The Annals of Statistics
2012, Vol. 40, No. 3, 1794–1815
DOI: 10.1214/12-AOS983

ASYMPTOTIC PROPERTIES OF COVARIATE-ADAPTIVE RANDOMIZATION

BY YANQING HU AND FEIFANG HU[1]

University of Virgina

Balancing treatment allocation for influential covariates is critical in clinical trials. This has become increasingly important as more and more biomarkers are found to be associated with different diseases in translational research (genomics, proteomics and metabolomics). Stratified permuted block randomization and minimization methods [Pocock and Simon *Biometrics* **31** (1975) 103–115, etc.] are the two most popular approaches in practice. However, stratified permuted block randomization fails to achieve good overall balance when the number of strata is large, whereas traditional minimization methods also suffer from the potential drawback of large within-stratum imbalances. Moreover, the theoretical bases of minimization methods remain largely elusive. In this paper, we propose a new covariate-adaptive design that is able to control various types of imbalances. We show that the joint process of within-stratum imbalances is a positive recurrent Markov chain under certain conditions. Therefore, this new procedure yields more balanced allocation. The advantages of the proposed procedure are also demonstrated by extensive simulation studies. Our work provides a theoretical tool for future research in this area.

1. Introduction. In clinical trials, covariates are factors that have a large impact on the responses of the patients. Typical covariates include gender, age, disease stage, different research center, etc. At the design stage it is often important to balance treatment allocation over covariates, as a well-balanced trial can lead to more efficient statistical comparison and more convincing results to the general audience [Kundt (2009)]. Balanced allocation is also particularly useful when the sample size is small or when interim analysis or subgroup analysis is desired [Toorawa et al. (2009)].

Stratified randomization is a popular way of achieving balance. It defines strata as different combinations of the covariates' levels and employs permuted block randomization within each stratum. This method is easy to implement and achieves good balance when the number of strata is small [Kalish and Begg (1985)]. However, the permuted block design is susceptible to selection bias [Matts and Lachin (1988)]. Moreover, it tends to cause severe allocation imbalance in the whole trial

Received May 2011; revised February 2012.
[1]Supported by NSF Grants DMS-09-07297 and DMS-09-06661.
MSC2010 subject classifications. Primary 60F15, 62G10; secondary 60F05, 60F10.
Key words and phrases. Balancing covariates, clinical trial, marginal balance, Markov chain, Pocock and Simon's design, stratified permuted block.

when there are too many strata, typically as a result of many covariates, or many levels within the individual covariates [Pocock (1982)]. Increasing numbers of strata, however, has become the trend, due to the need to conduct multicenter trials as well as the inclusion of newly identified biomarkers as covariates [Khan et al. (2010), Li et al. (2010), McIlroy et al. (2010), etc.].

Covariate-adaptive randomization (or minimization) has been proposed to address the above problem. The earliest work on minimization dates back to Taves (1974) and Pocock and Simon (1975). In particular, with I being the number of covariates and m_i the number of levels for covariate i, $i = 1, \ldots, I$, Pocock and Simon's (1975) procedure minimizes a weighted average of marginal imbalances $\sum_i w_i d_i(n)$, where $d_i(n)$ is a measure of imbalance among treatment groups with respect to the ith margin of the new patient. Simulation studies [Weir and Lees (2003), Toorawa et al. (2009), Kundt (2009)] found that this method reduces marginal imbalances as well as the overall imbalance. Wei (1978) generalized Taves's method by introducing a marginal urn model. Other works include Zelen (1974), Nordle and Brantmark (1977), Signorini et al. (1993) and Heritier, Gebski and Pillai (2005), which used a hierarchical decision rule and set priority among elements of strata, margins and overall trial. Despite the numerous works in the literature, "very little is known about the theoretical properties of covariate-adaptive designs" [Rosenberger and Sverdlov (2008)].

Model-based approach was introduced by Begg and Iglewicz (1980) and Atkinson (1982), and the theoretical work has been developed by Smith (1984a, 1984b). Smith considered the linear model $Ey_n = T_n\alpha + \sum_{j=1}^{p} z_{n,j}\beta_j$ with homogeneous errors and no interaction of any type, where y_n, T_n, $(z_{n,1}, \ldots, z_{n,p})$ are the response, assignment and covariate values of the nth patient, respectively, and $T_n = +1$ or -1 for treatment 1 or 2. Since α, the treatment effect, is the main interest of the trial, this method sequentially skews the allocation probability toward the treatment that would lead to a smaller variance of $\hat{\alpha}$ (the MLE of α). Under some appropriate allocation functions Smith derived the asymptotic normality of $\sum_{i=1}^{n} z_{i,j}T_i$ $(j = 1, \ldots, p)$. This asymptotic property was further applied to the construction of a conditional permutation test [Smith (1984b)].

Although the minimization approach [Pocock and Simon (1975), Wei (1978), etc.] and the model-based approach [Smith (1984a, 1984b), etc.] both lead to marginal and overall balance, they are rather different in nature. First, even if they use the same biased coin function, the two allocation rules are still not the same, unless in the trivial case of no covariates. Hence Smith's asymptotic result does not readily apply to Pocock and Simon's or Wei's procedure. Second, Smith's result depends on the homogeneous linear model. Therefore, once the data type has changed (such as binary or survival responses), model-based approach does not necessarily imply balance. Finally, minimization approach is more popular in practice [Taves (2010)]. In fact, as discussed by many authors such as Lagakos and Pocock (1984), Smith (1984b) and McEntegart (2003), balanced allocation enhances credibility of the trials for medical professions that are less statistically

sophisticated, and the simple comparisons of similar groups of patients are often more acceptable than a model-based approach adjusting for covariates.

In this paper we focus on the minimization approach that compares patient numbers at different levels. While the marginal procedures have good balance with respect to the margins and the whole trial, their performance within the individual strata is not as satisfactory [Signorini et al. (1993), Kundt (2009)]. Wei (1978) gave a short proof that if no interaction exists, marginal balances are sufficient to ensure unbiased estimation of treatment effect in an unadjusted analysis. In other words, when interactions do exist, ignorance of within-stratum imbalances may lead to biased estimation. Moreover, as the field of personalized medicine develops [Hu (2012)], subgroup analysis is often desired, and allocation balance within individual strata can improve the precision of such analysis.

To overcome the potential drawbacks of stratification and Pocock and Simon's (1975) method, we develop a new randomization procedure in this paper, which considers a weighted average of three types of imbalances (within-stratum, within-covariate-margin and overall). By adopting Efron's (1971) discrete allocation function, the next patient will be assigned with higher probability to a treatment that leads to a smaller value of the weighted average.

To study the theoretical properties of the new procedure, the main difficulties include the correlation structure of within-stratum imbalances as well as the discreteness of the allocation function. In the literature, a large number of adaptive designs adopt a continuous allocation function, and their properties are often obtained by a Taylor expansion of the allocation function, accompanied by a martingale approximation [Bai and Hu (1999), Hu and Zhang (2004), Zhang, Hu and Cheung (2006), etc.]. Since we use Efron's function, which is discrete at 0, the Taylor expansion is not feasible. We seek to take advantage of an alternative technique, namely "drift conditions," which was developed to study the stability of Markov chains on general state spaces. We show that the joint process of within-stratum imbalances under the new procedure is a *positive recurrent* Markov chain under some conditions, and thus preserves the order of $O_p(1)$ at the within-stratum level. Our simulations suggest that the within-stratum imbalances under Pocock and Simon's (1975) design have fast-increasing variances as sample size increases, implying a slower rate than $O_p(1)$.

In Section 2, the new procedure is described in general with I covariates. The theoretical results of the new procedure are given in Section 3. We further use simulations to study the different covariate-adaptive designs in Section 4 and conclude our paper with some observations in Section 5. The proofs of the theorems can be found in Section 6 and the supplemental article [Hu and Hu (2012)].

2. The new covariate-adaptive randomization procedure. This setting is similar to that of Pocock and Simon (1975), except that we only focus on two treatment groups, 1 and 2. Consider I covariates and m_i levels for the ith covariate, resulting in $m = \prod_{i=1}^{I} m_i$ strata. Let T_j be the assignment of the jth patient,

$j = 1, \ldots, n$, that is, $T_j = 1$ for treatment 1 and $T_j = 0$ for treatment 2. Let Z_j indicate the covariate profile of that patient, that is, $Z_j = (k_1, \ldots, k_I)$ if his or her ith covariate is at level k_i, $1 \le i \le I$ and $1 \le k_i \le m_i$. For convenience, we use $(k_1, \ldots, k_I)$ to denote the *stratum* formed by patients who possess the same covariate profile $(k_1, \ldots, k_I)$, and use $(i; k_i)$ to denote the *margin* formed by patients whose ith covariate is at level k_i.

The new procedure is defined as follows:

(1) The first patient is assigned to treatment 1 with probability 1/2.
(2) Suppose $(n-1)$ patients have been assigned to a treatment $(n > 1)$ and the nth patient falls within stratum $(k_1^*, \ldots, k_I^*)$.
(3) For the first $(n-1)$ patients:
   - let D_{n-1} be the difference between the numbers of patients in treatment group 1 and 2 as total, that is, the number in group 1 minus the number in group 2;
   - similarly, let $D_{n-1}(i; k_i^*)$ and $D_{n-1}(k_1^*, \ldots, k_I^*)$ be the differences between the numbers of patients in the two treatment groups on the margin $(i; k_i^*)$, and within the stratum $(k_1^*, \ldots, k_I^*)$, respectively;
   - these differences can be positive, negative or zero, and each one is used to measure the *imbalance* at the corresponding level (overall, marginal, or within-stratum).
(4) If the nth patient were assigned to treatment 1, then $D_n^{(1)} = D_{n-1} + 1$ would be the "potential" overall difference in the two groups; similarly,

$$D_n^{(1)}(i; k_i^*) = D_{n-1}(i; k_i^*) + 1$$

and

$$D_n^{(1)}(k_1^*, \ldots, k_I^*) = D_{n-1}(k_1^*, \ldots, k_I^*) + 1$$

would be the potential differences on margin $(i; k_i^*)$ and within stratum $(k_1^*, \ldots, k_I^*)$, respectively.
(5) Define an imbalance measure $Imb_n^{(1)}$ by

$$Imb_n^{(1)} = w_o\big[D_n^{(1)}\big]^2 + \sum_{i=1}^{I} w_{m,i}\big[D_n^{(1)}(i; k_i^*)\big]^2 + w_s\big[D_n^{(1)}(k_1^*, \ldots, k_I^*)\big]^2,$$

which is the weighted imbalance that would be caused if the nth patient were assigned to treatment 1. w_o, $w_{m,i}$ and w_s are nonnegative weights placed overall, within a covariate margin and within a stratum cell, respectively. Without loss of generality we can assume

$$w_o + w_s + \sum_{i=1}^{I} w_{m,i} = 1.$$

(6) In the same manner we can define $Imb_n^{(2)}$, the weighted imbalance that would be caused if the nth patient were assigned to treatment 2. In this case, the three types of potential differences are the existing ones minus 1, instead of plus 1.

(7) Conditional on the assignments of the first $(n-1)$ patients as well as the covariates' profiles of the first n patients, assign the nth patient to treatment 1 with probability

$$P(T_n = 1|\mathbf{Z}_n, \mathbf{T}_{n-1}) = \begin{cases} q, & \text{if } Imb_n^{(1)} > Imb_n^{(2)}, \\ p, & \text{if } Imb_n^{(1)} < Imb_n^{(2)}, \\ 0.5, & \text{otherwise,} \end{cases}$$

where $n > 1$, $0 < q < p < 1$, $p + q = 1$, $\mathbf{Z}_n = (Z_1, \ldots, Z_n)$ and $\mathbf{T}_{n-1} = (T_1, \ldots, T_{n-1})$.

REMARK 2.1. When $w_o = w_s = 0$, that is, only the marginal imbalances are considered, the proposed design reduces to a special case of Pocock and Simon's (1975) method; and when $w_{m,i} = w_o = 0$, it reduces to stratified randomization, where a separate biased coin is employed to determine the assignment within each stratum. However, we will explore procedures with other choices of weights, to see if they can lead to more balanced allocation from various perspectives.

REMARK 2.2. In the literature different views have been given as to the selection of the biasing probability p. Efron (1971) suggested $p = 2/3$, but his method does not consider covariates. The more recent papers, especially those involving covariate-adaptive randomization, suggested larger p's, such as 0.85, 0.90 and 0.95. See Weir and Lees (2003), Hagino et al. (2004), Toorawa et al. (2009), and Hu, Zhang and He (2009). One may also use other generators in step (7), for example, Wei's (1978) generator. The properties of the design will be different.

EXAMPLE 1. Suppose in a trial two covariates, gender and smoking behavior, are considered to be influential, each of which has two levels. Thus, the 4 strata $(1, 1)$, $(1, 2)$, $(2, 1)$, $(2, 2)$ represent male smokers, male nonsmokers, female smokers and female nonsmokers, respectively. Assume that the weights are $w_o = 1/3$, $w_{m,1} = w_{m,2} = 1/6$ and $w_s = 1/3$. The first 50 patients have been randomized and the 4 within-stratum differences among these 50 patients are -2, $+2$, $+1$ and -1. If the 51th patient is a *male smoker*, then the current imbalances are calculated as:

- overall: $D_{n-1} = -2 + 2 + 1 - 1 = 0$;
- margin of male: $D_{n-1}(1; 1) = -2 + 2 = 0$;
- margin of smokers: $D_{n-1}(2; 1) = -2 + 1 = -1$;
- stratum of male smokers: $D_{n-1}(1, 1) = -2$.

TABLE 1
An example illustrating the calculation under the new procedure

	$D_{50}(\cdot)$	$D_{51}^{(1)}(\cdot) = D_{50}(\cdot)+1$	$D_{51}^{(2)}(\cdot) = D_{50}(\cdot) - 1$
Overall	0	1	−1
Margin of male (1;1)	0	1	−1
Margin of smokers (2;1)	−1	0	−2
Stratum of male smoker (1,1)	−2	−1	−3

The potential imbalances if the new patient were assigned to treatment 1 or 2 are given in Table 1.

Therefore,

$$Imb_{51}^{(1)} = (1)^2 \cdot \tfrac{1}{3} + (1)^2 \cdot \tfrac{1}{6} + (0)^2 \cdot \tfrac{1}{6} + (-1)^2 \cdot \tfrac{1}{3} = 0.83,$$

$$Imb_{51}^{(2)} = (-1)^2 \cdot \tfrac{1}{3} + (-1)^2 \cdot \tfrac{1}{6} + (-2)^2 \cdot \tfrac{1}{6} + (-3)^2 \cdot \tfrac{1}{3} = 4.17.$$

Since $Imb_{51}^{(1)} = 0.83 < Imb_{51}^{(2)} = 4.17$, the coin will be biased toward treatment 1 with probability $p > 0.5$.

3. Theoretical properties of the new design. We now investigate the asymptotic properties of the proposed design. For the first n patients, we know that $D_n(k_1, \ldots, k_I)$ is the true difference of patient numbers within stratum $(k_1, \ldots, k_I)$. Furthermore, let

$$\mathbf{D}_n = \big[D_n(k_1, \ldots, k_I)\big]_{1 \le k_1 \le m_1, \ldots, 1 \le k_I \le m_I}$$

be an array of dimension $m_1 \times \cdots \times m_I$ which stores the current assignment differences in all strata. Also, assume that the covariates $Z_1, Z_2, \ldots$ are independently and identically distributed. Since $Z_n = (k_1, \ldots, k_I)$ can take $m = \prod_{i=1}^{I} m_i$ different values, it in fact follows an m-dimension multinomial distribution with parameter $\mathbf{p} = (p(k_1, \ldots, k_I))$, each element being the probability that a patient falls within the corresponding stratum. Obviously, $p(k_1, \ldots, k_I) \ge 0$ and $\sum_{k_1, \ldots, k_I} p(k_1, \ldots, k_I) = 1$.

First, we notice that $(\mathbf{D}_n)_{n \ge 1}$ is a Markov chain on the space $\mathbb{Z}^m$. In fact, by definition of the new procedure, $\mathbf{D}_n$ is a function f of $(\mathbf{D}_{n-1}, Z_n, T_n)$. Moreover, conditional on $\mathbf{D}_{n-1}$, (Z_n, T_n) is independent of $(\mathbf{D}_1, \ldots, \mathbf{D}_{n-2})$; therefore, $\mathbf{D}_n = f(\mathbf{D}_{n-1}, Z_n, T_n)$ is also conditionally independent of $(\mathbf{D}_1, \ldots, \mathbf{D}_{n-2})$.

We next explore the conditions under which $(\mathbf{D}_n)_{n \ge 1}$ is a *positive recurrent* chain, a desired property which indicates fast convergence rate. We will first investigate the special case of 2×2 strata, that is, only two covariates and two levels for each. This case enables us to obtain a finer result than the more general case, and at the same time also sheds light on how to set the conditions for the latter. With 2×2 strata, the weights on $Imb_n^{(1)}$ or $Imb_n^{(2)}$ reduce to w_o, $w_{m,1}$, $w_{m,2}$ and w_s.

THEOREM 3.1. *For the new design, consider* 2 *covariates and* 2 *levels for each.* w_o, $w_{m,1}$, $w_{m,2}$ *and* w_s *are nonnegative with* $w_o + w_{m,1} + w_{m,2} + w_s = 1$. *If the following two conditions hold*:

(A) $w_s > 0$,
(B) *define*

$$u_1 = w_o + w_{m,1} + w_{m,2} + w_s = 1,$$
$$u_2 = w_o + w_{m,1},$$
$$u_3 = w_o + w_{m,2},$$
$$u_4 = w_o;$$

the solution $\mathbf{x} = (x_1, x_2, x_3)$ *to the linear equation*

$$\begin{pmatrix} u_1 & u_2 & u_3 \\ u_2 & u_1 & u_4 \\ u_3 & u_4 & u_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} u_4 \\ u_3 \\ u_2 \end{pmatrix}$$

satisfies $|x_1| + |x_2| + |x_3| < 1$,

then $(\mathbf{D}_n)_{n\geq 1}$ *is a positive recurrent Markov chain with period* 2 *on* $\mathbb{Z}^4$.

REMARK 3.1. By Theorem 3.1 the chains $\mathbf{D}_{2n+1}$ and $\mathbf{D}_{2n}$ are two ergodic chains and converge to two limit distributions, respectively. Thus, $\mathbf{D}_{2n+1} = O_p(1)$ and $\mathbf{D}_{2n} = O_p(1)$, which implies that $\mathbf{D}_n = O_p(1)$. Accordingly, the imbalances at any level (within strata, on the margins, or overall) preserve the order of $O_p(1)$.

REMARK 3.2. In fact, u_1, u_2, u_3 and u_4 in the above theorem can be interpreted as the weights placed on individual strata: call the stratum in which the current patient falls a "target," then u_1 is the weight on the target itself; u_2 (u_3) on any stratum that is at the same level of covariate 1 (covariate 2) as the target; u_4 on any of the remaining strata.

COROLLARY 3.1. *In Theorem* 3.1, *if we further assume that* $w_{m,1} = w_{m,2} := w_m$, *then condition* (B) *is equivalent to*

$$(\mathrm{B}') \qquad w_m < C(w_o) := \frac{\sqrt{(1-w_o)^2 + 4(1+w_o)^2} - 1 - 3w_o}{4}.$$

REMARK 3.3. See Table 2 for certain values of $C(w_o)$. Since $C(w_o)$ is a decreasing and almost linear function of w_o on [0, 1], condition (B′) is much easier to verify than condition (B). For example, if the weight at the overall level $w_o = 0.20$, then the ones on the two margins need to be less than 0.23. Therefore, $(w_o, w_{m,1}, w_{m,2}, w_s) = (0.20, 0.22, 0.22, 0.36)$ is a legitimate weight set that ensures positive recurrence.

TABLE 2
Constraint on w_m as a function of w_o

w_o	0.00	0.20	0.40	0.60	0.80
$C(w_o)$	0.31	0.23	0.17	0.11	0.05

The next theorem deals with the general case of $m = \prod_{i=1}^{I} m_i$ strata. Using the basic equation $(x+1)^2 - (x-1)^2 = 4x$, the critical quantity $Imb_n^{(1)} - Imb_n^{(2)}$ in step (7) (Section 2) can be simplified as

$$\begin{aligned} &Imb_n^{(1)} - Imb_n^{(2)} \\ &\quad = 4\left\{ w_o D_{n-1} + \sum_{i=1}^{I} w_{m,i} D_{n-1}\big(i; k_i^*\big) + w_s D_{n-1}\big(k_1^*, \ldots, k_I^*\big) \right\} \\ &\quad := 4 \cdot \delta_{n-1}\big(k_1^*, \ldots, k_I^*\big). \end{aligned} \tag{3.1}$$

Therefore, the biasing probability p, q or $1/2$ is determined by the sign of $\delta_{n-1}(k_1^*, \ldots, k_I^*)$, which is a weighted average of current imbalances at different levels. Since D_{n-1} and $D_{n-1}(i; k_i^*)$ can both be expressed as a sum of certain $D_{n-1}(k_1, \ldots, k_I)$'s, we want to reformulate $\delta_{n-1}(k_1^*, \ldots, k_I^*)$ as a weighted average of imbalances within the individual strata.

As a motivating example, consider 3 covariates, gender (male or female), smoking behavior (smoker or nonsmoker) and clinical center (3 centers), with a total of 12 strata. Suppose for the new patient $Z_n = (1, 1, 1)$, that is, he falls into the stratum of "male smokers at center 1." Then for the remaining strata, the weights on $D_{n-1}(k_1, \ldots, k_I)$'s in the expression of $\delta_{n-1}(k_1^*, \ldots, k_I^*)$ are shown in Table 3.

TABLE 3
An example showing the weights of $D_{n-1}(k_1, \ldots, k_I)$'s in $\delta_{n-1}(k_1^*, \ldots, k_I^*)$

	Stratum	**Description**	**Weight**
1	(1,1,1)	male smokers at center 1	$w_o + w_{m,1} + w_{m,2} + w_{m,3} + w_s = 1$
2	(1,1,2)	male smokers at center 2	$w_o + w_{m,1} + w_{m,2}$
3	(1,1,3)	male smokers at center 3	$w_o + w_{m,1} + w_{m,2}$
4	(1,2,1)	male nonsmokers at center 1	$w_o + w_{m,1} + w_{m,3}$
5	(2,1,1)	female smokers at center 1	$w_o + w_{m,2} + w_{m,3}$
6	(1,2,2)	male nonsmokers at center 2	$w_o + w_{m,1}$
7	(1,2,3)	male nonsmokers at center 3	$w_o + w_{m,1}$
8	(2,1,2)	female smokers at center 2	$w_o + w_{m,2}$
9	(2,1,3)	female smokers at center 3	$w_o + w_{m,2}$
10	(2,2,1)	female nonsmokers at center 1	$w_o + w_{m,3}$
11	(2,2,2)	female nonsmokers at center 2	w_o
12	(2,2,3)	female nonsmokers at center 3	w_o

Generally, with respect to stratum $(k_1^*, \ldots, k_I^*)$ in which the new patient falls, we will divide the $m = \prod_{i=1}^{i} m_i$ strata into several categories and find out the corresponding weights in the expression of $\delta_{n-1}(k_1^*, \ldots, k_I^*)$. Let $\mathbb{I} = \{1, 2, \ldots, I\}$. For any stratum $(k_1, \ldots, k_I)$:

- if $(k_1, \ldots, k_I) = (k_1^*, \ldots, k_I^*)$, then the weight on $D_{n-1}(k_1, \ldots, k_I)$ is $w_o + \sum_{i=1}^{I} w_{m,i} + w_s = 1$;
- for any fixed i $(i \in \mathbb{I})$, if $k_i \neq k_i^*$ and $k_j = k_j^*$ for $j \in \mathbb{I}$ and $j \neq i$, then the weight on $D_{n-1}(k_1, \ldots, k_I)$ is $w_o + \sum_{j \neq i} w_{m,j}$, and there are $(m_i - 1)$ strata in this category;
- for any fixed $i_1 < i_2$ $(\{i_1, i_2\} \subset \mathbb{I})$, if $k_{i_1} \neq k_{i_1}^*$, $k_{i_2} \neq k_{i_2}^*$, and $k_j = k_j^*$ for $j \in \mathbb{I}$, $j \neq i_1$ and $j \neq i_2$, then the weight on $D_{n-1}(k_1, \ldots, k_I)$ is $w_o + \sum_{j \neq i_1, j \neq i_2} w_{m,j}$, and there are $(m_{i_1} - 1)(m_{i_2} - 1)$ strata in this category;
- for any fixed $i_1 < i_2 < \cdots < i_l$ $(\{i_1, \ldots, i_l\} \subset \mathbb{I})$, if $k_{i_t} \neq k_{i_t}^*$ and $k_j = k_j^*$ for $j \in \mathbb{I}$, $j \neq i_t$ and $1 \leq t \leq l$, then the weight on $D_{n-1}(k_1, \ldots, k_I)$ is $w_o + \sum_{j \neq i_t, 1 \leq t \leq l} w_{m,j}$, and there are $\prod_{t=1}^{l}(m_{i_t} - 1)$ strata in this category;
- if $k_i \neq k_i^*$ for all $i \in \mathbb{I}$, then the weight on $D_{n-1}(k_1, \ldots, k_I)$ is w_o, and there are $\prod_{i=1}^{l}(m_i - 1)$ strata in this category.

It is easily verified that

$$\begin{aligned} m &= \prod_{i=1}^{I} m_i \\ &= [(m_1 - 1) + 1][(m_2 - 1) + 1] \cdots [(m_I - 1) + 1] \\ &= 1 + \sum_{l=1}^{I} \sum_{1 \leq i_1 < i_2 < \cdots < i_l \leq I} \prod_{t=1}^{l} [m_{i_t} - 1], \end{aligned}$$

which is consistent with the counts listed above. Our general theorem in the following is closely related to the above weights and counts.

THEOREM 3.2. *For the new design, consider I covariates and m_i levels for the ith covariate, where $I \geq 1$, $1 \leq i \leq I$, and $m_i > 1$. w_o, w_s and $w_{m,i}$, $i = 1, \ldots, I$, are nonnegative with $w_o + \sum_{i=1}^{I} w_{m,i} + w_s = 1$. If*

$$\text{(C)} \quad u^* := \sum_{l=1}^{I} \sum_{1 \leq i_1 < i_2 < \cdots < i_l \leq I} \left\{ \left(w_o + \sum_{j \neq i_t, 1 \leq t \leq l} w_{m,j} \right) \prod_{t=1}^{l} [m_{i_t} - 1] \right\} < 1/2,$$

then $\mathbf{D}_n$ is a positive recurrent Markov chain on $\mathbb{Z}^m$.

To see the theorem in a more intuitive way, we will take a closer look at u^* in the special case of two covariates, as is shown in the following corollary.

COROLLARY 3.2. *In Theorem 3.2, if $I = 2$, then condition (C) is equivalent to*

$$(C') \qquad (m_1 m_2 - 1)w_o + (m_1 - 1)w_{m,2} + (m_2 - 1)w_{m,1} < 1/2.$$

REMARK 3.4. When $w_o = 0$ and $w_{m,1} = w_{m,2} = w_m$, condition (C′) further reduces to $w_m < [2(m_1 + m_2 - 2)]^{-1}$. For example, if $m_1 = m_2 = 5$, then $w_m < 1/16$ is required to satisfy condition (C).

REMARK 3.5. In both Theorems 3.1 and 3.2, $w_s > 0$ is required. Therefore, the theoretical results in these theorems do not apply to Pocock and Simon's (1975) design (with $w_s = 0$). The simulation result in Table 4 (Section 4) shows that the within-stratum imbalances under their method increase as the sample size increases, suggesting that they may not have the rate of $O_p(1)$. We hypothesize that the condition $w_s > 0$ is critical to ensure that $(\mathbf{D}_n)_{n\geq 1}$ is positive recurrent. These are further research problems.

To prove the above two theorems, we will use the technique of "drift conditions" [Meyn and Tweedie (1993)], which was developed for Markov chains on general state spaces. Applying their theory to our problem, in order to prove positive recurrence of $(\mathbf{D}_n)_{n\geq 1}$ we need to find a test function $V : \mathbb{Z}^m \to \mathbb{R}^+$, a bounded test set $\mathcal{C}$ on $\mathbb{Z}^m$ and two positive constants M_1 and M_2 such that

$$\Delta V(\mathbf{D}) := \sum_{\mathbf{D}' \in \mathbb{Z}^m} P(\mathbf{D}, \mathbf{D}')V(\mathbf{D}') - V(\mathbf{D}) \tag{3.2}$$

satisfies the following two conditions:

$$\Delta V(\mathbf{D}) \leq -M_1, \qquad \mathbf{D} \notin \mathcal{C}, \tag{3.3}$$

$$\Delta V(\mathbf{D}) \leq M_2, \qquad \mathbf{D} \in \mathcal{C}, \tag{3.4}$$

where $P(\mathbf{D}, \mathbf{D}')$ is the transition probability from $\mathbf{D}$ to $\mathbf{D}'$ on state space $\mathbb{Z}^m$ of the chain $(\mathbf{D}_n)_{n\geq 1}$. V is often a norm-like function on $\mathbb{Z}^m$. These drift conditions can roughly be interpreted as follows: so long as the average one-step movement ΔV tends to go back (with the magnitude uniformly greater than a positive constant M_1), that is, the chain is pulled back toward the finite set $\mathcal{C}$, positive recurrence can be ensured. For proofs of the theorems, see Section 6 and the supplemental article [Hu and Hu (2012)].

4. Simulation studies. We will compare the new procedure with stratified permuted block randomization and Pocock and Simon's (1975) minimization method. The simulations can be divided into three parts. First, we will simulate the case of 2×2 strata with a relatively large number of patients, to verify the convergence rate as stated in Theorem 3.1. Secondly, we are interested in the performances of different randomization methods when the number of strata is large

as compared to the sample size. An example of 500 patients and 10 covariates (each with 2 levels) will be studied. Finally, an example from Toorawa et al. (2009) will be considered, which is chosen because it resembles real situations in clinical trials.

4.1. 2×2 *strata*. For the three randomization procedures, we want to see whether the imbalances at any of the three levels (within-stratum, marginal and overall) stabilize, which indicates the rate of $O_p(1)$ at that specific level. The parameters are specified as follows:

- Multinomial probability $(p(1,1), p(1,2), p(2,1), p(2,2)) = (0.1, 0.2, 0.3, 0.4)$.
- Biasing probability $p = 0.85$ and $q = 0.15$ for Pocock and Simon's method (PS) as well as for the new procedure (NEW).
- Block size 4 for stratified randomization (STR-PB).
- Sample size $n = 200, 500, 1000$; number of simulated trials $N = 1000$.
- NEW: $(w_o, w_{m,1}, w_{m,2}, w_s) = (0.3, 0.1, 0.1, 0.5)$; conditions (A) and (B) are satisfied.
- PS: $(w_o, w_{m,1}, w_{m,2}, w_s) = (0, 0.5, 0.5, 0)$; Conditions (A) and (B) are NOT satisfied.

Table 4 shows the standard deviations (std's) of $D_n(\cdot)$'s at different levels (by symmetry of the designs, the theoretical mean of each $D_n(\cdot)$ is always 0). For simplicity, only the result of 2 strata and 2 margins are listed. Of the five columns, the first and the second give the std's of assignment differences within stratum $(1, 1)$ and $(2, 2)$; the third and fourth for the marginal differences of covariate 1 at level 1 and covariate 2 at level 2; and the last for the overall difference.

Table 4 suggests that all 5 standard deviations stabilize under NEW and STR-PB when the sample size increases. For example, under NEW the std's of $D_n(1, 1)$

TABLE 4
std's of $D_n(\cdot)$ of several methods under different sample sizes

Sample size		$\boldsymbol{D_n(1,1)}$	$\boldsymbol{D_n(2,2)}$	$\boldsymbol{D_n(1;1)}$	$\boldsymbol{D_n(2;2)}$	$\boldsymbol{D_n}$
STR-PB	200	0.92	0.89	1.30	1.27	1.83
	500	0.92	0.92	1.31	1.30	1.86
	1000	0.92	0.89	1.31	1.28	1.81
PS	200	3.16	3.27	1.15	1.13	1.30
	500	4.80	4.83	1.16	1.11	1.31
	1000	7.25	7.33	1.15	1.13	1.30
NEW	200	1.11	1.07	1.30	1.27	1.32
	500	1.14	1.10	1.33	1.28	1.22
	1000	1.03	1.10	1.20	1.24	1.27

are 1.11, 1.14 and 1.03; and those of D_n are 1.32, 1.22 and 1.27, which means that our new procedure preserves the rate of $O_p(1)$. The same conclusion can be reached for STR-PB. In fact, since the block size is 4, any within-stratum imbalance under STR-PB is bounded by 2. For PS, however, while the std's of marginal and overall differences stabilize, those of the within-stratum differences do not. For example, the std of $D_n(1, 1)$ increases from 3.16 to 4.80 and 7.25, much larger than those under the other two methods.

For the within-stratum imbalances, STR-PB is the best [0.92 for $D_n(1, 1)$], with NEW having slightly larger std's and PS the largest. For the marginal imbalances, PS is the best [around 1.15 for $D_n(1; 1)$], and the other two are about the same [around 1.30 for $D_n(1; 1)$]. For the overall imbalance, STR-PB is not as good as NEW and PS. Therefore, we see that even for 4 strata, STR-PB does not perform well for the overall imbalance.

4.2. 2^{10} *strata*. We simulate a hypothetical trial, which involves 500 patients, 10 covariates and 2 levels for each, that is, 1024 strata in total. The biased coin probabilities p and q for NEW and PS, the block size for STR-PB and the number of simulated trials N remain the same. The covariates are generated as follows: in addition to the independence assumption of covariates between patients, we further assume that within each patient the different covariates are independent and that each level within a fixed covariate is equally likely. Therefore, for the covariate profile $Z_i = (k_1, \ldots, k_I)$ of the ith patient, $k_1, \ldots, k_I$ are independently sampled from $\{1, 2\}$. For the weights, we use $w_o = 0$, $w_s = 0.5$ and $w_{m,i} = 0.5/10$.

Of the 1024 strata, on average 61.4% have no patient, and only 0.1% have 4 or more. Hence, if STR-PB is employed, most blocks are incomplete, which tends to cause large overall imbalance. Table 5 displays the mean absolute imbalances under each of the three randomization methods.

As seen in Table 5, STR-PB has an extremely large $\mathrm{E}|D_n|$ (17.07). In comparison, the other two methods have much smaller values of 0.76 and 0.98. So in this respect, PS has the best performance, and NEW is only slightly worse. In the second row, the mean absolute marginal imbalance is the average of the absolute differences over 20 margins as well as over the 1000 simulations, and the interpretation is the same as the overall imbalance. For the within-stratum imbalances,

TABLE 5
Mean $|D_n(\cdot)|$ *for* 2^{10} *strata and* 500 *patients*

	STR-PB	**PS**	**NEW**
Overall	17.07	0.76	0.98
Marginal	11.80	1.65	1.94
Within-strt. (2 pts)	0.66	0.98	0.50
Within-strt. (3 pts)	1.00	1.23	1.08

TABLE 6
Distribution of covariates

Sites	Small (2 sites)	1/120
	Medium (16 sites)	6/120
	Large (2 sites)	11/120
Other 3 covariates	Male; < 60; Moderate disease	10/20
	Male; ≥ 60; Moderate disease	2/20
	Male; < 60; Severe disease	2/20
	Male; ≥ 60; Severe disease	2/20
	Female; < 60; Moderate disease	1/20
	Female; ≥ 60; Moderate disease	1/20
	Female; < 60; Severe disease	1/20
	Female; ≥ 60; Severe disease	1/20

the table shows the result for strata with 2 or 3 patients. For example, under PS, 0.98 is the mean absolute difference over all strata with 2 patients as well as over the 1000 simulations. Under this criterion, PS is not recommended since the two means are 0.98 and 1.23, the largest among the three methods. STR-PB and NEW are quite similar, with means 0.66 versus 0.50 for strata with 2 patients, and 1.00 versus 1.08 with 3 patients. Hence, although our new procedure is not always the best, it ensures that no single type of the imbalances becomes too extreme.

4.3. *An example mimicking real clinical data*. We chose an example from Toorawa et al. (2009). The four covariates are site, gender, age and disease status, with 20, 2, 2 and 2 levels, respectively, resulting in 160 strata. The covariates' distribution is replicated in Table 6, where the marginal distribution of sites is independent of the joint distribution of the remaining three covariates.

120 patients enter the trial sequentially, and their covariates are independently simulated from the multinomial distribution in Table 6. We use the same p, q and block size as in the previous two examples. The weights are specified in the following way:

- NEW: $w_o = w_s = 1/3$ and $w_{m,i} = 1/12$, $i = 1, \ldots, 4$.
- PS: $w_o = w_s = 0$ and $w_{m,i} = 1/4$, $i = 1, \ldots, 4$.

Table 7 shows the distribution of 120 patients among 160 strata. In this case 24.3% of the strata have 1 patient; 11.4% contain 2 or 3 patients. If stratified ran-

TABLE 7
Distribution of patients among 160 *strata*

# of pts within stratum	**0**	**1**	**2**	**3**	**4 and more**
# of strata	95.4	38.8	12.7	5.6	7.6
Proportion	59.6%	24.3%	7.9%	3.5%	4.7%

TABLE 8
Comparison of absolute overall imbalance $|D_n|$

	STR-PB	**PS**	**NEW**
Mean	6.70	0.91	0.63
Median	6	0	0
95% quan	16	2	2

domization is employed, then the patients in the above 24.3% strata has to be randomized by equal probabilities. Moreover, the incomplete blocks in strata with 2 or 3 patients also pose a high risk of large overall imbalance.

The mean absolute imbalances at the three levels are compared, as shown in Tables 8, 9 and 10. Table 8 shows the result for the overall imbalance and lists the mean, median and 95% quantile of $|D_{120}|$. It is seen that NEW has mean, median and 95% quantile of 0.63, 0 and 2, respectively, whereas PS has slightly higher values. The three quantities are extremely high under STR-PB, which are not recommended for this case.

Table 9 gives the mean absolute marginal imbalances. For the covariates of gender, age and disease, the table explicitly lists the mean values on these 6 margins, as each of them only has two levels. For example, over the 1000 simulations, on average the absolute differences of patients in the two treatment groups within all male are 5.52, 1.10 and 1.59 under STR-PB, PS and NEW, respectively. Therefore, in this respect PS has the best performance; NEW is slightly worse, but still tolerable; STR-PB is the worst, since its mean is as high as 5.52. Similar conclusions can be reached for the other 5 margins. Moreover, for the margins relating to "site," since there are a total of 20 margins, we are unable to show the result on each margin due to the space limit. Hence, these 20 margins are further catego-

TABLE 9
Comparison of mean absolute marginal imbalances $E|D_n(i;k_i)|$

		STR-PB	**PS**	**NEW**
Gender	male	5.52	1.10	1.59
	female	3.86	1.06	1.55
Age	< 60	4.84	1.08	1.57
	≥ 60	4.40	1.11	1.23
Disease	moderate	5.01	1.10	1.56
	severe	4.35	1.18	1.52
20 sites	2 small	1.45	0.94	1.02
	16 median	1.44	1.21	1.32
	2 large	1.47	1.33	1.52

TABLE 10
Comparison of absolute within-stratum imbalances $|D_n(k_1, \ldots, k_I)|$: Distribution and mean

# of pts' within strt.	$\|D_n(k_1, \ldots, k_I)\|$	STR-PB	PS	NEW
2	prob(= 0)	0.68	0.57	0.69
	prob(= 2)	0.32	0.43	0.31
	mean	**0.64**	**0.86**	**0.62**
3	prob(= 1)	1.00	0.85	0.94
	prob(= 3)	0.00	0.15	0.06
	mean	**1.00**	**1.30**	**1.12**

rized into three groups of small, median and large sizes, and the mean values in the table are further averaged over the margins within the groups. For example, 1.32 is the mean absolute imbalance over the 16 median-sized sites as well as over the 1000 simulations. In terms of imbalances on margins defined by site, PS is still the best, and STR-PB has similar performance to NEW. This is because each margin of site contains only 8 strata, hence the "accumulating effect" of within-stratum imbalances under STR-PB is not as strong.

Table 10 displays the distribution and absolute mean of within-stratum imbalances for strata with 2 or 3 patients. For example, of all the strata which contain 2 patients, the absolute difference is either 0 or 2, and the distribution is 0.69 to 0 and 0.31 to 2 under NEW, leading to an average of 0.62. According to this criterion, NEW has the lowest mean, STR-PB has a slightly larger value and PS has mean as large as 0.86. For strata containing 3 patients, since the block size is 4 for STR-PB, it is impossible to get an absolute value of 3. Hence, the mean absolute imbalance is 1, the minimum among the three methods.

In summary, our new method maintains good balance from all three perspectives and should be favored. We also performed the simulations under other parameter values. Some of them include: (1) changing the weights w_o, w_s, and $w_{m,i}$, as well as the block size; (2) 2×100 strata, representing few covariates but many levels at least for one covariate; (3) $3 \times 4 \times 5 \times 6$ strata, representing a few covariates and a few levels for each. In all the above settings, our new procedure shows advantages over the other two methods.

5. Conclusion. In this paper we propose a new covariate-adaptive design that minimizes a weighted average of three types of imbalances (within-stratum, within-covariate-margin and overall). Simulation results show that the proposed method provides better allocation balance from different perspectives, while stratified randomization and Pocock and Simon's (1975) marginal method have large imbalances either as a whole, or within-stratum.

The new procedure can also be generalized in several ways. In this paper we only considered balanced allocation (1:1), whereas in some problems unequal ratios [Hu and Rosenberger (2006)] are also desired. For example, if the two groups

are an innovation versus a placebo, and a pilot study has shown some effect of the innovation, then it is more ethical to assign more patients to the innovation. If one treatment is much more costly than the other, then assigning more patients to the latter would be more economical. Sometimes, the randomization has to be adapted to covariates as well as responses. Zhang et al. (2007) proposed "covariate-adjusted response-adaptive randomization," whose allocation ratio depends on both covariate profiles and responses of patients. One may modify our proposed procedure to accommodate these situations. On the other hand, some trials (e.g., some Phase II trials) involve the comparison of more than two treatments [Pocock and Simon (1975), Hu and Rosenberger (2006), etc.]. We can generalize the proposed procedure to clinical trials for comparing three or more treatments. We leave these as future research topics.

For Efron's (1971) biased coin design (without involving covariates), it is well known that the imbalance is a positive recurrent Markov chain. Markaryan and Rosenberger (2010) studied some exact properties of Efron's (1971) biased coin design. However, to our best knowledge, there is no theoretical result about the imbalance of covariate-adaptive randomization in literature, due to the complex of the problem and the lack of technical tools. In this paper, we introduced the technique of "drift conditions" in Markov chains to study the theoretical properties of covariate-adaptive randomization. This technique could provide a possible way of studying the properties of general covariate-adaptive designs as well as covariate-adjusted response-adaptive designs.

The inference under covariate-adaptive randomization is also an important issue. By simulation studies, several authors have raised concerns about the conservativeness of the unadjusted analysis (such as two-sample t-test) under covariate-adaptive randomization and suggested that all covariates that are used in the randomization should be included in the analysis [Birkett (1985), Forsythe (1987), etc.]. Shao, Yu and Zhong (2010) studied the theoretical relationship between different randomization designs and different inference methods. To make the problem more tractable, the authors focused on a simple homogeneous linear model. They found that if the underlying response-covariate model can be correctly specified, then the usual regression analysis is valid and has the highest power as compared to other types of analysis, no matter what randomization is employed. These results also apply to the proposed randomization procedure in this paper.

If the model specification is not feasible and only a two-sample t-test can be used, then the test under stratified randomization tends to have a conservative type I error rate due to the overestimation of $\mathsf{Var}(\bar{Y}_1 - \bar{Y}_2)$. Shao, Yu and Zhong (2010) used a bootstrap method to correct the variance estimation. The resulting bootstrap t-test restores the type I error rate, and is more powerful than the traditional t-test under simple randomization. Similar bootstrap adjustment can be used as an inference method for the new randomization procedure. We leave this as a future research project.

6. Sketch of proofs.

PROOF OF THEOREM 3.1. With 2×2 strata, the within-stratum imbalances $\mathbf{D}_n$ and the multinomial probabilities $\mathbf{p}$ are both matrices of 2×2. Let $\tilde{\mathbf{D}}_n = (D_{n,1}, D_{n,2}, D_{n,3}, D_{n,4}) := (D_n(1,1), D_n(1,2), D_n(2,1), D_n(2,2))$, that is, $\tilde{\mathbf{D}}_n$ is simply the vector form of $\mathbf{D}_n$. $\tilde{\mathbf{p}} = (p_1, \ldots, p_4)$ can be defined in the same way. By the above notation, any stratum can be represented by the 2-index form (k_1, k_2), or the single-index form (r) $(1 \le r \le 4)$. The quantity $\delta_{n-1}(k_1^*, k_2^*)$ in (3.1) then reduces to

$$\begin{aligned}
&\delta_{n-1}(k_1^*, k_2^*) \\
&\quad = (w_o + w_{m,1} + w_{m,2} + w_o) D_{n-1}(k_1^*, k_2^*) + (w_o + w_{m,1}) D_{n-1}(k_1^*, k_2) \\
&\qquad + (w_o + w_{m,2}) D_{n-1}(k_1, k_2^*) + w_o D_{n-1}(k_1, k_2) \\
&\quad = u_1 D_{n-1}(k_1^*, k_2^*) + u_2 D_{n-1}(k_1^*, k_2) + u_3 D_{n-1}(k_1, k_2^*) + u_4 D_{n-1}(k_1, k_2),
\end{aligned} \tag{6.1}$$

where $k_1 \neq k_1^*$, $k_2 \neq k_2^*$ and $u_1 = 1$. Let $\tilde{\boldsymbol{\delta}}_n = (\delta_{n,1}, \delta_{n,2}, \delta_{n,3}, \delta_{n,4}) := (\delta_n(1,1), \delta_n(1,2), \delta_n(2,1), \delta_n(2,2))$. Then, according to (6.1), $\tilde{\mathbf{D}}_n$ and $\tilde{\boldsymbol{\delta}}_n$ are linked by

$$\tilde{\boldsymbol{\delta}}_n = \tilde{\mathbf{D}}_n \begin{pmatrix} u_1 & u_2 & u_3 & u_4 \\ u_2 & u_1 & u_4 & u_3 \\ u_3 & u_4 & u_1 & u_2 \\ u_4 & u_3 & u_2 & u_1 \end{pmatrix} := \tilde{\mathbf{D}}_n \mathbf{U}. \tag{6.2}$$

For any $\tilde{\mathbf{D}}_n \in \mathbb{Z}^4$, we define a test function

$$V(\tilde{\mathbf{D}}_n) = \sum_{r=1}^{4} \frac{[D_{n,r}]^2}{p_r},$$

that is, the sum of squared within-stratum differences adjusted for the corresponding multinomial probabilities. The test set $\mathcal{C}$ is defined as $\mathcal{C} = \{\tilde{\mathbf{D}}_n : \max_r \|\tilde{D}_{n,r}\| \le K\}$ ($K > 0$ is to be determined). V and $\mathcal{C}$ are the key elements in proving positive recurrence, according to the drift conditions (3.3) and (3.4).

For the ease of representation, in the rest of the proof we will simply use the notation $\mathbf{D}$ and $\boldsymbol{\delta}$ for $\tilde{\mathbf{D}}_n$ and $\tilde{\boldsymbol{\delta}}_n$, respectively, unless specified otherwise. Under the new allocation rule, it can be derived that the one-step movement $\Delta V(\mathbf{D})$, defined in (3.2), has the form

$$\Delta V(\mathbf{D}) = 2(q - p) \sum_{r=1}^{4} \operatorname{sgn}(\delta_r) D_r + 4,$$

where D_r and δ_r are the rth element of vectors $\mathbf{D}$ and $\boldsymbol{\delta}$, respectively, and $\operatorname{sgn}(x) = 1, -1, 0$ for $x > 0, < 0$ or $= 0$. For derivation of $\Delta V(\mathbf{D})$, see Section 1 of the supplemental article [Hu and Hu (2012)].

We need to show that $\Delta V(\mathbf{D})$ satisfies drift conditions (3.3) and (3.4). In fact, since the test set $\mathcal{C}$ is bounded, (3.4) is trivially true. Since $q - p < 0$, (3.3) is equivalent to finding $M_1' > 2/(p-q)$ such that

$$\Delta W(\mathbf{D}) := \sum_{r=1}^{4} \operatorname{sgn}(\delta_r) D_r > M_1' \qquad \text{for } \mathbf{D} \notin \mathcal{C}. \tag{6.3}$$

Intuitively, when u_2, u_3 and u_4 are small, δ_r is expected to be close to D_r so that they have the same sign. Thus, a larger proportion of the strata have $\operatorname{sgn}(\delta_r)D_r = \operatorname{sgn}(D_r)D_r = |D_r|$ and $\Delta W(\mathbf{D})$ tends to be positive. In the trivial case that $u_2 = u_3 = u_4 = 0$, that is, $\mathbf{D} = \boldsymbol{\delta}$, we have $\Delta W(\mathbf{D}) = \sum_{r=1}^{4} |D_r| > K$, so (6.3) holds by letting $M_1' = K = 2.1/(p-q)$. Therefore, in the following we can assume that $\max\{u_2, u_3, u_4\} > 0$.

For any $\mathbf{D} \in \mathcal{C}^c \subset \mathbb{Z}^4$, call the pair of (D_r, δ_r) a "*match*" if $\delta_r \neq 0$ and $\delta_r D_r \geq 0$. Hence, for a match $\operatorname{sgn}(\delta_r)D_r = |D_r|$. Furthermore, define $M(\mathbf{D}, \boldsymbol{\delta})$ as the number of matches in (D_r, δ_r)'s, $r = 1, \ldots, 4$. Obviously, $0 \leq M(\mathbf{D}, \boldsymbol{\delta}) \leq 4$. It can be shown that $M(\mathbf{D}, \boldsymbol{\delta}) = 0$ is impossible for $\mathbf{D} \in \mathcal{C}^c$. Therefore, for $M(\mathbf{D}, \boldsymbol{\delta}) = i$, $i = 1, 2, 3, 4$, if we can find $d_i > 0$, such that $\Delta W(\mathbf{D}) > d_i K$, then (6.3) is true by letting $M_1' = K \min_i d_i$ and $K = 2.1/[(p-q)\min_i d_i]$.

In fact, finding d_4 for $M(\mathbf{D}, \boldsymbol{\delta}) = 4$ is quite trivial ($d_4 = 1$). We will show how to find d_3 for $M(\mathbf{D}, \boldsymbol{\delta}) = 3$ below. When $M(\mathbf{D}, \boldsymbol{\delta}) = 3$, we know that $a_1 = \max\{u_2, u_3, u_4\} \neq 0$ and $a_2 = \min\{1 - u_2, 1 - u_3, 1 - u_4\} \neq 0$ (since $w_s \neq 0$). Without loss of generality assume D_1 and δ_1 do not match, which means $\delta_1 D_1 \leq 0$. Thus $|\delta_1 - D_1| \geq |D_1|$. By (6.2), $\delta_1 - D_1 = u_2 D_2 + u_3 D_3 + u_4 D_4$, which implies $|u_2 D_2 + u_3 D_3 + u_4 D_4| \geq |D_1|$. Then

$$\begin{aligned}
\Delta W(\mathbf{D}) &\geq -|D_1| + |D_2| + |D_3| + |D_4| \\
&\geq -\big(u_2|D_2| + u_3|D_3| + u_4|D_4|\big) + |D_2| + |D_3| + |D_4| \\
&\geq a_2\big(|D_2| + |D_3| + |D_4|\big) \\
&\geq a_2\big[(1/2)\big(|D_2| + |D_3| + |D_4|\big) + (1/2)a_1^{-1}|D_1|\big] \\
&\geq (a_2/2)\min\big\{1, a_1^{-1}\big\} \cdot \max\big\{|D_1|, |D_2|, |D_3|, |D_4|\big\} \\
&> (a_2/2)\min\big\{1, a_1^{-1}\big\} \cdot K := d_3 K.
\end{aligned}$$

The ways of finding d_2 and d_1 for $M(\mathbf{D}, \boldsymbol{\delta}) = 2$ and 1 are similar, but require more work. In particular, condition (B) in Theorem (3.1) is needed to verify the case of $M(\mathbf{D}, \boldsymbol{\delta}) = 1$. In Section 2 of the supplemental article [Hu and Hu (2012)], we show how to find d_i for $i = 4, 3, 2, 1$ and explain why $M(\mathbf{D}, \boldsymbol{\delta}) \neq 0$.

Corollary 3.1 is obtained by solving the linear equation for $\mathbf{x}$ in Theorem 3.1 under the assumption that $w_{m,1} = w_{m,2}$ and then substituting the solution in $|x_1| + |x_2| + |x_3| < 1$. For detailed proof of Corollary 3.1, see Section 3 of the supplemental article [Hu and Hu (2012)]. □

PROOF OF THEOREM 3.2. The main steps are similar to those in Theorem 3.1. Let $\tilde{\mathbf{D}}_n = (D_{n,1}, \ldots, D_{n,m})$ be the vector version of $\mathbf{D}_n = (D_n(k_1, \ldots, k_I))$, where the m strata can be arbitrarily ordered and indexed by $1, \ldots, m$. Similary, let $\tilde{\boldsymbol{\delta}}_n$ and $\tilde{\mathbf{p}}$ be the vector forms of array $(\delta_n(k_1, \ldots, k_I))$ and array $(p(k_1, \ldots, k_I))$, respectively, using the same order as in $(D_n(1), \ldots, D_n(m))$. By the above notation, any stratum can be represented by the I-index form $(k_1, \ldots, k_I)$, or the single-index form (r) $(1 \le r \le m)$. As in the 2×2 case, let $\tilde{\boldsymbol{\delta}}_n := \tilde{\mathbf{D}}_n \mathbf{U}$. Then by the definition of $\tilde{\boldsymbol{\delta}}_n$ as well as the description of weights before Theorem 3.2, for any two strata $(r) = (k_1^*, \ldots, k_I^*)$ and $(s) = (k_1, \ldots, k_I)$, the element u_{rs} in the matrix of $\mathbf{U}$ is determined as follows: for any fixed $i_1 < i_2 < \cdots < i_l$ $(\{i_1, \ldots, i_l\} \in \mathbb{I})$, if $k_{i_t} \ne k_{i_t}^*$ and $k_j = k_j^*$ for $j \in \mathbb{I}$, $j \ne i_t$ and $1 \le t \le l$, then

$$u_{rs} = w_o + \sum_{j \ne i_t, 1 \le t \le l} w_{m,j}.$$

So $u_{rs} = u_{sr}$, and for any r, $\sum_{s=1,\ldots,m, s \ne r} u_{rs} = u^*$, as defined in Theorem 3.2.

The test function V and the test set $\mathcal{C}$ are still defined as before, except that in this case the dimension of $\tilde{\mathbf{D}}_n$ is m instead of 4. Use the simple notation $\mathbf{D}$ and $\boldsymbol{\delta}$ for $\tilde{\mathbf{D}}_n$ and $\tilde{\boldsymbol{\delta}}_n$, respectively. In the same manner, to verify the drift conditions it is equivalent to find $M_1' > 2/(p-q)$ such that

$$\Delta W(\mathbf{D}) := \sum_{r=1}^{m} \operatorname{sgn}(\delta_r) D_r > M_1' \qquad \text{for } \mathbf{D} \notin \mathcal{C}. \tag{6.4}$$

For any fixed $\mathbf{D} \in \mathcal{C}^c \subset \mathbb{Z}^m$, suppose for (D_r, δ_r)'s, $r = 1, \ldots, m$, there are m_0 mismatched pairs. Without loss of generality assume that the mismatched pairs occur in the 1st, 2nd, ... and the m_0th strata. By the definition of a mismatched pair, D_r and $\delta_r = \sum_{s=1}^{r-1} u_{rs} D_s + D_r + \sum_{s=l+1}^{m} u_{rs} D_s$ have different signs, $r = 1, \ldots, m_0$. Therefore,

$$\begin{aligned} |D_r| &\le \left| \sum_{s=1}^{r-1} u_{rs} D_s + \sum_{s=l+1}^{m} u_{rs} D_s \right| \\ &\le \sum_{s=1,\ldots,m_0, s \ne r} u_{rs} |D_s| + \sum_{s=m_0+1}^{m} u_{rs} |D_s|. \end{aligned} \tag{6.5}$$

First, we notice that $m_0 \ne m$; otherwise, by summing (6.5) over $r = 1$ to m, we have $\sum_{r=1}^{m} |D_r| \le u^* \sum_{r=1}^{m} |D_r|$ which is impossible for $\mathbf{D} \in \mathcal{C}^c$ and $u^* < 1/2$. Second, suppose $m_0 \ne 0$. By summing (6.5) over $r = 1$ to $r = m_0$, we have

$$\sum_{r=1}^{m_0} \left(1 - \sum_{s=1,\ldots,m_0, s \ne r} u_{rs} \right) |D_r| \le \sum_{r=m_0+1}^{m} \left(\sum_{s=1}^{m_0} u_{rs} \right) |D_r|.$$

Combined with the fact that $1-u^* \le (1-\sum_{s=1,\ldots,m_0,k\neq r} u_{rs})$ for $r=1,\ldots,m_0$ and $(\sum_{s=1}^{m_0} u_{rs}) \le u^*$ for $r=m_0+1,\ldots,m$, it is seen that

$$\sum_{r=1}^{m_0}(1-u^*)|D_r| \le \sum_{r=m_0+1}^{m} u^*|D_r|.$$

Then

$$\begin{aligned}\Delta W(\mathbf{D}) &\ge -|D_1|-|D_2|-\cdots-|D_{m_0}|+|D_{m_0+1}|+|D_{m_0+2}|+\cdots+|D_m| \\ &\ge -\frac{u^*}{1-u^*}\sum_{r=m_0+1}^{m}|D_r|+\sum_{r=m_0+1}^{m}|D_r| = \left(1-\frac{u^*}{1-u^*}\right)\sum_{r=m_0+1}^{m}|D_r|.\end{aligned}$$

Since $0 \le u^* < 1/2$, we have $0 < 1-\frac{u^*}{1-u^*} \le 1$. Hence, the above inequality is also true for $m_0=0$. If $\max_{m_0+1\le r\le m}|D_r| > K$, then $\Delta W(\mathbf{D}) > (1-\frac{u^*}{1-u^*})K$; otherwise $\max_{1\le r\le m_0}|D_r| > K$ and $\sum_{r=m_0+1}^{m}|D_r| \ge \frac{1-u^*}{u^*}\sum_{r=1}^{m_0}|D_r|$, which means $\Delta W(\mathbf{D}) > (1-\frac{u^*}{1-u^*})\frac{1-u^*}{u^*}K > (1-\frac{u^*}{1-u^*})K$. Thus, if we define $M_2' = (1-\frac{u^*}{1-u^*})K$ and $K = \frac{2.1}{p-q}(1-\frac{u^*}{1-u^*})^{-1}$, then

$$\Delta W(\mathbf{D}) > M_2' > 2/(p-q). \qquad \square$$

Acknowledgments. Special thanks go to anonymous referees, the Associate Editor and the Editor for the constructive comments, which led to a much improved version of the paper.

SUPPLEMENTARY MATERIAL

Additional proofs (DOI: 10.1214/12-AOS983SUPP; .pdf). We provide additional proofs that are omitted in Section 6. They include: (1) derivation of $\Delta V(\mathbf{D})$; (2) the appropriate choice of d_i when $M(\mathbf{D},\boldsymbol{\delta}) = i$ $(i=4,3,2,1)$; (3) proof of Corollary 3.1.

REFERENCES

ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67. MR0655670

BAI, Z. D. and HU, F. (1999). Asymptotic theorems for urn models with nonhomogeneous generating matrices. *Stochastic Process. Appl.* **80** 87–101. MR1670107

BEGG, C. B. and IGLEWICZ, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics* **36** 81–90.

BIRKETT, N. J. (1985). Adaptive allocation in randomized controlled trials. *Control. Clin. Trials* **6** 146–155.

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660

FORSYTHE, A. B. (1987). Validity and power of tests when groups have been balanced for prognostic factors. *Comput. Statist. Data Anal.* **5** 193–200.

HAGINO, A., HAMADA, C., YOSHIMURA, I., OHASHI, Y., SAKAMOTO, J. and NAKAZATO, H. (2004). Statistical comparison of random allocation methods in cancer clinical trials. *Control. Clin. Trials* **25** 572–584.

HERITIER, S., GEBSKI, V. and PILLAI, A. (2005). Dynamic balancing randomization in controlled clinical trials. *Stat. Med.* **24** 3729–3741. MR2221964

HU, F. (2012). Statistical issues to trial design and personalized medicine. *Clinical Investigation* **2** 121–124.

HU, Y. and HU, F. (2012). Supplement to "Asymptotic properties of covariate-adaptive randomization." DOI:10.1214/12-AOS983SUPP.

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

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

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

KALISH, L. A. and BEGG, C. B. (1985). Treatment allocation methods in clinical trials: A review. *Stat. Med.* **4** 129–144.

KHAN, O., FOTHERINGHAM, S., WOOD, V., STIMSON, L., ZHANG, C., PEZZELLA, F., DUVIC, M., KERR, D. J. and THANGUE, N. B. L. (2010). HR23B is a biomarker for tumor sensitivity to HDAC inhibitor-based therapy. *Proc. Natl. Acad. Sci. USA* **107** 6532–6537.

KUNDT, G. (2009). Comparative evaluation of balancing properties of stratified randomization procedures. *Methods Inf. Med.* **48** 129–134.

LAGAKOS, S. W. and POCOCK, S. J. (1984). Randomization and stratification in cancer clinical trials: An international survey. In *Cancer Clinical Trials*: *Methods and Practice* (M. E. Buyse, M. J. Staquet and R. J. Sylvester, eds.). Oxford Univ. Press, Oxford.

LI, Y., SHEU, C. C., YE, Y., ANDRADE, M. D., WANG, L., CHANG, S. C., AUBRY, M. C., AAKRE, J. A., ALLEN, M. S., CHEN, F., CUNNINGHAM, J. M., DESCHAMPS, C., JIANG, R., LIN, J., MARKS, R. S., PANKRATZ, V. S., SU, L., LI, Y., SUN, Z., TANG, H., VASMATZIS, G., HARRIS, C. C., SPITZ, M. R., JEN, J., WANG, R., ZHANG, Z. F., CHRISTIANI, D. C., WU, X. and YANG, P. (2010). Genetic variants and risk of lung cancer in never smokers: A genome-wide association study. *Lancet Oncology* **11** 321–330.

MARKARYAN, T. and ROSENBERGER, W. F. (2010). Exact properties of Efron's biased coin randomization procedure. *Ann. Statist.* **38** 1546–1567. MR2662351

MATTS, J. P. and LACHIN, J. M. (1988). Properties of permuted-block randomization in clinical trials. *Control. Clin. Trials* **9** 327–344.

MCENTEGART, D. J. (2003). The pursuit of balance using stratified and dynamic randomization techniques: An overview. *Drug Information Journal* **37** 293–308.

MCILROY, M., MCCARTAN, D., EARLY, S., GAORA, P., PENNINGTON, S., HILL, A. D. K. and YOUNG, L. S. (2010). Interaction of developmental transcription factor HOXC11 with steroid receptor coactivator SRC-1 mediates resistance to endocrine therapy in breast cancer. *Cancer Research* **70** 1585–1594.

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London. MR1287609

NORDLE, O. and BRANTMARK, B. (1977). A self-adjusting randomization plan for allocation of patients into two treatment groups. *Clin. Pharmacol. Ther.* **22** 825–830.

POCOCK, S. J. (1982). Statistical aspects of clinical trial design. *The Statistician* **31** 1–18.

POCOCK, S. J. and SIMON, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31** 103–115.

ROSENBERGER, W. F. and SVERDLOV, O. (2008). Handling covariates in the design of clinical trials. *Statist. Sci.* **23** 404–419. MR2483911

SHAO, J., YU, X. and ZHONG, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* **97** 347–360. MR2650743

SIGNORINI, D. F., LEUNG, O., SIMES, R. J., BELLER, E., GEBSKI, V. J. and CALLAGHAN, T. (1993). Dynamic balanced randomization for clinical trials. *Stat. Med.* **12** 2343–2350.

SMITH, R. L. (1984a). Properties of biased coin designs in sequential clinical trials. *Ann. Statist.* **12** 1018–1034. MR0751289

SMITH, R. L. (1984b). Sequential treatment allocation using biased coin designs. *J. Roy. Statist. Soc. Ser. B* **46** 519–543. MR0790636

TAVES, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clin. Pharmacol. Ther.* **15** 443–453.

TAVES, D. R. (2010). The use of minimization in clinical trials. *Contemp. Clin. Trials* **31** 180–184.

TOORAWA, R., ADENA, M., DONOVAN, M., JONES, S. and CONLON, J. (2009). Use of simulation to compare the performance of minimization with stratified blocked randomization. *Pharm. Stat.* **8** 264–278.

WEI, L. J. (1978). An application of an urn model to the design of sequential controlled clinical trials. *J. Amer. Statist. Assoc.* **73** 559–563. MR0514157

WEIR, C. J. and LEES, K. R. (2003). Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Stat. Med.* **22** 705–726.

ZELEN, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases* **27** 365–375.

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

ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *Ann. Statist.* **35** 1166–1182. MR2341702

DEPARTMENT OF STATISTICS
UNIVERSITY OF VIRGINIA
HALSEY HALL, CHARLOTTESVILLE
VIRGINIA 22904-4135
USA
E-MAIL: yh2s@virginia.edu
fh6e@virginia.edu