Published in final edited form as:

Stat Med. 2014 October 15; 33(23): 4043–4052. doi:10.1002/sim.6213.

Managing competing demands in the implementation of response-adaptive randomization in a large multi-center phase III acute stroke trial

Wenle Zhao and Valerie Durkalski

Department of Public Health Science, Medical University of South Carolina, 135 Cannon Street, Charleston, SC 29425, USA

SUMMARY

It is well known that competing demands exist between the control of important covariate imbalance and protection of treatment allocation randomness in confirmative clinical trials. When implementing a response-adaptive randomization algorithm in confirmative clinical trials designed under a frequentist framework, additional competing demands emerge between the shift of the treatment allocation ratio and the preservation of the power. Based on a large multicenter phase III stroke trial, we present a patient randomization scheme that manages these competing demands by applying a newly developed minimal sufficient balancing design for baseline covariates and a cap on the treatment allocation ratio shift in order to protect the allocation randomness and the power. Statistical properties of this randomization plan are studied by computer simulation. Trial operation characteristics, such as patient enrollment rate and primary outcome response delay, are also incorporated into the randomization plan.

Keywords

response-adaptive randomization; covariate balancing; allocation ratio; allocation randomness; minimal sufficient balance

1. INTRODUCTION

The use of response-adaptive randomization (RAR) in clinical trials has received increasing attention in the last two decades [1, 2]. With RAR, the treatment allocation probabilities for the current patient to be randomized can be altered based on observed response data from previously randomized patients within the trial. As a component of Bayesian adaptive methods, RAR has gained noticeable application in trials designed under a Bayesian framework [3–5]. However, so far, very few confirmative trials designed under a frequentist framework using RAR have been reported. This is important to explore since under a frequentist framework, there is in most cases a loss in power when the allocation ratio is not balanced.

Motivations for using RAR in clinical trial treatment allocation mainly come from efficiency and ethics considerations [3]. If the purpose of the RAR is to maximize the efficiency (in terms of power), the increase in power of RAR over a fixed equal allocation randomization for large confirmative trials designed under a frequentist framework will be minimal. Therefore, the ethical benefit in this setting becomes the primary motivation for using RAR, in which, RAR is proposed as assigning a higher proportion of patients to the so-far better performing arm. However, the shift of the treatment allocation ratio from a balanced to an unbalanced allocation can result in a loss of power and creates competing demands between the ethical and efficiency considerations.

In addition to the competing demands between ethics and efficiency, the randomization scheme needs to consider the balance between treatment assignment randomness and baseline covariate balancing. The impacts of the treatment allocation ratio and the preservation of assignment randomness have been highlighted in the Michigan extracorporeal membrane oxygenation (ECMO) trial [6]. Using the randomized play-the-winner (RPW) rule [7], the Michigan ECMO trial ended with all 11 patients assigned to the ECMO arm surviving, and one patient assigned to the conventional treatment arm who died. This trial has been met with strong controversy [8, 9], partially because of the 11:1 final allocation ratio as the result of the RAR algorithm. The doubly adaptive biased coin (DABC) design was developed to control the variation of the treatment allocation ratio [10–12]. Recently, a variety of RAR designs with optimized target allocation ratios have been proposed with the aim of either maximizing the power or minimizing the total failures [13–15]. However, these two goals may compete with each other as it has been shown that they may not be achieved simultaneously [16]. Meanwhile, covariate-adjusted response adaptive (CARA) algorithms have been developed, with the aim of controlling imbalance in baseline prognostic factors during the RAR process [17–22]. Some of these studies used computer simulations to mimic real clinical trials in order to study the statistical and operational characteristics of the proposed RAR algorithm [22, 23]. However, there is a paucity of literature on the implementation of RAR in actual phase III trials designed under a frequentist framework. The competing demands under RAR are valid concerns and until these concerns are addressed, the actual value of using RAR in large phase III clinical trials will continue to be challenged [24–26].

In this paper, we examine the competing demands of ethics, allocation randomness and power when using RAR with covariate balancing. We implement an approach to manage these demands in the setting of an existing large multi-center phase III acute stroke trial with a binary outcome designed under a frequentist framework. Section 2 describes the setting of the trial and the primary goals of the randomization procedure. Section 3 presents the design of the randomization algorithm and its statistical properties. The implementation of the randomization algorithm in the trial is covered in Section 4, followed by a discussion on the benefits and costs of RAR in Section 5.

2. BACKGROUND

The Stroke Hyperglycemia Insulin Network Effort (SHINE) is an ongoing multicenter phase III clinical trial in acute stroke funded by the National Institute of Neurological Diseases and

Stroke (NCT01369069) [27, 28]. In this study, approximately 1400 acute ischemic stroke patients with elevated blood glucose levels are being recruited from approximately 60 clinical sites in the US and will be randomized to either a sliding scale insulin treatment arm or the experimental insulin infusion arm. Investigators in charge of treating the enrolled patients will be unblinded. The primary efficacy outcome is Success or Failure as determined by the 90-day modified Rankin scale (mRS) and the baseline stroke severity measured by the National Institutes of Health Stroke Scale (NIHSS) [29]. Success outcome is defined as mRS = 0 for patients with mild stroke (NIHSS 3–7), mRS = 1 for moderate stroke patients (NIHSS 8–14), and mRS = 2 for severe stroke patients (NIHSS 15–22). Three baseline covariates are believed to have important impact on the primary outcome and therefore require balancing. They are the baseline stroke severity category (mild, moderate and severe), the use of thrombolysis (yes, no), and clinical site.

The patient randomization procedure in the SHINE trial is implemented in three stages and has three primary goals. The first goal is to protect the randomness of treatment allocation in order to prevent potential selection biases. The second goal is to prevent serious imbalances in the distributions of the three important baseline covariates in order to ensure the compatibility of the two treatment groups. The third goal is to promote a response-adaptive treatment allocation ratio, so that patients will have a higher than 50% chance to be assigned to the better performing treatment arm if a difference in the success rates between the two arms is observed from previously enrolled patients. When trying to achieve these goals, it is also required that the power reduction due to the unbalanced allocation ratio under RAR is minimal. The randomization algorithm is implemented in three stages as described in detail in Section 4. The rationale for the three stages is based on the specific trial setting. The SHINE trial has an anticipated enrollment rate of approximately 25 patients per month and a primary outcome that is measured 90 days post randomization. Although this may be considered a relatively short-term outcome, a burn-in period is required in order to obtain the necessary information on the covariates as well as the primary outcome that will be used specifically for the pre-defined adaptation of the allocation ratio.

3. METHOD

3.1 Allocation randomness and covariate balancing

Randomization in patient treatment allocation serves as a basis for frequentist inferences in clinical trials [30, 31]. The randomness of treatment allocation reduces selection bias. This benefit is maximized when the assignment is purely random (i.e., 50% chance of being assigned to either arm in a two-arm trial). When covariate balancing is included in the randomization algorithm in order to reduce accidental biases, the randomness of treatment assignment will be reduced [32].

Although the necessity of balancing important prognostic variables in clinical trials remains a topic of debate [33], balanced baseline variables between treatment arms remains at the top of the randomization objective list for many researchers. The absolute difference between the sizes of two treatment arms has been used as the imbalance control limit for the permuted block design [34], the biased coin design [35], the big stick design [36], and the maximal procedure design [37]. Some other randomization designs, such as the urn design

[38], use the ratio of the absolute arm size difference to the total sample size as the imbalance measurement. A small p -value, i.e. less than 0.05, for a baseline covariate balance test often triggers suspicion of selection bias or wrong doing [39]. The recently published minimal sufficient balancing strategy aims to maximize the treatment allocation randomness while containing the baseline covariate imbalances within a pre-specified limit [40]. Because the baseline covariate balance is more likely to be challenged based on the p -value of the imbalance test [39], we use the imbalance test p -value as the imbalance measure and set the control limit (p^*) to range from 0.1 to 0.3 depending on the importance of the covariate.

Let $i = 1, 2$, and 3 refer to baseline covariates NIHSS, thrombolysis, and site respectively, and n_{ijk} be the observed number of patients in the j^{th} category of the baseline covariate i previously randomized into treatment arm k (A, B). Let E_{ijk} be the expected number corresponding to n_{ijk} under the hypothesis of independence. Let the marginal totals be $n_{1j} = n_{1jA} + n_{1jB}$, ($j = 1, 2, 3$), $n_{2j} = n_{2jA} + n_{2jB}$, ($j = 1, 2$), $n_A = n_{11A} + n_{12A} + n_{13A}$, $n_B = n_{11B} + n_{12B} + n_{13B}$. Chi-square tests are used to measure the imbalance in the categorical covariates, and the one-sample test for a binomial proportion is used to measure the imbalance within each site. The chi-square test statistics are:

$$\chi_{NIHSS}^2 = \sum_{j=1}^3 \left[\left(n_{1jA} - E_{1jA} \right)^2 / E_{1jA} + \left(n_{1jB} - E_{1jB} \right)^2 / E_{1jB} \right], \quad (df=2) \quad (1)$$

$$\chi_{tPA}^2 = \sum_{j=1}^2 \left[\left(n_{2jA} - E_{2jA} \right)^2 / E_{2jA} + \left(n_{2jB} - E_{2jB} \right)^2 / E_{2jB} \right], \quad (df=1) \quad (2)$$

Where:

$$E_{1jk} = \frac{n_{1j} \times n_k}{n_A + n_B}, \quad (j=1, 2, 3; k=A, B) \quad (3)$$

$$E_{2jk} = \frac{n_{2j} \times n_k}{n_A + n_B}, \quad (j=1, 2; k=A, B) \quad (4)$$

The p -values for these two tests, p_{NIHSS} and p_{tPA} , can be obtained based on the test statistics and their respective degrees of freedom.

When the number of sites is large, the chi square test for the $R \times C$ contingency table will not work because many cells will be smaller than 6 during the trial [41]. Therefore, the imbalance within a site is measured by the one-sample test for a binomial proportion based on the difference between the observed allocation ratio within the site and the observed overall or target allocation ratio. Assume the current patient is in site j . Let $n_j = n_{3jA} + n_{3jB}$ be the number of patients previously randomized at site j and $n = n_A + n_B$ be the total number of patients previously randomized in the study. When $n_j \geq 20$, the large sample normal approximation to the binomial distribution is valid [41] and the test statistic is

$$z_{Site} = \frac{\left(\frac{n_{3jA}}{n_j} - \frac{n_A}{n}\right)}{\sqrt{\frac{n_A}{n} \times \frac{n_B}{n} \times \frac{1}{n_j}}} \quad (5)$$

The p -value for this test, p_{Site} , can be obtained based on the standard normal distribution. When $n_j < 20$, the exact method can be used, and the p -value of the test is:

$$p_{Site} = \begin{cases} 2 \sum_{i=0}^{n_{3jA}} \binom{n_j}{i} \times \left(\frac{n_A}{n}\right)^i \times \left(\frac{n_B}{n}\right)^{n_{3jA}-i} & \text{if } \frac{n_{3jA}}{n_j} \leq \frac{n_A}{n} \\ 2 \sum_{i=n_{3jA}}^{n_j} \binom{n_j}{i} \times \left(\frac{n_A}{n}\right)^i \times \left(\frac{n_B}{n}\right)^{n_{3jA}-i} & \text{if } \frac{n_{3jA}}{n_j} > \frac{n_A}{n} \end{cases} \quad (6)$$

Unlike the minimization methods proposed by Pocock and Simon [42] and Taves [43], which have been criticized for the lack of randomness [44], the minimal sufficient balancing strategy uses biased coin assignments only when imbalances exceed a pre-specified control limit (p^*). If the p -value of the imbalance test is less than its control limit, a vote is registered toward the treatment arm aimed to reduce the imbalance. Otherwise, no vote is submitted. Based on the results from each test (NIHSS, tPA and site), if one arm receives more votes than the other arm, a biased coin assignment is applied. Otherwise, the target allocation ratio is used for the calculation of the treatment allocation probability.

A treatment assignment is considered deterministic if its conditional allocation probability is 1 for one arm and 0 for the other. The randomness of treatment assignments can be quantified by several different measurements, including the probability of deterministic assignments, the probability of purely random assignments, and the probability of correct guess defined by Blackwell and Hodges [45]. The values of these properties are affected by the number of covariates, the imbalance control limit p^* , and the biased coin probability p_{bc} . Shown in Figure 1 are computer simulation results of the proportion of purely random assignments under minimal sufficient balancing for a two-arm trial with three categorical covariates, balanced allocation and a sample size of 500. Based on input from clinical investigators, the distribution of NIHSS among the three levels was set at 0.3 (mild), 0.4 (moderate), and 0.3 (severe); and the proportion of tPA use was set at 0.4 (no) and 0.6 (yes). The imbalance control limit p^* was set to 0.3. A higher biased coin probability will be more efficient in imbalance correction and will be less frequently needed thus resulting in a higher proportion of purely random assignments. For example, with $p_{bc} = 0.85$, out of 500 patients, a median of 422 patients will receive purely random assignments. As a random process, the actual number of patients receiving purely random assignment may vary, as shown in Figure 1.

As p_{bc} increases, the probability of making a correct guess for those biased coin assignments increases, however the overall number of biased coin assignments decreases. The average chance of making a correct guess for treatment assignments is 52.8% when $p_{bc} = 0.55$, and 55.2% when $p_{bc} = 1.0$. These are substantially lower than the correct guess probability for permuted block randomization, that is 70.9%, 68.0%, and 66.2% for block size of 4, 6, and 8 respectively [32]. Between the randomness and balance properties, simple randomization

and the permuted block method each emphasize one property and overlook the other property [46]. The minimal sufficient balancing provides a practical strategy to manage the competing demands between randomness and balance. For this purpose, it is recommended to use a biased coin probability of 0.7. If RAR is incorporated into the randomization algorithm, a higher p_{bc} value (0.85 or 0.9) should be considered in order to provide a sufficient proportion (84.4% or 87.0%) of treatment assignments for the RAR component of the randomization scheme.

Alternatively, should the original version of the Pocock and Simon minimization method be used in the above setting, computer simulation shows that 49.0% of assignments will be purely random. If a threshold of 3 is included in the minimization algorithm so that a deterministic assignment will be applied only when the imbalance sum is 3 or more, the expected proportion of purely random assignments will increase to 62.8%. When a biased coin probability of 0.85 is used to replace the deterministic assignments, the expected proportion of purely random assignments will be reduced to 53.8%.

3.2 Balancing ethics and trial efficiency

Primary motivations of using response-adaptive randomization to shift treatment allocation ratios in clinical trials are potential advantages in ethics, efficiency, and economics [11, 47–48]. Similar motivations also have been used to justify the use of fixed unequal randomization ratios in clinical trial practices. While the use of a fixed unequal allocation ratio is justified based on observed response information from previous studies or the need to acquire more secondary information from one arm (i.e, safety of experimental treatment), the use of RAR allows the randomization algorithm to dynamically adjust the treatment allocation ratio based on response information obtained within the study. For this reason, RAR could be a better choice over fixed unequal randomization. However, the potential reduction in power caused by the allocation ratio shifting away from a balanced ratio remains the same for both fixed unequal randomization and RAR.

Consider a trial comparing two independent binomial proportions testing the hypothesis $H_0 : p_A = p_B$ versus $H_1 : p_A \neq p_B$ for the specific alternative $|p_A - p_B| = \delta$, with significance level α and sample size $n = n_A + n_B$, the power of the trial is [41, page 385]:

$$Power = \Phi \left[\frac{|p_A - p_B| - z_{1-\frac{\alpha}{2}} \sqrt{\bar{p}\bar{q} \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}} \right] \quad (7)$$

Where, $q_A = 1 - p_A$, $q_B = 1 - p_B$, $\bar{p} = \frac{p_A n_A + p_B n_B}{n}$, and $\bar{q} = 1 - \bar{p}$. Let $|p_A - p_B|$ be the effect size and $r = \frac{n_A}{n_B}$ be the treatment allocation ratio, we have $n_A = \frac{rn}{1+r}$, $n_B = \frac{n}{1+r}$, $\bar{p} = \frac{rp_A + p_B}{1+r}$, and $\bar{q} = \frac{rq_A + q_B}{1+r}$. Formula (7) can then be expressed as:

$$Power = 1 - \beta = \Phi \left[\frac{|p_A - p_B| - z_{1-\frac{\alpha}{2}} \sqrt{\frac{(rp_A + p_B)(rq_A + q_B)}{nr}}}{\sqrt{\left(\frac{p_A q_A}{r} + p_B q_B\right) \left(\frac{1+r}{n}\right)}} \right] \quad (8)$$

Based on formula (8), Figure 2 shows the sensitivity of the power on the changes of the randomization allocation ratio. It is clear that as the allocation ratio shifts toward the better performing arm ($r > 1$), the power is lower than the power under balanced allocation ($r = 1$).

It is important to ensure that this effect is either minimized or taken into account when estimating the sample size in the trial design stage. By fixing the power, the required sample size is:

$$n = \frac{(1+r)}{(p_A - p_B)^2} \left[z_{1-\frac{\alpha}{2}} \sqrt{\frac{(rp_A + p_B)(rq_A + q_B)}{r(1+r)}} + z_{1-\beta} \sqrt{\frac{p_A q_A}{r} + p_B q_B} \right]^2 \quad (9)$$

The expected number of total failures associated with this required sample size is:

$$n_{Failure} = \bar{q}n = \frac{rq_A + q_B}{(p_A - p_B)^2} \left[z_{1-\frac{\alpha}{2}} \sqrt{\frac{(rp_A + p_B)(rq_A + q_B)}{r(1+r)}} + z_{1-\beta} \sqrt{\frac{p_A q_A}{r} + p_B q_B} \right]^2 \quad (10)$$

Figure 3 shows the impact of the randomization allocation ratio on the expected number of failures under two conditions, one with a fixed sample size and the other with a fixed power.

With a fixed sample size, there is a gradual decrease in the total number of failures as the allocation ratio increases. However, the seemingly ethical benefit of reducing the number of failures is at the cost of power as seen in Figure 2. Under the fixed power scenario, shifting the allocation ratio from 1:1 in either direction results in an increase in the required sample size, and therefore increases the total number of failures.

From this point of view, the expected ethical benefit of shifting treatment allocation ratio by either fixed unbalanced randomization or RAR is questionable. For example, in a trial to detect 7% difference between two independent binomial proportions (32% for the treatment arm and 25% for the control) with significance level of $\alpha = 0.05$, the power will be 82.8% with 700 patients in each of the two treatment groups, and the expected number of total failures in the study will be 1001. Should the allocation ratio be shifted to 2:1, with 933 patients randomized in the better performing arm and 467 patients in the other arm, the expected number of total failures will be 985, and the power will be reduced to 77.6% at the same time. In order to keep the original power at 82.8%, an additional 188 patients will be required, leading to a total sample size of 1588. The expected number of total failures in this trial will be 1117, that is 116 higher than the original 1001 under the balanced allocation. This fact reflects the conflict between the potential individual ethics and collective ethics within a clinical trial. Let $p_{r:1}$ and $p_{1:1}$ represent the average chance of success for each

patient enrolled in the study under $r:1$ and $1:1$ allocation ratios respectively. The ethics benefit for the unequal allocation ratio can be measured as:

$$\bar{p}_{r:1} - \bar{p}_{1:1} = \frac{rp_A + p_B}{1+r} - \frac{p_A + p_B}{2} = \frac{(r-1)(p_A - p_B)}{2(r+1)} \quad (11)$$

In the scenario discussed above with $r = 2$, formula (11) represents the benefit of a 1.17% increase in the trial's average success rate, at the cost of enrolling an additional 188 patients and expecting 116 of them to have failure outcomes. This result makes the selection of the 2:1 allocation ratio more difficult to justify.

In RAR research, not only is the expected value of the target allocation ratio important, but also the variance of the allocation ratio. Hu and Zhang proposed a Doubly-adaptive Biased Coin Design (DBCD) procedure for trials comparing two treatments with binomial outcomes [12]. With DBCD, the probability of assigning subject i to treatment A is:

$$\Pr(T_i = A) = \frac{\rho(\rho/x)^\gamma}{\rho(\rho/x)^\gamma + (1-\rho)((1-\rho)/(1-x))^\gamma} \quad (12)$$

where $x = \frac{n_A}{n_A + n_B}$ is the observed allocation proportion of arm A, $\rho = \frac{p_A^\lambda}{p_A^\lambda + p_B^\lambda}$ is the target allocation proportion of arm A, γ is a nonnegative integer effecting the asymptotic variance of a RAR procedure, and λ is parameter effecting the target allocation ratio. The variance of the allocation proportion is:

$$\text{Var}(\rho) = \frac{q_A q_B}{(q_A + q_B)^3} \left(p_A + p_B + \frac{2}{1+2\lambda} \right) \quad (13)$$

Suppose $p_A = 0.32$, $p_B = 0.25$, and $\lambda = 2$, the target allocation proportion will be $\rho = 0.621$ with a variance of $\sigma_\rho^2 = 0.169$. This indicates a wide range of possible unbalanced allocation ratios, which may significantly reduce the power.

To manage the competing demands between power and ethics, we include two parameters in the RAR algorithm, an allocation ratio cap $r^* > 1$ to protect power, and a maximal effect size to adapt, δ^* . When the absolute difference between the two observed response rates exceeds δ^* , the probability of assigning patients to the better performing arm will be capped:

$$\Pr(T_i = A) = \begin{cases} \frac{1}{2} + \frac{(r^*-1)(p_A - p_B)}{2(1+r^*)\delta^*} & \text{if } |p_A - p_B| \leq \delta^* \\ \frac{1}{2} + \frac{(r^*-1)}{2(1+r^*)} \text{sign}(p_A - p_B) & \text{if } |p_A - p_B| > \delta^* \end{cases} \quad (14)$$

Selection of r^* and δ^* can be made based on the desired power of the trial and the outcome response sensitivity. The allocation ratio cap used for the calculation of treatment allocation probabilities is not equal to the final allocation ratio, because the observed response rates of the two arms could change from time to time, and not all assignments are made based on the

capped allocation ratio. However, with the allocation ratio cap for each assignment, the final allocation ratio will not exceed the cap. This ensures that the trial will maintain the desired power to detect the treatment effect if such an effect exists.

4. IMPLEMENTATION

The strategies for managing competing demands between treatment allocation randomness and covariate balancing, and between patient ethics and power provide a general framework for implementing covariate adjusted response adaptive (CARA) randomization. In practice, two other issues must be taken into account when setting up a randomization plan for a large multicenter trial. One is the uneven patient accrual rate during the study period, and the other is the response delay. For trials involving a large number of sites with a multi-year long patient recruitment period, it is not uncommon for the trial to start with few sites, and gradually add sites during the first year. The slow early enrollment phase can involve amendments to the study protocol, study treatment procedure adjustments, case report form changes and individual site protocol learning curves. These changes often result in larger variability in patient outcomes during the initial stages of the trial. In addition, outcome delay is common in trials treating neurological diseases like stroke. For example, the primary outcome for the SHINE trial is the functional outcome at 3 months as measured by the modified Rankin Scale (mRS) Score. In order to have stable estimates on response rates, the RAR component of the randomization algorithm should begin once an adequate proportion of patients have been through the study protocol and have available outcome data.

Based on these considerations, the SHINE randomization plan is implemented in three periods. Period 1 begins the trial enrollment and since there is no available data to contribute to the RAR or covariate balancing, it functions as a ‘burn-in period’ and simply controls overall treatment assignment balance. Period 2 begins once there is sufficient accumulating information on the distribution of treatment assignment for the pre-specified prognostic variables. At this stage in the trial, there still may not be adequate outcome information so the RAR aspect of the randomization algorithm continues in a burn-in phase while the covariate balancing can commence and begin to focus on preventing serious imbalances in the specified baseline covariates. Period 3 begins once there is sufficient information on the primary outcome and incorporates a response adaptive allocation probability that is used when baseline covariate imbalances are under control. The actual timing of these periods is study dependent and should be planned based on expected enrollment rates and timing of the primary outcome.

In Period 1, patients are randomized with the block urn design (BUD) [49], which has the following conditional allocation probabilities:

$$\Pr(T_i=A)=\frac{MTI+\min(n_{i-1,A},n_{i-1,B})-n_{i-1,A}}{2\left(MTI+\min(n_{i-1,A},n_{i-1,B})\right)-(i-1)} \quad (15)$$

Here $n_{i-1,A}$ and $n_{i-1,B}$ are the total number of patients previously randomized into treatment arms A and B respectively before the randomization for subject i . MTI is the maximal tolerable imbalance. The block urn design consistently controls imbalance like the permuted block randomization, so that at any given time point, the absolute imbalance between the two arms will not exceed the MTI . However, the block urn design offers a higher randomness when $MTI > 1$. For example, with $MTI = 3$ the block urn design has an expected 5.9% of deterministic assignments, compared to 25% for permuted block design under the same MTI [49].

Period 2 begins controlling imbalances in the pre-specified baseline covariates with the minimal sufficient balancing method [35]. When an eligible patient is ready for randomization, imbalances in the distribution of the baseline NIHSS category, the thrombolysis category, and site corresponding to the current patient are checked with formulas (1–6). A biased coin probability of 0.7 is used for the treatment assignment for subject i in favor of reducing imbalances if the number of serious imbalances toward one treatment arm is more than the number of serious imbalances toward the other treatment arm. Based on computer simulation results, it is anticipated that three out of every four patients in this period will receive purely random assignments, and none of the three baseline covariates will have a serious imbalance indicated by a p -value less than 0.3 in the imbalance test.

When Period 3 begins, the system first checks the imbalances for the three baseline covariates. A biased coin assignment is used in order to reduce imbalances if it is deemed being required based on the minimal sufficient balancing strategy. Otherwise, the target allocation ratio, based on the observed response rates of the two treatment arms, is directly used as the conditional allocation ratio for the treatment assignments. A proper biased coin probability p_{bc} , such as 0.85 or 0.9, can be selected so that a sufficient majority of patients in this period will receive treatment assignments based on the RAR algorithm. When RAR is started, the target allocation ratio no longer equals 1:1. It is important to point out that, the target allocation ratio change does not affect the definition of baseline covariate imbalances. Formulas (1–6) are defined to assess the treatment distribution within a covariate subgroup to the overall treatment distribution. Therefore, during Period 3, the imbalance test is conducted in the same way as Period 2.

With this randomization plan, it is anticipated that the treatment allocation randomness will be well protected with an estimate of more than three-fourths of patients receiving treatment assignment based solely on the target allocation ratio. Imbalances with p -value less than 0.3 in the three baseline covariates will be prevented. The power of the trial will be protected, and those patients enrolled in Period 3 will have more than 50% chance of being assigned to the better performing arm.

5. DISCUSSION

Response adaptive randomization has rarely been implemented in large, multicenter confirmative trials. As its potential benefits for individual patients within the trial remain an attractive feature, other issues related to or affected by the randomization algorithm must be

simultaneously considered. Among them, treatment allocation randomness and baseline covariate balance must be properly protected. It is also important to point out that the ethical benefit of using RAR in trials designed under a frequentist framework is obtained at the cost of power. Under the condition of fixed power, the overall ethical benefit of RAR measured by the reduction of expected total number of failures may not exist as shown above.

Impact on the type I error is another important issue associated with RAR [50]. Consider a trial with the primary outcome affected by a time-dependent factor, so that those patients enrolled in the early phase of the trial have a lower success rate than those enrolled in the later stage of the trial. With RAR, the treatment group showing a higher success rate will benefit more from those later patients than the other treatment group. This will trigger an inflation of the type I error. If a negative time trend exists in the outcome, so that the success rate is reduced over time, the use of RAR will decrease the type I error of the trial. Computer simulation under the trial setting described above shows that if the success rate starts at 25% at the beginning of the trial and linearly increases to 35% at the end, the type I error will be inflated from 0.05 to 0.06. On the other hand, if the success rate drops from 25% to 20%, the type I error will be deflated from 0.05 to 0.04. Therefore, when RAR is implemented and a strong monotonic increase or decrease pattern is observed in patient outcomes, the interpretation of the trial results may be biased. Further work is needed in this area.

The use of RAR could create additional complexities for trial operations. For example, the change in target allocation ratio triggered by RAR could pose difficulties in managing study drug distribution. Nevertheless, with so many concerns, we believe it is possible to implement RAR with covariate balancing in a real trial and maintain a balance between randomness and baseline covariate balancing and ethics and power. Lessons and experiences learned from the use of the proposed randomization scheme for the SHINE trial will provide valuable knowledge for future trialists.

Acknowledgments

This study is supported by two National Institute of Neurological Diseases and Stroke NINDS grants: SHINE Grant U01 NS069498; NETT Grant U01 NS059041. The authors thank Dr. Yanqiu Weng for his valuable input on the simulation programming code.

References

1. Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statist Med.* 1995; 14:231–246.
2. Hu, F.; Rosenberger, WF. *The theory of response-adaptive randomization in clinical trials.* New York: Wiley; 2006.
3. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science.* 2004; 19:175–187.
4. Berry, SM.; Carlin, BP.; Lee, JJ.; Müller, P. *Bayesian adaptive methods for clinical trials.* Boca Raton: Chapman & Hall; 2011.
5. Meurer WJ, Lewis RJ, et al. An overview of the adaptive designs accelerating promising trials into treatments (ADAPT-IT) project. *Ann Emerg Med.* 2012; 60:451–457. [PubMed: 22424650]
6. Bartlett RH, et al. Extracorporeal circulation in neonatal respiration failure: a prospective randomized study. *Pediatrics.* 1985; 76:479–487. [PubMed: 3900904]

7. Wei LJ, Durham SD. The randomized play-the-winner rule in medical trials. *J Am Stat Assoc.* 1978; 73:840–843.
8. Ware JH. Investigating therapies of potentially great benefit: ECMO (with discussion). *Statistical Science.* 1989; 4:298–340.
9. Mike V, Krauss AN, Ross GS. Neonatal extracorporeal membrane oxygenation (ECMO): clinical trials and the ethics of evidence. *J Med Ethics.* 1993; 19:212–218. [PubMed: 8308876]
10. Eisele JR. The doubly adaptive biased coin design for sequential clinical trials. *J Stat Plan Inference.* 1994; 38:249–261.
11. Hu F, Rosenberger F. Optimality variability power: evaluating response-adaptive randomization procedures for treatment comparison. *J Am Stat Assoc.* 2003; 98:671–678.
12. Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Annals of Statistics.* 2004; 32:268–301.
13. Rosenberger WF, Huc F. Maximizing power and minimizing treatment failures in clinical trials. *Clin Trials.* 2004; 1:141–7. [PubMed: 16281886]
14. Sverdlov O, Tymofyeyev Y, Wong WK. Optimal response-adaptive randomized designs for multi-armed survival trials. *Stat Med.* 2011; 30:2890–910. [PubMed: 21823146]
15. Biswas A, Bhattacharya R. Near efficient target allocations in response-adaptive randomization. *Stat Methods Med Res.* 2012;10.1177/0962280212468378
16. Gu X, Lee J. A simulation study for comparing testing statistics in response-adaptive randomization. *BMC Medical Research Methodology.* 2010; 10:48. [PubMed: 20525382]
17. Thall P, Wathen K. Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Stat Med.* 2005; 24:1947–1964. [PubMed: 15806621]
18. Zhang L, Hu F. A new family of covariate-adjusted response adaptive designs and their properties. *Appl Math J Chinese Univ.* 2009; 24:1–13.
19. Ning J, Huang X. Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. *Stat Med.* 2010; 29:1761–1768. [PubMed: 20658546]
20. Magirr D. Block response-adaptive randomization in clinical trials with binary endpoints. *Pharmaceut Statist.* 2011; 10:341–346.
21. Bandyopadhyay U, Bhattacharya R. An urn based covariate adjusted response adaptive allocation design. *Stat Methods in MedRes.* 2012; 21:135–148.
22. Yuan Y, Huang X, Liu S. A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial. *Stat Med.* 2011; 30 :1218–29. [PubMed: 21432894]
23. Lewis RJ, Viele K, Broglio K, Berry SM, Jones AE. An Adaptive, Phase II, Dose-finding Clinical Trial Design to Evaluate L-Carnitine in the Treatment of Septic Shock Based on Efficacy and Predictive Probability of Subsequent Phase III Success. *Crit Care Med.* 2013 Mar 19. Epub ahead of print.
24. Korn EL, Freidlin B. Outcome-adaptive randomization: is it useful? *J Clin Oncol.* 2011; 29:771–776. [PubMed: 21172882]
25. Lee J, Chen N, Yin G. Worth adapting? Revisiting the usefulness of outcome-adaptive randomization. *Clin Cancer Res.* 2012; 18:1–10.
26. Lipsky AM, Lewis RJ. Response-adaptive decision-theoretic trial design: operating characteristics and ethics. *Stat Med.* 2013 Epub ahead of print. 10.1002/sim.5807
27. Southerland AM, Johnston KC. Considering hyperglycemia and thrombolysis in the Stroke Hyperglycemia Insulin Network Effort (SHINE) trial. *Ann N Y Acad Sci.* 2012; 1268:72–8. [PubMed: 22994224]
28. Bruno A, Durkalski VL, Hall CE, Juneja R, Barsan WG, Janis S, Meurer WJ, Fansler A, Johnston KC. The Stroke Hyperglycemia Insulin Network Effort (SHINE) trial protocol: a randomized, blinded, efficacy trial of standard vs. intensive hyperglycemia management in acute stroke. *Int J Stroke.* 2013 Mar 19. Epub ahead of print. 10.1111/ijvs.12045
29. Saver JL. Optimal end points for acute stroke therapy trials: best ways to measure treatment effects of drugs and devices. *Stroke.* 2011; 42:2356–62. [PubMed: 21719772]
30. Rosenberger, WF.; Lachin, JM. *Randomization in Clinical Trials Theory and Practice.* Wiley Interscience; 2002.

31. Rosenkranz GK. The impact of randomization on the analysis of clinical trials. *Stat Med.* 2011; 30:3475–87. [PubMed: 21953285]
32. Zhao W, Weng Y, Wu Q, Palesch Y. Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. *Pharm Stat.* 2012; 11:39–48. [PubMed: 21544929]
33. Senn S. Seven myths of randomisation in clinical trials. *Stat Med.* 2012 Dec 17. Epub ahead of print. 10.1002/sim.5713
34. Hill AB. The clinical trial. *British Medical Bulletin.* 1951; 71:278–282. [PubMed: 14879025]
35. Efron B. Forcing a sequential experiment to be balanced. *Biometrika.* 1971; 58:403–417.
36. Soares JF, Wu CF. Some restricted randomization rules in sequential designs. *Communications in Statistics – Theory and Methods.* 1983; 12:2017–2034.
37. Berger VM, Ivanova A, Knoll MD. Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Stat Med.* 2003; 22:3017–28. [PubMed: 12973784]
38. Wei LJ. A class of designs for sequential clinical trials. *J Am Stat Asso.* 1977; 72:382–386.
39. Berger, Vance. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials.* John Wiley & Sons, Ltd; 2005. p. 74-75.
40. Zhao W, Hill MD, Palesch Y. Minimal sufficient balance – a new strategy to balance baseline covariates and preserve randomness of treatment allocation. *Stat Methods Med Res.* 2012 Jan 26. Epub ahead of print.
41. Rosner, B. *Fundation of Biostatistics.* 4. Duxbury Press; 1995.
42. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics.* 1975; 31:103–115. [PubMed: 1100130]
43. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clin Pharmacol Ther.* 1974; 15:443–453. [PubMed: 4597226]
44. Berger V. Minimization, by its nature, precludes allocation concealment, and invites selection bias. *Contemp Clin Trials.* 2010; 31:406. [PubMed: 20457277]
45. Blackwell D, Hodges JL. Design for the control of selection bias. *Annals of Mathematical Statistics.* 1957; 28:449–460.
46. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet.* 2002; 16:359, 966–70.
47. Dumville JC, Hahn S, Miles JN, Torgerson DJ. The use of unequal randomisation ratios in clinical trials: a review. *Contemp Clin Trials.* 2006; 27:1–12. [PubMed: 16236557]
48. Rosenberger WF. Randomized play-the-winner clinical trials: review and recommendations. *Control Clin Trials.* 1999; 20:328–42. [PubMed: 10440560]
49. Zhao W, Weng Y. Block urn design - a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemp Clin Trials.* 2011; 32:953–61. [PubMed: 21893215]
50. Simon R, Simon NR. Using Randomization Tests to Preserve Type I Error With Response-Adaptive and Covariate-Adaptive Randomization. *Stat Probab Lett.* 2011; 81:767–772. [PubMed: 21769160]

Figure 1.

Treatment assignment randomness under Minimal Sufficient Balancing

Computer simulation settings: Baseline covariates: three categorical covariates. Imbalance control limit: p -value < 0.3 . Sample size: 500. Simulation: 1000 per scenario.

Figure 2.
Impact of allocation ratio on power under fixed sample size

Figure 3.
Impact of allocation ratio on expected number of total failures under two conditions