CR	0.896	0.500(0.027)	0.389(0.016)	386(16)
$(0.65, 0.7, 0.75, 0.6)$	DBCD	0.928	0.599(0.027)	0.369(0.010)	293(15)
$(0.65, 0.7, 0.75, 0.6)$	CR	0.922	0.600(0.026)	0.389(0.016)	296(16)

Table 12. Performance of DBCD targeting the intuitively ethical allocation proportion under H_1 when four treatments are under study.

(p_1, p_2, p_3, p_0)	Design	Power	$\hat{p}_0$	ρ_0	Failure
$(0.35, 0.4, 0.45, 0.3)$	DBCD	0.911	0.299(0.028)	0.317(0.026)	556(17)
$(0.35, 0.4, 0.45, 0.3)$	CR	0.910	0.300(0.025)	0.389(0.016)	566(16)
$(0.45, 0.5, 0.55, 0.4)$	DBCD	0.894	0.399(0.029)	0.333(0.021)	469(17)
$(0.45, 0.5, 0.55, 0.4)$	CR	0.896	0.400(0.026)	0.389(0.016)	476(17)
$(0.55, 0.6, 0.65, 0.5)$	DBCD	0.896	0.499(0.029)	0.342(0.018)	380(17)
$(0.55, 0.6, 0.65, 0.5)$	CR	0.896	0.500(0.027)	0.389(0.016)	386(16)
$(0.65, 0.7, 0.75, 0.6)$	DBCD	0.926	0.599(0.028)	0.349(0.015)	290(16)
$(0.65, 0.7, 0.75, 0.6)$	CR	0.922	0.600(0.026)	0.389(0.016)	296(16)

and

$$\boldsymbol{\rho}_2(\boldsymbol{\theta}) = \left(\frac{p_0}{p_0 + p_M}, \frac{p_M}{p_0 + p_M} \right).$$

We detected similar conclusions to the numerical studies with two experimental treatment arms above. While keeping the same power level, our methods can always assign more people to better treatment and reduce the number of failures by up to 10 patients.

4. CONCLUSION

Clinical trials are complex and involve a variety of design features related to efficiency and ethics. ASD and RAR have been proposed to achieve different aims [15, 17, 22, 23]. The desire to reduce development costs and the time-to-market of new treatments has led to the development of ASD. DBCD is a well-known RAR design with a variety of favorable properties. However, there has been limited theoretical and numerical study of the combination of ASD and DBCD, which hinders the development and application of this procedure. In this paper, we proposed a versatile approach and studied this complex procedure's theoretical and numerical properties. Our methods can lead to less failure without sacrificing power than traditional designs while controlling the type I error rate.

[73] also tried implementing DBCD in seamless clinical trials. However, their methods depend on the method in [43] to construct the test statistics and control the type I error rate. As a result, strictly speaking, their methods can only be used for normal responses, and other future investigations of the procedure will require new challenging theoretical proof. This is a severe limitation in practice. The current paper proposed more versatile approaches based on the closure principle combined with the combination test and methods to address multiplicity problems, which FDA and pharmaceutical industry will more readily accept. More importantly, many existing approaches based on this framework, such as its combination with sequential monitoring and other endpoints, can be directly used in future trials. We leave these for future research. Fundamentally, we proposed a totally different and more flexible approach to implement DBCD in seamless clinical trials, which will significantly promote the procedure in clinical trials.

It is also worth discussing the benefit-cost tradeoff of the adaptive designs. First, exploring the seamless phase II/III design is often desirable in pharmaceutical companies for various reasons. For example, the regulatory agencies often require the comparison of a new dose in addition to the proposed two-arm clinical trial design, so a seamless phase II/III design often becomes one of some natural choices. RAR might make the design more complex compared to a fixed design. However, with the development of technology such as central data monitoring, interactive voice response services, and interactive web response service, the complexity of implementing advanced designs such as RAR is much reduced. Second, the evaluation of the reduction of total failures depends on the disease characteristics. For lethal diseases like the Ebola virus, failure means quick death, and any savings could be worth it.

This paper opens the door to future research topics. First, there are two families of RAR designs, DBCD and urn models [64, 65, 72]. It is worth exploring the seamless clinical trials with urn models. Second, research on adaptive randomization design and ASD under the Bayesian framework includes but is not limited to [3, 6, 7, 12, 28, 29, 54, 66, 67]. We can comprehensively compare our methods with the bayesian approaches. Third, [18, 28, 57] investigated the ASD with different types of study endpoints in the two phases. All these factors can be explored for the proposed design. We leave all these for future research topics.

SUPPLEMENTARY MATERIAL

Proof of Theorem 2.1: The rigorous proof for applying the closure principle [37] with the combination test and Simes

test in a seamless Phase II/III clinical trial with complete randomization to control the type I error rate has been obtained and discussed in [11, 13, 35]. (We offer some explanation for this procedure here; details can be seen in the above papers.) First, the closure principle [37] was proposed to construct multiple test procedures to strongly control the family-wise error rate. Then, the randomness of M for the combination test is addressed by the *conditional invariance principle* (see, for example, [11, 13]. According to the invariance principle, the p-values $P_{1,I}$ and $P_{2,I}$ are independent and also independent of the choice of M, so the combination test will be used to control the Type I error rate for our proposed adaptation rules.) Based on [25, 71], under (A1)–(A6), the joint distribution of $(Z_1, \ldots, Z_K)$ from this paper is asymptotically the same as that from complete randomization. So our method can asymptotically control the type I error rate.

ACKNOWLEDGMENTS

This work was partly supported by grant DMS-2014951 (Dr. Hongjian Zhu) from the National Science Foundation (USA) and Weatherhead Foundation (Dr. Dejian Lai). This publication was neither originated nor managed by AbbVie, and it does not communicate results of AbbVie-sponsored Scientific Research. Thus, it is not in scope of the AbbVie Publication Procedure (PUB-100).

Accepted 29 January 2023

REFERENCES

[1] Andersen, J. S. (1996). Clinical trial designs—made to order. *Journal of Biopharmaceutical Statistics* **6**(4) 515–522. https://doi.org/10.1198/016214503000000576. MR2011680

[2] Antognini, A. B. and Giovagnoli, A. (2010). Compound optimal allocation for individual and collective ethics in binary clinical trials. *Biometrika* **97**(4) 935–946. https://doi.org/10.1214/aos/1079120137. MR2051008

[3] Atkinson, A. C. and Biswas, A. (2005). Bayesian Adaptive Biased-Coin Designs for Clinical Trials with Normal Responses. *Biometrics* **61**(1) 118–125. https://doi.org/10.1214/08-AOS655. MR2543702

[4] Barnes, P. J., Pocock, S. J., 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.

[5] Bauer, P. and Köhne, K. (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics* **50**(4) 1029–1041. https://doi.org/10.1002/sim.3578. MR2675244

[6] Berry, D. A. (2004). Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science* **19**(1) 175–187. https://doi.org/10.1111/j.0006-341X.2002.00823.x. MR1945019

[7] Berry, D. A. (2012). Adaptive clinical trials in oncology. *Nature Reviews Clinical Oncology* **9**(4) 199–207.

[8] Biswas, A., Mandal, S. and Bhattacharya, R. (2011). Multitreatment optimal response-adaptive designs for phase III clinical trials. *Journal of the Korean Statistical Society* **40**(1) 33–44. https://doi.org/10.1081/BIP-200062857. MR2190575

[9] Bowden, J. and Glimm, E. (2008). Unbiased estimation of selected treatment means in two-stage trials. *Biometrical Journal* **50**(4) 515–527. https://doi.org/10.1002/sim.5757. MR3073825

[10] 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. https://doi.org/10.1002/sim.4430. MR2880482

[11] Brannath, W., Gutjahr, G. and Bauer, P. (2012). Probabilistic foundation of confirmatory adaptive designs. *Journal of the American Statistical Association* **107**(498) 824–832.

[12] 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. https://doi.org/10.1198/016214502388618852. MR1951257

[13] Brannath, W., Koenig, F. and Bauer, P. (2007). Multiplicity and flexibility in clinical trials. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* **6**(3) 205–216. https://doi.org/10.1093/biomet/ass002. MR2931269

[14] 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. https://doi.org/10.1093/biomet/63.3.655. MR0468056

[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. https://doi.org/10.1002/sim.2389. MR2221962

[16] Chen, Y. J., 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.

[17] Chow, S. C. and Chang, M. (2012) *Adaptive Design Methods in Clinical Trials, Second Edition*. Chapman and Hall/CRC. https://doi.org/10.1002/sim.6974. MR3545616

[18] Chow, S. C., Lu, Q. and Tse, S. K. (2007). Statistical analysis for two-stage seamless design with different study endpoints. *Journal of Biopharmaceutical Statistics* **17**(6) 1163–1176. https://doi.org/10.1111/j.0006-341X.2001.00909.x. MR1863454

[19] Cohen, A. and Sackrowitz, H. B. (1989). Two stage conditionally unbiased estimators of the selected mean. *Statistics & Probability Letters* **8**(3) 273–278.

[20] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**(272) 1096–1121. https://doi.org/10.1002/bimj.200410119. MR2145117

[21] Gao, L., Zhu, H. and Zhang, L. (2020). Sequential monitoring of response-adaptive randomized clinical trials with sample size re-estimation. *Journal of Statistical Planning and Inference* **205** 129–137. https://doi.org/10.1093/biomet/77.3.507. MR1087840

[22] Hampson, L. V. and Jennison, C. (2015). Optimizing the data combination rule for seamless phase II/III clinical trials. *Statistics in Medicine* **34**(1) 39–58. https://doi.org/10.1002/bimj.200510231. MR2247049

[23] Hu, F. and Rosenberger, W. F. (2006) *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley and Sons, Inc., New York.

[24] Hu, F. and Rosenberger, W. F. (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* **98**(463) 671–678. https://doi.org/10.1093/biomet/73.3.751. MR0897872

[25] Hu, F. and Zhang, L.-X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics* **32**(1) 268–301. https://doi.org/10.1002/sim.3863. MR2756565

[26] Hu, F., Zhang, L.-X. and He, X. (2009). Efficient randomized-adaptive designs. *The Annals of Statistics* 2543–2560. https://doi.org/10.1002/sim.3436. MR2522318

[27] 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.

[28] Huang, X., Ning, J., Li, Y., Estey, E., Issa, J. P. and Berry,

D. A. (2009). Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Statistics in Medicine* **28**(12) 1680–1689. https://doi.org/10.1016/j.jspi.2004.05.006. MR2200477

[29] Inoue, L. Y., Thall, P. F. and Berry, D. A. (2002). Seamlessly expanding a randomized phase II trial to phase III. *Biometrics* **58**(4) 823–831. https://doi.org/10.1016/j.jspi.2007.05.045. MR2427293

[30] Ivanova, A. and Rosenberger, W. F. (2000). A comparison of urn designs for randomized clinical trials of K > 2 treatments. *Journal of Biopharmaceutical Statistics* **10**(1) 93–107.

[31] Kelly, P. J., Stallard, N. and Todd, S. (2005). An adaptive group sequential design for phase II/III clinical trials that select a single treatment from several. *Journal of Biopharmaceutical Statistics* **15**(4) 641–658.

[32] Kimani, P. K., 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.

[33] Koopmeiners, J. S., Feng, Z. and Pepe, M. S. (2012). Conditional estimation after a two-stage diagnostic biomarker study that allows early termination for futility. *Statistics in Medicine* **31**(5) 420–435. https://doi.org/10.2307/2531495. MR1010517

[34] Lehmacher, W. and Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* **55**(4) 1286–1290.

[35] Liu, Q., Proschan, M. A. and Pledger, G. W. (2002). A unified theory of two-stage adaptive designs. *Journal of the American Statistical Association* **97**(460) 1034–1041. https://doi.org/10.1080/03610929908832376. MR1707106

[36] Magirr, D., Jaki, T. and Whitehead, J. (2012). A generalized Dunnett test for multi-arm multi-stage clinical studies with treatment selection. *Biometrika* **99**(2) 494–501. https://doi.org/10.1198/016214506000000906. MR2345540

[37] Marcus, R., Eric, P. and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**(3) 655–660.

[38] 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.

[39] Prowell, T. M., Theoret, M. R. and Pazdur, R. (2016). Seamless oncology-drug development. *New England Journal of Medicine* **374**(21) 2001–2003. https://doi.org/10.1016/j.jspi.2016.09.002. MR3574510

[40] Robertson, D. S., Prevost, A. T. and Bowden, J. (2016). Unbiased estimation in seamless phase II/III trials with unequal treatment effect variances and hypothesis-driven selection rules. *Statistics in Medicine* **35**(22) 3907–3922. https://doi.org/10.1177/0962280214550759. MR3592738

[41] Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N. and Ricks, M. L. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57**(3) 909–913.

[42] Rout, C., Rocke, D., Levin, J., Gouws, E. and Reddy, D. (1993). A reevaluation of the role of crystalloid preload in the prevention of hypotension associated with spinal anesthesia for elective cesarean section. *Anesthesiology* **79**(2) 262–269. https://doi.org/10.1002/cjs.11609. MR4349636

[43] Sampson, A. R. and Sill, M. W. (2005). Drop-the-losers design: Normal case. *Biometrical Journal* **47**(3) 257–268. https://doi.org/10.1002/sim.4218. MR2828933

[44] Schaid, D. J., Wieand, S. A. M. and Therneau, T. M. (1990). Optimal two-stage screening designs for survival comparisons. *Biometrika* **77**(3) 507–513.

[45] 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.

[46] Shen, L. (2001). An improved method of evaluating drug effect in a multiple dose clinical trial. *Statistics in Medicine* **20**(13) 1913–1929. https://doi.org/10.1214/105051605000000746. MR2209345

[47] Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**(3) 751–754. https://doi.org/10.1016/j.jspi.2008.11.003. MR2508003

[48] Stallard, N. (2010). A confirmatory seamless phase II/III clinical trial design incorporating short-term endpoint information. *Statistics in Medicine* **29**(9) 959–971. https://doi.org/10.1214/10-AOS796. MR2676888

[49] Stallard, N. and Friede, T. (2008). A group-sequential design for clinical trials with treatment selection. *Statistics in Medicine* **27**(29) 6209–6227. https://doi.org/10.1002/cjs.11140. MR2968398

[50] Stallard, N. and Todd, S. (2003). Sequential designs for phase III clinical trials incorporating treatment selection. *Statistics in Medicine* **22**(5) 689–703. https://doi.org/10.1080/10543400701645322. MR2414569

[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., 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.

[53] Tamura, R. N., Faries, D. E., Andersen, J. S. and Heiligenstein, J. H. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association* **89**(427) 768–776.

[54] Thall, P. F. and Wathen, J. K. (2007). Practical Bayesian adaptive randomisation in clinical trials. *European Journal of Cancer* **43**(5) 859–866.

[55] Thall, P. F., Simon, R. and Ellenberg, S. S. (1988). Two-stage selection and testing designs for comparative clinical trials. *Biometrika* **75**(2) 303–310.

[56] Thall, P. F., Simon, R. and Ellenberg, S. S. (1989). A two-stage design for choosing among several experimental treatments and a control in clinical trials. *Biometrics* **45** 537–547.

[57] Todd, S. and Stallard, N. (2005). A new clinical trial design combining phases 2 and 3: Sequential designs with treatment selection and a change of endpoint. *Drug Information Journal* **39**(2) 109–118.

[58] Troendle, J. F. and Yu, K. F. (1999). Conditional estimation following a group sequential clinical trial. *Communications in Statistics - Theory and Methods* **28**(7) 1617–1634.

[59] Tymofyeyev, Y., Rosenberger, W. F. and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association* **102**(477) 224–234.

[60] US Food and Drug Administration (2006). Critical Path Opportunities Report and List.

[61] US Food and Drug Administration (2018). Expansion cohorts: Use in first-in-human clinical trials to expedite development of oncology drugs and biologics guidance for industry.

[62] 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.

[63] Wason, J., Stallard, N., Bowden, J. and Jennison, C. (2017). A multi-stage drop-the-losers design for multi-arm clinical trials. *Statistical Methods in Medical Research* **26**(1) 508–524.

[64] Wei, L. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association* **73**(364) 840–843.

[65] Yu, J. and Lai, D. (2021). Interim analysis of sequential estimation-adjusted urn models with sample size re-estimation. *Canadian Journal of Statistics* **49**(4) 1075–1092.

[66] Yuan, Y., Huang, X. and Liu, S. (2011). A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial. *Statistics in Medicine* **30**(11) 1218–1229.

[67] Zang, Y. and Lee, J. J. (2014). Adaptive clinical trial designs in

oncology. *Chinese Clinical Oncology* **3**(4).

[68] Zeymer, U., Suryapranata, H., Monassier, J. P., Opolski, G., Davies, J., Rasmanis, G., Linssen, G., Tebbe, U., Schroder, R., Tiemann, R., Machnig, T., and Neuhaus, K. L. (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 (ESCAM1) trial. *J Am Coll Cardiol* **38** 1644–1650.

[69] Zhang, L.-X., Hu, F. and Cheung, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *The Annals of Applied Probability* **16**(1) 340–369.

[70] 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.

[71] Zhu, H. and Hu, F. (2010). Sequential monitoring of response-adaptive randomized clinical trials. *Annals of Statistics* **38**(4) 2218–2241.

[72] Zhu, H. and Hu, F. (2012). Interim analysis of clinical trials based on urn models. *Canadian Journal of Statistics* **40**(3) 550–568.

[73] Zhu, H., Piao, J., Lee, J. J., Hu, F. and Zhang, L. (2020). Response adaptive randomization procedures in seamless phase II/III clinical trials. *Journal of Biopharmaceutical Statistics* **30**(1) 3–17.

Hongjian Zhu. Virtual Office, Sugar Land, Texas, Statistical Innovation Group, AbbVie Inc, USA. E-mail address: hongjian.zhu@abbvie.com

Jun Yu. Virtual Office, Sugar Land, Texas, Medical Affairs and Health Technology Assessment Statistics, AbbVie Inc, USA. E-mail address: jun.yu@abbvie.com

Dejian Lai. Houston, Texas, Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, USA. E-mail address: dejian.lai@uth.tmc.edu

Li Wang. North Chicago, Illinois, Statistical Innovation Group, AbbVie Inc, USA. E-mail address: li.wang1@abbvie.com