Statistics and Its Interface Volume 11 (2018) 455–462

Response-adaptive randomization using power function of hypothesis testing

Zhongqiang Liu*,† and Xuefeng Han

Response-adaptive randomization (RAR) procedures have received extensive attention in clinical trials due to their considerations of ethics and efficiency. In the framework of RAR procedures, various target allocation proportions have been proposed and studied in literature. In this paper, we develop a family of RAR procedures using power function of corresponding hypothesis testing (RAR-P), and obtain the asymptotic properties under widely satisfied conditions. The proposed procedures are: (i) easy to understand and implement; (ii) applicable in more situations (continuous and discrete responses); (iii) more ethical than classical RAR procedures under some situations; and (iv) capable of monitoring efficiency and re-estimating sample size in clinical trials. Finally, we investigate performance of the proposed procedures through simulation studies.

AMS 2000 subject classifications: Primary 62L05, 62L20; secondary 62K99.
Keywords and phrases: Power function, Doubly adaptive biased coin design, Allocation proportion function, Efficiency and ethics.

1. INTRODUCTION

In clinical trials, it is extensively accepted to use response-adaptive randomization based on updated response outcomes of patients. The purpose is not only to efficiently identify clinical benefits of treatments under investigation, but also to increase the success probability of the treatments. During the conduct of clinical trials, to preliminarily identify clinical effectiveness of treatments under study, an intuitive and natural idea is to consider the power function of hypothesis testing.

Motivated by efficiency and ethical concerns, many response-adaptive randomization procedures have been introduced and studied in literature [8]. In the development of response-adaptive randomization, to achieve the desired allocation proportion, Eisele [4] proposed the "doubly adaptive biased coin design (DBCD)". After that, Eisele and Woodroofe [5] studied the asymptotic properties of DBCDs under somewhat restrictive conditions on allocation functions. The strong consistency, a law of the iterated logarithm and asymptotic normality of the DBCDs are derived under some widely satisfied conditions by Hu and Zhang [9]. To obtain asymptotically best randomization procedures, Hu et al. [7] derived a lower bound on the asymptotic variance of the allocation proportions for general response-adaptive randomization procedures.

It is well-known that Neyman allocation can maximize the power of tests while comparing two treatments, however, it may be inappropriate for ethical constraints. In order to reduce the number of patients who are assigned to the inferior treatment without loss of power, in the case of binary outcomes, Rosenberger et al. [13] proposed an optimal allocation, which can minimize the expected number of treatment failures under fixed power. To evaluate the performance of randomization procedures with continuous outcomes, Zhang and Rosenberger [17] proposed an allocation proportion as the counterpart of the optimal allocation of binary outcomes. However, due to specific assumptions, application of the above two designs has certain limitations.

In clinical trials, when new patients' response outcomes are available, an intuitive idea is to test whether there is significant difference between two treatment effects. Since power function can be thought of as the probability of accepting the difference between two treatment effects, we consider using power function to identify the better performing treatment.

Response-adaptive randomization procedures have a dual goal of estimating the treatment effect and randomizing patients with a higher probability of receiving the superior treatment. These are competing objectives, and no procedure in the literature is "perfect" with respect to both objectives [6]. Since the value of power function is a composite indicator which reflects how strongly response outcomes are in accordance with the alternative hypothesis, to achieve the competing objectives, we try to construct allocation proportion function by using power function. Obviously, power functions are available under both parametric and nonparametric tests, therefore, the proposed RAR procedures are applicable in most clinical trials.

The paper is organized as follows. Section 2 introduces the design scheme of the RAR procedures using power function of hypothesis testing (RAR-P). Theoretical results of the proposed procedures are presented in Section 3. To validate theoretical results and investigate the performance of

*Corresponding author.
†Partially supported by National Natural Science Foundation of China (Nos. 11701152 and 11731012).

the proposed procedures, Some simulation studies are carried out in Section 4. An illustration of redesigning a real clinical trial is shown in Section 5. Section 6 ends with some concluding remarks.

2. RAR PROCEDURE USING POWER FUNCTION OF HYPOTHESIS TESTING

Consider a clinical trial with two treatments 0 and 1. Let I_i be an indicator variable denoting the assignment of patient i, $i = 1, \cdots, N$, where N is the total number of patients and there is no early stopping. If patient i is assigned to treatment j, $I_i = j$, $j = 0, 1$. Let Y_{ij} be the response of patient i under treatment j. For patient i, if $I_i = j$, we observe Y_{ij}, $j = 0, 1$. Then we can denote the observed response outcome as $Y_{ij} = Y_{i0}(1 - I_i) + Y_{i1}I_i$.

Patients enter the trial sequentially, for different patients, suppose $\boldsymbol{Y}_i = (Y_{i0}, Y_{i1})$, $i = 1, \cdots, N$ are independently and identically distributed random vectors, with

$$Y_{1j} \sim f_j(\cdot, \theta_j), \quad j = 0, 1,$$

where θ_0 and θ_1 are the population parameters measuring the effects of treatment 0 and 1, respectively. For example, θ_0 and θ_1 are the expectations of two treated populations. Similar to the assumptions in the theory of response adaptive randomization [8], we assume that $\boldsymbol{Y}_i$ is independent of $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_{i-1}, I_1, \cdots, I_{i-1}, I_i$, however, I_i depends on all previous treatment assignments $(I_1, \cdots, I_{i-1})$ and responses $(\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_{i-1})$. Throughout this paper, we assume a larger response value (Y_{ij}) indicates a favourable clinical situation, if a smaller response value indicates a favourable situation, we consider $-Y_{ij}$.

2.1 Power function

The power function of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(X \in R)$.

Since we assume a larger response value represents a favourable clinical situation, i.e., the expected change is in one direction, the hypothesis test should reflect this as being one-sided, that is,

$$H_0 : \theta_1 = \theta_0 \quad \text{versus} \quad H_a : \theta_1 > \theta_0. \tag{1}$$

Let $\widehat{\theta}_0$ and $\widehat{\theta}_1$ be the maximum likelihood estimator of θ_0 and θ_1, respectively. Under regularity conditions in Theorem 6.7 [1], the Wald statistic for (1) has an asymptotic normal distribution, that is, as $n \to \infty$,

$$T_W = \frac{\widehat{\theta}_1 - \widehat{\theta}_0 - (\theta_1 - \theta_0)}{\sqrt{Var(\widehat{\theta}_1) + Var(\widehat{\theta}_0)}} \to N(0, 1) \quad \text{in distribution.}$$

Let Φ be the cumulative distribution function of $N(0, 1)$, and let z_τ be the τth quantile of $N(0, 1)$. For given statistical significance level α, the power function of the Wald test under (1) is

$$\beta(\theta_0, \theta_1) = 1 - \Phi\left(z_{1-\alpha} - \frac{\theta_1 - \theta_0}{\sqrt{Var(\widehat{\theta}_1) + Var(\widehat{\theta}_0)}}\right).$$

If we are interested in testing change in the other direction, that is,

$$H_0 : \theta_1 = \theta_0 \quad \text{versus} \quad H_a : \theta_1 < \theta_0,$$

then the power function is

$$\beta(\theta_0, \theta_1) = \Phi\left(z_\alpha - \frac{\theta_1 - \theta_0}{\sqrt{Var(\widehat{\theta}_1) + Var(\widehat{\theta}_0)}}\right).$$

When there is no obvious direction to compare two treatments, or considering the treatment difference in the opposite direction might turn up, two sided hypothesis test should be used as follows

$$H_0 : \theta_1 = \theta_0 \quad \text{versus} \quad H_a : \theta_1 \neq \theta_0,$$

and the power function is

$$\begin{aligned} \beta(\theta_0, \theta_1) &= \Phi\left(z_{\alpha/2} - \frac{\theta_1 - \theta_0}{\sqrt{Var(\widehat{\theta}_1) + Var(\widehat{\theta}_0)}}\right) \\ &+ \Phi\left(\frac{\theta_1 - \theta_0}{\sqrt{Var(\widehat{\theta}_1) + Var(\widehat{\theta}_0)}} - z_{1-\alpha/2}\right). \end{aligned}$$

It is worth mentioning that, under two sided hypothesis tests, given the same significance level, one may need to pay for a considerable larger sample size to detect a similar difference as that of one-sided tests.

In terms of quantification, power function gives an idea of how strongly the response results accord with the alternative hypothesis, namely, power function can be viewed as the degree of superiority of one treatment to the other in comparative clinical trials. Based on the intrinsic characteristics of power function, next we consider using power function to construct allocation proportion function in the allocation stage.

2.2 Conceptual framework of RAR-P

The response-adaptive randomization procedure using power function of hypothesis testing (RAR-P) is defined as follows:

(i) To start, allocate m patients to both treatment 0 and 1 by some restricted randomization procedures. For example, truncated binomial design or permuted block design [12]. When the responses of the first n $(n \geq 2m)$ patients are observed, use n_0 and n_1 to denote the number of observations on treated populations Y_0 and Y_1, respectively, where $n_0 + n_1 = n$.

During the conduct of clinical trials, as the parameters in power functions are unknown, they need to be replaced by corresponding estimators. After the responses of the first n patients are observed, compute the test statistic T and denote it as

$$T_n = \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{Var(\overline{Y}_1) + Var(\overline{Y}_0)}},$$

where $\overline{Y}_1$ and $\overline{Y}_0$ are the sample means of Y_1 and Y_0, respectively. For hypothesis testing $H_0 : \theta_1 = \theta_0$ versus $H_a : \theta_1 > \theta_0$ and given significance level α, calculate the power function by the following formula

$$\beta_n = 1 - \Phi\left(z_{1-\alpha} - \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{Var(\overline{Y}_1) + Var(\overline{Y}_0)}}\right).$$

(ii) Based on the above power function, define the allocation proportion function as follows

$$\rho_n = \begin{cases} 0.5, & \text{if } \beta_n \le 2\alpha, \\ \phi_n(\beta_n), & \text{if } 2\alpha < \beta_n \le p_0, \\ \phi_n(p_0), & \text{otherwise.} \end{cases} \tag{2}$$

where $\{\phi_n\}$ is a sequence of monotonically increasing continuous functions on $[0,1]$ satisfying $\phi_n(0.5) = 0.5$ and $\lim_{n\to\infty} \phi_n = \phi$. p_0 is a predetermined tuning parameter and $p_0 > 0.5$.

According to the definition of power function, for level α test, when the null hypothesis is true, the value of power function should be less than or equal to the significance level α. Therefore, in the allocation proportion function, we let the allocation proportion be 0.5 when the value of power function is less than or equals 2α, namely, we implement complete randomization when the null hypothesis tends to be true.

When we use the allocation proportion function (2), note that as β_n becomes larger, more patients will be allocated to the better performing treatment. To avoid extreme allocation cases and ensure a certain number of patients in both treatment groups, we set boundaries in the allocation proportion function (2). By selecting different tuning parameter p_0, the upper bound of the allocation proportion function can be adjusted, which will influence the targets of allocation proportions. In clinical trial practice, we shall choose an appropriate tuning parameter p_0 according to early studies and experiences.

Note that we assume a larger response value indicates a favourable clinical situation, power can be viewed as the competing probability of comparing two treatments. Therefore, we construct allocation functions as follows

$$\phi_n(\beta_n) = \frac{\beta_n^{\tau_n}}{\beta_n^{\tau_n} + (1-\beta_n)^{\tau_n}}, \tag{3}$$

where $\tau_n = n/2N$, n is the current sample size when a new patient is coming and N is the total sample size of the trial set in advance. Under level 0.05 test, by choosing tuning parameter $p_0 = 0.8$, we can obtain a sequence of allocation proportion functions as follows

$$\rho_n = \begin{cases} 0.5, & \text{if } \beta_n \le 0.1, \\ \phi_n(\beta_n), & \text{if } 0.1 < \beta_n \le 0.8, \\ \phi_n(0.8), & \text{otherwise.} \end{cases} \tag{4}$$

where functions $\phi_n(\cdot)$ are defined in (3). The characteristics of (4) are: i) when sample size n is small, the parameter τ_n is close to 0, then the value of allocation proportion function is close to 0.5, at this moment, the proposed procedure allocates patients almost as complete randomization; ii) as the sample size increases, power function will become accurate and play a more significant role in the allocation procedure, namely, the proposed procedure will allocate patients to the better performing treatment with larger probability according to power function.

(iii) Conditional on the assignments and responses of the first n patients, assign the $(n+1)$th patient to treatment 1 with probability

$$\begin{aligned} &P(I_{n+1} = 1 | I_1, \cdots, I_n, Y_{10}, \cdots, Y_{n_0 0}, Y_{11}, \cdots, Y_{n_1 1}) \\ &= g(n_1/n, \rho_n), \end{aligned}$$

where $g(x,y) : [0,1]\times[0,1] \to [0,1]$ is the allocation function proposed by Hu and Zhang [9] as follows

$$g(x,y) = \begin{cases} \frac{y(y/x)^\gamma}{y(y/x)^\gamma + (1-y)[(1-y)/(1-x)]^\gamma}, & \text{if } 0 < x < 1, \\ 1-x, & \text{if } x = 0 \text{ or } x = 1. \end{cases} \tag{5}$$

where $\gamma \ge 0$, which can adjust the degree of randomness of the randomization procedure.

To clarify the proposed randomization procedure under the other direction of tests, we would like to introduce the allocation proportion function under hypothesis testing $H_0 : \theta_1 = \theta_0$ versus $H_a : \theta_1 < \theta_0$.

Given significance level α, when responses of the first n patients are observed, power function can be calculated by

$$\beta_n = \Phi\left(z_\alpha - \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{Var(\overline{Y}_1) + Var(\overline{Y}_0)}}\right).$$

Note that if the alternative hypothesis H_a is true, the value of power function will become larger as the sample size increases. From the perspective of ethics, we shall allocate less patients to treatment 1, at this moment, the allocation proportion functions should be defined as follows

$$\rho_n = \begin{cases} 0.5, & \text{if } \beta_n \le 2\alpha, \\ 1 - \phi_n(\beta_n), & \text{if } 2\alpha < \beta_n \le p_0, \\ 1 - \phi_n(p_0), & \text{otherwise.} \end{cases} \tag{6}$$

where $\{\phi_n\}$ is a sequence of monotonically increasing continuous functions on $[0,1]$ satisfying $\phi_n(0.5) = 0.5$ and $\lim_{n\to\infty}\phi_n = \phi$. p_0 is a predetermined tuning parameter and $p_0 > 0.5$.

When two-sided hypothesis tests are needed, although the critical regions are different, by comparing sample mean of two treated populations, we can still construct allocation proportion functions the same as that for one-sided hypothesis tests.

2.3 An illustrative example

To illustrate how the proposed design is implemented, we give an example in the context of a real clinical trial.

In order to compare the efficacy of a fixed-dose triple combination (FDTC) of antihypertensive drugs with that of a free combination of three antihypertensives in patients with uncontrolled hypertension, Mazza et al. [11] conducted a clinical trial. A total of 184 eligible patients were enrolled between October 2015 and June 2016 in the Hypertension Centre of the Rovigo General Hospital. In this study, half of the patients ($n_1 = 92$) were treated with fixed-dose triple combination antihypertensive therapy, and the other ($n_0 = 92$) were treated with a free combination of three antihypertensives.

Continuous variable was expressed as mean and standard deviation in final statistical analyses. Finally, Mazza et al. [11] reported their conclusion: FDTC of perindopril/indapamide/amlodipine was more effective at reducing systolic blood pressure and pulse pressure in previously treated patients with uncontrolled hypertension, and well tolerated.

If the clinical goal is to control the blood pressure of hypertensive patients only, we can focus on whether there is significant reduction in systolic blood pressure during the clinical trial. Assume reduction values of systolic blood pressure in both treatment groups follow normal distribution, i.e., $Y_1 \sim N(15.3, 8)$, $Y_0 \sim N(13.1, 8)$. To simulate allocation procedure, next we use Monte Carlo method to generate random samples.

To accrue some data on patients' responses to each treatment, we begin the procedure with permuted block design (block size is 4) for the first 20 patients ($m = 10$). Suppose 100 patients have been assigned to a treatment, record the number of patients who are assigned to treatment 1 (FDTC) and denote it as n_1. When the first 100 patients' responses are available, for significance level 0.05, according to the formula of power function, we can figure out the value of power function $\beta_{100} = 0.756$. Note that $n = 100$, $N = 184$, it is easy to know $\tau_n = 100/368$. By using the allocation proportion function in Eq. (4), we have

$$\rho_n = \phi_n(0.756) = \frac{0.756^{\tau_n}}{0.756^{\tau_n} + 0.244^{\tau_n}} = 0.576.$$

Based on the above results, applying Hu and Zhang's allocation function in (5) with $\gamma = 2$, we can obtain that the 101th patient shall be assigned to treatment 1 with probability

$$g(n_1/n, \rho_n) = g(54/100, 0.576) = 0.645.$$

Repeat the above steps till the last patient is assigned to a treatment.

3. THEORETICAL RESULTS

In this section, we present asymptotic properties of the response-adaptive randomization using power function of hypothesis testing. Intuitively, if one treatment is superior to the other, the allocation function will skew the allocation towards the superior treatment, the random skewing will become more and more notable till the value of power function is greater than the predetermined upper bound (p_0). As a result from Theorem 4.1 and Theorem 4.3 [9], we have the following theorem.

Theorem 3.1. *For some $\epsilon > 0$, $E\|Y_{ij}\|^{2+\epsilon} < \infty$, $j = 0, 1$. Consider hypothesis testing problem*

$$H_0 : \theta_1 = \theta_0 \quad \textit{versus} \quad H_a : \theta_1 > \theta_0,$$

the allocation function g is given in (5), and the allocation proportion function based on power functions is given in (2). If the alternative hypothesis H_a is true, then as $\min(n_0, n_1) \to \infty$,

$$\rho_n \to \phi(p_0) \ \ a.s., \ \ \frac{n_1}{n} \to \phi(p_0) \ \ a.s.$$

and

$$\sqrt{n}\left(\frac{n_1}{n} - \phi(p_0)\right) \to N(0, v) \ \ \textit{in distribution},$$

where $v = \frac{1}{2\gamma+1}\phi(p_0)\big(1 - \phi(p_0)\big)$.

Proof. Under the alternative hypothesis $H_a : \theta_1 > \theta_0$, the values of Y_1 tend to be greater than the values of Y_0, then the corresponding power functions are

$$\beta_n = 1 - \Phi\left(z_{1-\alpha} - \frac{\overline{Y}_1 - \overline{Y}_0}{\sqrt{Var(\overline{Y}_1) + Var(\overline{Y}_0)}}\right).$$

Since the alternative hypothesis H_a is true, as $\min(n_0, n_1) \to \infty$, β_n have a distribution that is more concentrated on 1, i.e., $\beta_n \to 1$ a.s. Note that we assume $\lim_{n\to\infty}\phi_n = \phi$, and by the continuity of function (2) at point p_0, we have

$$\rho_n = \phi_n(\beta_n) \to \phi(p_0) \ \text{ a.s.}$$

Let $\mathcal{F}_n = \sigma(I_1, \cdots, I_n, Y_{10}, \cdots, Y_{n_0 0}, Y_{11}, \cdots, Y_{n_1 1})$ be the sigma-algebra generated by the first n treatment assignments and responses. According to the intrinsic properties of allocation function $g(x, y)$, for each $n > 2m$, there exists constant $c > 0$, such that

$$P(I_{n+1}=1|\mathcal{F}_n)=g(n_1/n,\rho_n)>c,$$

then we have

$$\sum_{n=2m+1}^{\infty}P(I_{n+1}=1|\mathcal{F}_n)\geq\sum_{n=2m+1}^{\infty}c=+\infty,$$

which implies $\{I_n=1,\text{i.o.}\}=\{n_1\to\infty\}$ almost surely by the generalized Borel-Cantelli lemma. Then, by Theorem 4.1 of Hu and Zhang [9], we can obtain

$$\frac{n_1}{n}\to\phi(p_0)\ \text{ a.s.}$$

The asymptotic normality of n_1/n is a direct result of Theorem 4.3 of Hu and Zhang [9], and the asymptotic variance v can be easily derived by calculating Fisher's information. □

Theorem 3.1 is the large sample properties of the proposed randomization procedure. When H_a is true, i.e., one treatment is better than the other, the proposed randomization procedure will assign about $\phi(p_0)$ proportion to the better treatment. Theorem 3.1 suggests, the parameter γ in allocation function (5) directly affects the allocation variability of the randomization procedure, which will decrease as γ increases. Given specific sequence of monotonically increasing functions and fixed tuning parameter, the proposed randomization procedure can target an explicit allocation proportion.

4. SIMULATION STUDIES

In this section, simulations are conducted to study the performance of the proposed randomization procedure. We compare different randomization procedures from the following aspects: (1) type I error rate, which is the probability of rejecting null hypothesis given that it is true; (2) the average proportion of patients assigned to treatment 1, which reflects ethical gains; (3) the power, which is the probability of rejecting null hypothesis when the alternative hypothesis is true; (4) allocation variability (AVar), which is the variability of final allocation proportions; and (5) total expected response (TER), which is the mean of all patients' response outcomes.

For continuous responses, we compare the RAR-P with three available randomization procedures: (i) complete randomization (CR); (ii) response-adaptive randomization targeting the Neyman allocation (RAR I); (iii) response-adaptive randomization targeting the optimal allocation (RAR II) proposed by Zhang and Rosenberger [17].

It is worth mentioning that RAR II is designed for clinical trials with continuous outcomes, hence, we do not use it in the case of binary responses. Since we assume a larger response value indicates a favourable clinical situation, the optimal allocation proportion of RAR II shall be $\rho=\sigma_1\sqrt{\mu_1}/(\sigma_1\sqrt{\mu_1}+\sigma_0\sqrt{\mu_0})$, where $Y_0\sim N(\mu_0,\sigma_0^2)$, $Y_1\sim N(\mu_1,\sigma_1^2)$.

Table 1. Type I error rate and allocation proportion

Procedure	N	Type I error	N_1/N
CR	100	0.0497	0.5005
	200	0.0506	0.5000
	500	0.0517	0.4997
RAR I	100	0.0575	0.4993
	200	0.0550	0.4999
	500	0.0525	0.5002
RAR II	100	0.0562	0.5000
	200	0.0539	0.4995
	500	0.0528	0.5000
RAR-P	100	0.0487	0.4691
	200	0.0498	0.4684
	500	0.0502	0.4673

In RAR I, RAR II and RAR-P, we apply Hu and Zhang's allocation function (5) with $\gamma=2$. The first 20 patients ($m=10$) are assigned to treatment 0 or treatment 1 by permuted block design (block size is 4). In particular, for RAR-P, we choose the tuning parameter $p_0=0.8$, and use the allocation proportion functions (4) to allocate patients.

In the following simulations, total sample size N is 100, 200 and 500, respectively. The final number of patients in treatment group 0 and 1 are denoted as N_0 and N_1, respectively. All simulation results are based on 10,000 repetitions.

4.1 Case 1: normal responses

Firstly, we simulate four randomization procedures to show Type I error rates. Assume that the responses of patients under two treatments follow the same distribution $N(1,1)$. The simulated Type I error rates (under significance level 0.05) and allocation proportions are reported in Table 1. The results in Table 1 suggest, compared with the other three randomization procedures, RAR-P have valid Type I error rates (not inflated).

Secondly, suppose the responses of patients under treatment 0 and treatment 1 follow distribution $N(1,1)$ and $N(1.5,1)$, respectively. In the final tests, Student's t-tests are applied. Power, allocation proportion of patients under treatment 1 and corresponding allocation variability (AVar), total expected response (TER) are reported in Table 2.

The results in Table 2 show the power of RAR-P is relatively small when the sample size is not large, however, as the sample size increases, RAR-P performs better on power compared with the other three procedures. From the perspective of ethics, RAR-P has a larger expected allocation proportion and a larger expected response (desirable). The results of allocation variability show the allocation proportion of RAR-P converges as the sample size increases, which confirms Theorem 3.1. Furthermore, by simulating different symmetrically distributed responses, we get similar results as that in Table 2.

Table 2. Power, allocation proportion, allocation variability and total expected normal responses

Procedure	N	Power	N_1/N	AVar	TER
CR	100	0.7932	0.4999	0.0500	1.2480
	200	0.9690	0.5000	0.0352	1.2496
	500	0.9999	0.5000	0.0228	1.2504
RAR I	100	0.7975	0.5000	0.0468	1.2505
	200	0.9681	0.4992	0.0326	1.2496
	500	0.9998	0.4998	0.0202	1.2502
RAR II	100	0.7976	0.5529	0.0535	1.2779
	200	0.9656	0.5519	0.0370	1.2766
	500	0.9999	0.5512	0.0231	1.2758
RAR-P	100	0.7931	0.5489	0.0776	1.2739
	200	0.9693	0.5971	0.0486	1.2992
	500	1.0000	0.6234	0.0114	1.3119

Table 3. Power, allocation proportion, allocation variability and total expected binary responses

Procedure	N	Power	N_1/N	AVar	TER
CR	100	0.5770	0.4999	0.0501	0.3998
	200	0.8733	0.4999	0.0356	0.3997
	500	0.9976	0.5003	0.0223	0.4003
RAR I	100	0.5885	0.5310	0.0665	0.4069
	200	0.8725	0.5302	0.0640	0.4055
	500	0.9973	0.5269	0.0465	0.4052
RSIHR	100	0.5838	0.5759	0.0716	0.4157
	200	0.8719	0.5764	0.0703	0.4155
	500	0.9978	0.5776	0.0756	0.4154
RAR-P	100	0.5842	0.5367	0.0673	0.4076
	200	0.8733	0.5778	0.0570	0.4156
	500	0.9985	0.6190	0.0189	0.4241

4.2 Case 2: binary responses

To simulate the most commonly used discrete responses, we consider that patients' responses under treatment 0 and treatment 1 follow distribution $Bernoulli(0.3)$ and $Bernoulli(0.5)$, respectively. Here we still simulate four randomization procedures: (i) complete randomization (CR); (ii) response-adaptive randomization targeting the Neyman allocation (RAR I); (iii) optimal allocation for binary responses proposed by Rosenberger et al. [13], which will be denoted as "RSIHR" in the simulated results; and (iv) response-adaptive randomization using power function of hypothesis testing (RAR-P).

In the case of binary responses, for the design RSIHR, the optimal allocation proportion for treatment 1 is $\sqrt{p_1}/(\sqrt{p_0}+\sqrt{p_1})$, where p_0 and p_1 are the success probability of treatment 0 and 1, respectively; for the proposed design RAR-P, we still use the allocation proportion function (4). All simulated results for binary responses are reported in Table 3.

The results in Table 3 display when the sample size is large, since power function of hypothesis testing is relatively accurate at the moment, compared with the other three randomization procedures, RAR-P can perform better on power and total expected success rates. Of course, RAR I and RSIHR can do better on power as expected when the sample size is not large. Besides, for RAR-P, the downward trend of allocation variability indicates that the allocation proportion will converge as the sample size increases.

5. REDESIGNING A REAL CLINICAL TRIAL

In this section, we consider redesigning a real clinical trial conducted by Dworkin et al. [3]. The real trial was also used for illustration by Zhang and Rosenberger [17], Biswas et al. [2]. It was a randomized, placebo-controlled trial with an objective to evaluate the efficacy and safety of pregabalin in the treatment of postherpetic neuralgia (PHN). In the trial, 173 patients were randomized to treatments: 89 to pregabalin (treatment 1) and 84 to placebo (treatment 0). The primary efficacy measure was pain reduction, as recorded by patients in a daily diary using the 11-point numerical pain rating scale (0= no pain, 10= worst possible pain). Therefore, a lower score indicates a favorable situation. After an 8 week duration of the trial, it was observed that pregabalin-treated patients experienced a higher decrease in pain score than patients treated with placebo. In these patients, there was still greater improvement in the endpoint mean pain scores in the patients treated with pregabalin than patients treated with placebo (endpoint mean scores 3.60 vs 5.29, p = 0.0001).

Here we redesign the trial based on the endpoint mean scores, i.e., 3.60 (with SD=2.25) for pregabalin and 5.29 (with SD=2.20) for placebo as the true ones. To investigate the performances of four randomization procedures, we use Normal and Double Exponential (DE) distributions to simulate the ordinal pain scores, respectively.

Since a lower score (response) indicates a favourable clinical situation, the optimal allocation proportion of RAR II shall be $\rho = \sigma_1\sqrt{\mu_0}/(\sigma_1\sqrt{\mu_0}+\sigma_0\sqrt{\mu_1})$, where $Y_0 \sim N(\mu_0, \sigma_0^2)$, $Y_1 \sim N(\mu_1, \sigma_1^2)$. To simulate RAR-P correctly, we shall use allocation proportion function (6).

The responses of 173 treated patients are simulated by distributions: (i) $N(5.29, 2.20^2)$ under placebo and $N(3.60, 2.25^2)$ under pregabalin; and (ii) $DE(5.29, 2.20/\sqrt{2})$ under placebo and $DE(3.60, 2.25/\sqrt{2})$ under pregabalin. All simulation results in Table 4 are obtained on the basis of 10,000 repetitions.

We observe in Table 4 that whether two treated responses are from normal distribution or double exponential distribution, the proposed procedure (RAR-P) can allocate more patients to the better treatment and has smaller TER, namely, it outperforms the other three procedures from the perspective of ethics. Simultaneously, RAR-P has the smallest allocation variability. As for power, there is no significant difference among the four procedures.

Table 4. Power, allocation proportion, allocation variability and total expected responses

	Procedure	N	Power	N_1/N	AVar	TER
Normal	CR	173	0.9997	0.5003	0.0375	4.4437
	RAR I	173	0.9994	0.5060	0.0351	4.4344
	RAR II	173	0.9997	0.5544	0.0365	4.3537
	RAR-P	173	0.9994	0.6213	0.0200	4.2365
DE	CR	173	0.9995	0.4997	0.0384	4.4467
	RAR I	173	0.9995	0.5076	0.0503	4.4311
	RAR II	173	0.9996	0.5555	0.0518	4.3513
	RAR-P	173	0.9994	0.6206	0.0206	4.2420

6. CONCLUDING REMARKS

In this paper, we develop a family of RAR procedures using power function of hypothesis testing. Since power function indicates the evidence of hypothesis testing, the proposed randomization procedures are intuitively attractive. Note that we keep updating the values of power function in the randomization procedures, when the values of power function reach a certain level, we can concentrate on maximizing the total expected response which reflects ethical gains. Hence, the proposed procedures can be fitted in the framework of Zhang and Rosenberger [17] for parametric models, namely, under parametric models, the new procedures may be optimal when the sample size reaches a certain level. Under nonparametric models, to implement the proposed randomization procedures, we can apply the power function proposed by Rosner and Glynn [15].

According to the basic idea of adaptive randomization, we construct a family of allocation proportion functions to make more patients be assigned to the better treatment. As discussed by Tymofyeyev et al. [16], it is important to have a boundary of the allocation proportion in practice. To make the allocation not go to an extreme, we set a boundary in the allocation proportion function. By adjusting tuning parameter, the proposed randomization procedures can target different allocation proportions, therefore, the proposed procedures have flexibility.

Clinical trials have multiple objectives, different trials may emphasize on different objectives. Efficiency is critical for demonstrating efficacy. Randomization mitigates certain biases. Ethics is an essential component in any human experimentation, and dictates the treatment of patients in the trial [14]. In the proposed randomization procedures, these considerations can be compromised by adjusting the tuning parameter. Since investigators understand the background of a trial, they can choose an appropriate tuning parameter beforehand. For example, when ethical concerns dominate the trial, investigators shall choose a larger tuning parameter p_0 to allow more patients to be allocated to the better treatment, this is especially important when the disease being combatted is life-threatening.

To make the design more tractable, we focus on comparing two treatments in this paper. However, comparing multiple treatments are often encountered in clinical trial practice. Generalizing the proposed procedures to multi-treatments is an open question, we leave this for future research.

In fact, we assume immediate availability of patients' response outcomes. However, it is not difficult to incorporate delayed responses into the proposed procedures. In practice, we can always update the values of power function whenever new patients' responses become available. This issue has been well-studied for classical response-adaptive designs in literature [10]. Corresponding results can be extended to the proposed procedures under similar conditions.

ACKNOWLEDGEMENTS

We thank the anonymous referees and editors for their insightful suggestions and comments which helped improve this paper. The research work was supported in part by National Natural Science Foundation of China (Nos. 11701152 and 11731012), Zhejiang Provincial Natural Science Foundation (No. LY17A010016), Doctoral Research Fund of Henan Polytechnic University (No. 672103/001/147) and China Postdoctoral Science Foundation (No. 2017M620240).

Received 6 April 2017

REFERENCES

[1] Boos, Dennis D. and Stefanski, L. A. (2013). *Essential statistical inference*. Springer, New York. MR3024617

[2] Biswas A., Bhattacharya R. and Zhang, L. (2007). Optimal response-adaptive designs for continuous responses in phase III trials. *Biometrical Journal* **49** 928–940. MR2416453

[3] Dworkin R. H., Corbin A. E., Young J. P., Sharma U., LaMoreaux L., Bockbrader H., Garofalo E. A. and Poole R. M. (2003). Pregabalin for the treatment of postherpetic neuralgia: a randomized, placebo-controlled trial. *Neurology* **60** 1274–1283.

[4] Eisele, J. R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference* **38** 249–261. MR1256599

[5] Eisele, J. R. and Woodroofe, M. B. (1995). Central limit theorems for doubly adaptive biased coin designs. *Annals of Statistics* **23** 234–254. MR1331666

[6] Flournoy, N., Haines, L. M. and Rosenberger, W. F. (2013). A graphical comparison of response-adaptive randomization procedures. *Statistics in Biopharmaceutical Research* **5** 126–141.

[7] Hu, F., Rosenberger, W. F. and Zhang, L. X. (2006). Asymptotically best response-adaptive randomization procedures. *Journal of Statistical Planning and Inference* **136** 1911–1922. MR2255603

[8] Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, New York. MR2245329

[9] Hu, F. and Zhang, L. X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trails. *Annals of Statistics* **32** 268–301. MR2051008

[10] Hu, F., Zhang, L. X., Cheung, S. H. and Chan, W. S. (2008). Doubly adaptive biased coin designs with delayed responses. *Canadian Journal of Statistics* **36** 541–559. MR2532251

[11] Mazza, A., Lenti, S., Schiavon, L. Sacco, A. P., Dell'Avvocata, F., Rigatelli, G. and Ramazzina, E. (2017). Fixed-dose triple combination of antihypertensive drugs improves blood pressure control: from clinical trials to clinical practice. *Advances in Therapy* **34** 975–985.

[12] Rosenberger, W. F. and Lachin, J. M. (2016). *Randomization in clinical trials. Theory and practice*, 2nd ed. Wiley, New York. MR3443072
[13] 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
[14] Rosenberger, W. F. and Sverdlov, O. (2008). Handling covariates in the design of clinical trials. *Statistical Science* **23** 404–419. MR2483911
[15] Rosner, B. and Glynn, R. J. (2009). Power and sample size estimation for the Wilcoxon rank sum test with application to comparisons of C statistics from alternative prediction models. *Biometrics* **65** 188–197. MR2665860
[16] Tymofyeyev, Y., Rosenberger, W. F. and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association* **102** 224–234. MR2345540
[17] Zhang, L. J. and Rosenberger, W. F. (2006). Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* **62** 562–569. MR2236838

Zhongqiang Liu
School of Mathematics and Information Science
Henan Polytechnic University
Jiaozuo 454000
China
School of Mathematical Sciences
Zhejiang University
Hangzhou 310027
China
E-mail address: zhongqiang@hpu.edu.cn

Xuefeng Han
School of Mathematics and Information Science
Henan Polytechnic University
Jiaozuo 454000
China
E-mail address: hanxuefeng@hpu.edu.cn