Research Article

Balancing Treatment Allocation over Continuous Covariates: A New Imbalance Measure for Minimization

Yanqing Hu[1] and Feifang Hu[2]

[1] *Department of Statistics, West Virginia University, Hodges Hall, Morgantown, WV 26506, USA*
[2] *Department of Statistics, University of Virginia, Halsey Hall, Charlottesville, VA 22904, USA*

Correspondence should be addressed to Feifang Hu, fh6e@virginia.edu

In many clinical trials, it is important to balance treatment allocation over covariates. Although a great many papers have been published on balancing over discrete covariates, the procedures for continuous covariates have been less well studied. Traditionally, a continuous covariate usually needs to be transformed to a discrete one by splitting its range into several categories. Such practice may lead to loss of information and is susceptible to misspecification of covariate distribution. The more recent papers seek to define an imbalance measure that preserves the nature of continuous covariates and set the allocation rule in order to minimize that measure. We propose a new design, which defines the imbalance measure by the maximum assignment difference when all possible divisions of the covariate range are considered. This measure depends only on ranks of the covariate values and is therefore free of covariate distribution. In addition, we developed an efficient algorithm to implement the new procedure. By simulation studies we show that the new procedure is able to keep good balance properties in comparison with other popular designs.

1. Introduction

Balanced allocation among treatment groups is often desirable in many clinical trials. A well-balanced design enhances the trials by increasing the credibility of the trial, precision of subgroup or interim analysis, and robustness to model misspecification [1]. Moreover, ignorance of balance at the design stage may lead to loss of statistical efficiency, especially in small trials [2]. In the existence of covariates (or prognostic factors), well-balanced allocation does not only mean similar group sizes but also similar distributions of covariate values across treatment groups [3–6]. With discrete covariates, a great many papers have been published, which include stratified permuted block designs, marginal procedures (or minimization) [7, 8], and hierarchical models [9–11]. Y. Hu and F. Hu [12] are among the few who explored the theoretical properties of such procedures.

With respect to balancing over continuous covariates, a traditional way is to discretize a continuous covariate by splitting its range into several small intervals, each of which defines a distinct category [4, 13]. However, such methods have several potential problems. First, there is no standard way of defining such intervals: usually the boundaries are set by experts' opinion from a nonstatistical perspective, or they are heavily dependent on the model assumption of the covariate distribution. Weir and Lees [13] studied the randomization procedure in an acute stroke clinical trial in which 12 covariates were involved. The five continuous covariates were discretized before the traditional minimization was applied. Based on the historical information of the covariate distribution, the covariates age and mean arterial pressure were discretized by their quintiles; plasma glucose level was first discretized into two groups by the critical value 8 mmol/L and the one with larger values was further split by quartiles; the remaining two covariates, time of delay from stroke onset and Glasgow coma scale, were categorized whose critical values were set by clinical experience. Therefore, we see that discretization methods are rather covariate-specific and may depend on the distribution of the covariates. Second, there is a dilemma as to the selection of the number of categories (denoted by m). While a small m may lead to loss of information, a large m can result in a severe overall imbalance, as is in the case of stratified designs when the number of strata is large [7].

An alternative that saves the step of discretization is to define a certain "imbalance measure" for continuous covariates among different treatment groups and then allocate the treatments sequentially to minimize that specific measure [5, 6, 14–16]. With two treatment groups A and B, many authors defined imbalance measures as the mean difference of covariate values between the two groups: some of the measures may further be adjusted for standard deviations within the two groups [5, 14], others normalized the mean difference by calculating the two-sample t-statistic and the corresponding P value [15], that is, a larger P value indicates more balanced allocation. An alternative to the above mean comparison methods is model-based approach [17–20], which relates different allocation to different design matrices in a linear model, which in turn influence the maximum likelihood estimator of the treatment effect $\hat{\alpha}_n$; it then sequentially favors an assignment that would lead to a smaller $var(\hat{\alpha}_n)$. This approach only implies balance under a homogeneous linear model.

The methods discussed above, however, usually depend on the assumption of normal distributions of the covariates. Among the methods that apply to more general distributions, Stigsby and Taves [16] considered the difference of rank sums as their imbalance measure; Su [6] used a weighted average of the difference in overall patient numbers and the difference in quartiles of the continuous covariates. Rosenberger and Sverdlov [21] found the empirical distributions of the covariates in the two groups and calculated their Kolmogorov-Smirnov distance to measure the imbalance for a series of randomization procedures.

In this paper, we propose a new minimization procedure which carefully examines the assignment differences over all intervals and set their maximum value as the actual imbalance measure. The new procedure utilizes only the rank information of the covariates and is therefore robust against various distributions. Moreover, we developed an efficient algorithm whose computation is even less intensive than discretization. By simulation studies, we show that our new procedure leads to well-balanced allocation in terms of overall patient numbers and covariate distributions. For simplicity, we first consider only one continuous covariate and then extend the idea to more general cases.

In Section 2, by giving a motivating example, we analyze the downside of simple discretization and then highlight the idea underlying the new procedure. In Section 3, details of the procedure are described, and the key part of the algorithm is explained. Simulation
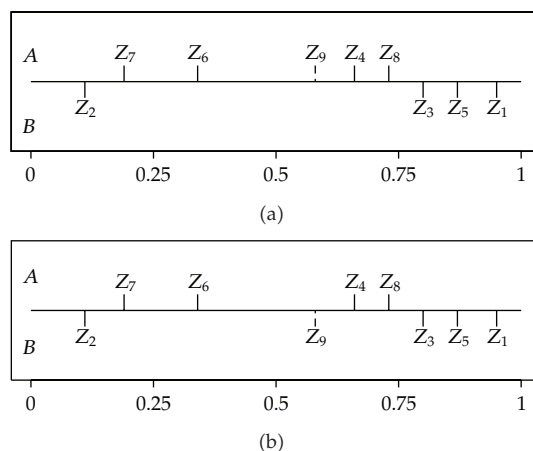
Figure 1: Two allocation sequences for the same covariate values.

studies are carried out in Section 4, comparing our procedure with discretization methods as well as some other designs. In Section 5, we conclude the paper by discussing possible extension of the current procedure.

2. Motivation

We will first focus on two treatments, A and B and only one continuous covariate Z. For simplicity, assume Z takes value over interval $[0, 1]$; any other type of Z can be transformed to $[0, 1]$ by a linear function if Z is bounded, or a nonlinear function such as $e^x/(e^x + 1)$ otherwise.

To see the difficulty in defining "balance" for a continuous covariate, we use an example as shown in Figure 1. The first 8 patients have been randomized. The 9th has just arrived, who is tentatively assigned to treatment A (a) and B (b). The figure shows the assignment as well as the covariate values of the patients. First, we point out that discretization could cause complications since different ways of splitting may lead to different assignments. For example, if the covariate range $[0, 1]$ in Figure 1 is split into m intervals of equal lengths, then the choices of $m = 1$, $m = 2$, and $m = 4$ result in different preferences of the assignment: if $m = 1$, that is, only overall patient numbers are considered, then the 9th patient could be assigned to treatment A or treatment B, since in either way the absolute difference is 1; if $m = 2$, then treatment A would be favored, producing balanced patient numbers in the second category, that is, over the interval $[0.5, 1.0]$ where the 9th patient belongs; if $m = 4$, then treatment B would be preferred instead, since it produces allocation of $2 : 1$ in the third category, that is, over the interval $[0.5, 0.75)$, rather than $3 : 0$ when treatment A is assigned. Second, we emphasize that balance in the mean values does not necessarily lead to balance in distributions. Taking the upper panel in Figure 1, for instance, the assignment of treatment A would not cause a significant mean difference in the two treatment groups. However, in terms of covariate distribution, a severe imbalance would exist, since the 5 patients (7th, 6th, 9th, 4th, and 8th) in group A have covariate values at the center of the range and the 4 patients (2nd, 3rd, 5th, and 1st) in group B at the two ends.

In light of the above discussion, we propose a new imbalance measure, which is defined as the maximum (absolute) assignment difference when all possible ways of

discretization are considered. By doing so, we want to make sure that the assignment difference over any interval does not become too extreme. Thus, if the 9th patient is tentatively assigned to A, we need to examine the differences of patient numbers in the two groups in all neighborhoods of Z_9 and use the maximum of them as the resulting degree of imbalance. The corresponding imbalance caused by the assignment of B can be calculated in the same manner. Then with a probability greater than $1/2$, the 9th patient should be assigned to a treatment that would cause a smaller maximum difference.

3. Procedure

Consider a sequential trial with two treatments A and B (control and test). For the first n patients that have arrived, let $\mathbf{T}_n = (T_1, \ldots, T_n)$ be the allocation sequence with $T_i = A$ if the ith patient is assigned to treatment A and $T_i = B$ otherwise. Suppose we need to balance treatment allocation over one single continuous covariate Z with cumulative distribution function F, whose range U is an interval on the real line. The two endpoints of U can be finite or infinite. Let $\mathbf{Z}_n = (Z_1, \ldots, Z_n)$, where Z_i is the covariate value of the ith patient. Suppose $Z_1, \ldots, Z_n$ are independent. Define

$$\Delta(I, \mathbf{Z}_n, \mathbf{T}_n) = N_A(I, \mathbf{Z}_n, \mathbf{T}_n) - N_B(I, \mathbf{Z}_n, \mathbf{T}_n) \tag{3.1}$$

as the difference of patient numbers in treatment groups A and B over the interval $I \subset U$, given the covariate information $\mathbf{Z}_n$ and allocation sequence $\mathbf{T}_n$. Suppose the first n patients have been randomized and the $(n+1)$th patient has just arrived, that is, $\mathbf{Z}_{n+1} = (Z_1, \ldots, Z_{n+1})$ and $\mathbf{T}_n = (T_1, \ldots, T_n)$ are known and T_{n+1} is to be determined. Define

$$D_{n+1}^{(k)} = \sup_I \{|\Delta(I, \mathbf{Z}_{n+1}, \mathbf{T}_{n+1})| : Z_{n+1} \in I \subset U, \ T_{n+1} = k\}, \tag{3.2}$$

where $k = A$ or B. Thus, $D_{n+1}^{(k)}$ is the potential maximum absolute difference it would cause if the new patient was assigned to treatment k. Note that the interval I in (3.2) needs to contain the new covariate value Z_{n+1}, because the difference over any other interval will not be affected by the the arrival of Z_{n+1} and is therefore not of our interest. Note also that $D_{n+1}^{(k)}$ is a function of $(\mathbf{Z}_{n+1}, \mathbf{T}_n, k)$.

Then, assign the $(n+1)$th patient to treatment A with the following probability:

$$P(T_{n+1} = A \mid \mathbf{Z}_{n+1}, \mathbf{T}_n) = \begin{cases} p, & \text{if } D_{n+1}^{(A)} < D_{n+1}^{(B)}, \\ q, & \text{if } D_{n+1}^{(A)} > D_{n+1}^{(B)}, \\ 0.5, & \text{otherwise}, \end{cases} \tag{3.3}$$

where $0 < q < p < 1$ and $p + q = 1$.

We will show how the 9th patient is randomized according to our new procedure. The critical part lies in the calculation of $D_9^{(A)}$ and $D_9^{(B)}$. For the former, that is, the 9th patient is temporarily assigned to treatment A (see the upper panel in Figure 1), then the maximum absolute difference is 5, attained over intervals which exclusively contain Z_7, Z_6, Z_9, Z_4, and Z_8, that is, 5 patients over these intervals are assigned to A and 0 is to B, so $D_9^{(A)} = 5$. Similarly,

$D_9^{(B)} = 3$. Therefore, since $D_9^{(A)} > D_9^{(B)}$, with probability $p > 1/2$ the 9th patient will be assigned to treatment B.

We would like to point out that in order to calculate the imbalance $D_{n+1}^{(k)}$, which is defined as a supremum, it is sufficient to examine intervals whose endpoints belong to the set $\{Z_1, \ldots, Z_{n+1}\}$. Hence the total number of such intervals has the order $O(n^2)$. Moreover, the difference of patient numbers over any of these intervals is only related to the ranks of $\{Z_1, \ldots, Z_{n+1}\}$, whose joint distribution places an equal probability $1/[(n+1)!]$ on any permutation of $[1, 2, \ldots, (n+1)]$. To support the above argument, we can reexamine the upper panel in Figure 1: for any interval $I = [a, b]$, where $Z_2 < a \le Z_7$ and $Z_5 \le b \le Z_1$, the difference of patient numbers over I is exactly the same as that over interval $[Z_7, Z_5]$, which is $+3 = 5 - 2$; furthermore, so long as the relative positions of $\{Z_1, \ldots, Z_{n+1}\}$ remain the same, this difference of $+3$ does not change. Nor does $D_{n+1}^{(k)}$. Therefore, we come to the conclusion that the new procedure is free of the underlying distribution F.

In fact, the computation time of $D_{n+1}^{(k)}$ can be reduced by examining an even smaller number of intervals, that is, $(n+2)$ instead of $O(n^2)$. Before demonstrating this, we need a few more notations and definitions. Since our new procedure is distribution-free, simply assume that the covariate Z is from uniform $[0, 1]$. Suppose $\mathbf{Z}_{n+1}$ and $\mathbf{T}_n$ have been observed. $\Delta(I, \mathbf{Z}_n, \mathbf{T}_n)$ in (3.1), defined as difference of patient numbers in groups A and B over interval I, will simply be written as $\Delta_n I$. For the ease of representation, let $Z_L = 0$ and $Z_R = 1$. Define two sets $\mathcal{S}_L$ and $\mathcal{S}_R$ as

$$
\begin{aligned}
\mathcal{S}_L &= \{\Delta_n[Z_i, Z_{n+1}) : Z_i \le Z_{n+1}, i = L, 1, \ldots, n\}, \\
\mathcal{S}_R &= \{\Delta_n(Z_{n+1}, Z_j] : Z_{n+1} \le Z_j, j = R, 1, \ldots, n\}.
\end{aligned}
\tag{3.4}
$$

That is, Z_i is any point from the set $\{0, Z_1, \ldots, Z_n\}$, that is, to the left of or equal to Z_{n+1}, and $[Z_i, Z_{n+1})$ is a left-closed and right-open interval. For instance, in Figure 1, $n = 8$ and $\mathcal{S}_L = \{0, Z_2, Z_7, Z_6\}$ from left to right. The interpretation of Z_j and $(Z_{n+1}, Z_j]$ is similar.

Proposition 3.1. *Let* $C_{L_1} = \min \mathcal{S}_L$, $C_{L_2} = \max \mathcal{S}_L$, $C_{R_1} = \min \mathcal{S}_R$, *and* $C_{R_2} = \max \mathcal{S}_R$. *Then,*

$$
\begin{aligned}
D_{n+1}^{(A)} &= \max\{|C_{L_1} + C_{R_1} + 1|, |C_{L_1} + C_{R_2} + 1|, |C_{L_2} + C_{R_1} + 1|, |C_{L_2} + C_{R_2} + 1|\}, \\
D_{n+1}^{(B)} &= \max\{|C_{L_1} + C_{R_1} - 1|, |C_{L_1} + C_{R_2} - 1|, |C_{L_2} + C_{R_1} - 1|, |C_{L_2} + C_{R_2} - 1|\}.
\end{aligned}
\tag{3.5}
$$

The proof of Proposition 3.1 is given in the Appendix. Proposition 3.1, together with the definitions of $\mathcal{S}_L$ and $\mathcal{S}_R$, suggests the following.

(1) The calculation of $D_{n+1}^{(A)}$ and $D_{n+1}^{(B)}$ only requires the examination of $(n+2)$ intervals, instead of $O(n^2)$.

(2) For two consecutive intervals $[Z_i, Z_{n+1})$ and $[Z_{i'}, Z_{n+1})$ in $\mathcal{S}_L$, where $Z_i < Z_{i'}$, $i \neq L$ and $i' \neq L$, we have $\Delta_n[Z_i, Z_{n+1}) = \Delta_n[Z_{i'}, Z_{n+1}) \pm 1$, depending on the assignment A or B for the patient at Z_i (the same argument applies to intervals in $\mathcal{S}_R$).

The above two observations form the basis of the algorithm, which was developed for the new procedure in the simulation studies (Section 4). We found that the computation time of the new procedure was even less than discretization methods.

4. Simulation Studies

Suppose $N = 60$ patients enrolled. As mentioned in Section 1, for continuous covariates, it is desirable to keep similarity between treatment groups in two aspects: the group sizes and the distributions of the covariates. Therefore, the new procedure was first compared with several other procedures in terms of the following two criteria: (1) the mean absolute difference $E|N_A(N) - N_B(N)|$ of all patient numbers in the two groups, shown as ED_{all}; (2) the mean Kolmogorov-Smirnov distance (K-S) between the empirical distributions of covariate Z in groups A and B, shown as ED_{ks}, which basically measures the similarity between two distributions. In addition, we used a new criterion: (3) the "maximum imbalance" defined by us as:

$$\sup_I \{|\Delta(I, \mathbf{Z}_N, \mathbf{T}_N)| : I \subset [0,1]\}, \tag{4.1}$$

shown as ED_{max}, which is the maximum absolute difference over all possible intervals after all patients have been assigned to a treatment. We will show that criterion (3) acts as a compromise between criterion (1) and criterion (2).

Since the procedures we compared are all distribution-free, the independent covariate values $Z_1, \ldots, Z_N$ were simply generated from Unif(0,1). All procedures use the strategy of minimization, but each has a different imbalance measure. More specifically, under a certain imbalance measure D, we calculate $D^{(A)}$ or $D^{(B)}$, defined as the imbalance that would occur if the new patient was assigned to treatment A or B. Depending on whether $D^{(A)} - D^{(B)}$ is positive, negative, or zero, the allocation probability toward the treatment A is q, p, or $1/2$, where $0 < q < p < 1$ and $p + q = 1$. In the simulation, we used $p = 2/3$ and $p = 1$, with the latter corresponding to deterministic allocation unless there is a tie.

The following procedures were studied.

(1) Efron's design (EFRON) [22]. Let N_A and N_B be the patient numbers in the two groups. Define the imbalance measure D as $|N_A - N_B|$. This method solely focuses on the balance of patient numbers.

(2) Kolmogorov-Smirnov measure (K-S). Let $\widehat{F}_k$ be the empirical distribution function of the covariate values in group k, $k = A, B$. Define imbalance measure D as the Kolmogorov-Smirnov distance between $\widehat{F}_A$ and $\widehat{F}_B$. This method solely focuses on the balance of distributions.

The above two methods are rarely used as a way of balancing over a continuous covariate, since each of them is designed to meet only one criterion. In our simulations, they served as two controls to evaluate other procedures.

(3) Discretization (DSCRT). In practice, in order to discretize a continuous covariate Z with cumulative distribution function F, the range is often split by the quantiles of F at probabilities $1/m$, $2/m$, $\ldots$, $(m-1)/m$. This is equivalent to splitting $[0,1]$ into m intervals of equal length $1/m$ for $Z \sim \text{unif}[0,1]$.

In our simulations, we tried $m = 2, 4, 8$. Within each category Efron's design was applied.

(4) The new procedure (MAX-IMB).

Table 1: Comparison of ED_{all}, ED_{ks}, and ED_{max} for $p = 2/3$.

	ED_{all} (s.e.)	ED_{ks} (s.e.)	ED_{max} (s.e.)
EFRON	1.28 (.023)	.212 (.0009)	9.03 (.031)
K-S	6.65 (.072)	.137 (.0005)	10.22 (.058)
DSCRT			
$m = 2$	2.17 (.029)	.178 (.0007)	8.52 (.030)
$m = 4$	2.94 (.036)	.161 (.0006)	8.18 (.030)
$m = 8$	3.76 (.042)	.159 (.0007)	8.40 (.034)
MAX-IMB	2.36 (.029)	.159 (.0006)	7.38 (.025)
RANK-SUM	3.25 (.037)	.205 (.0009)	9.82 (.039)
WGT-AVE			
$c_0 = 2$	1.67 (.028)	.183 (.0008)	8.48 (.030)
$c_0 = 4$	2.40 (.036)	.172 (.0008)	8.47 (.031)
$c_0 = 6$	3.06 (.041)	.167 (.0007)	8.62 (.034)

(5) Stigsby and Taves' rank sum (RANK-SUM) [16]. Let $(R_1, \ldots, R_N)$ be the ranks of $(Z_1, \ldots, Z_N)$. Suppose patients $i_1, \ldots, i_{N_1}$ are in group A and patients $j_1, \ldots, j_{N_2}$ are in group B. The imbalance measure D is defined by $|\sum_{k=1}^{N_1} R_{i_k} - \sum_{k=1}^{N_2} R_{j_k}|$.

(6) Su's weighted average (WGT-AVE) [6]. Let Q_1^k, Q_2^k, Q_3^k be the quartiles of the covariate values in group k, $k = A, B$. The "qualitative" imbalance measure D_L is defined by

$$w_0 I\{|N_A - N_B| > c_0\} + w_1 I\left\{\max_{i=1,2,3} \frac{|Q_i^A - Q_i^B|}{\max\{Q_i^A, Q_i^B\}} > c_1\right\}, \tag{4.2}$$

where w_0 and w_1 are two weights placed on the two items, c_0 and c_1 are two upper limits, and I is the indicator function; the "quantitative" imbalance measure D_T is defined by

$$\frac{w_0 |N_A - N_B|}{c_0} + \frac{w_1 \left[\max_{i=1,2,3}\left(|Q_i^A - Q_i^B| / \max\{Q_i^A, Q_i^B\}\right)\right]}{c_1}. \tag{4.3}$$

Therefore, let $D_L^{(k)}$ be the the qualitative imbalance resulted from the tentative assignment of treatment k to the new patient, $k = A, B$. If $D_L^{(A)} - D_L^{(B)}$ is positive (negative), the allocation probability toward the treatment A will be $q(p)$; if $D_L^{(A)} - D_L^{(B)} = 0$, use measure D_T to determine the probability p, q, or $1/2$.

w_0, w_1, c_0, and c_1 can be changed freely from a subjective point of view. In the simulations, we fixed $w_0 = w_1 = 1$ and $c_1 = 10\%$, but tried $c_0 = 2, 4, 6$.

The results for $p = 2/3$ and $p = 1$ under 5000 repetitions are shown in Tables 1 and 2.

We first focus on Table 1 ($p = 2/3$). The 1st and 2nd columns suggest that the "best" ED_{all} and the "best" ED_{ks} that can be achieved are 1.28 by EFRON and 0.137 by K-S, respectively, at the expense of large imbalance under the other criterion. For DSCRT when m increases from 2 to 4 and 8, ED_{all} increases from 2.17 to 2.94 and 3.76, whereas ED_{ks} decreases from 0.178 to 0.161 and 0.159. Therefore, we see that there is a trade-off between the balance of group sizes and the balance of covariate distributions. Similar trend can be

Table 2: Comparison of ED_{all}, ED_{ks}, and ED_{max} for $p = 1$.

	ED_{all} (s.e.)	ED_{ks} (s.e.)	ED_{max} (s.e.)
EFRON	0.00 (.0000)	.209 (.0010)	8.78 (.030)
K-S	13.14 (.1396)	.096 (.0002)	15.08 (.120)
DSCRT			
$m = 2$	0.49 (.0121)	.171 (.0007)	8.03 (.027)
$m = 4$	0.93 (.0150)	.140 (.0005)	6.93 (.021)
$m = 8$	1.45 (.0190)	.119 (.0004)	6.16 (.018)
MAX-IMB	1.19 (.0170)	.108 (.0003)	4.90 (.010)
RANK-SUM	2.82 (.0317)	.196 (.0008)	9.24 (.035)
WGT-AVE			
$c_0 = 2$	0.52 (.0183)	.136 (.0005)	6.82 (.023)
$c_0 = 4$	1.59 (.0414)	.132 (.0004)	7.16 (.034)
$c_0 = 6$	2.56 (.0541)	.132 (.0004)	7.63 (.044)

observed for WGT-AVE when c_0 increases from 2 to 4 and 6. In terms of these two criteria, the new procedure (MAX-IMB), with ED_{all} 2.36 and ED_{ks} 0.159, has better performance than DSCRT with $m = 4$ and $m = 8$ and WGT-AVE with $c_0 = 4$ and $c_0 = 6$; it also has lower ED_{all} and ED_{ks} than RANK-SUM.

In terms of ED_{max} (the 3rd column), MAX-IMB has the minimum value 7.38 since it sequentially minimizes this criterion. On the contrary, DSCRT minimizes the imbalance of patient numbers over the selected intervals, but ignores the imbalance over others. As a result, on average, the maximum imbalance under DSCRT is higher than that under MAX-IMB. In a sense, ED_{max} serves as a tool which detects any allocation imbalance that is ignored by DSCRT. Since the new procedure examines both "global" imbalance, that is, over the whole range, and "local" imbalance, that is, over any small interval, it can be regarded as a compromise between achieving balance in overall group sizes and achieving balance in covariate distributions.

Similar conclusion can be drawn for $p = 1$ (see Table 2), that is, the allocation is deterministic except the case of a tie. From $p = 2/3$ to $p = 1$, the decrease in ED_{all} is most significant under DSCRT, from $(2.17, 2.94, 3.76)$ to $(0.49, 0.93, 1.45)$. This is because when $p = 1$ only covariate values are random, and so long as the numbers of patients over the selected intervals are even (e.g., 34 over $[0, 0.5]$ and 26 over $(0.5, 1]$ with $m = 2$), DSCRT can always achieve perfect overall balance. Moreover, even if the patient numbers are odd (e.g., 35 over $[0, 0.5]$ and 25 over $(0.5, 1]$), there are still chances that the allocation differences are $+1$ and -1 or in the reversed way, again resulting perfect overall balance. Other procedures are more complex and the decrease in ED_{all} is less significant. As a result, when $p = 1$, MAX-IMB is only uniformly better than DSCRT with $m = 8$, not $m = 4$.

We also compared the above procedures by other commonly used measures including the mean absolute difference of sample means (ED_{mean}) and the mean absolute difference of sample standard deviations (ED_{std}) of the covariate values in the two treatment groups. Furthermore, Lin and Su [23] introduced another criterion, the area between the empirical cumulative distribution functions of the covariate values in the two treatment groups (normalized by the difference of the maximum and the minimum values), denoted as ED_{area}, and pointed out that this criterion has better performance than Kolmogorov-Smirnov distance in capturing the difference in two distributions. We thus included this criterion in the simulation. Since the measurements of mean, standard deviation, and area under

Table 3: $Z \sim \mathrm{Unif}[0,1]$. Comparison of ED_{area}, ED_{mean}, and ED_{std} for $p = 2/3$.

	$ED_{\mathrm{mean}} \times 10^2$ (s.e.)	$ED_{\mathrm{std}} \times 10^2$ (s.e.)	$ED_{\mathrm{area}} \times 10^2$ (s.e.)
EFRON	5.96 (.063)	2.79 (.030)	8.30 (.051)
K-S	2.29 (.028)	1.77 (.020)	5.02 (.022)
DSCRT			
$m = 2$	3.60 (.039)	2.72 (.029)	6.33 (.030)
$m = 4$	3.22 (.036)	1.92 (.021)	5.56 (.027)
$m = 8$	3.80 (.042)	1.86 (.020)	5.75 (.033)
MAX-IMB	3.75 (.041)	2.10 (.023)	6.04 (.032)
RANK-SUM	5.85 (.063)	2.76 (.029)	8.27 (.050)
WGT-AVE			
$c_0 = 2$	4.26 (.050)	2.42 (.026)	6.78 (.040)
$c_0 = 4$	3.53 (.042)	2.34 (.025)	6.14 (.034)
$c_0 = 6$	3.26 (.040)	2.25 (.025)	5.88 (.032)

Table 4: $Z \sim \mathrm{Norm}[0,1]$. Comparison of ED_{area}, ED_{mean}, and ED_{std} for $p = 2/3$.

	$ED_{\mathrm{mean}} \times 10^2$ (s.e.)	$ED_{\mathrm{std}} \times 10^2$ (s.e.)	$ED_{\mathrm{area}} \times 10^2$ (s.e.)
EFRON	20.66 (.220)	14.79 (.158)	6.78 (.035)
K-S	9.47 (.112)	11.45 (.124)	4.61 (.018)
DSCRT			
$m = 2$	14.08 (.152)	14.64 (.155)	5.62 (.024)
$m = 4$	12.32 (.133)	12.39 (.135)	5.00 (.020)
$m = 8$	13.44 (.149)	11.60 (.126)	5.00 (.023)
MAX-IMB	14.17 (.154)	12.61 (.137)	5.28 (.023)
RANK-SUM	20.24 (.218)	14.67 (.157)	6.77 (.035)
WGT-AVE			
$c_0 = 2$	15.46 (.176)	13.75 (.147)	5.80 (.029)
$c_0 = 4$	13.21 (.152)	13.53 (.143)	5.39 (.025)
$c_0 = 6$	12.59 (.147)	13.36 (.143)	5.22 (.023)

a distribution function depend on the underlying distribution of the covariate, we did simulation studies under a uniform distribution $Z \sim \mathrm{Unif}[0,1]$ and under a normal distribution $Z \sim N(0,1)$ and show the results in Tables 3 and 4, respectively.

From Table 3 under the uniform distribution, it is seen that K-S has the best performance, since its ED_{mean}, ED_{std}, and ED_{area} (2.29, 1.77, and 5.02) are the lowest among all procedures. This is expected, since K-S solely minimizes the distance between the two distributions. Once the distributions are closest, so are the summary statistics of means and standard deviations as well as the area between the distributions. However, K-S is likely to produce severe imbalance of group sizes, as shown in Tables 1 and 2. EFRON has the worst performance under the three criteria since it completely ignores the covariate distributions.

Among the three choices of m under DSCRT, roughly speaking $m = 4$ performs best: its ED_{mean} and ED_{area} (3.22 and 5.56) are the lowest and its ED_{std} (1.92) slightly higher than that under $m = 8$. Moreover, under these three criteria, DSCRT with $m = 4$ is uniformly better than MAX-IMB, RANK-SUM, and WGT-AVE. However, the good performance of DSCRT with $m = 4$ is based on the correct identification of quartiles of the true covariate distribution,

which may not be feasible before the collection of data. In contrast, other methods do not require such information.

Comparing MAX-IMB, RANK-SUM, and WGT-AVE, RANK-SUM has the highest values under the three criteria; MAX-IMB has comparable performance to WGT-AGE with $c_0 = 4$, with the former having slightly higher ED_{mean} and the latter slightly higher ED_{std} and ED_{area}. Similar conclusion can be reached for the normal distribution (see Table 4). The result for $p = 1$ under the three criteria ED_{mean}, ED_{std}, and ED_{area} resembles that for $p = 2/3$, the only difference being that the best choice of m under DSCRT is $m = 8$ instead of $m = 4$. In fact, we also did simulations under different sample sizes ($N = 30$ and 150) and the results are quite consistent.

5. Discussion and Conclusions

In this paper, we propose a new minimization procedure that balances treatment allocation over continuous covariates. For any new patient, it examines the imbalances in the neighborhoods of his or her covariate value and bias the allocation probability towards the treatment that would result in a smaller value of the maximum imbalance. The new method only depends on the ranks of the covariates and is therefore distribution-free. Our simulation studies have shown that it is able to maintain relatively good balance in terms of group sizes and covariate distributions across treatment groups.

In addition, the new procedure does not require the specification of any critical values, which is usually needed for discretization methods in order to define categories. For the latter methods, if quantiles of the covariate distribution F are used for the critical values, then lack of knowledge about F may lead to wrong guesses of the quantiles. The new procedure saves this step by considering all possible divisions of the range. Nevertheless, only the assignment differences over $(n+2)$ intervals have to be examined to calculate the new imbalance measure, and the corresponding algorithm is computationally efficient.

Borrowing the idea of Pocock and Simon's design [7], our method can easily be generalized to two or more continuous covariates or a mix of discrete and continuous covariates. Suppose that for a total of L covariates $Z^1, \ldots, Z^L$, the first L_1 is continuous and the rest are discrete. When the $(n + 1)$th patient is enrolled, for any continuous covariate Z^i, $i = 1, \ldots, L_1$, we define $D_{n+1,i}^{(k)}$, $k = A, B$ by (3.2), which is the the maximum imbalance measure with respect to the ith covariate; for any discrete covariate Z^j, $j = L_1 + 1, \ldots, L$, observe the category the new patient belongs to, tentatively assign him to treatment k, and define $D_{n+1,j}^{(k)}$ as the absolute difference of patient numbers in the two treatment groups with respect to that specific category. For example, if the jth covariate is gender and the new patient is a male, then $D_{n+1,j}^{(k)}$ is calculated among all males. Define

$$D_{n+1,\mathrm{mix}}^{(k)} = \sum_{i=1}^{L_1} w_i D_{n+1,i}^{(k)} + \sum_{j=L_1+1}^{L} w_j D_{n+1,j}^{(k)}, \tag{5.1}$$

where w_i's and w_j's are the weights placed on the covariates and can be assigned by the importance of the different covariates. Depending on whether $D_{n+1,\mathrm{mix}}^{(A)}$ is greater than, less than, or equal to $D_{n+1,\mathrm{mix}}^{(B)}$, assign the $(n + 1)$th patient to treatment A with probability q, p or $1/2$. From (5.1), it is seen that $D_{n+1,\mathrm{mix}}^{(k)}$ is similar to Pocock and Simon's weighted average of

marginal imbalances. The only difference is that for those continuous covariates we redefine the marginal imbalances by the new measure proposed in the current paper, so that the negative effect caused by discretization can be mitigated. Since the marginal imbalances for discrete covariates in $D^{(k)}_{n+1,\text{mix}}$ remain the same as in Pocock and Simon's, it is expected that the good balance properties for discrete covariates in their design can be preserved when $D^{(k)}_{n+1,\text{mix}}$ is used in the minimization. We did simulations for the case of two continuous covariates and the new procedure again showed improvement over other procedures.

In the case that an unequal allocation ratio such as $r_A : r_B$ is desired, one can easily generalize the proposed metric by redefining $\Delta(I, \mathbf{Z}_n, \mathbf{T}_n)$ in (3.1) as

$$\frac{N_A(I, \mathbf{Z}_n, \mathbf{T}_n)}{r_A} - \frac{N_B(I, \mathbf{Z}_n, \mathbf{T}_n)}{r_B}. \tag{5.2}$$

By doing so, it can be ensured that the allocation ratio over any interval is close to $r_A : r_B$. In practice, one can also modify the maximum imbalance measure by adding weights to different intervals. The weight of each interval can be assigned as a function of the number of patients within the interval, so that the procedure remains distribution-free. But the algorithm to implement such a procedure will be more complicated. We will leave these as future research topics.

Appendix

Proof of Proposition 3.1

We will use the basic fact that for a sequence $\{a_1, \ldots, a_n\}$ and a constant b,

$$\max_{i=1,\ldots,n} |a_i + b| = \max \left[\left| \max_i a_i + b \right|, \left| \min_i a_i + b \right| \right]. \tag{A.1}$$

Following the notations and argument in Section 3,

$$D^{(A)}_{n+1} = \max_{\substack{i=L,1,\ldots,n, \\ j=R,1,\ldots,n}} \left\{ \left| \Delta\left([Z_i, Z_j], \mathbf{Z}_{n+1}, \mathbf{T}_{n+1}\right) \right| : Z_i \leq Z_{n+1} \leq Z_j, T_{n+1} = A \right\}. \tag{A.2}$$

But $[Z_i, Z_j] = [Z_i, Z_{n+1}) \cup \{Z_{n+1}\} \cup (Z_{n+1}, Z_j]$ and $\Delta(\{Z_{n+1}\}, \mathbf{Z}_{n+1}, \mathbf{T}_{n+1}) = 1$ if $T_{n+1} = A$. Thus,

$$D^{(A)}_{n+1} = \max_{i=L,1,\ldots,n} \max_{j=R,1,\ldots,n} \left\{ \left| \Delta_n[Z_i, Z_{n+1}) + 1 + \Delta_n(Z_{n+1}, Z_j] \right| : Z_i \leq Z_{n+1} \leq Z_j \right\}. \tag{A.3}$$

Note that by definition of the notation $\Delta_n I$, $\Delta_n[Z_i, Z_{n+1}) =: \Delta([Z_i, Z_{n+1}), \mathbf{Z}_n, \mathbf{T}_n)$, which further equals $\Delta([Z_i, Z_{n+1}), \mathbf{Z}_{n+1}, \mathbf{T}_{n+1})$ since the interval $[Z_i, Z_{n+1})$ does not contain Z_{n+1}.

Now for any fixed i, apply (A.1) to the constant $\Delta_n[Z_i, Z_{n+1}) + 1$ and the sequence $\Delta_n(Z_{n+1}, Z_j]$, $j = R, 1, \ldots, n$, and we have

$$
\begin{aligned}
D_{n+1}^{(A)} &= \max_{i=L,1,\ldots,n} \{\max[|C_{R_2} + \Delta_n[Z_i, Z_{n+1}) + 1|, |C_{R_1} + \Delta_n[Z_i, Z_{n+1}) + 1|] : Z_i \leq Z_{n+1}\} \\
&= \max\left\{ \max_{i=L,1,\ldots,n} \{|C_{R_1} + 1 + \Delta_n[Z_i, Z_{n+1})| : Z_i \leq Z_{n+1}\}, \right. \\
&\qquad\qquad \left. \max_{i=L,1,\ldots,n} \{|C_{R_2} + 1 + \Delta_n[Z_i, Z_{n+1})| : Z_i \leq Z_{n+1}\} \right\} \\
&=: \max\{I, II\}.
\end{aligned}
\tag{A.4}
$$

Applying (A.1) again to I with constant $C_{R_1} + 1$ and sequence $\Delta_n[Z_i, Z_{n+1})$, we have

$$
I = \max[|C_{L_1} + (C_{R_1} + 1)|, |C_{L_2} + (C_{R_1} + 1)|].
\tag{A.5}
$$

Similarly,

$$
II = \max[|C_{L_1} + (C_{R_2} + 1)|, |C_{L_2} + (C_{R_2} + 1)|].
\tag{A.6}
$$

Therefore,

$$
D_{n+1}^{(A)} = \max\{|C_{L_1} + C_{R_1} + 1|, |C_{L_2} + C_{R_1} + 1|, |C_{L_1} + C_{R_2} + 1|, |C_{L_2} + C_{R_2} + 1|\}.
\tag{A.7}
$$

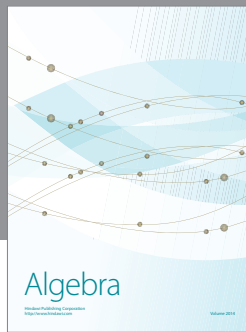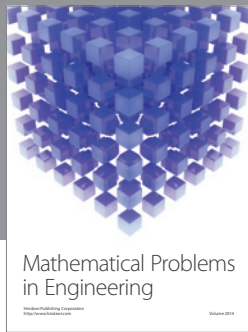The derivation of $D_{n+1}^{(B)}$ is similar.

Acknowledgments

References

[1] D. McEntegart, "The pursuit of balance using stratified and dynamic randomization techniques: an overview," *Drug Information Journal*, vol. 37, pp. 293–308, 2003.

[2] J. Ciolino, W. Zhao, R. Martin, and Y. Palesch, "Quantifying the cost in power of ignoring continuous covariate imbalances in clinical trial randomization," *Contemporary Clinical*, vol. 32, no. 2, pp. 250–259, 2011.

[3] G. Kundt, "Comparative evaluation of balancing properties of stratified randomization procedures," *Methods of Information in Medicine*, vol. 48, pp. 129–134, 2009.

[4] A. Hagino, C. Hamada, I. Yoshimura, Y. Ohashi, J. Sakamoto, and H. Nakazato, "Statistical comparison of random allocation methods in cancer clinical trials," *Controlled Clinical Trials*, vol. 25, no. 6, pp. 572–584, 2004.

[5] A. Endo, F. Nagatani, C. Hamada, and I. Yoshimura, "Minimization method for balancing continuous prognostic variables between treatment and control groups using Kullback-Leibler divergence," *Contemporary Clinical Trials*, vol. 27, no. 5, pp. 420–431, 2006.

[6] Z. Su, "Balancing multiple baseline characteristics in randomized clinical trials," *Contemporary Clinical Trials*, vol. 32, no. 4, pp. 547–550, 2011.

[7] S. J. Pocock and R. Simon, "Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial," *Biometrics*, vol. 31, no. 1, pp. 103–115, 1975.

[8] L. J. Wei, "An application of an urn model to the design of sequential controlled clinical trials," *Journal of the American Statistical Association*, vol. 73, no. 363, pp. 559–563, 1978.

[9] M. Zelen, "The randomization and stratification of patients to clinical trials," *Journal of Chronic Diseases*, vol. 27, pp. 365–375, 1974.

[10] D. F. Signorini, O. Leung, R. J. Simes, E. Beller, V. J. Gebski, and T. Callaghan, "Dynamic balanced randomization for clinical trials," *Statistics in Medicine*, vol. 12, no. 24, pp. 2343–2350, 1993.

[11] S. Heritier, V. Gebski, and A. Pillai, "Dynamic balancing randomization in controlled clinical trials," *Statistics in Medicine*, vol. 24, no. 24, pp. 3729–3741, 2005.

[12] Y. Hu and F. Hu, "Asymptotic properties of covariate-adaptive randomization," *The Annals of Statistics*. In press.

[13] C. J. Weir and K. R. Lees, "Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial," *Statistics in Medicine*, vol. 22, no. 5, pp. 705–726, 2003.

[14] F. K. Hoehler, "Balancing allocation of subjects in biomedical research: a minimization strategy based on ranks," *Computers and Biomedical Research*, vol. 20, no. 3, pp. 209–213, 1987.

[15] J. W. Frane, "A method of biased coin randomization, its implementation, and its validation," *Drug Information Journal*, vol. 32, no. 2, pp. 423–432, 1998.

[16] B. Stigsby and D. R. Taves, "Rank-Minimization for balanced assignment of subjects in clinical trials," *Contemporary Clinical Trials*, vol. 31, no. 2, pp. 147–150, 2010.

[17] C. B. Begg and B. Iglewicz, "A treatment allocation procedure for sequential clinical trials," *Biometrics*, vol. 36, no. 1, pp. 81–90, 1980.

[18] A. C. Atkinson, "Optimum biased coin designs for sequential clinical trials with prognostic factors," *Biometrika*, vol. 69, no. 1, pp. 61–67, 1982.

[19] R. L. Smith, "Properties of biased coin designs in sequential clinical trials," *The Annals of Statistics*, vol. 12, no. 3, pp. 1018–1034, 1984.

[20] R. L. Smith, "Sequential treatment allocation using biased coin designs," *Journal of the Royal Statistical Society. Series B*, vol. 46, no. 3, pp. 519–543, 1984.

[21] W. F. Rosenberger and O. Sverdlov, "Handling covariates in the design of clinical trials," *Statistical Science*, vol. 23, no. 3, pp. 404–419, 2008.

[22] B. Efron, "Forcing a sequential experiment to be balanced," *Biometrika*, vol. 58, pp. 403–417, 1971.

[23] Y. Lin and Z. Su, "Balancing continuous and categorical baseline covariates in sequential clinical trials using the area between empirical cumulative distribution functions," *Statistics in Medicine*, vol. 31, no. 18, pp. 1961–1971, 2012.