SEQUENTIAL MONITORING OF RESPONSE-ADAPTIVE RANDOMIZED CLINICAL TRIALS

BY HONGJIAN ZHU AND FEIFANG HU[1]

University of Virginia

Clinical trials are complex and usually involve multiple objectives such as controlling type I error rate, increasing power to detect treatment difference, assigning more patients to better treatment, and more. In literature, both response-adaptive randomization (RAR) procedures (by changing randomization procedure sequentially) and sequential monitoring (by changing analysis procedure sequentially) have been proposed to achieve these objectives to some degree. In this paper, we propose to sequentially monitor response-adaptive randomized clinical trial and study it's properties. We prove that the sequential test statistics of the new procedure converge to a Brownian motion in distribution. Further, we show that the sequential test statistics asymptotically satisfy the *canonical joint distribution* defined in Jennison and Turnbull (2000). Therefore, type I error and other objectives can be achieved theoretically by selecting appropriate boundaries. These results open a door to sequentially monitor response-adaptive randomized clinical trials in practice. We can also observe from the simulation studies that, the proposed procedure brings together the advantages of both techniques, in dealing with power, total sample size and total failure numbers, while keeps the type I error. In addition, we illustrate the characteristics of the proposed procedure by redesigning a well-known clinical trial of maternal-infant HIV transmission.

1. Introduction. Clinical trials usually involve multiple competing objectives such as maximizing the power of detecting clinical difference among treatments, minimizing total sample size and protecting more people from possibly inferior treatments. To achieve these objectives, two different techniques have been proposed in literature: (i) the analysis approach—by analyzing the observed data sequentially [sequential monitoring, Jennison and Turnbull (2000)], and (ii) the design approach—by changing the allocation probability sequentially [response-adaptive randomization, Hu and Rosenberger (2006)]. In this paper, we discuss how to combine the two procedures in one clinical trial in order to utilize both of their advantages.

In experiments where data accumulates sequentially, it is natural to conduct a sequential analysis. Sequential techniques originated from a methodology of long

history based on Brownian motion. Wald's classic work about the sequential probability ratio test (SPRT) [Wald (1947)] led to the application of sequential analysis in numerous fields of statistics. Armitage (1957, 1975) introduced sequential methods to clinical studies, which required monitoring results on a patient-by-patient basis. Pocock (1977) proposed sequential monitoring of clinical trials based on a group basis. Since then, many authors have done important work on group sequential studies. These work are summarized in Jennison and Turnbull (2000) and Proschan, Lan and Wittes (2006).

The main advantages of sequential monitoring were listed in Jennison and Turnbull (2000). First, it is ethical to monitor clinical trials sequentially because we could ensure that patients are not exposed to dangerous treatments and we could stop trials as soon as possible if needed. Second, administratively one needs to ensure that the protocol is not violated and the assumption, which the clinical trial is based on, is correct and valid. Third, sequential monitoring can decrease sample size and cost. With all the above advantages, sequential monitoring has now become a standard technique in conducting clinical trials.

The idea of response-adaptive randomization (RAR) can be traced back to Thompson (1933) and Robbins (1952). The play-the-winner rule [Zelen (1969)] and the randomized play-the-winner rule [Wei and Durham (1978)] were proposed to reduce number of patients in the inferior treatments. Hu and Rosenberger (2003) proved theoretically that adaptive randomization can be used to increase statistical efficiency in some clinical trials. In literature, many papers showed its efficient and ethical advantages over fixed designs [Hu and Rosenberger (2006)]. With modern technology and high capability of collecting data, it becomes easier and easier to implement adaptive designs in sequential experiments. Some clinical trials have already implemented the response-adaptive designs [Rout et al. (1993), Tamura et al. (1994), Andersen (1996), etc.].

Bayesian adaptive designs have also been proposed and studied in literature. Berry (2005) provided some comprehensive introduction of Bayesian designs in clinical trials. Recently, Cheng and Shen (2005) proposed to sequentially monitor a Bayesian adaptive design using decision-theoretic approaches and allowing the maximum sample size to be sequentially adjusted by the observed data. Lewis, Lipsky and Berry (2007) proposed a Bayesian decision-theoretic group sequential design for a disease with two possible outcomes based on a quadratic loss function. Wathen and Thall (2008) studied Bayesian adaptive model selection for optimizing group sequential clinical trials. In this paper, we focus on sequential monitoring of response-adaptive randomized clinical trials.

Traditionally, sequential monitoring deals with fixed designs (usually with equal allocation). No systematic study is available about sequential monitoring a sequential experiment using response-adaptive randomization, except a simulation study by Coad and Rosenberger (1999). They found that the expected number of treatment failures can be further reduced by combining the triangular test with

the randomized play-the-winner rule. In this paper, we will study both theoretical properties and finite sample properties of combining sequential monitoring with response-adaptive randomization.

Sequential monitoring procedures use responses to stop or continue a clinical trial. Response-adaptive randomization procedures sequentially estimate the parameters and update the allocation probability of the next patient. To monitor a response-adaptive randomized clinical trial sequentially, one needs to study the two sequential procedures simultaneously. This is conceptually difficult because: (1) the number of patients assigned to each treatment is a random variable at each time point; (2) both the treatment assignments (probabilities) and the estimators of parameters (test statistics) depend on the responses at each time point. These problems arise from the sequential updating of estimators of the parameters and the allocation probability function, which leads to difficulties in finding the joint distribution of sequential test statistics. We overcome above difficulties by (i) approximating these different processes by martingale processes at each time point simultaneously; (ii) then using continuous Gaussian approximation to study these martingale processes simultaneously.

In this paper, we discuss sequential monitoring of doubly adaptive biased coin design proposed by Hu and Zhang (2004) for comparing two treatments. Under widely satisfied conditions, we show that the sequential test statistics converge to (i) a standard Brownian motion in distribution under null hypothesis; and (ii) a drifted Brownian motion in distribution under alternative hypothesis. For a standard Brownian motion, the critical value for fixed type I error rate has been well studied in literature. Therefore, the problem of controlling type I error is theoretically solved. Further, we show that the sequential test statistics satisfy the *canonical joint distribution* defined in Jennison and Turnbull (2000) asymptotically. Hence, one can apply the group sequential methods in the book to response-adaptive randomized clinical trials.

Simulation results support our theoretical founds in terms of type I error and display that sequential monitoring of response-adaptive randomization procedure could increase power and decrease total failure number. Also compared to complete randomization, sequential monitoring of response-adaptive randomization procedure could stop earlier, and thus reduce the actual sample size. In other words, the proposed procedure achieves the goals of both RAR and sequential monitoring. We also redesign an experiment evaluating the effect of zidovudine treatment in reducing the risk of maternal-infant HIV transmission performed by Connor et al. (1994). The proposed procedure can be used to decrease the number of HIV infected people and increase the power comparing to the complete randomization.

In Section 2, we introduce the notation, describe the framework and state the main theorem. In Sections 3 and 4, we use both generated data and real data to compare the proposed procedure with other randomization procedures. Conclusions are in Section 5 and technical proofs are given in the Appendix.

2. Sequential monitoring of response-adaptive randomization procedures.

2.1. *Notation and framework.* We first describe the framework for the randomized adaptive designs. In this article, we consider clinical trials with two treatments 1 and 2. Let $\mathbf{T}_i = (T_{i,1}, T_{i,2}) = (1, 0)$, $i = 1, \ldots, n$, if the ith patient is assigned to treatment 1, and $(0, 1)$ otherwise, where n is the sample size. $\mathbf{N}(n) = (N_1(n), N_2(n))$, where $N_j(n) = \sum_{i=1}^{n} T_{ij}$, $j = 1, 2$, is the number of patients in treatment j. Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)'$, where $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2})$, $i = 1, \ldots, n$, is a random matrix of responses variable and $\mathbf{X}_{ij}$, $j = 1, 2$, are d-dimensional random vectors. Here, only one element of $\mathbf{X}_i$, say $\mathbf{X}_{ij}$, can be observed if the ith patient is assigned to treatment j. We assume that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent and identical distributed with unknown parameter $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_j$ is the corresponding d_j-dimensional parameter vector $(\theta_{j1}, \ldots, \theta_{jd_j})$ of treatment j ($j = 1, 2$). To simplify the notation, we assume that the parameter vectors of both treatments have the same dimension ($d_1 = d_2 = d$). Without loss of generality, we also assume that $\boldsymbol{\theta}_j = E(\mathbf{X}_{ij})$. Otherwise, we can transform $\mathbf{X}$ and treat the transformation as responses to make the former equation hold if such transformation exists. Such transformation usually exists asymptotically. See Gwise, Hu and Hu (2008) and Hu and Zhang (2004) for further discussion.

Let $[nt]$ denote the largest integer that is smaller than or equal to nt for $t \in [0, 1]$. Then $\mathbf{N}([nt]) = (N_1([nt]), N_2([nt]))$ and $N_j([nt]) = \sum_{i=1}^{[nt]} T_{ij}$, $j = 1, 2$. Note that $t = N/n$ when N is the number of patients who have already been enrolled. We introduce the so-called information time t in order to formulate this problem into the Skorohod topology [Ethier and Kurts (1986)]. After $N = [nt]$ patients have been assigned and the responses observed, we use the modified sample means $\hat{\boldsymbol{\theta}}_{[nt]} = (\hat{\boldsymbol{\theta}}_{[nt],1}, \hat{\boldsymbol{\theta}}_{[nt],2})$ to estimate the parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, that is,

$$(2.1) \qquad \hat{\boldsymbol{\theta}}_{[nt],1} = \frac{\sum_{i=1}^{[nt]} T_{i,1} X_{i1} + \boldsymbol{\theta}_{0,1}}{N_1([nt]) + 1} \quad \text{and} \quad \hat{\boldsymbol{\theta}}_{[nt],2} = \frac{\sum_{i=1}^{[nt]} T_{i,2} X_{i2} + \boldsymbol{\theta}_{0,2}}{N_2([nt]) + 1}.$$

Here, we add 1 in the denominator to prevent discontinuity, and add $\boldsymbol{\theta}_{0,j}$, say 0.5, to estimate $\boldsymbol{\theta}_j$ when no patient has been assigned to the treatment j, $j = 1, 2$.

Let $\boldsymbol{\rho} = (\rho_1, \rho_2)$ be the target allocation proportion. Usually $\boldsymbol{\rho}$ is obtained based on some optimal criteria and depends on unknown parameter $\boldsymbol{\theta}$. The selection of $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\theta})$ has been studied by Hayre (1979), Jennison and Turnbull (2000) and Tymofyeyev, Rosenberger and Hu (2007). In practice, the parameters are unknown. Therefore, we have to first estimate them according to previous treatment assignments and responses so that we can target the allocation proportion. We consider a general family of doubly adaptive biased coin design (DBCD) [Eisele and Woodroofe (1995)] here.

Doubly adaptive biased coin design: (i) assign the first $2n_0$ patients to treatment 1 and 2 by some restricted randomization procedures [permuted block or truncated binomial randomization, see Rosenberger and Lachin (2002)]; (ii) when the lth

($l > 2n_0$) patient arrives and all the responses on the previous $l - 1$ patients are available, we compute $\hat{\boldsymbol{\theta}}_{l-1}$ and $\hat{\boldsymbol{\rho}}_{l-1} = \boldsymbol{\rho}(\hat{\boldsymbol{\theta}}_{l-1})$; (iii) then assign the lth patient to treatment 1 with probability

$$g\big(N_1(l-1)/(l-1), \rho_1(\hat{\boldsymbol{\theta}}_{l-1})\big),$$

where $g(s, r): [0, 1] \times [0, 1] \to [0, 1]$ is the allocation function. Hu and Zhang (2004) proposed ($\gamma \geq 0$):

$$g^{(\gamma)}(0, r) = 1,$$

(2.2) $$g^{(\gamma)}(1, r) = 0,$$

$$g^{(\gamma)}(s, r) = \frac{r(r/s)^\gamma}{r(r/s)^\gamma + (1-r)((1-r)/(1-s))^\gamma}.$$

The design has drawn much attention since it was proposed and its advantages and properties can be found in Hu and Rosenberger (2003), Rosenberger and Hu (2004) and Tymofyeyev, Rosenberger and Hu (2007).

   To compare two treatments in clinical trials, one consider a general hypothesis test:

$$H_0 : h(\boldsymbol{\theta}_1) = h(\boldsymbol{\theta}_2) \quad \text{versus} \quad H_1 : h(\boldsymbol{\theta}_1) \neq h(\boldsymbol{\theta}_2),$$

where h is a $\Re^d \to \Re$ function of parameters. In this paper, we assume $h(\boldsymbol{\theta}_j)$ is continuous and twice differentiable in a small neighborhood of $\boldsymbol{\theta}_j$, $j = 1, 2$. If one would like to test the above hypothesis at time point $t \in (0, 1]$, it is natural to construct the test statistic as

(2.3) $$Z_t\bigg(\frac{\mathbf{N}([nt])}{[nt]}, \hat{\boldsymbol{\theta}}([nt])\bigg) = \frac{h(\hat{\boldsymbol{\theta}}_1([nt])) - h(\hat{\boldsymbol{\theta}}_2([nt]))}{\sqrt{\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}.$$

Here, $\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt])))$ and $\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt])))$ are some consistent estimators of the variances of $h(\hat{\boldsymbol{\theta}}_1([nt]))$ and $h(\hat{\boldsymbol{\theta}}_2([nt]))$, respectively. There is no covariance term on the denominator since the two terms on the numerator are asymptotically independent [Hu, Rosenberger and Zhang (2006)]. Without loss of generality, we also assume that for some functions v_1 and v_2

$$[nt]\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_j([nt]))) = v_j\bigg(\frac{\mathbf{N}([nt])}{[nt]}, \hat{\boldsymbol{\theta}}([nt])\bigg)(1 + o(1)) \qquad \text{a.s. } j = 1, 2.$$

It is easy to see that both $v_j(\mathbf{y}, \mathbf{z})$ and $Z_t(\mathbf{y}, \mathbf{z})$ are $\Re^{2+2d} \to \Re$ function, where $\mathbf{y}$ is a two-dimensional vector and $\mathbf{z}$ is a $2d$-dimensional vector. Examples of using this formulation are discussed in Section 2.3.

2.2. *Main results.* Based on the notation in Section 2.1, we observe the random processes $(\mathbf{T}_1, \ldots, \mathbf{T}_{[nt]})$, $(\mathbf{X}_1, \ldots, \mathbf{X}_{[nt]})$, $\mathbf{N}([nt])$, $\hat{\boldsymbol{\theta}}_{[nt]}$, $\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}_{[nt]})$ and Z_t at time point t. When a response-adaptive randomization procedure is used, these random processes have the following characteristics different from those in fixed designs:

(1) The allocation $(\mathbf{N}([nt]))$ at any time t is a random vector instead of a constant in fixed designs.
(2) The allocation $(\mathbf{N}([nt]))$ and $(\mathbf{T}_1, \ldots, \mathbf{T}_{[nt]})$ are not independent with the responses $(\mathbf{X}_1, \ldots, \mathbf{X}_{[nt]})$ and the parameter estimator vector $\hat{\boldsymbol{\theta}}_{[nt]}$.
(3) The elements $\hat{\boldsymbol{\theta}}_1[nt]$ and $\hat{\boldsymbol{\theta}}_2[nt]$ depend on each other at any given time $t \in (0, 1]$.

These differences directly lead to difficulties in deriving the joint distributions of sequential testing statistics.

To sequentially monitor a clinical trial, we need to figure out how to control the type I error. The answer to this question relies on the derivation of the asymptotical joint distribution of the sequential statistics and right choices of the boundaries. Before we give the main theorem, we need the following conditions for the response $\mathbf{X}$, target allocation $\rho(\boldsymbol{\theta})$, allocation function g and the function $v_j(\mathbf{y}, \mathbf{z})$, $j = 1, 2$.

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

(A2) $g(s, r)$ is jointly continuous and twice differentiable at (ρ_1, ρ_1);

(A3) $g(r, r) = r$ for all $r \in (0, 1)$ and $g(s, r)$ is strictly decreasing in s and strictly increasing in r on $(0, 1) \times (0, 1)$;

(A4) $\boldsymbol{\rho}(\mathbf{z})$ is a continuous function and twice continuously differentiable in a small neighborhood of $\boldsymbol{\theta}$;

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

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

REMARK 2.1. All the conditions are widely satisfied. An example of a design which satisfies these conditions is DBCD in Hu and Zhang (2004). Condition (A1) is used to ensure the consistency of the procedure and asymptotic normality of the allocation proportions. Condition (A3) forces the actual allocation proportion to approach the theoretically targeted one. Conditions (A4), (A5) and (A6) are satisfied in all the examples in Chapter 5 of Hu and Rosenberger (2006).

THEOREM 2.1. *Let* $B_t = \sqrt{t} Z_t$ *in the space* $D_{[0,1]}$ *with Skorohod topology. Assume conditions* (A1)–(A6) *are satisfied. Then we have the following two results*:

(i) *Under* H_0, B_t *is asymptotically a standard Brownian motion in distribution.*

(ii) *Under H_1, $B_t - \sqrt{n}\mu t$ is asymptotically a standard Brownian motion in distribution*, *where*

$$\mu = \frac{(h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2))}{\sqrt{v_1(\boldsymbol{\rho}, \boldsymbol{\theta}) + v_2(\boldsymbol{\rho}, \boldsymbol{\theta})}}.$$

Based on Theorem 2.1, we can obtain the asymptotical distribution of the sequence of test statistics $\{Z_{t_1}, \ldots, Z_{t_K}\}$, where $0 \leq t_1 \leq t_2 \leq \cdots \leq t_K \leq 1$. Because $Z_{t_i} = (\sqrt{t_i})^{-1} B_{t_i}$, we have asymptotically:

(i) $\{Z_{t_1}, \ldots, Z_{t_K}\}$ is multivariate normal;
(ii) $E Z_{t_i} = \mu \sqrt{n t_i}$; and
(iii) $\mathrm{Cov}(Z_{t_i}, Z_{t_j}) = \sqrt{[n t_i]/[n t_j]}, 0 \leq t_i \leq t_j \leq 1$.

Therefore, the sequence of test statistics $\{Z_{t_1}, \ldots, Z_{t_K}\}$ has the asymptotical *canonical joint distribution* defined in Jennison and Turnbull (2000).

REMARK 2.2. Based on the canonical joint distribution of the sequence of test statistics $\{Z_{t_1}, \ldots, Z_{t_K}\}$, we can see that the doubly adaptive biased coin design has a simple form of information time, which is just the proportion of the sample size enrolled. This is because the DBCD consistently allocates same proportion of patients to different treatments from the beginning to the end asymptotically. We conjecture that this simple form of information time is true for most response-adaptive randomization procedures.

Based on Theorem 2.1, we can easily choose the correct critical values for the asymptotic Brownian process, so that the inflation of the type I error will be avoided. Moreover, we can also make use of all the well-known properties of Brownian process to do further analysis on the process of sequentially monitoring a response-adaptive randomization procedure. Because $\{Z_{t_1}, \ldots, Z_{t_K}\}$ satisfies the *canonical joint distribution* asymptotically, we can apply the sequential techniques in Chapters 2, 3, 4, 5, 6, 7 of Jennison and Turnbull (2000) to response-adaptive randomized clinical trials. We may also apply different types of spending functions to monitor a response-adaptive randomized clinical trial sequentially. Here, we will use α spending functions proposed by Lan and DeMets (1983).

Any increasing function $\alpha(t)$ defined on $[0, 1]$ with $\alpha(0) = 0$ and $\alpha(1) = \alpha$ is called a α spending function. We spend $\alpha(t_i) - \alpha(t_{i-1})$ of the total type I error rate at time point t_i, so that $\alpha(t_i)$ has been spent after this point. For time $t_i, i = 1, 2, \ldots$, we can sequentially obtain the boundaries. This method does not require the predetermined number of looks and equally spaced looks. We can perform the interim monitor anytime during the trial. Such a procedure is usually preferred by *Data* and *Safety Monitoring Boards* (DSMB). Proschan, Lan and Wittes (2006) provided three special spending functions. The first one approximates the O'Brien–Fleming boundaries [O'Brien and Fleming (1979)]

$$\alpha_1(t) = 2\{1 - \Phi(z_{\alpha/2}/t^{1/2})\}.$$

The second one is the linear spending function:

$$\alpha_2(t) = \alpha t.$$

The third one approximates the Pocock boundaries [Pocock (1982)]:

$$\alpha_3(t) = \alpha \ln\{1 + (e - 1)t\}.$$

The O'Brien–Fleming-like function spends little of the type I error at early looks. Consequently, the boundary for the last look is very close to what it would have been without sequential monitoring. Conversely, the Pocock-like function rejects the null hypothesis easier with smaller boundaries for early looks and then has to use a reasonably large critical value at the end to keep the type I error. The linear function is between these two. Therefore, the three functions above represent three typical types of spending function. Finally, it is worth mentioning that these three spending functions are corresponding to the process Z_t.

2.3. *Examples*. Here, we use two examples to illustrate how to sequentially monitor the response-adaptive randomization procedures based on Theorem 2.1.

EXAMPLE 1 (Continuous responses from normal populations). Suppose the responses of the two treatments are from two normal distributions $Y_{i1} \sim N(\mu_1, \sigma_1^2)$ and $Y_{i2} \sim N(\mu_2, \sigma_2^2)$, $i = 1, \ldots, n$. We would like to compare μ_1 and μ_2. In this case, $\boldsymbol{\theta}_1 = (\mu_1, \sigma_1^2 + \mu_1^2)$, $\boldsymbol{\theta}_2 = (\mu_2, \sigma_2^2 + \mu_2^2)$, $\mathbf{X}_{ij} = (Y_{ij}, Y_{ij}^2)$ and $h(\boldsymbol{\theta}_j) = \theta_{j1} = \mu_j$, $j = 1, 2$. Then the hypotheses are

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2.$$

Let target allocation proportion be the Neyman allocation [Jennison and Turnbull (2000)] with

$$(2.4) \qquad \rho_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2} \quad \text{and} \quad \rho_2 = 1 - \rho_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2}.$$

We can use other target allocation proportions, for example, the optimal allocation proportion [Zhang and Rosenberger (2006)] and the D_A-optimal allocation proportion [Gwise, Hu and Hu (2008)]. The sequential statistics $Z_t(\mathbf{y}, \mathbf{z})$ is a function from $\mathfrak{R}^6$ to $\mathfrak{R}$:

$$Z_t(\mathbf{y}, \mathbf{z}) = \frac{z_{11} - z_{21}}{\sqrt{(z_{12} - z_{11}^2)/([nt]y_1) + (z_{22} - z_{21}^2)/([nt]y_2)}},$$

where $\mathbf{y} = \mathbf{N}([nt])/[nt]$ and $\mathbf{z} = \hat{\boldsymbol{\theta}} = (\hat{\theta}_{11}([nt]), \hat{\theta}_{12}([nt]), \hat{\theta}_{21}([nt]), \hat{\theta}_{22}([nt]))$. It is easy to see that $h(\hat{\boldsymbol{\theta}}_1([nt])) = \hat{\mu}_1([nt])$ and $h(\hat{\boldsymbol{\theta}}_2([nt])) = \hat{\mu}_2([nt])$. Also the natural variance estimators are

$$\widehat{\text{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt]))) = \frac{\hat{\sigma}_1^2([nt])}{N_1([nt])} \quad \text{and} \quad \widehat{\text{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt]))) = \frac{\hat{\sigma}_2^2([nt])}{N_2([nt])},$$

where $\hat{\sigma}_1^2([nt])$ and $\hat{\sigma}_2^2([nt])$ are the usual unbiased estimators of σ_1^2 and σ_2^2 based on the first $[nt]$ responses ($N_1([nt])$ from treatment 1 and $N_2([nt])$ from treatment 2), respectively. Therefore,

$$v_1(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{\sigma_1^2}{\rho_1} \quad \text{and} \quad v_2(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{\sigma_2^2}{\rho_2}.$$

The test statistic is then

$$(2.5) \qquad Z_t = \frac{\hat{\mu}_1([nt]) - \hat{\mu}_2([nt])}{\sqrt{\hat{\sigma}_1^2([nt])/N_1([nt]) + \hat{\sigma}_2^2([nt])/N_2([nt])}}.$$

Then based on Theorem 2.1, the joint distribution of $B_t = \sqrt{t}Z_t$ is asymptotically a standard Brownian process under H_0. Under H_1, $B_t - \sqrt{n}\mu t$ is asymptotically a standard Brownian motion in distribution, where

$$\mu = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/\rho_1 + \sigma_2^2/(1 - \rho_1)}}.$$

EXAMPLE 2 (Binary responses). Assume $Y_{i1} \sim \text{Bin}(1, p_1)$ and $Y_{i2} \sim \text{Bin}(1, p_2)$, $i = 1, \ldots, n$, and we would like to compare p_1 and p_2. In this case, $\boldsymbol{\theta}_1 = (p_1)$, $\boldsymbol{\theta}_2 = (p_2)$, $\mathbf{X}_{ij} = (Y_{ij})$ and $h(\boldsymbol{\theta}_j) = \theta_{j1}$, $j = 1, 2$. The hypotheses are

$$H_0 : p_1 = p_2 \quad \text{versus} \quad H_1 : p_1 \neq p_2.$$

Three common target allocations are: (i) Neyman allocation,

$$(2.6) \qquad \begin{aligned} \rho_1 &= \frac{\sqrt{p_1(1 - p_1)}}{\sqrt{p_1(1 - p_1)} + \sqrt{p_2(1 - p_2)}} \qquad \text{and} \\ \rho_2 &= \frac{\sqrt{p_2(1 - p_2)}}{\sqrt{p_1(1 - p_1)} + \sqrt{p_2(1 - p_2)}}; \end{aligned}$$

(ii) optimal allocation proposed by Rosenberger et al. (2001),

$$(2.7) \qquad \rho_1 = \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}} \quad \text{and} \quad \rho_2 = \frac{\sqrt{p_2}}{\sqrt{p_1} + \sqrt{p_2}};$$

(iii) Urn allocation [Wei and Durham (1978)],

$$(2.8) \qquad \rho_1 = \frac{q_2}{q_1 + q_2} \quad \text{and} \quad \rho_2 = \frac{q_1}{q_1 + q_2}.$$

Neyman allocation is a commonly discussed allocation which is related to the efficiency issue in the field of response-adaptive randomization procedures. We studied sequential monitoring of response-adaptive designs with Neyman allocation in order to show that our proposed procedure is able to achieve various objects.

In this case, $Z_t(\mathbf{y}, \mathbf{z})$ is a function from $\Re^4$ to $\Re$:

$$Z_t(\mathbf{y}, \mathbf{z}) = \frac{z_{11} - z_{21}}{\sqrt{z_{11}(1 - z_{11})/([nt]y_1) + z_{21}(1 - z_{21})/([nt]y_2)}},$$

where $\mathbf{y} = (N_1([nt])/[nt], N_2([nt])/[nt])$, $\mathbf{z} = (\hat{\theta}_{11}([nt]), \hat{\theta}_{21}([nt]))$, $h(\hat{\boldsymbol{\theta}}_1([n \times t])) = \hat{p}_1([nt])$ and $h(\hat{\boldsymbol{\theta}}_2([nt])) = \hat{p}_2([nt])$. The corresponding variance estimators are

$$\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt]))) = \frac{\hat{p}_1([nt])(1 - \hat{p}_1([nt]))}{N_1([nt])}$$

and

$$\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt]))) = \frac{\hat{p}_2([nt])(1 - \hat{p}_2([nt]))}{N_2([nt])}.$$

Therefore,

$$v_1(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{p_1(1 - p_1)}{\rho_1} \quad \text{and} \quad v_2(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{p_2(1 - p_2)}{\rho_2}.$$

The test statistic is

(2.9)
$$\begin{aligned} Z_t = &\left(\hat{p}_1([nt]) - \hat{p}_2([nt])\right) \\ &\times \left(\sqrt{\hat{p}_1([nt])\frac{1 - \hat{p}_1([nt])}{N_1([nt])} + \hat{p}_2([nt])\frac{1 - \hat{p}_2([nt])}{N_2([nt])}}\right)^{-1}. \end{aligned}$$

Then $B_t = \sqrt{t}Z_t$ converges to a standard Brownian process in distribution under H_0. Under H_1, $B_t - \sqrt{n}\mu t$ is asymptotically a standard Brownian motion in distribution, where

$$\mu = \frac{p_1 - p_2}{\sqrt{p_1(1 - p_1)/\rho_1 + p_2(1 - p_2)/(1 - \rho_1)}}.$$

Theorem 2.1 can be applied to different situations such as the examples considered in Chapter 5 of Hu and Rosenberger (2006). In Examples 1 and 2, now assume we would like to look at the process at three points: $t_1 = 0.2$, $t_2 = 0.5$ and $t_3 = 1$. Then we can use the corresponding critical values from the three spending functions [Proschan, Lan and Wittes (2006)] in the last subsection for Z_t to keep the overall type I error 0.05: O'Brien–Fleming-like boundaries (4.877, 2.963, 1.969), linear boundaries (2.576, 2.377, 2.141) and Pocock-like boundaries (2.438, 2.333, 2.225).

TABLE 1
Example 1 with Neyman allocation, $\mu_1 = \mu_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 2$

Critical values	Randomization	Type I error	$\hat{\rho}_1$ (s.e.)
B–F-like	DBCD	0.055	0.333 (0.020)
B–F-like	CR	0.052	0.500 (0.022)
Linear	DBCD	0.048	0.333 (0.020)
Linear	CR	0.053	0.500 (0.023)
Pocock-like	DBCD	0.051	0.332 (0.020)
Pocock-like	CR	0.052	0.500 (0.023)

3. Simulation study. In Section 2, we obtained the asymptotical distribution of the test statistic Z_t. In this section, we will use the two examples in Section 2 to study the finite sample properties of the proposed procedure.

In Examples 1 and 2, we use the doubly adaptive biased coin design with Hu and Zhang's allocation function in (2.2) and $\gamma = 2$ is used. In Tables 1–5, we use the same total sample size 500. The first 50 patients ($n_0 = 25$) are randomly assigned to treatments 1 and 2 by using permuted block randomization. Then, for the lth ($l > 50$) patient, the unknown parameters are estimated by using (2.1) based on the first $l - 1$ responses with $\theta_{0,1} = \theta_{0,2} = 0.5$. For normal responses in Example 1, we estimate σ_1^2 and σ_2^2 by using the standard unbiased estimators based on the first $l - 1$ responses.

For simplicity, we look at the test at three time points [$n_1 = 100$ ($t_1 = 0.2$), $n_2 = 250$ ($t_2 = 0.5$) and $n = 500$ ($t_3 = 1$)]. Then the three sets of spending function boundaries in Section 2.3 are used to ensure $\alpha = 0.05$. For each spending function, the first row in the table is for DBCD and the second row is for complete randomization (denoted as CR in the tables). All the simulations are based on 5000 replications.

In Table 1, we simulate Example 1 with two normal responses $N(1, 1)$ and $N(1, 2)$ by using the Neyman allocation (2.4). We find that the type I error of sequentially monitoring the response-adaptive randomization procedure and com-

TABLE 2
Example 2 with optimal allocation, $p_1 = p_2 = 0.5$

Critical values	Randomization	Type I error	$\hat{\rho}_1$ (s.e.)
B–F-like	DBCD	0.051	0.500 (0.016)
B–F-like	CR	0.046	0.500 (0.023)
Linear	DBCD	0.055	0.500 (0.019)
Linear	CR	0.061	0.500 (0.023)
Pocock-like	DBCD	0.056	0.500 (0.019)
Pocock-like	CR	0.050	0.500 (0.022)

TABLE 3
Example 1 with Neyman allocation, $\mu_1 = 1$, $\mu_2 = 1.4$, $\sigma_1 = 1$, $\sigma_2 = 2$

Critical values	Randomization	Power	$\hat{\rho}_1$ (s.e.)	N_1	N_2	N_3
B–F-like	DBCD	0.847	0.333 (0.021)	2	1013	3222
B–F-like	CR	0.807	0.500 (0.024)	1	842	3193
Linear	DBCD	0.812	0.332 (0.027)	594	1429	2035
Linear	CR	0.765	0.500 (0.028)	477	1380	1970
Pocock-like	DBCD	0.792	0.332 (0.028)	741	1443	1774
Pocock-like	CR	0.738	0.500 (0.028)	544	1309	1835

plete randomization are both well kept at the 0.05 level. We also report the mean and standard deviation of actual allocation proportion ($\hat{\rho}_1$) for treatment 1 [$N(1, 1)$]. We find that the mean agrees with Neyman allocation and the standard deviation is reasonably small for DBCD. This indicates that the DBCD is able to target the theoretical targeted allocation proportion very well. In Table 2, we simulate the Example 2 with two binary responses $p_1 = p_2 = 0.5$ and the target allocation is the optimal allocation (2.7). We obtain the same conclusion as Table 1. We have also done simulations for some other cases, and similar results are obtained. These numerical results indicate that sequential monitoring of the response-adaptive randomization will not inflate the type I error with the appropriate boundaries based on Theorem 2.1.

Next, we show other advantages of the sequential monitoring of the response-adaptive randomization procedure. In Table 3, we simulate Example 1 with two normal responses $N(1, 1)$ and $N(1.4, 2)$ using Neyman allocation (2.4) as the target allocation that maximizes the power. The power of the sequential monitoring of the response-adaptive randomization procedure is about 5%–8% higher than sequentially monitoring the complete randomization. N_i in the table is the number of rejections at the ith look. Rejection at the first two looks means stopping the trial earlier. DBCD with sequential monitoring obviously stops the trial earlier than complete randomization.

TABLE 4
Example 2 with urn allocation, $p_1 = 0.5$, $p_2 = 0.625$

Critical values	Randomization	Power	$\hat{\rho}_1$ (s.e.)	N_1	N_2	N_3	Total failures (s.e.)
B–F-like	DBCD	0.811	0.426 (0.033)	4	839	3214	211 (13)
B–F-like	CR	0.811	0.500 (0.024)	1	839	3215	217 (13)
Linear	DBCD	0.762	0.421 (0.041)	503	1396	1912	206 (14)
Linear	CR	0.767	0.500 (0.029)	521	1300	2016	212 (14)
Pocock-like	DBCD	0.749	0.421 (0.042)	609	1325	1809	205 (14)
Pocock-like	CR	0.738	0.501 (0.029)	603	1312	1773	211 (15)

TABLE 5
Example 2 with optimal allocation, $p_1 = 0.5$, $p_2 = 0.625$

Critical values	Randomization	Power	$\hat{\rho}_1$ (s.e.)	N_1	N_2	N_3	Total failures (s.e.)
B–F-like	DBCD	0.810	0.471 (0.017)	4	863	3185	214 (12)
B–F-like	CR	0.805	0.501 (0.024)	4	795	3229	218 (13)
Linear	DBCD	0.768	0.468 (0.022)	520	1354	1964	210 (14)
Linear	CR	0.762	0.500 (0.029)	474	1367	1971	214 (14)
Pocock-like	DBCD	0.754	0.469 (0.023)	673	1309	1787	210 (14)
Pocock-like	CR	0.749	0.500 (0.03)	602	1351	1793	213 (15)
1.96	DBCD	0.805	0.472 (0.015)	NA	NA	NA	217 (11)
1.96	CR	0.802	0.500 (0.022)	NA	NA	NA	221 (11)

In Table 4, we simulate Example 2 with two binary responses $p_1 = 0.5$ and $p_2 = 0.625$ using the urn allocation (2.8) as the target allocation that assigns more people to the better treatment. If we reject the null hypothesis at the first two looks, we assign all the remaining patients to the estimated better treatment and count the total failure number. We do this only for the comparison in the simulation study. In a real clinical trial, we stop the trial if the null hypothesis is rejected at an interim look. From the mean total failure number, the DBCD with sequential monitoring has lower failure numbers than complete randomization for each type of spending function. N_1, N_2, and N_3 show that our methods stop the trial a little earlier and the power is almost the same.

In Table 5, we simulate Example 2 with two binary responses $p_1 = 0.5$ and $p_2 = 0.625$ using the optimal allocation (2.7) used to maximize the power while keeping the total failure number. We deal with the remaining patients in the same way as in Table 4 if we reject the null hypothesis at the first two looks. We find that sequential monitoring of the response-adaptive randomization procedure can achieve the aim of optimal allocation. Its power is larger and its failure number is less than the complete randomization procedure. In this table, we also do the simulation without sequential monitoring. That is, we only look at the test once at the end of the trial and the critical value is 1.96 for the nominal significance level 0.05. We report it at the last two rows. It is obvious that sequential monitoring can reduce the total failures.

Based on the simulation results, we can see the advantages of sequentially monitoring response-adaptive randomized clinical trials: (i) controlling type I error well; (ii) reducing the total number of failures; (iii) increasing power; and (iv) stopping the trail earlier (reducing total sample size).

4. Re-designing the HIV transmission trial. Maternal-infant transmission is the primary means by which infants are infected by HIV virus. Connor et al. (1994) reported a trial to evaluate the drug AZT (Zidovudine treatment) in reducing the risk of maternal-infant HIV transmission. In this clinical trial, 477 HIV-infected

TABLE 6
Re-designed the HIV trial with full sample size

Target allocation	Critical values	$\hat{\rho}_1$ (s.e.)	Power	Total failures (s.e.)
CR	linear	0.500 (0.039)	0.999	60.1 (11.1)
CR	1.96	0.501 (0.023)	0.999	80.7 (8.2)
Urn allocation	linear	0.751 (0.062)	0.996	52.3 (9.2)
Optimal allocation	linear	0.527 (0.021)	0.997	56.4 (10.8)

pregnant women were enrolled from April 1991 to December 1993 and assigned to the Zidovudine treatment group and placebo group with a 50–50 randomization scheme. This experiment was a randomized, double-blind and placebo-controlled trial. 239 were allocated to the treatment group and 238 to the placebo group. At the end of the trial, 8.3% of the infant from the treatment group were infected by the HIV virus, while 25.5% from the placebo group were infected.

In Table 6, we redesign the study by sequential monitoring of both complete randomization (the first two rows in the table) and response-adaptive randomization [DBCD (2.2) with $\gamma = 2$] (the last three rows in the table). We assume the success rate for the treatment group is $p_1 = 0.917$ and that for the placebo group is $p_2 = 0.745$ (as reported in the original paper). We look at the test at the three same time points as mentioned in the last section, $n_1 = 95$ ($t_1 = 0.2$), $n_2 = 143$ ($t_2 = 0.5$) and $n = 239$ ($t_3 = 1$). The boundary we use is the linear spending function (2.576, 2.377, 2.141) except the second row in the table where we do the equal allocation without sequential monitoring. We report the actual allocation proportion for the treatment group, power and the total HIV-infected number. As before, if we reject the null hypothesis at the first two looks, we will assign all the remaining patients to the estimated better treatment. We find that the sequential monitoring technique will decrease the HIV-infected number dramatically from the first two rows. Response-adaptive randomization technique will also reduce the HIV-infected number compared to the complete randomization. Sequential monitoring DBCD while targeting at the urn allocation has the least HIV-infected number, which agrees with the aim of urn allocation.

In Table 7, we reduce the full sample size to 245 (to achieve power 0.95 for complete randomization) and keep all the other settings unchanged. We obtain the same conclusion about the HIV-infected number as in Table 6. We also find that targeting optimal allocation with DBCD has slightly higher power than targeting equal allocation when sequential monitoring is used. Targeting urn allocation with DBCD has slightly less power but the HIV-infected number in this way is the least. Overall, sequential monitoring of the response-adaptive randomization procedure is better than that of complete randomization, since it reduces the HIV-infected number and remains good power.

TABLE 7
Re-designed the HIV trial with sample size n = 245

Target allocation	Critical values	$\hat{\rho}_1$ (s.e.)	Power	Total failures (s.e.)
CR	B–F-like	0.500 (0.036)	0.947	40.1 (7.0)
CR	linear	0.501 (0.042)	0.942	36.6 (7.5)
CR	1.96	0.500 (0.032)	0.958	43.1 (5.8)
Urn allocation	B–F-like	0.745 (0.068)	0.920	30.7 (5.9)
Urn allocation	linear	0.747 (0.074)	0.885	29.3 (6.1)
Optimal allocation	B–F-like	0.528 (0.023)	0.952	36.8 (6.7)
Optimal allocation	linear	0.529 (0.025)	0.945	32.8 (7.3)

5. Conclusion remarks. Now sequential monitoring becomes a standard technique in clinical trials. To apply response-adaptive randomization in clinical trials, it is important to know how to sequentially monitor adaptive randomized trials. In this paper, we overcome this hurdle and show the advantages of sequential monitoring response-adaptive randomized clinical trials both theoretically and numerically. We use a Gaussian process in the Skorohod topology to describe the relationship between the allocation and parameter estimators. One of the main contributions of this paper is to show that sequential statistics can be asymptotically approximated by a Brownian process in distribution under both null and alternative hypotheses. Further, we find that the sequential test statistics satisfy the *canonical joint distribution* asymptotically. Consequently, the results of this paper not only solve the problem of preserving a preset type I error but may lead to many area of potential future research.

We have studied how to sequentially monitor a clinical trial based on doubly adaptive biased coin design proposed by Eisele and Woodroofe (1995) and Hu and Zhang (2004). Another important family of response-adaptive randomization procedure is based on urn models, which include randomized play-the-winner rule [Wei and Durham (1978)], generalized Friedman's urn models [Athreya and Karlin (1968), Bai and Hu (2005)], drop-the-loser rule [Ivanova (2003)], sequential estimation-adjusted urn models [Zhang, Hu and Cheung (2006)], etc. The technique used in this paper opens a door to study the properties of sequential monitoring of clinical trials based on these urn models or the efficient randomized adaptive designs [Hu, Zhang and He (2009)]. We leave this for future study.

In this paper, we have used α-spending function to calculate the critical boundaries. Because the sequential test statistics satisfy the *canonical joint distribution* asymptotically, we can implement all the sequential techniques introduced in Jennison and Turnbull (2000) based on this canonical form. Also we can use the optimal spending functions in Anderson (2007), or the beta spending functions in DeMets (2006). We also leave the details for future research.

APPENDIX: PROOFS

First, we introduce some further notation. For a function $\boldsymbol{\eta}(\mathbf{u}, \mathbf{w}) : \mathfrak{R}^L \times \mathfrak{R}^M \to \mathfrak{R}^2$, we denote the partial derivative matrices as

$$\nabla_u(\boldsymbol{\eta}) = \left(\frac{\partial \eta_k}{\partial u_i}; i = 1, \ldots, L, k = 1, 2 \right)_{L \times 2}$$

and

$$\nabla_w(\boldsymbol{\eta}) = \left(\frac{\partial \eta_k}{\partial w_j}; j = 1, \ldots, M, k = 1, 2 \right)_{M \times 2}.$$

Let $H = \nabla_r(g(r, s), 1 - g(r, s))|_{(\rho_1, \rho_1)}$ and $E = \nabla_s(g(r, s), 1 - g(r, s))|_{(\rho_1, \rho_1)}$ be the partial derivative matrices of the allocation function g. Further, let $V = \text{diag}(\text{var}(\mathbf{X}_{11})/\rho_1, \text{var}(\mathbf{X}_{12})/\rho_2)$, $\Sigma_3 = (\nabla(\boldsymbol{\rho})|_{\boldsymbol{\theta}})' V \nabla(\boldsymbol{\rho})|_{\boldsymbol{\theta}}$, $\Sigma_1 = \text{diag}(\boldsymbol{\rho}) - \boldsymbol{\rho}' \boldsymbol{\rho}$ and $\Sigma_2 = E' \Sigma_3 E$. In Hu and Zhang ([2004](#)), they studied the asymptotic properties of $\mathbf{N}(n)$, $\hat{\boldsymbol{\rho}}(n)$ and $\hat{\boldsymbol{\theta}}(n)$ at the end of the trial. Based on their results, one can do the corresponding statistical inference after observing all responses of the clinical trial. To monitor the response-adaptive randomized trial sequentially, we need to know the theoretical properties of the process $\mathbf{N}([nt])$ and $\hat{\boldsymbol{\theta}}([nt])$ for any given $t \in (0, 1]$. To do this, we start with Lemma A.1.

LEMMA A.1.   *Let W_{1t} and W_{2t} be two independent standard two-dimensional Brownian processes. $\mathbf{N}([nt])$, $\hat{\boldsymbol{\theta}}([nt])$, $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ are defined as in Section 2. Under the conditions of Theorem 2.1, we have*

$$(A.1) \qquad n^{-1/2}([nt]) \left(\frac{\mathbf{N}([nt])}{[nt]} - \boldsymbol{\rho}, \hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta} \right) \to (G_t, W_{2t} V^{1/2})$$

in distribution in the space $D_{[0,1]}$ with the Skorohod topology, where the Gaussian process

$$(A.2) \quad G_t = \int_0^t (dW_{1x}) \Sigma_1^{1/2} \left(\frac{t}{x} \right)^H + \int_0^t (dW_{2x}) \Sigma_2^{1/2} \left[\int_x^t \frac{1}{y} \left(\frac{t}{y} \right)^H dy \right],$$

which is the solution of the stochastic differential equation

$$dG_t = (dW_{1t}) \Sigma_1^{1/2} + \frac{W_{2t} \Sigma_2^{1/2}}{t} dt + \frac{G_t}{t} H dt \qquad \text{with } G_0 = 0,$$

and a^H is the matrix power function defined as

$$a^H = e^{H \ln a} = \sum_{j=0}^{\infty} \frac{(\ln a)^j}{j!} H^k.$$

PROOF.   It is worth noting that the response-adaptive design in Theorem 2.1 satisfies all the conditions of Hu and Zhang (2004). So all the results in Hu and Zhang (2004) are valid. We will prove this lemma by using the weak convergence of the martingale [cf. Theorem 4.1 of Hall and Heyde (1980)]. To do this, we first approximate the process $(\frac{\mathbf{N}([nt])}{[nt]} - \boldsymbol{\rho}, \hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta})$ by a martingale and then prove the following two facts: (1) Lindeberg condition holds for the approximated martingale process; and (2) the limiting covariance of $n^{-1/2}([nt])(([nt])^{-1}\mathbf{N}([nt]) - \boldsymbol{\rho}, \hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta})$ agrees with that of $(G_t, W_{2t}V^{1/2})$.

Now, we use the martingale approximation of $\mathbf{N}(n) - n\boldsymbol{\rho}$ and $\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}$ from Hu and Zhang (2004). Let $\mathcal{F}_m = \sigma(\mathbf{T}_1, \ldots, \mathbf{T}_m, \mathbf{X}_1, \ldots, \mathbf{X}_m)$ be the σ-field generated by the previous m stages. Then under $\mathcal{F}_{m-1}$, $\mathbf{T}_m$ and $\mathbf{X}_m$ are independent, and

$$E[T_{m1}|\mathcal{F}_{m-1}] = g\left(\frac{N_1(m-1)}{m-1}, \hat{\rho}_1(m-1)\right).$$

Let $\mathbf{Q}_n = \sum_{m=1}^n \Delta\mathbf{Q}_m$, where $\Delta\mathbf{Q}_m = (\Delta\mathbf{Q}_{m,1}, \Delta\mathbf{Q}_{m,2}) = (\Delta Q_{m,1k}, \Delta Q_{m,2k};$ $k = 1, \ldots, d)$ and $\Delta Q_{m,jk} = T_{m,j}(X_{m,jk} - \theta_{jk})/\rho_j, j = 1, 2$. Then $\mathbf{Q}_n = O(\sqrt{n\log\log n})$ a.s. is a sequence of martingales and we can prove

$$(A.3) \qquad \hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta} = \frac{\mathbf{Q}_n}{n} + O\left(\frac{\log\log n}{n}\right) \qquad \text{a.s.}$$

Let $\mathbf{M}_n = \sum_{m=1}^n \Delta\mathbf{M}_m$, where $\Delta\mathbf{M}_m = \mathbf{T}_m - E[\mathbf{T}_m|\mathcal{F}_{m-1}]$, and $B_{n,m}$ as defined in Hu and Zhang (2004), then

$$\mathbf{N}(n) - n\boldsymbol{\rho} = \sum_{m=1}^n \Delta\mathbf{M}_m B_{n,m} + \sum_{m=1}^n \Delta\mathbf{Q}_m \nabla(\boldsymbol{\rho})\Big|_{\boldsymbol{\theta}} E \sum_{k=m}^n \frac{1}{k} B_{n,k} + o(n^{-1/2-\delta/3})$$

$$:= \mathbf{U}_n + o(n^{-1/2-\delta/3})$$

almost surely, where $\mathbf{U}_n$ is a sum of martingale differences.

We can approximate the process $\mathbf{N}([nt]) - [nt]\boldsymbol{\rho}$ and $\hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta}$ (for any point $t \in (0, 1]$) similarly as $\mathbf{N}(n) - n\boldsymbol{\rho}$ and $\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}$. We obtain

$$(A.4) \qquad \hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta} = \frac{\mathbf{Q}_{[nt]}}{[nt]} + O\left(\frac{\log\log[nt]}{[nt]}\right) \qquad \text{a.s.}$$

and

$$\mathbf{N}([nt]) - [nt]\boldsymbol{\rho}$$

$$= \sum_{m=1}^{[nt]} \Delta\mathbf{M}_m B_{[nt],m} + \sum_{m=1}^{[nt]} \Delta\mathbf{Q}_m \nabla(\boldsymbol{\rho})\Big|_{\boldsymbol{\theta}} E \sum_{k=m}^{[nt]} \frac{1}{k} B_{[nt],k} + o(([nt])^{-1/2-\delta/3})$$

$$:= \mathbf{U}_{[nt]} + o(([nt])^{-1/2-\delta/3})$$

almost surely.

Hu and Zhang ([2004](#)) proved that both martingales $\mathbf{Q}_n$ and $\mathbf{U}_n$ satisfy the Lindberg conditions. Similarly, we can show that both martingales $\mathbf{Q}_{[nt]}$ and $\mathbf{U}_{[nt]}$ also satisfy the Lindberg conditions. Now we just have to calculate the covariance matrix of the martingales $\mathbf{Q}_{[nt]}$ and $\mathbf{U}_{[nt]}$. First, based on the results of Hu and Zhang ([2004](#)), we have

$$\hat{\boldsymbol{\rho}}(n) - \boldsymbol{\rho} = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{and} \quad n^{-1}\mathbf{N}(n) - \boldsymbol{\rho} = O\left(\sqrt{\frac{\log\log n}{n}}\right)$$

almost surely. Therefore, for any $t \in (0, 1]$, we have

(A.5)

$$\hat{\boldsymbol{\rho}}([nt]) - \boldsymbol{\rho} = O\left(\sqrt{\frac{\log\log[nt]}{[nt]}}\right) \quad \text{and}$$

$$([nt])^{-1}\mathbf{N}([nt]) - \boldsymbol{\rho} = O\left(\sqrt{\frac{\log\log[nt]}{[nt]}}\right)$$

almost surely. Now, we can calculate $\text{Var}[\Delta\mathbf{M}_{[nt]}|\mathcal{F}_{[nt]-1}]$, $\text{Var}[\Delta\mathbf{Q}_{[nt]}|\mathcal{F}_{[nt]-1}]$ and $\text{Cov}[\Delta\mathbf{M}_{[nt]}, \Delta\mathbf{Q}_{[nt]}|\mathcal{F}_{[nt]-1}]$.

First, $\Delta\mathbf{M}_{[nt]} = \mathbf{T}_{[nt]} - E[\mathbf{T}_{[nt]}|\mathcal{F}_{[nt]-1}]$ is a binary random vector. Based on conditions (A2), (A3) and (A.5), we have

(A.6)
$$\text{Var}[\Delta\mathbf{M}_{[nt]}|\mathcal{F}_{[nt]-1}] = \Sigma_1 + o(1)$$

almost surely. Similarly, we can show

(A.7)
$$\text{Var}[\Delta\mathbf{Q}_{[nt]}|\mathcal{F}_{[nt]-1}] = V + o(1)$$

and

(A.8)
$$\text{Cov}[\Delta\mathbf{M}_{[nt]}, \Delta\mathbf{Q}_{[nt]}|\mathcal{F}_{[nt]-1}] = o(1)$$

almost surely.

Based on results (A.6), (A.7) and (A.8), it follows that for any $0 < s < t < 1$,

$$\text{Cov}[\mathbf{Q}_{[ns]}, \mathbf{Q}_{[nt]}] = \text{Cov}\left(\sum_{m=1}^{[ns]} \Delta\mathbf{Q}_m, \sum_{m=1}^{[nt]} \Delta\mathbf{Q}_m\right)$$

$$= ns(V + o(1)) = nsV + o(n),$$

$$\text{Cov}[\mathbf{U}_{[ns]}, \mathbf{U}_{[nt]}] = n \wedge_{11}(s, t) + o(n),$$

$$\text{Cov}[\mathbf{Q}_{[ns]}, \mathbf{U}_{[nt]}] = \text{Cov}\left[\sum_{m=1}^{[ns]} \Delta\mathbf{Q}_m, \sum_{m=1}^{[nt]} \Delta\mathbf{M}_m B_{[nt],m}\right.$$

$$\left. + \sum_{m=1}^{[nt]} \Delta\mathbf{Q}_m \nabla(\boldsymbol{\rho})\Big|_{\boldsymbol{\theta}} E \sum_{k=m}^{[nt]} \frac{1}{k} B_{[nt],k}\right]$$

$$= \text{Cov}\left[\sum_{m=1}^{[ns]} \Delta\mathbf{Q}_m, \sum_{m=1}^{[nt]} \Delta\mathbf{M}_m B_{[nt],m}\right]$$

$$+ \text{Cov}\left[\sum_{m=1}^{[ns]} \Delta\mathbf{Q}_m, \sum_{m=1}^{[nt]} \Delta\mathbf{Q}_m \nabla(\rho)\Big|_{\boldsymbol{\theta}} E \sum_{k=m}^{[nt]} \frac{1}{k} B_{[nt],k}\right]$$

$$= \text{Cov}\left[\sum_{m=1}^{[ns]} \Delta\mathbf{Q}_m, \sum_{m=1}^{[nt]} \Delta\mathbf{Q}_m \nabla(\rho)\Big|_{\boldsymbol{\theta}} E \sum_{k=m}^{[nt]} \frac{1}{k} B_{[nt],k}\right]$$

$$= \left(V\nabla(\rho)|_{\boldsymbol{\theta}} E + o(1)\right) \sum_{m=1}^{[ns]} \left(\sum_{k=m}^{[nt]} \frac{1}{k} B_{[nt],k}\right)$$

$$= nV\nabla(\rho)|_{\boldsymbol{\theta}} E \int_0^s dx \left[\int_x^t \frac{t}{y}\left(\frac{t}{y}\right)^H dy\right] + o(n)$$

$$= n \wedge_{21}(s,t) + o(n),$$

and similarly,

$$\text{Cov}\left[\mathbf{Q}_{[nt]}, \mathbf{U}_{[ns]}\right] = n \wedge_{12}(s) + o(n),$$

where

$$\wedge_{11}(s,t) = \int_0^s \left(\frac{s}{x}\right)^{H'} \Sigma_1 \left(\frac{t}{x}\right)^H dx$$

$$+ \int_0^s dx \left[\int_x^s \frac{1}{y}\left(\frac{s}{y}\right)^H dy\right]' \Sigma_2 \left[\int_x^t \frac{1}{y}\left(\frac{t}{y}\right)^H dy\right],$$

$$\wedge_{21}(s,t) = V\nabla(\rho)|_{\boldsymbol{\theta}} E \int_0^s dx \left[\int_x^t \frac{t}{y}\left(\frac{t}{y}\right)^H dy\right],$$

$$\wedge_{12}(s) = \int_0^s dx \left[\int_x^s \frac{s}{y}\left(\frac{s}{y}\right)^H dy\right] E'\nabla(\rho)|_{\boldsymbol{\theta}}' V.$$

Therefore, the asymptotic covariance function of $n^{-1/2}(\mathbf{U}_{[nt]}, \mathbf{Q}_{[nt]})$ agrees with that of $(G_t, W_{2t} V^{1/2})$. So by weak convergence of the martingale [cf. Theorem 4.1 of Hall and Heyde (1980)], we have

$$n^{-1/2}([nt])\left(\frac{\mathbf{N}([nt])}{[nt]} - \boldsymbol{\rho}, \hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta}\right) \to (G_t, W_{2t} V^{1/2})$$

in distribution in the space $D_{[0,1]}$ with the Skorohod topology. $\square$

PROOF OF THEOREM 2.1.   We assume for $j = 1, 2$

$$[nt]\widehat{\text{Var}}(h(\hat{\boldsymbol{\theta}}_j([nt]))) = [nt]v_j\big(\mathbf{N}([nt])/[nt], \hat{\boldsymbol{\theta}}([nt])\big)\big(1 + o_P(1)\big)$$

and

$$[nt]\operatorname{Var}(h(\hat{\boldsymbol{\theta}}_j([nt]))) = [nt]v_j(\boldsymbol{\rho}, \boldsymbol{\theta}),$$

where v is a continuous function. We also assume

$$[nt]v_j(\mathbf{N}([nt])/[nt], \hat{\boldsymbol{\theta}}([nt])) = [nt]v_j(\boldsymbol{\rho}, \boldsymbol{\theta}) + O\left(\sqrt{\frac{\log\log[nt]}{[nt]}}\right) \qquad \text{a.s.,}$$

which holds for most circumstances, since

$$\mathbf{N}([nt])/[nt] = \boldsymbol{\rho} + O\left(\sqrt{\frac{\log\log[nt]}{[nt]}}\right) \qquad \text{a.s.}$$

and

$$\hat{\boldsymbol{\theta}}([nt]) = \boldsymbol{\theta} + O\left(\sqrt{\frac{\log\log[nt]}{[nt]}}\right) \qquad \text{a.s.}$$

So

$$[nt]\widehat{\operatorname{Var}}(h(\hat{\boldsymbol{\theta}}_j([nt]))) = [nt]\operatorname{Var}(h(\hat{\boldsymbol{\theta}}_j([nt]))) + O_P\left(\sqrt{\frac{\log\log[nt]}{[nt]}}\right).$$

That is, $[nt]\widehat{\operatorname{Var}}(h(\hat{\boldsymbol{\theta}}_j([nt])))$ converges to $[nt]\operatorname{Var}(h(\hat{\boldsymbol{\theta}}_j([nt])))$, $j = 1, 2$, in probability. By Slutsky's theorem, the sequential statistics

$$B_t\left(\frac{\mathbf{N}([nt])}{[nt]}, \hat{\boldsymbol{\theta}}([nt])\right) = \sqrt{t}\frac{h(\hat{\boldsymbol{\theta}}_1([nt])) - h(\hat{\boldsymbol{\theta}}_2([nt]))}{\sqrt{\widehat{\operatorname{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \widehat{\operatorname{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}$$

and

$$B_t^*(\hat{\boldsymbol{\theta}}([nt])) = \sqrt{t}\frac{h(\hat{\boldsymbol{\theta}}_1([nt])) - h(\hat{\boldsymbol{\theta}}_2([nt]))}{\sqrt{\operatorname{Var}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \operatorname{Var}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}$$

have the same distribution asymptotically. So we only need to prove the sequential statistics B_t^* converges to Brownian motion in distribution. Now

$$h(\hat{\boldsymbol{\theta}}_j) - h(\boldsymbol{\theta}_j) = (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)(\partial h(\boldsymbol{\theta}_j)/\partial\boldsymbol{\theta}_j)' + o(\|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|^{1+\delta})$$

$$= (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)(\partial h(\boldsymbol{\theta}_j)/\partial\boldsymbol{\theta}_j)' + o(n^{-1/2-\delta/3}) \qquad \text{a.s., } j = 1, 2.$$

It is easy to see that

$$\operatorname{Var}[\hat{\boldsymbol{\theta}}_j([nt])] = \operatorname{Var}[\hat{\boldsymbol{\theta}}_j(n)]/t + o(n^{-1}) \qquad \text{a.s., } j = 1, 2.$$

Here, we define

$$C = \sqrt{\operatorname{Var}[h(\hat{\boldsymbol{\theta}}_1([nt]))] + \operatorname{Var}[h(\hat{\boldsymbol{\theta}}_2([nt]))]}\sqrt{\operatorname{Var}[h(\hat{\boldsymbol{\theta}}_1([ns]))] + \operatorname{Var}[h(\hat{\boldsymbol{\theta}}_2([ns]))]}$$

and

$$\mathbf{D} = \left(\partial h(\boldsymbol{\theta}_1)/\partial \boldsymbol{\theta}_1, -\partial h(\boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_2\right).$$

Then

$$B_t^*(\hat{\boldsymbol{\theta}}([nt])) = \sqrt{t}\,\frac{h(\hat{\boldsymbol{\theta}}_1([nt])) - h(\hat{\boldsymbol{\theta}}_2([nt]))}{\sqrt{\mathrm{Var}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \mathrm{Var}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}$$

$$= \sqrt{t}\,\frac{h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2) + (\hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta})\mathbf{D}' + o(n^{-1/2-\delta/3})}{\sqrt{\mathrm{Var}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \mathrm{Var}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}.$$

By the conclusion of Lemma A.1:

$$n^{-1/2}([nt])\left(\hat{\boldsymbol{\theta}}([nt]) - \boldsymbol{\theta}\right) \to (W_{2t}V^{1/2})$$

in distribution in the space $D_{[0,1]}$ with the Skorohod topology. Under H_0, we have

$$B_t^* = \sqrt{t}\,\frac{n^{1/2}([nt])^{-1}W_{2t}V^{1/2}\mathbf{D}'}{\sqrt{\mathrm{Var}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \mathrm{Var}(h(\hat{\boldsymbol{\theta}}_2([nt])))}} + o(n^{-\delta/3})$$

almost surely. So the sequential statistics B_t^* converges to a Gaussian process in distribution. In order to prove that B_t^* converges to a "Brownian process" in distribution, it is enough to show $EB_t^* \to 0$ and for any $0 < s < t < 1$,

$$\mathrm{cov}(B_t^*, B_s^*) \to s,$$

$$\mathrm{cov}(B_t^*, B_s^*) = \frac{n\sqrt{ts}}{[nt][ns]}\,\frac{\mathrm{cov}(W_{2t}, W_{2s})\mathbf{D}V\mathbf{D}'}{C} + o(n^{-\delta/3})$$

$$= \frac{n\sqrt{t}s^{3/2}}{[nt][ns]}\,\frac{\mathbf{D}V\mathbf{D}'}{C} + o(n^{-\delta/3})$$

$$= \frac{n\sqrt{t}s^{3/2}}{[nt][ns]}\,\frac{\mathbf{D}(n\,\mathrm{Var}[\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}] + o(1))\mathbf{D}'}{C} + o(n^{-\delta/3})$$

$$= \frac{n^2\sqrt{t}s^{3/2}}{[nt][ns]}\,\frac{\partial h(\boldsymbol{\theta}_1)/\partial \boldsymbol{\theta}_1\,\mathrm{Var}[\hat{\boldsymbol{\theta}}_1(n) - \boldsymbol{\theta}_1]\,\partial h(\boldsymbol{\theta}_1)/\partial \boldsymbol{\theta}_1'}{C}$$

$$\quad + \frac{n^2\sqrt{t}s^{3/2}}{[nt][ns]}\,\frac{\partial h(\boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_2\,\mathrm{Var}[\hat{\boldsymbol{\theta}}_2(n) - \boldsymbol{\theta}_2]\,\partial h(\boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_2'}{C} + o(1)$$

$$= \frac{n^2\sqrt{t}s^{3/2}}{[nt][ns]}\,\frac{\mathrm{Var}[h(\hat{\boldsymbol{\theta}}_1(n))] + \mathrm{Var}[h(\hat{\boldsymbol{\theta}}_2(n))]}{C} + o(1)$$

$$= \frac{n^2ts^2}{[nt][ns]} + o(1)$$

$$\to s \qquad \text{a.s.}$$

It is easy to see that $EB_t^* \to 0$. This completes the proof and shows that B_t is asymptotical Brownian process in distribution.

Under H_1, the sequential statistics

$$B_t^*(\hat{\boldsymbol{\theta}}([nt])) = \sqrt{t}\frac{h(\hat{\boldsymbol{\theta}}_1([nt])) - h(\hat{\boldsymbol{\theta}}_2([nt])) - (h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2))}{\sqrt{\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}$$
$$+ \sqrt{t}\frac{h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2)}{\sqrt{\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}.$$

With similar proof, the first term converges to a standard Brownian motion in distribution asymptotically. Because

$$[nt]\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_j([nt]))) = v_j\left(\frac{\mathbf{N}([nt])}{[nt]}, \hat{\boldsymbol{\theta}}([nt])\right)(1 + o(1)) \qquad \text{a.s. } j = 1, 2,$$

we have that

$$\sqrt{t}\frac{h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2)}{\sqrt{\widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_1([nt]))) + \widehat{\mathrm{Var}}(h(\hat{\boldsymbol{\theta}}_2([nt])))}}$$

converges to

$$(A.9) \qquad t\frac{\sqrt{n}(h(\boldsymbol{\theta}_1) - h(\boldsymbol{\theta}_2))}{\sqrt{v_1(\boldsymbol{\rho}, \boldsymbol{\theta}) + v_2(\boldsymbol{\rho}, \boldsymbol{\theta})}} = \sqrt{n}\mu t$$

in probability. Therefore, under H_1, by Slutsky's theorem, $B_t^* - \sqrt{n}\mu t$ converges to a standard Brownian motion asymptotically. □

REFERENCES

ANDERSEN, J. (1996). Clinical trials designs—made to order. *J. Biopharm. Statist.* **6** 515–522.

ANDERSEN, K. M. (2007). Optimal spending functions for asymmetric group sequential designs. *Biom. J.* **49** 337–345. MR2380516

ARMITAGE, P. (1957). Restricted sequential procedures. *Biometrika* **44** 9–26. MR0085685

ARMITAGE, P. (1975). *Sequential Medical Trials*. Blackwell, Oxford. MR0370997

ATHREYA, K. B. and KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39** 1801–1817. MR0232455

BAI, Z. D. and HU, F. (2005). Asymptotics in randomized urn models. *Ann. Appl. Probab.* **15** 914–940. MR2114994

BERRY, D. A. (2005). Introduction to Bayesian methods III: Use and interpretation of Bayesian tools in design and analysis. *Clinical Trials* **2** 295–300.

COAD, D. S. and ROSENBERGER, W. F. (1999). A comparison of the randomized play-the-winner rule and the triangular test for clinical trials with binary responses. *Stat. Med.* **18** 761–769.

CHENG, Y. and SHEN, Y. (2005). Bayesian adaptive designs for clinical trials. *Biometrika* **92** 633–646. MR2202651

CONNOR, E. M., SPERLING, R. S., GELBER, R., KISELEV, P., SCOTT, G., O'SULLIVAN, M. J., VANDYKE, R., BEY, M., SHEARER, W., JACOBSON, R. L., JIMENEZ, E., O'NEILL, E., BAZIN, B., DELFRAISSY, J. F., CULNAME, M., COOMBS, R., ELKINS, M., MOYE, J., STRATTON, P. and BALSLEY, J. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type I with zidovudine treatment. *The New England Journal of Medicine* **331** 1173–1180.

DEMETS, D. L. (2006). Futility approaches to interim monitoirng by data monitoring committees. *Clinical Trials* **3** 522–529.

EISELE, J. and WOODROOFE, M. (1995). Central limit theorems for doubly adaptive biased coin designs. *Ann. Statist.* **23** 234–254. MR1331666

ETHIER, S. N. and KURTS, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York. MR0838085

GWISE, T. E., HU, J. and HU, F. (2008). Optimal biased coins for two-arm clinical trials. *Stat. Interface* **1** 125–135. MR2425350

HAYRE, L. S. (1979). Two-population sequential tests with three hypotheses. *Biometrika* **66** 465–474. MR0556733

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

HU, F. and ROSENBERGER, W. F. (2003). Optimality, variability, power: Evaluating response-adaptive randomization procedures for treatment comparisons. *J. Amer. Statist. Assoc.* **98** 671–678. MR2011680

HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, New York. MR2245329

HU, F., ROSENBERGER, W. F. and ZHANG, L. (2006). Asymptotically best response-adaptive randomization procedures. *J. Statist. Plann. Inference* **136** 1911–1922. MR2255603

HU, F. and ZHANG, L. X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Ann. Statist.* **32** 268–301. MR2051008

HU, F., ZHANG, L. X. and HE, X. (2009). Efficient randomized adaptive designs. *Ann. Statist.* **37** 2543–2560. MR2543702

IVANOVA, A. V. (2003). A play-the-winner type urn model with reduced variability. *Metrika* **58** 1–13. MR1999248

JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods With Applications to Clinical Trials.* Chapman and Hall, Boca Raton, FL.. MR1710781

LAN, K. and DEMETS, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70** 659–663. MR0725380

LEWIS, R. J., LIPSKY, A. M. and BERRY, D. A. (2007). Bayesian decision-theoretic group sequential clinical trial design based on a quadratic loss function: A frequentist evaluation. *Clinical Trials* **4** 5–14.

O'BRIEN, P. C. and FLEMING, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35** 549–556.

POCOCK, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64** 191–199.

POCOCK, S. J. (1982). Interim analyses for randomized clinical trials: The group sequential approach. *Biometrics* **38** 153–162.

PROSCHAN, M. A., LAN, K. and WITTES, J. T. (2006). *Statistical Monitoring of Clinical Trials, A Unified Approach*. Springer, New York.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535. MR0050246

ROSENBERGER, W. F. and HU, F. (2004). Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials* **1** 141–147.

ROSENBERGER, W. F. and LACHIN, J. M. (2002). *Randomization in Clinical Trials: Theory and Practice.* Wiley, New York. MR1914364

ROSENBERGER, W. F., STALLARD, N., IVANOVA, A., HARPER, C. N. and RICKS, M. L. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57** 909–913. MR1863454

ROUT, C. C., ROCKE, D. A., 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.

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. *J. Amer. Statist. Assoc.* **89** 768–776.

THOMPSON, W. R. (1933). On the likelihood that one unknown probability exceeds another in the review of the evidence of the two samples. *Biometrika* **25** 275–294.

TYMOFYEYEV, Y., ROSENBERGER, W. F. and HU, F. (2007). Implementing optimal allocation in sequential binary response experiments. *J. Amer. Statist. Assoc.* **102** 224–234. MR2345540

WALD, A. (1947). *Sequential Analysis*. Wiley, New York. MR0020764

WATHEN, J. K. and THALL, P. F. (2008). Bayesian adaptive model selection for optimizing group sequential clinical trials. *Statistics in Medicine* **27** 5586–5604.

WEI, L. J. and DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** 840–843.

ZELEN, M. (1969). Play the winner and the controlled clinical trial. *J. Amer. Statist. Assoc.* **64** 131–146. MR0240938

ZHANG, L. and ROSENBERGER, W. F. (2006). Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* **62** 562–569. MR2236838

ZHANG, L. X., HU, F. and CHEUNG, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *Ann. Appl. Probab.* **16** 340–369. MR2209345

DEPARTMENT OF STATISTICS
UNIVERSITY OF VIRGINA
KERCHOF HALL, CHARLOTTESVILLE
VIRGINIA 22904-4135
USA
E-MAIL: hz5n@virginia.edu
              fh6e@virginia.edu