Published in final edited form as:

Stat Med. 2011 May 20; 30(11): 1218–1229. doi:10.1002/sim.4218.

A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial

Ying Yuan*, Xuelin Huang, and Suyu Liu

Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.

Abstract

We propose a Bayesian response-adaptive covariate-balanced randomization design for multiple-arm comparative clinical trials. The goal of the design is to skew the allocation probability to more efficacious treatment arms, while also balancing the distribution of the covariates across the arms. In particular, we first propose a new covariate-adaptive randomization method based on a prognostic score that naturally accommodates continuous and categorical prognostic factors and automatically assigns imbalance weights to covariates according to their importance in response prediction. We then incorporate this covariate-adaptive design into a group sequential response-adaptive randomization scheme. The resulting response-adaptive covariate-balanced randomization design combines the advantages of both covariate-adaptive and response-adaptive randomizations and meets the design goal. We illustrate the proposed design through its application to a phase II leukemia clinical trial, and evaluate its operating characteristics through simulation studies.

Keywords

adaptive randomization; prognostic factor; balance covariates; response-adaptive

1. Introduction

In clinical trials, response-adaptive randomization designs utilize accumulating information on the previous subjects' responses to skew the treatment assignment probabilities and assign more patients to the better treatment arms. Such designs are useful to mitigate the ethical problem of randomly assigning an equal number of subjects to each treatment in a clinical trial when some treatment arms may be inferior to others. Considerable research has been done on the response-adaptive randomization designs from both frequentist and Bayesian perspectives. See, for example, Thompson [1], Efron [2], Wei and Durham [3], Eisele [4], Berry and Eick [5], Rosenberger et al. [6], and Thall, Inoue and Martin [7], among others. A recent treatment of response-adaptive randomization designs can be found in Hu and Rosenberger [8].

Response-adaptive randomization designs have the advantage of assigning fewer patients to inferior treatment arms. However, these designs lack a mechanism to actively control the imbalance of prognostic factors, i.e., covariates that substantially affect the study outcome, across treatment arms. This may not be a serious issue under large samples as in the limit the

randomization automatically balances prognostic factors among treatment groups. However, for trials with small or moderate sample sizes, the imbalance of the prognostic factors can be substantial when using response-adaptive randomization designs, and thus causes difficulties to the inference after randomization. For example, in the presence of imbalanced prognostic factors, a direct comparison of marginal efficacy among the treatment arms is biased [9].

Our research is motivated by a phase II two-arm randomized clinical trial for acute myeloid leukemia (AML), which is being conducted at M. D. Anderson Cancer Center. A maximum of 100 patients will be enrolled and randomized to receive treatment A or B. The outcome of interest is a binary variable indicating whether the patients achieve complete remission (CR) after the treatment. The challenge of designing this trial is that the randomization procedure needs to be response-adaptive such that more patients will be allocated to the more effective treatment arm, and meanwhile the imbalance of the covariates must be controlled within a reasonable range.

Without considering response, various methods have been proposed to balance covariate distributions across treatment arms during randomization. For a small set of discrete covariates, stratified randomization is an effective method to achieve balance with respect to the covariates across treatment arms. This method, however, breaks down when there is a large number of covariates. Covariate-adaptive randomization designs have been developed to address this issue. In particular, Pocock and Simon [10] proposed a minimization design to balance prognostic factors in randomization. Wei [11] discussed the use of an urn model for covariate-adaptive randomization. Atkinson [12] proposed optimal biased coin designs for clinical trials by employing the D-optimality criterion with a linear model. Signorini et al. [13] and Heritier, Gebski and Pillai [14] proposed covariate-adaptive randomization procedures that balance interactions between factors when such interactions exist. Scott et al. [15] and McEntegart [16] provided comprehensive reviews on covariate-adaptive randomization.

We propose a randomization procedure that is response-adaptive and which also actively balances the covariates across treatment arms. Specifically, we develop a new prognostic-score-based covariate-adaptive randomization method that accommodates both categorical and continuous covariates. We then incorporate this method into a group sequential response-adaptive randomization design such that the resulting design skews the allocation probability to the better treatment arm, and also controls the imbalance of the prognostic factors across the arms. The proposed design differs from existing randomization designs that target a similar goal of being response-adaptive and balancing covariates. For example, Atkinson and Biswas [17] proposed an adaptive biased-coin design based on the optimum design theory. That design assumes a normal linear regression model with constant variance, whereas the methodology we proposed considers the logistic model for binary outcomes. Ning and Huang [18] proposed a response-adaptive randomization design that controls covariate imbalance. That approach requires polytomizing continuous covariates and focuses on 2-arm trials. In contrast, the proposed approach can naturally handle continuous variables and accommodate k -arm ($k \geq 2$) trials.

The rest of the article is organized as follows. In Section 2, we propose a new response-adaptive covariate-balanced randomization method based on a prognostic score and the group sequential design. In Section 3, we apply the proposed design to a leukemia clinical trial and use simulation studies to examine its operating characteristics. We conclude with a brief discussion in Section 4.

2. Method

Our response-adaptive covariate-balanced design is the combination of a response-adaptive randomization scheme selected from the literature and a newly proposed covariate-adaptive randomization method. For clarity of exposition, we first describe the new covariate-adaptive randomization method, and then describe how to incorporate this method into the response-adaptive randomization scheme.

2.1. Balancing prognostic factors through a prognostic score

The minimization method [10] is one of the methods most commonly used to balance prognostic factors. This method takes into account the distribution of prognostic factors of patients already recruited to the trial and allocates a new patient to the treatment arm that would, with a high probability, minimize the imbalance of those factors among the treatment arms. Pocock and Simon's method possesses good operating characteristics as far as balancing the prognostic factors among the treatment arms. Such a balance leads to a minimum variance unbiased estimation of the treatment effect under a homoscedastic linear regression model with marginal covariate effects and no treatment-covariate interactions. However, there are several difficulties in the implementation of the minimization method. First, continuous prognostic factors need to be categorized, and it is not always clear how many categories and what cutoff values should be used in this process. Second, as pointed out by Pocock and Simon [10], when some prognostic factors are considered more important than others, it is desirable to assign larger weights to the more important factors when determining the overall imbalance during a randomization procedure. Unfortunately, there is little guidance as to how these weights should be chosen to reflect the relative importance of the prognostic factors.

To address these problems, we propose a new covariate-adaptive randomization method based on a prognostic score, which is defined as follows. Let x denote a vector of prognostic factors that can be continuous or categorical, y denote the binary outcome variable, and z denote the treatment arm indicator. We assume a standard logistic response model

$$\text{logit}(\text{pr}(y=1|x, z)) = \alpha + x\beta' + \gamma z, \quad (1)$$

where α , β and γ are unknown parameters; and define the prognostic score as

$$w(x) = x\beta'.$$

A useful feature of the prognostic score is that the distribution of y conditional on covariates x is equal to the distribution of y conditional on the single variable prognostic score $w(x)$, or mathematically,

$$\text{pr}(y=1|w(x), z) = \text{pr}(y=1|x, z).$$

Therefore, to balance out the effect of prognostic factors across treatment arms, we actually only need to balance the distribution of the prognostic score during the randomization. By focusing on the single variable of the prognostic score rather than multiple prognostic factors simultaneously, the adaptive randomization procedure is conceptually simplified and a better balance can be achieved. Furthermore, the prognostic score automatically accommodates continuous and categorical prognostic factors, and assigns weights to prognostic factors according to their importance in predicting the response.

An evaluation of the prognostic score requires the estimation of the regression parameter β . At the beginning of the randomization, there are no or very few observations, making the estimation of the prognostic score impossible or very unstable. This difficulty can be overcome by utilizing historical data, which are often available. In practice, such prior information is routinely used to determine which prognostic factors need to be balanced before conducting the randomization. Under the Bayesian framework, we elicit an informative prior of β based on historical data, and continuously update the posterior mean of β using the observed data during the ongoing trial. The updating of $\hat{\beta}$ can be done continuously (after each patient) or after each group of patients, say, of equal size $m > 1$. If historical data are unavailable, we can use equal randomization at the beginning of the trial and then switch to the prognostic-factor based adaptive randomization after a certain amount of data are observed.

During randomization, we assign an incoming patient to the treatment arm such that the imbalance of the prognostic score across the treatment arms is minimized. This is the strategy advocated by Pocock and Simon [10]. To this end, we first define a measure of imbalance for the prognostic score. We propose using the Kolmogorov-Smirnov (KS) statistic as a measure of imbalance between two treatment arms. By minimizing the KS statistic during the randomization, we make the distributions of the prognostic score as close together as possible across treatment arms. Let w_k denote the vector of prognostic scores for patients assigned to the k th treatment arm, and $S_{kk'}$ denote the KS statistic based on w_k and $w_{k'}$ for $k \neq k'$. Then the overall imbalance among K treatment arms is measured by

$$S = \sum_{k=1}^{K-1} \sum_{k'=k+1}^K S_{kk'}.$$

Let $S^{(k)}$ denote the value of S if the incoming new patient is assigned to the k th treatment arm, and R_k denote the rank of $S^{(k)}$ in the set $\{S^{(k)}, k = 1, \dots, K\}$ in the ascending order. In the case of ties, a random ordering is used among the ties. We assign the new patient to the k th treatment arm with a probability

$$\pi_k^c = \begin{cases} \phi & \text{if } R_k = 1 \\ \frac{1-\phi}{K-1} & \text{if } R_k \neq 1 \end{cases} \quad (2)$$

where ϕ is a constant satisfying $1/K < \phi \leq 1$. The superscript of π_k^c indicates that it is a covariate-adaptive randomization probability. The allocation rule (2) biases treatment assignment in favor of balancing the prognostic score, thus in the long run it leads to the balanced allocation. The same allocation rule has been employed in the minimization design to achieve covariate balance [10]. Although more sophisticated allocation schemes are certainly possible, for example, we may take π_k^c proportional to the value of $S^{(k)}$ or R_k , Pocock and Simon [10] pointed out that the simple rule of (2) has very good operating characteristics and more sophisticated allocation schemes are typically unnecessary.

There are several important differences between the proposed prognostic-score-based randomization and the Pocock-Simon minimization. First, the prognostic-score-based method is a model-based approach that accommodates both continuous and discrete covariates, while the minimization is a model-free method which is more robust but requires polytomizing covariates. In addition, the Pocock-Simon's method is completely response-free, whereas the proposed prognostic-score-based randomization, although aims at minimizing covariate imbalances, does utilize response data and relies on the logistic

regression model, which is assumed to be correctly specified. Furthermore, the prognostic-score approach focuses on balancing a linear combination of the covariates that predicts the outcome as defined in the logistic regression model (1), rather than balancing each individual covariate. In contrast, the minimization targets to balance all prognostic factors at the same time. Therefore, in the case that balancing each prognostic factor marginally is of main interest, the Pocock-Simon minimization may be more appropriate.

2.2. Response-adaptive covariate-balanced randomization

The proposed prognostic-factor based randomization design is useful for minimizing the bias incurred by a possible imbalance of the prognostic factors. In the long run, it tends to allocate an equal number of subjects to the treatment arms. However, it is ethically desirable to skew the allocation probability to favor the better treatment arms. The approach we use to achieve this, described below, is similar to the response-adaptive group sequential designs by Jennison and Turnbull [19] and Karrison, Huo and Chappell [20].

Patients are enrolled in sequential groups of size $\{n_j\}$, $j = 1, \dots, J$, where n_j is the sample size of the sequential group j . Typically, before conducting the trial, researchers have little prior information regarding the superiority of the treatment arms. This is in contrast to the fact that physicians often have good knowledge on which prognostic factors are predictive of response. Therefore, initially, for the first J^* groups, e.g., $J^* = 1$, patients are allocated to K treatment arms with an equal probability $1/K$. The response information observed from these patients then can be used to skew the allocation probability in subsequent groups as follows: Conditional on the observed response data (without considering the covariate information) from all previous groups, patients in the j th group, $j > J^*$, are allocated to the k th treatment arm according to the posterior probability that treatment k is superior to all others,

$$\lambda_k = \text{pr}(p_k = \max \{p_\ell, 1 \leq \ell \leq K\} | \text{data}) \quad (3)$$

where p_k denotes the response rate of the k th treatment. The randomization criterion (3) is a generalization of the response-adaptive randomization design proposed by Thompson [1] for the case of two treatment arms. It has been used by Thall, Inoue and Martin [7] to adaptively randomize patients in a five-arm lymphocyte infusion trial. Clearly, the allocation criterion (3) assigns more patients to better treatment arms. Under the beta-binomial response model, the allocation probability can be easily evaluated based on the posterior distributions of the p_k 's, which follow Beta distributions.

One drawback of the randomization criterion (3) is that the resulting randomization probability λ_k is quite variable. To stabilize it, we apply the square root transformation on λ_k , and obtain the following stabilized randomization probability:

$$\pi_k^r = \frac{\sqrt{\lambda_k}}{\sum_{k=1}^K \sqrt{\lambda_k}}.$$

where the superscript of π_k^r indicates that it is a response-adaptive randomization. Although other forms of transformation, e.g., any positive transformation on λ_k , are possible, the square root transformation is often adequate and yields good operating characteristics in stabilizing the randomization probability. In the above group sequential design, π_k^r may vary across groups $j = 1, \dots, J$, but is constant for patients within any sequential group j .

Next, we describe how to incorporate the covariate-adaptive randomization proposed in Section 2.1 into the above group sequential response-adaptive randomization design. The idea is to vary the allocation probability for patients within a group according to the patient's individual prognostic score. Instead of π_k^r , we assign a patient to the k th treatment arm with the following probability,

$$\pi_k = \frac{\pi_k^c \pi_k^r}{\sum_{\ell=1}^K \pi_\ell^c \pi_\ell^r}, \quad (4)$$

which has an appealing feature that when π_k^r 's (or π_k^c 's) are the same across treatment arms, the randomization becomes simple covariate-adaptive randomization, i.e., $\pi_k = \pi_k^c$ (or response-adaptive randomization, i.e., $\pi_k = \pi_k^r$). The allocation probability (4) combines the covariate- and response-adaptive randomization probabilities in a multiplicative manner, but other forms of combination can be entertained. For example, any monotone function of the numerator of (4) will serve the purpose of combining the covariate- and response-adaptive randomization probabilities. In our design, we update $\hat{\beta}$ and π_k^r at the same time for each sequential group, but in principle we can update them separately and as frequently as we like by changing the size of the sequential group. When the size of the sequential group is one, we update these estimates after every patient. During randomization, we also impose the following futility and superiority early stopping rules:

- Futility: if $\text{pr}(p_k < p_{\text{ctl}} | \text{data}) > 0.99$ where p_{ctl} denote the response rate for the control arm, that is, there is strong evidence that treatment k is inferior to the control, we drop treatment arm k .
- Superiority: if $\text{pr}(p_k = \max\{p_\ell, 1 \leq \ell \leq K\} | \text{data}) > 0.99$, that is, there is strong evidence that treatment k is superior to all of the other treatments, we terminate the trial early and claim the superiority of treatment k .

3. Application

We applied the proposed design to the leukemia cancer trial that motivated our research. Through consultation with the PI, five potentially important prognostic factors were identified. The first one was patient age. The second one was patient cytogenetics (chromosomal deletion, duplication, or other chromosomal changes in the leukemia cells), by which patients can be classified into favorable, intermediate or poor risk cytogenetic groups. The other three possibly important factors are indicators of hematologic function, including the platelet count, concentration of hemoglobin, and white blood cell count. We estimated the effects of these factors on the outcome (i.e., CR or not) from historical data, as described in section 3.1.

3.1. Elicitation of the prior

We utilized the historical data to derive an informative prior for the unknown parameters in the logistic model (1). The historical data were collected from 1,374 AML patients who had been treated at M. D. Anderson Cancer Center in the years 1980 to 1999. We fitted the data using a linear logistic model with uniform noninformative priors on the regression parameters. Of the five potential prognostic factors, age (as a continuous variable) and cytogenetic groups were significantly associated with the response (see Table I). Cytogenetic groups was a three-level categorical variable (poor/intermediate/favorable) coded by two dummy variables *cyt1* (intermediate or not) and *cyt2* (favorable or not). Therefore, in the trial, we focused on balancing age and cytogenetic groups. We discounted

the posterior distribution of β by inflating the posterior variance by a factor of 68.7 such that the historical data was equivalent to one group (i.e., 20) patients. The resulting posterior distribution was used as the prior of β in the randomization, and is listed below:

$$\begin{aligned}\beta_1 &\sim N(-0.02, 0.03^2) \text{ for age} \\ \beta_2 &\sim N(-2.90, 4.11^2) \text{ for cyt1} \\ \beta_3 &\sim N(-1.88, 3.98^2) \text{ for cyt2}\end{aligned}$$

3.2. Operating characteristics

We used simulation studies to assess the operating characteristics of the proposed response-adaptive covariate-balanced (RC) design, and compared it with other randomization designs, including the equal randomization (EQ) design, the Bayesian response-adaptive randomization (RA) design of Thompson [1], the covariate-adaptive randomization (CA) design of Pocock and Simon [10], and the prognostic-score-adaptive randomization (PS) design discussed in Section 2.1. For ease of comparison, in the RA design, we updated the allocation probability for each sequential group of 20 patients rather than for each patient; and in the CA design, we used a rule that is similar to rule (4), i.e., the rule (a) of Pocock and Simon [10], to determine the allocation probabilities.

We generated data from the following model

$$\text{logit}(\text{pr}(y=1|x, z)) = \alpha + \beta_1 \text{age} + \beta_2 \text{cyt1} + \beta_3 \text{cyt2} + \beta_4 \text{treatment}, \quad k=1, 2.$$

where treatment is a indicator variable with 1 denoting the new treatment and 0 denoting the control. We generated the continuous variable of age from $N(52, 17^2)$, the binary indicator variables, cyt1 and cyt2, from Bernoulli distributions with success probabilities of 0.26 and 0.63, and set the values of β_1 , β_2 and β_3 to -0.02 , -2.90 and -1.88 , respectively. The distributions of the covariates and their regression parameters were chosen to match that of the historical data. As the CA design cannot handle continuous covariates directly, for that design, we categorized age into three categories according to tertiles. In the simulation, we varied the values of α and β_4 to generate different marginal response rates for the two treatment arms (Table II). For example, by setting $\alpha = 0.45$ and $\beta_4 = 0$, we obtained $p_1 = 0.1$ and $p_2 = 0.1$, and setting $\alpha = 0.45$ and $\beta_4 = 1.54$ yielded $p_1 = 0.1$ and $p_2 = 0.3$. The maximum total sample size is 100 and the size of each sequential group is 20. At the end of the study, the null hypothesis of equal efficacy between the two arms is rejected if the posterior probability $\text{pr}(p_2 > p_1 | \text{data}) > 0.975$ (or < 0.025), where p_2 and p_1 are the response rates for the treatment and control arms, respectively. This Bayesian style test is asymptotically equivalent to the chi-square test. A total of 10,000 independent simulations was performed for each configuration and allocation method. Note that when stopping rules are applied, the actually used sample size varies under different designs, which makes the comparison between designs difficult. To facilitate the comparison, we carried out simulations both with and without early stopping. The simulation code was written using R [22]. We used the function MCMClogit from the MCMCpack package to fit logistic response model (1) to obtain the posterior distributions of the regression parameters.

Table II shows the simulation results without early stopping based on the fixed sample size of $n = 100$. For each design, we present the average number of patients assigned to the inferior treatment arm (ITN), the expected success lost (ESL), the imbalance of covariates in terms of prognostic score, the standard error of the estimate of the treatment effect (i.e., β_4) based on the logistic regression model including all prognostic factors, and the power/type I

error rate of testing the null hypothesis of $p_1 = p_2$. The ESL is the difference between the expected number of successes had all patients received the superior treatment and the expected number of successes using the allocation rule, and equals $ITN \times (p_2 - p_1)$ [21]. The imbalance between the arms is measured by the KS statistic of the prognostic score. A larger value represents a more severe imbalance between the two arms. The percentage of significant imbalance (the p -value of the KS statistic is less than 0.05) is also reported.

The simulation results show that the proposed RC design successfully combines the advantages of the response-adaptive and the covariate-adaptive designs. Like the response-adaptive design, RA, the RC design effectively skewed the allocation probability toward the superior arm. It allocated substantially fewer patients to the inferior treatment arm compared to the EQ and the covariate-adaptive designs, CA and PS. For example, when $p_1 = 0.1$ and $p_2 = 0.2, 0.3$ and 0.4 , the number of patients assigned to the inferior treatment (i.e., ITN) and the expected success lost under the RC design were approximately 23%, 40% and 53% less than those under the EQ (or CA and PS) design, respectively. The variation of the ITN under the proposed RC design was larger than the covariate-adaptive designs (i.e., CA and PS), but less than the response-adaptive design (i.e., RA).

On the other hand, in terms of balancing the covariates, the performance of the RC design was comparable to the “pure” covariate-adaptive designs, CA and PS, and substantially better than the RA and EQ designs. For instance, when $p_1 = 0.1$ and $p_2 = 0.3$, the percentage of significantly imbalanced covariates under the RC design was 0.1%, which is close to that of the CA (0.6%) and PS (0.0%) designs, and substantially better than that of the RA (4.6%) and EQ (5.0%) designs. A better balance of covariates under the RC design often translated into a lower type I error rate (when the efficacy of the two treatment arms is the same) or a higher statistical power (when the efficacy of the two arms is different). For example, when $p_1 = p_2 = 0.3$, the type I error rate of the RA was 11.1%, while that of the RC design was only 5.0%. When $p_1 = 0.1$ and $p_2 = 0.3$, the power of the RC and RA designs were 69.6% versus 59.5%, respectively. In addition, in terms of estimating the treatment effect, the RC was less efficient than the covariate-adaptive designs (i.e., CA and PS), but substantially more efficient than the response-adaptive design (i.e., RA). For instance, when $p_1 = 0.3$; $p_2 = 0.5$, standard errors of the estimate of the treatment effect were 0.5, 1.6 and 3.1 for the CA, RC and RA, respectively.

In Figures 1 and 2, we show the average covariate imbalance and the randomization assignment probability when patients sequentially enter the trial under $p_1 = 0.3$ and $p_2 = 0.3, 0.4, 0.5$, and 0.6 . During the randomization process, the PS and RC designs had uniformly lower covariate imbalances than the other methods (Figure 1). In all scenarios, the two curves at the bottom are for the PS and RC designs. The RA design had the highest covariate imbalance. This imbalance was more severe when the efficacy level of arm 2 was substantially higher than that of arm 1 (Figure 1(d)). In these cases the RA design tended to assign a small number of patients to the inferior arm (i.e., arm 1), which resulted in the high covariate imbalance. Figure 2 displays the randomization assignment probability for treatment arm 2 during the randomization. Both the RA and the proposed RC design effectively skewed the allocation probability to the more efficacious arm, treatment arm 2, after the first sequential group. Without considering response information, the other designs performed essentially equal randomizations.

Table III shows simulation results with early stopping. We applied the stopping rules at the end of each sequential group of 20 patients, which resulted in a total of four interim analyses. In the presence of early stopping, the actual sample sizes used in trials vary under different designs. Therefore, in addition to the summary statistics that are similar to those listed in Table II, we also reported the average sample size across 10,000 simulated trials.

The simulation results are similar to those achieved without stopping rules. Compared to the RA design, the proposed RC design has a substantially lower percentage of significantly imbalanced covariates. Minimizing the covariate imbalance helps to control the type I errors, and helps to retain power and save sample size. Consequently, the necessary sample size under the RC design was smaller than that under the RA design. For example, when $p_1 = 0.1$ and $p_2 = 0.4$, the necessary sample size under the RC design was 60.0, while that under the RA design was 66.1. Moreover, compared to the covariate-adaptive designs (i.e., CA and PS), the RC design allocated fewer patients to the inferior arm. For example, in the case where $p_1 = 0.1$ and $p_2 = 0.4$, the number of patients assigned to the inferior treatment was 27.9 and 28.6, respectively, under the CA and PS designs, while that under the proposed RC design was only 20.8.

4. Discussion

We have proposed a Bayesian response-adaptive covariate-balanced randomization design for multiple-arm clinical trials. We first proposed a novel covariate-adaptive randomization method based on a prognostic score that naturally accommodates continuous and categorical prognostic factors and automatically assigns imbalance weights to the covariates according to their importance in response prediction. We then incorporated this covariate-adaptive design into a group sequential response-adaptive randomization design. The resulting design combines the advantages of covariate-adaptive and response-adaptive randomizations. It allocates more patients to efficacious arms, while also balancing the covariates across the treatment arms during the randomization process, as demonstrated in the simulation studies.

We have investigated the finite-sample operating characteristics of the proposed design using simulation studies. It is also of interest to investigate the theoretical properties of the proposed response-adaptive covariate-balanced design, such as the convergence and asymptotic distribution of the allocation proportions. Considerable research has been conducted on the theoretical properties of the covariate-adaptive and response-adaptive randomization methods individually [8]. It is not immediately clear whether these theories directly apply to the proposed design, which combines the covariate-adaptive and response-adaptive randomization. This is a topic of our future research.

The proposed design simultaneously considers response and covariate information when conducting the randomization. It belongs to a general class of covariate-adjusted response-adaptive (CARA) randomization designs, which have been an area of active research in the last decade [9]. A general CARA randomization design is defined as a randomization procedure in which the treatment assignment probability of the $(m + 1)$ th patient depends on the history of previous m patients' treatment assignments, responses, covariates, and the covariate vector of the $(m + 1)$ th patient. The method proposed in this paper utilizes all these data to determine randomization probabilities (unlike Pocock-Simon's procedure which does not depend on responses). The CARA randomization procedures may serve different study objectives. For instance, the goals may be: skewing the allocation probability in the direction of the treatment which is clinically "best" given a patient's covariate profile; targeting allocations which maximize inferential aspects of the study design when the primary outcome follows some heteroscedastic or nonlinear model; and skewing the allocation probability in the direction of the better treatment, while controlling covariate imbalances. The proposed prognostic score balanced randomization design addresses the last goal.

Acknowledgments

We would like to thank the referees, Associate Editor and Editor for very helpful comments that substantially improved this paper. This research was partially supported by National Cancer Institute grant R01CA154591-01A1.

References

1. Thompson WR. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*. 1933; 25:275–294.
2. Efron B. Forcing a sequential experiment to be balanced. *Biometrika*. 1971; 58:403–417.
3. Wei LJ, Durham S. The randomized pay-the-winner rule in medical trials. *Journal of the American Statistical Association*. 1978; 73:840–843.
4. Eisele JR. The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*. 1994; 38:249–261.
5. Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. *Statistics in Medicine*. 1995; 14:231–224. [PubMed: 7724909]
6. Rosenberger WF, Stallard N, Ivanova A, Harper CH, Ricks ML. Optimal adaptive designs for binary response trials. *Biometrics*. 2001; 57:909–913. [PubMed: 11550944]
7. Thall PF, Inoue LYT, Martin TG. Adaptive decision making in a lymphocyte infusion trial. *Biometrics*. 2002; 58:560–568. [PubMed: 12229990]
8. Hu, F.; Rosenberger, WF. *The Theory of Response-Adaptive Randomization in Clinical Trials*. New York: Wiley-Interscience; 2006.
9. Rosenberger WF, Sverdlov O. Handling covariates in the design of clinical trials. *Statistical Science*. 2008; 23:404–419.
10. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975; 31:103–115. [PubMed: 1100130]
11. Wei LJ. An application of an urn model to the design of sequential controlled clinical trials. *Journal of the American Statistical Association*. 1978; 73:559–563.
12. Atkinson AC. Optimal biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*. 1982; 69:61–67.
13. Signorini DF, Leung O, Simes RJ, Beller E, Gebiski VJ. Dynamic balanced randomization for clinical trials. *Statistics in Medicine*. 1993; 12:2343–2350. [PubMed: 8134737]
14. Heritier S, Gebiski V, Pillai A. Dynamic balancing randomization in controlled clinical trials. *Statistics in Medicine*. 2005; 24:3729–3741. [PubMed: 16320277]
15. Scott NW, Mcpherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: A review. *Controlled Clinical Trials*. 2002; 23:662–674. [PubMed: 12505244]
16. McEntegart D. The pursuit of balance using stratified and dynamic randomization techniques: An overview. *Drug Information Journal*. 2003; 37:293–308.
17. Atkinson AC, Biswas A. Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal response. *Statistics in Medicine*. 2005; 24:2477–2492. [PubMed: 16007571]
18. Ning J, Huang X. Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. *Statistics in Medicine*. 2010; 29:1761–1768. [PubMed: 20658546]
19. Jennison, C.; Turnbull, B. *Group Sequential Methods with Applications to Clinical Trials*. London: Chapman & Hall/CRC; 2000.
20. Karrison T, Huo D, Chappell R. Group sequential, response-adaptive designs for randomized clinical trials. *Controlled Clinical Trials*. 2003; 24:506–522. [PubMed: 14500050]
21. Coad DS. A comparative study of some data-dependent allocation rules for Bernoulli data. *The Journal of Statistical Computation and Simulation*. 1992; 40:219–231.
22. R Development Core Team. *R Foundation for Statistical Computing*. Austria: Vienna; 2010. R: a language and environment for statistical computing.

Figure 1.

Covariate imbalance between two treatment arms as patients sequentially enter the trial under the equal (EQ), response-adaptive (RA), covariate-adaptive (CA), prognostic-score-adaptive (PS), and response-adaptive covariate-balanced (RC) randomization designs. For ease of viewing, the labels in the legends are arranged in the same vertical order as corresponding curves. In (a) and (b), the two largely overlapping curves in the bottom are for the RC and PS, with the lower one being the PS.

Figure 2.

Randomization probability of arm 2 as patients sequentially enter the trial under the equal (EQ), response-adaptive (RA), covariate-adaptive (CA), prognostic-score-adaptive (PS), and response-adaptive covariate-balanced (RC) randomization designs.

Estimates of historical data based on a logistic regression model with five covariates: age, the platelet count (plt), concentration of hemoglobin (hgb), white blood cell count (wbc), and cytogenetic groups (coded by two dummy variables cyt1 and cyt2).

Table 1

Estimate	age	cyt1	cyt2	hgb	wbc	plt
Posterior mean	-0.0216	-2.90	-1.88	0.03	-0.002	0.07
Posterior SD	0.004	0.50	0.48	0.04	0.001	0.07
95% CI	(-0.029, -0.014)	(-3.98, -2.03)	(-2.93, -1.02)	(-0.04, 0.10)	(-0.004, 0.001)	(-0.07, 0.20)

Table 2

Simulation results without early stopping, including inferior treatment number (ITN), expected success lost (ESL), imbalance and power (%), for equal (EQ), response-adaptive (RA), covariate-adaptive (CA), prognostic-score-adaptive (PS), and response-adaptive covariate-balanced (RC) randomization designs. The number in parentheses under the “ITN” column is the standard deviation of the ITN, and the number in parentheses under the “Imbalance” column is the percentage of significantly imbalanced covariates.

p_1	p_2	Methods	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^d (%)	p_1	p_2	Methods	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^d (%)
0.1	0.1	EQ	50.0 (5.0)	0.0	0.17 (4.7)	1.8	2.0	0.3	0.5	EQ	50.0 (5.0)	10.0	0.17 (4.9)	0.5	49.4
		RA	50.1 (22.0)	0.0	0.19 (5.0)	8.1	8.2			RA	24.7 (15.2)	4.9	0.21 (5.1)	3.1	45.6
		CA	50.0 (0.8)	0.0	0.14 (0.8)	1.5	3.0			CA	50.0 (0.9)	10.0	0.14 (0.7)	0.5	50.9
		PS	50.0 (7.0)	0.0	0.11 (0.0)	1.7	4.6			PS	50.1 (7.4)	10.0	0.10 (0.0)	0.5	48.1
		RC	50.0 (12.1)	0.0	0.11 (0.1)	4.3	6.8			RC	33.7 (11.3)	6.7	0.12 (0.0)	1.6	48.9
0.1	0.2	EQ	50.0 (5.0)	5.0	0.17 (4.9)	1.5	30.0	0.3	0.6	EQ	49.9 (5.0)	15.0	0.17 (4.7)	0.5	82.8
		RA	30.9 (18.4)	3.1	0.20 (4.7)	7.1	28.1			RA	18.1 (10.2)	5.4	0.23 (4.5)	3.2	73.5
		CA	50.0 (0.9)	5.0	0.14 (0.7)	1.2	29.2			CA	50.0 (0.9)	15.0	0.14 (0.5)	0.5	85.4
		PS	49.9 (7.4)	5.0	0.11 (0.0)	1.5	27.6			PS	50.1 (7.5)	15.0	0.10 (0.0)	0.5	83.2
		RC	38.4 (11.8)	3.8	0.12 (0.0)	4.5	27.9			RC	27.0 (10.1)	8.1	0.13 (0.1)	1.9	82.2
0.1	0.3	EQ	50.0 (4.9)	10.0	0.17 (5.0)	1.4	71.1	0.5	0.5	EQ	50.0 (5.0)	0.0	0.17 (5.0)	0.5	5.2
		RA	20.5 (12.3)	4.1	0.22 (4.6)	7.0	59.5			RA	49.4 (21.9)	0.0	0.19 (5.2)	1.4	10.8
		CA	50.0 (0.9)	10.0	0.14 (0.6)	1.3	73.3			CA	50.0 (0.9)	0.0	0.14 (0.7)	0.5	4.4
		PS	50.0 (7.4)	10.0	0.11 (0.0)	1.4	72.2			PS	50.0 (7.4)	0.0	0.10 (0.0)	0.5	3.5
		RC	30.0 (10.8)	6.0	0.13 (0.1)	5.2	69.6			RC	49.9 (12.4)	0.0	0.11 (0.0)	0.6	4.5
0.1	0.4	EQ	49.9 (5.0)	15.0	0.17 (4.8)	1.5	94.2	0.5	0.6	EQ	50.0 (5.0)	5.0	0.17 (5.0)	0.5	13.7
		RA	15.3 (7.5)	4.6	0.24 (4.8)	7.2	84.5			RA	37.3 (20.2)	3.7	0.20 (5.2)	1.3	17.9
		CA	50.0 (0.9)	15.0	0.14 (0.5)	1.2	95.3			CA	50.0 (0.9)	5.0	0.14 (0.5)	0.5	12.9
		PS	49.9 (7.4)	15.0	0.11 (0.0)	1.3	95.5			PS	49.9 (7.4)	5.0	0.10 (0.0)	0.5	11.6
		RC	23.7 (8.9)	7.1	0.14 (0.1)	5.9	93.1			RC	42.5 (12.2)	4.2	0.11 (0.0)	0.7	13.7
0.3	0.3	EQ	50.0 (5.0)	0.0	0.17 (4.7)	0.5	5.2	0.5	0.7	EQ	49.9 (5.0)	10.0	0.17 (5.1)	0.5	44.6
		RA	50.0 (22.1)	0.0	0.19 (5.1)	3.2	11.1			RA	25.7 (12.6)	5.1	0.21 (4.6)	1.2	45.5
		CA	50.0 (0.9)	0.0	0.14 (0.5)	0.5	3.8			CA	50.0 (3.3)	10.0	0.14 (0.6)	0.5	46.9
		PS	50.0 (7.4)	0.0	0.10 (0.0)	0.5	3.5			PS	50.0 (6.0)	10.0	0.10 (0.0)	0.5	44.2
		RC	50.0 (12.4)	0.0	0.11 (0.0)	1.2	5.0			RC	34.3 (10.0)	6.9	0.12 (0.0)	0.7	47.2

p_1	p_2	Methods	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^a (%)	p_1	p_2	Methods	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^a (%)
0.3	0.4	EQ	50.0 (5.0)	5.0	0.17 (4.9)	0.5	16.4	0.5	0.8	EQ	50.0 (5.0)	15.0	0.17 (4.7)	0.6	84.7
		RA	35.8 (20.0)	3.6	0.20 (4.9)	2.9	18.5			RA	17.2 (9.2)	5.2	0.23 (4.4)	1.3	80.3
		CA	50.0 (0.9)	5.0	0.14 (0.6)	0.5	15.2			CA	50.0 (0.9)	15.0	0.14 (0.5)	0.6	86.8
		PS	50.0 (7.4)	5.0	0.10 (0.0)	0.5	13.8			PS	49.9 (7.5)	15.0	0.10 (0.0)	0.6	86.8
		RC	41.3 (12.1)	4.1	0.11 (0.0)	1.3	16.1			RC	26.3 (10.0)	7.9	0.13 (0.1)	0.8	85.5

^aType I error when $p_1 = p_2$

Table 3

Simulation results with early stopping, including total sample size, inferior treatment number (ITN), expected success lost (ESL), imbalance and power (%), for equal (EQ), response-adaptive (RA), covariate-adaptive (CA), prognostic-score-adaptive (PS), and response-adaptive covariate-balanced (RC) randomization designs with an early-stopping rule. The number in parentheses under the "ITN" column is the standard deviation of the ITN, and the number in parentheses under the "Imbalance" column is the percentage of significantly imbalanced covariates.

P_1	P_2	Methods	Sample size	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^a (%)	P_1	P_2	Methods	Sample size	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^a (%)
0.1	0.1	EQ	98.1	49.1 (7.1)	0.0	0.17 (5.1)	1.9	7.8	0.3	0.5	EQ	83.6	41.8 (13.6)	8.4	0.20 (5.0)	1.2	53.1
		RA	99.3	49.5 (14.2)	0.0	0.19 (5.1)	3.7	8.6			RA	86.8	24.6 (13.2)	4.9	0.22 (4.6)	1.7	45.7
		CA	98.4	49.2 (4.5)	0.0	0.14 (0.5)	1.7	6.6			CA	85.0	42.5 (12.6)	8.5	0.16 (0.6)	1.1	53.3
		PS	98.5	49.2 (8.2)	0.0	0.11 (0.0)	1.8	6.5			PS	85.3	42.6 (13.6)	8.5	0.12 (0.0)	1.1	52.0
		RC	99.0	49.3 (12.6)	0.0	0.11 (0.0)	3.1	7.4			RC	86.3	32.6 (12.9)	6.5	0.13 (0.0)	1.4	49.7
		EQ	91.2	45.6 (11.0)	4.6	0.18 (4.7)	2.6	33.0	0.3	0.6	EQ	68.4	34.2 (15.0)	10.3	0.22 (4.8)	1.5	83.7
0.1	0.2	RA	95.2	30.4 (13.7)	3.0	0.21 (4.7)	3.9	28.7			RA	73.7	17.8 (11.0)	5.3	0.25 (4.7)	1.8	73.9
		CA	91.7	45.8 (9.9)	4.6	0.15 (0.6)	2.4	32.5			CA	69.1	34.5 (14.3)	10.4	0.19 (0.8)	1.3	86.1
		PS	92.8	46.3 (11.4)	4.6	0.12 (0.0)	2.3	30.6			PS	68.8	34.5 (15.4)	10.3	0.14 (0.0)	1.4	85.7
		RC	93.5	37.7 (13.2)	3.8	0.12 (0.0)	3.4	29.9			RC	70.4	25.0 (11.3)	7.5	0.15 (0.0)	1.6	82.7
		EQ	75.1	37.5 (14.6)	7.5	0.21 (4.8)	3.1	74.4	0.5	0.5	EQ	97.1	48.6 (8.3)	0.0	0.17 (5.3)	0.7	8.3
		RA	82.4	20.2 (11.7)	4.0	0.23 (4.3)	4.3	61.6			RA	96.8	48.7 (14.7)	0.0	0.20 (5.2)	0.8	12.2
0.1	0.3	CA	75.1	37.5 (14.0)	7.5	0.18 (0.8)	2.8	75.5			CA	98.0	49.0 (5.8)	0.0	0.15 (0.5)	0.5	6.6
		PS	75.8	38.0 (14.7)	7.6	0.14 (0.0)	3.0	75.3			PS	98.2	49.2 (8.5)	0.0	0.10 (0.0)	0.5	5.7
		RC	78.3	27.9 (11.9)	5.6	0.14 (0.0)	4.0	71.4			RC	98.0	48.9 (13.1)	0.0	0.11 (0.0)	0.6	6.7
		EQ	56.1	28.1 (13.9)	8.4	0.24 (4.5)	3.4	94.9	0.5	0.6	EQ	95.2	47.6 (9.5)	4.8	0.18 (4.8)	0.8	15.9
		RA	66.1	15.0 (9.0)	4.5	0.27 (4.5)	4.6	85.3			RA	95.0	37.2 (14.9)	3.7	0.20 (4.8)	0.7	17.7
		CA	55.8	27.9 (13.5)	8.4	0.21 (0.8)	2.9	95.9			CA	96.2	48.1 (7.7)	4.8	0.15 (0.7)	0.8	15.7
0.3	0.3	PS	57.1	28.6 (14.1)	8.6	0.16 (0.0)	3.2	95.6			PS	96.5	48.2 (10.0)	4.8	0.11 (0.0)	0.8	14.3
		RC	60.0	20.8 (9.7)	6.3	0.17 (0.0)	4.2	94.6			RC	96.1	42.0 (13.4)	4.2	0.11 (0.0)	0.8	14.9
		EQ	96.7	48.4 (8.4)	0.0	0.17 (5.0)	0.9	8.6	0.5	0.7	EQ	86.6	43.2 (12.9)	8.6	0.19 (5.0)	0.4	47.5
		RA	96.8	48.5 (15.0)	0.0	0.19 (4.8)	1.5	11.3			RA	86.8	25.1 (13.3)	5.0	0.22 (4.9)	0.4	46.6
		CA	97.5	48.8 (6.3)	0.0	0.15 (0.7)	0.8	6.7			CA	86.8	43.4 (11.9)	8.7	0.16 (0.8)	0.4	49.8
		PS	97.9	48.9 (8.8)	0.0	0.11 (0.0)	0.7	6.0			PS	87.5	43.7 (13.1)	8.7	0.12 (0.0)	0.4	47.9
		RC	97.8	48.7 (13.4)	0.0	0.11 (0.0)	1.1	6.7			RC	86.9	33.1 (12.9)	6.6	0.13 (0.0)	0.4	48.9

p_1	p_2	Methods	Sample size	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^d (%)	p_1	p_2	Methods	Sample size	ITN	ESL	Imbalance	$sd(\beta_4)$	Power ^d (%)
0.3	0.4	EQ	93.4	46.7 (10.3)	4.7	0.18 (5.0)	1.0	20.4	0.5	0.8	EQ	67.2	33.5 (14.6)	10.0	0.22 (4.8)	1.7	86.0
		RA	94.6	35.6 (14.7)	3.6	0.20 (5.0)	1.5	19.1			RA	69.6	17.0 (10.3)	5.1	0.25 (4.6)	1.5	81.0
		CA	94.7	47.3 (8.5)	4.7	0.15 (0.8)	0.8	18.6			CA	67.3	33.6 (13.9)	10.1	0.19 (0.9)	1.5	88.1
		PS	95.2	47.6 (10.5)	4.8	0.11 (0.0)	0.9	17.4			PS	68.0	34.0 (15.1)	10.2	0.14 (0.0)	1.6	87.6
		RC	95.2	41.0 (13.3)	4.1	0.11 (0.0)	1.2	17.6			RC	67.8	24.0 (11.1)	7.2	0.15 (0.0)	1.3	87.2

^dType I error when $p_1 = p_2$