Statistica Sinica Preprint No: SS-2025-0213

Title	Explicit Form of the Asymptotic Covariance Matrix of the Normalized Within-Stratum Imbalances Following Minimization with Independent Factors with Application to the Log-Rank Test
Manuscript ID	SS-2025-0213
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0213
Complete List of Authors	Olga M Kuznetsova, Victoria P Johnson and Michael Gekhtman
Corresponding Authors	Olga M Kuznetsova
E-mails	olga_kuznetsova@merck.com
Notice: Accepted author version.	

Statistica Sinica: Newly accepted Paper

Statistica Sinica

EXPLICIT FORM OF THE ASYMPTOTIC COVARIANCE MATRIX OF THE NORMALIZED WITHIN-STRATUM IMBALANCES FOLLOWING MINIMIZATION WITH INDEPENDENT FACTORS WITH APPLICATION TO THE LOG-RANK TEST

Olga M. Kuznetsova[1], Victoria P. Johnson[2], Michael Gekhtman[3]

[1]Merck & Co., Inc., Rahway, NJ, USA; ORCID ID: https://orcid.org/0000-0002-7037-3641

[2]Statistics and Data Science Innovation Hub, GlaxoSmithKline, Collegeville, PA, USA; ORCID ID: https://orcid.org/0000-0002-3045-8560

[3]Department of Mathematics, University of Notre Dame, Notre Dame, IN; ORCID ID: https://orcid.org/0000-0003-1731-7015

The Pocock and Simon's covariate-adaptive randomization (minimization) is employed to dynamically balance the treatment groups in prognostic factors when the number of strata is large. The asymptotic covariance matrix of the within-stratum imbalances plays an important role in the inferential properties of the robust score test and the unstratified logrank test following a covariate-adaptive allocation procedure. However, the explicit form of this asymptotic covariance matrix is known only for case of the equal prevalence of the strata. In this work, based on empirical observations from

Statistica Sinica: Newly accepted Paper

extensive simulations, the previously unknown explicit form of the asymptotic covariance matrix is provided for the minimization with unequal prevalence and independent covariates. It was determined that the covariance matrix is proportional to the one arising from a simpler probabilistic model intuitively connected to minimization. The coefficient of proportionality V depends on the bias of the biased coin used with minimization. It is proven theoretically that when $V \leq 1$, the maximum eigenvalue of the asymptotic covariance matrix does not exceed 1. This means that the unstratified log-rank test and the robust score test following such minimization is valid or conservative under a misspecified model. These results are supported by simulations of a clinical trial with the treatment groups balanced using minimization on two factors: the region with 5 levels and the risk group with 4 levels.

Key words and phrases: Asymptotic covariance matrix, log-rank test, minimization, Pocock and Simon covariate-adaptive randomization, unequal prevalence of the strata.

1. Introduction

Covariate-adaptive randomization procedures are often used in clinical trials to ensure that the treatment groups are balanced with respect to important prognostic factors. The most common approach is to employ stratified randomization, where the study population is broken into strata formed by the combinations of the levels of the covariates. A restricted randomization procedure is then used to randomize subjects within each stratum independently from randomization of subjects in other strata (Zelen (1974); Rosenberger and Lachin (2016)). Stratified randomization, however, cannot balance a large number of factors or use factors with a large number of lev-

Statistica Sinica: Newly accepted Paper

els. In this case the number of strata becomes too large and the strata too small (Therneau (1973)). Instead, one of the dynamic allocation procedures is used in this situation (Taves (1974); Pocock and Simon (1975); Rosenberger and Sverdlov (2008); Hu, Hu, Ma and Rosenberger (2014); Scott, McPherson, Craig, Ramsay and Campbell (2002)). The Pocock and Simon covariate-adaptive procedure (Pocock and Simon (1975)), a generalization of the minimization proposed by Taves (1974) that adds a random element to every allocation, is a popular dynamic allocation procedure (Taves (2010)). Based on the covariates of the subject who arrived for randomization, it finds the treatment arm that would result in the smallest overall imbalance. This arm is assigned to the subject with high probability p (typically, 0.8–0.95) called bias (as in the biased coin randomization (Efron (1971))). In this paper the equal randomization to two treatment arms is considered.

Historically the properties of the tests following minimization were studied through simulations (Birkett (1985); Forsythe (1987); Weir and Lee (2003); Kahan and Morris (2012); Barbachano and Coad (2013); Kuznetsova and Tymofyeyev (2012); Luo, Li, Xu and Tu (2016)). However, lately the inference following covariate-adaptive allocation is a topic of fast-developing research. The validity of the t-test under linear or generalized linear model

Statistica Sinica: Newly accepted Paper

was demonstrated by Shao and Yu (2010, 2013); Ma, Hu and Zhang (2015); Bugni, Canay and Shaikh (2018); Ye (2018). The validity of the inference methods following covariate-adaptive randomization was further studied in Ye, Shao, Yi and Zhao (2023); Ye, Shao and Yi (2024); Ma, Tu and Liu (2022); Liu, Tu and Ma (2022).

Ye and Shao (2020) developed a comprehensive theory for log-rank-type and partial likelihood score tests (Peto et al. (1976); Cox (1972, 1975)) under covariate-adaptive randomization. For these tests, the asymptotic properties of the within-stratum imbalances in treatment assignments normalized by the square root of the stratum size play very important role. Ye and Shao (2020) introduced a classification of the randomization procedures according to these properties. The first class described (called Type 1 by Shao (Shao (2021))) includes randomization procedures for which the normalized within-stratum imbalances converge to 0 as the sample size increases. The stratified permuted blocks (Zelen (1974)) or biased coin (Efron (1971)) randomization belong to this class. The second class (Type 2 per Shao (2021)) (Table 1) includes procedures with independent within-strata imbalances for which the normalized imbalances converge in distribution to the standard normal variable with variance not exceeding that for the simple randomization. The example of such procedures is the stratified

Statistica Sinica: Newly accepted Paper

urn model randomization (Wei (1978a,b)). Ye and Shao (2020) demonstrated that under a misspecified model, the robust score test (Lin and Wei, 1989) and the log-rank test are valid or conservative for these two classes of covariate-adaptive procedures. Ye and Shao (2020) also proposed a variance adjustment to the partial likelihood score test that makes the test valid and robust under model misspecification.

Minimization does not belong to either of the two classes. Indeed, its within-stratum imbalances are not independent as the probability of the next treatment assignment depends on the treatment assignments of subjects already enrolled in different strata. Also, the normalized within-stratum imbalances do not converge to 0 as the sample size increases (Hu and Zhang (2020)). Johnson, Gekhtman and Kuznetsova(2024) expanded on the methodological framework built by Ye and Shao (2020) to demonstrate that under model misspecification the robust score test and the log-rank test are also valid or conservative for a class of allocation procedures they called Type 2a. This class includes allocation procedures with the following properties: as the study size increases, 1) the vector of the normalized within-stratum imbalances converges in distribution to a multivariate Gaussian vector and 2) the covariance matrix of the normalized within-stratum treatment imbalances converges to a non-negative-definite

Statistica Sinica: Newly accepted Paper

matrix with the maximum eigenvalue that does not exceed 1 (Table 1). However, the form of the asymptotic covariance matrix of the normalized within-stratum imbalances following minimization remains unknown in the general case. This hinders the study of the inferential properties of the log-rank and score tests following minimization that were established for other types of covariate-adaptive procedures by Ye and Shao (2020). Zhao et al. (2022) proved that the asymptotic covariance matrix of the normalized within-stratum imbalances can be estimated through the consistent Monte Carlo estimator when the sample is large enough to estimate the prevalence of the strata in the study population. The authors pointed out that this approach could be considered a non-parametric bootstrap (Shao and Tu (2012)). They also stated that the direct computation of the asymptotic covariance matrix is challenging, because it requires solving Poisson equations associated with an induced Markov chain. Very importantly, they proved the joint asymptotic normality of the within-stratum normalized imbalances.

For the particular case of equal strata prevalence, Johnson, Gekhtman and Kuznetsova (2024) provided the empirical formula for the correlation matrix for the normalized within-stratum treatment imbalances. They estimated the variance for the normalized within-stratum treatment imbal-

Statistica Sinica: Newly accepted Paper

ances through simulations for 2 to 7 factors and established that for large enough bias ($p > 0.8$), the maximum eigenvalue of the covariance matrix of the normalized within-stratum treatment imbalances does not exceed 1. Thus, under these conditions, minimization is a Type 2a procedure and the properties established by Ye and Shao (2020) hold. Johnson, Gekhtman and Kuznetsova (2024) also saw in limited simulations of the unequal prevalence minimization that with the high enough bias the maximum eigenvalue does not exceed 1. However, no explicit form of either the asymptotic covariance or the asymptotic correlation matrix was offered for this case.

In this paper, based on empirical observations supported by simulations results, we provide the explicit form of the asymptotic covariance matrix of the normalized within-stratum imbalances following minimization with any number of independent factors and unequal prevalence of the strata. This is a novel result. We conjecture that the asymptotic covariance matrix is proportional with the coefficient V to the conditional covariance matrix of a simpler probabilistic model that shares some of the features of minimization. This is supported by empirical evidence, but no theoretical proof is offered. The coefficient V depends on the bias of the biased coin used in minimization. It was observed that with high enough bias $V \leq 1$, and thus, per Johnson, Gekhtman and Kuznetsova (2024), such minimization belongs

Statistica Sinica: Newly accepted Paper

to Type 2a. In this case, the log-rank and robust score tests following minimization are valid or conservative.

In section 2 we provide the notations and background on minimization. In section 3 we describe the asymptotic covariance matrix for the normalized within-stratum imbalances in treatment assignments for the Pocock and Simon covariate-adaptive allocation with independent factors and unequal strata prevalence. In section 4 we supplement these results with the simulations of the Type I error and power for a hypothetical multi-regional study of overall survival analyzed using the correct model of the fully stratified log-rank test as well as partially stratified and unstratified log-rank tests. A discussion completes the paper.

2. Notations and background

2.1 Notations

Consider a population that has M known prognostic factors (baseline covariates) $k = 1, \ldots, M$, where factor k has $n_k > 1$ levels. Denote by $\boldsymbol{Z}$ the vector of these k baseline covariates. The population is broken into $M_s = n_1 \times \cdots \times n_M$ strata formed by the combinations of the factor levels; these strata correspond to the categories of $\boldsymbol{Z}$. Denote the stratum $\boldsymbol{z}$ as $\boldsymbol{z} = (i_1, \ldots, i_M)$, where $i_1, \ldots, i_M$ are the levels of the factors that formed

Statistica Sinica: Newly accepted Paper

the stratum $\boldsymbol{z}$. We will also use the notation where the strata are numbered from 1 to M_s.

Consider a clinical trial with N subjects from the described population randomized in a 1:1 ratio to Treatment 1 and Treatment 0. Denote by $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iM}),\ i = 1, \ldots, N$, the vector of baseline covariates of the i-th subject; this vector determines the stratum to which the i-th subject belongs.

Denote by N_z the size of the stratum $\boldsymbol{z}$; $\sum_{z=1}^{M_s} N_z = N$. Let $w_z^N = \frac{N_z}{N}$ be the fraction of subjects in the stratum $\boldsymbol{z}$, $\sum_{z=1}^{M_s} w_z^N = 1$. Suppose that subjects are sampled independently from the target population with the fraction of subjects in stratum $\boldsymbol{z}$ denoted by w_z, and thus for all z, $w_z^N \xrightarrow[N\to\infty]{P} w_z$, where $\sum_{z=1}^{M_s} w_z = 1$.

Denote by m_z the size of the stratum $\boldsymbol{z}$ after m allocations and by $D_m(\boldsymbol{z})$ the imbalance in treatment assignments within stratum $\boldsymbol{z}$, that is the difference in the number of subjects allocated to Treatment 1 and Treatment 0, after m allocations. Let us denote by $d(\boldsymbol{z})$ the imbalance within stratum $\boldsymbol{z}$ normalized by the square root of its size (normalized imbalance) after m allocations: $d_m(\boldsymbol{z}) = \frac{D_m(\boldsymbol{z})}{\sqrt{m_z}}$. Then

$$d_N(\boldsymbol{z}) = \frac{D_N(\boldsymbol{z})}{\sqrt{N_z}} = \frac{D_N(\boldsymbol{z})}{\sqrt{N}\sqrt{w_z^N}}$$

Denote by $D_m(j;h)$ the marginal imbalance in treatment assignments,

Statistica Sinica: Newly accepted Paper

that is the difference in the number of subjects allocated to Treatment 1 and Treatment 0 among the subjects with $i_j = h$, $h = 1, \ldots, n_j$, after m allocations. Then

$$D_m(j;h) = \sum_{z:i_j=h} D_m(z)$$

When the marginal imbalance at the end of randomization $D_N(j;h)$ is low, the two treatment arms have very similar percentages of subjects with $i_j = h$. Hu and Hu (2012) and Hu, Ye and Zhang (2024) consider a generalization of the Pocock and Simon minimization where the treatment arm imbalance in the number of subjects, all marginal imbalances, and all within-stratum imbalances are included in the calculation of the overall imbalance. However, in this paper we follow the classical Pocock and Simon covariate-adaptive procedure where only marginal imbalances are included in the calculation of the overall imbalance. The Pocock and Simon minimization chooses the treatment assignment for the next subject to minimize the marginal imbalances $D_m(j;h)$ but not the imbalances within individual strata $D_m(\boldsymbol{z})$. A popular way to define the overall imbalance after m allocations is as the sum of the squared marginal imbalances across all levels of all covariates (Freedman and White (1976)):

$$\text{Imb}(m) = \sum_{j=1}^{M} \sum_{h=1}^{n_j} [D_m(j;h)]^2 \tag{2.1}$$

When a new subject arrives for randomization, the treatment arm that

Statistica Sinica: Newly accepted Paper

would result in the lowest overall imbalance (2.1) after the allocation is assigned to the patient with probability $p > 1/2$ (bias). If both arms would result in the same overall imbalance, Treatment 1 is assigned with probability 0.5.

Denote by

$$\mathrm{Cov}(\boldsymbol{d}_N(z)) = [\mathrm{cov}(d_N(i_1,\dots,i_M), d_N(j_1,\dots,j_M)]\,,$$

$i_1 = 1,\dots,n_1;\dots;i_M = 1,\dots,n_M; j_1 = 1,\dots,n_1;\dots;j_M = 1,\dots,n_M$, the covariance matrix of the vector of normalized imbalances $\boldsymbol{d}_N(z)$, $z = 1,\dots,M_s$, and by

$$\mathrm{Cor}(\boldsymbol{d}_N(z)) = [\mathrm{cor}(\boldsymbol{d}_N(i_1,\dots,i_M), d_N(j_1,\dots,j_M)]\,,$$

$i_1 = 1,\dots,n_1;\dots;i_M = 1,\dots,n_M; j_1 = 1,\dots,n_1;\dots;j_M = 1,\dots,n_M$, the correlation matrix of the vector of normalized imbalances $\boldsymbol{d}_N(z)$, $z = 1,\dots,M_s$.

For the case when all strata z have the same prevalence $w_z = 1/(n_1 \times \cdots \times n_M)$ examined by Johnson, Gekhtman and Kuznetsova (2024), the following notations are used. In this case σ_z is the same for all strata and the asymptotic covariance matrix for the normalized within-stratum imbalances $\mathrm{Cov} = \sigma_z^2\,\mathrm{Cor}$, where Cor is the asymptotic correlation matrix of the normalized imbalances. The terms of the asymptotic correlation

Statistica Sinica: Newly accepted Paper

matrix Cor can be written as

$$\text{cor}((i_1,\ldots,i_M),(j_1,\ldots,j_M)) = c_I, I \subset \{1,\ldots,M\},$$

where I denotes the set of the factors that have the same levels for the two strata $(i_1,\ldots,i_M)$ and $(j_1,\ldots,j_M)$. In these notations, $c_I = 1$ for $I = \{1,\ldots,M\}$.

2.2 Background

Hu, Ye and Zhang (2024) showed that when individual strata are assigned 0 weight in the calculation of the overall imbalance (as in the classical version of the Pocock and Simon covariate-adaptive procedure considered in this paper), for every stratum z, $d_N(z) = \frac{D_N(z)}{\sqrt{N_z}} \xrightarrow[N\to\infty]{D} N(0,\sigma_z)$, where $\xrightarrow[N\to\infty]{D}$ denotes convergence in distribution as $N \to \infty$. This implies that the diagonal elements of the covariance matrix of the normalized imbalances $\text{Cov}(\boldsymbol{d}_N(z))$ converge to σ_z^2 as $N \to \infty$, however, the formula for σ_z is not known. As follows from Zhao et al. (2022), the vector of the normalized within-stratum imbalances converges in distribution to a multivariate normal vector. Hu, Ye and Zhang (2024) also showed that the marginal imbalances $D_m(j;h)$ are bounded in probability for all $j = 1,\ldots,M$, $h = 1,\ldots,n_j$.

The three classes (Type 1, Type2, and Type 2a) of covariate-adaptive allocation procedures introduced by Zhao (2021) and Johnson, Gekhtman

Statistica Sinica: Newly accepted Paper

Table 1: Three types of covariate-adaptive allocation procedures

Type	Definition
Type 1	$\frac{D_N(\boldsymbol{z})}{\sqrt{N_z}} = o_p(1)$ for every z
Type 2	$\frac{D_N(z)}{\sqrt{N_z}} \xrightarrow[N\to\infty]{D} N(0,\nu)$, where $\nu \leq 1$ and $D_N(z_1)$ and $D_N(z_2)$ are independent for all $z_1 \neq z_2$
Type 2a	$\boldsymbol{d}_N(z)$ converges in distribution to a normal vector with covariance Cov, where Cov is a non-negative definite $M_s \times M_s$ matrix with the maximum eigenvalue ≤ 1

and Kuznetsova (2024) are defined in Table 1 using the notations above.

Per Johnson, Gekhtman and Kuznetsova (2024), the maximum eigenvalue of the asymptotic covariance matrix of the within-stratum normalized imbalances not exceeding 1 is a sufficient condition for minimization being a Type 2a procedure.

While the form for the asymptotic covariance matrix for $\boldsymbol{d}_N(z)$ in general case is unknown, Johnson, Gekhtman and Kuznetsova (2024) described its form for the case of equal prevalence of the strata. Based on empirical evidence supported by simulations for > 2 factors and theoretical derivations for 2 factors, they concluded that for $I \neq \{1, \ldots, M\}$,

$$c_I = \frac{M - 1 - \sum_{i \in I} n_i}{\prod_{i=1}^{M} n_i - \sum_{i=1}^{M} n_i + M - 1} \tag{2.2}$$

Statistica Sinica: Newly accepted Paper

and the maximum eigenvalue of the correlation matrix Cor is

$$\lambda_{\max} = \frac{\prod_{i=1}^{M} n_i}{\prod_{i=1}^{M} n_i - \sum_{i=1}^{M} n_i + M - 1} \tag{2.3}$$

By estimating σ_z^2 through simulations in examples that included up to 7 factors, they established empirically that with a bias of ≥ 0.8, $\sigma_z^2 \lambda_{\max} < 1$, and thus under equal strata prevalence minimization belongs to the Type 2a class. To expand the search for the form of the asymptotic covariance matrix in the unequal strata prevalence case, in this paper the asymptotic covariance matrix is described for a general case of independent factors with unequal strata prevalence.

3. Asymptotic covariance matrix for the normalized within-stratum imbalances for the Pocock and Simon covariate-adaptive allocation with independent covariates

3.1 Supporting urn model

It was revealed through extensive simulations that for independent covariates the asymptotic covariance matrix of the within-stratum imbalances is proportional to the conditional covariance matrix of the imbalances that arises from a simple probabilistic model. Specifically, consider drawing N balls from an urn with a very large number of balls. Each ball in the urn is

Statistica Sinica: Newly accepted Paper

independently assigned M covariates $Z_k, k = 1, \ldots, M$, where covariate k has n_k levels, and one of the two treatments: Treatment 1 or Treatment 0. The covariates form $M_s = n_1 \times \cdots \times n_M$ strata. The prevalence of level i_k of the covariate Z_k in the urn is w_{k,i_k} and the prevalence of either treatment is 1/2. Thus, the probability of drawing a ball from stratum $z = (i_1, \ldots, i_M)$ is $w_z = \prod_{k=1}^{m} w_{k,i_k}$. Let us denote by $N_1(z)$ and $N_0(z)$ the number of balls drawn from stratum z that have treatment assignments of 1 and 0, respectively. Let us define the vector of within-stratum imbalances normalized by the square root of the expected stratum size: $Y_z = \frac{N_1(z) - N_0(z)}{\sqrt{N w_z}}$.

To capture the balancing property of minimization, we impose the following condition on marginal imbalances in the urn model:

Condition A. For all $k = 1, \ldots, M$, for all $j = 1, \ldots, n_k$ if $k = 1$ and all $j = 1, \ldots, n_k - 1$ if $k > 1$,

$$\sum_{z: z_k = j} (N_1(z) - N_0(z)) = 0$$

The set of $n_1 + \cdots + n_k - M + 1$ equations in Condition A means that the marginal imbalances are all 0.

Statement. We conjecture, as supported by simulation results in Tables 2-3 and Tables A1-A14 in the Supplementary Materials, that the asymptotic covariance matrix of the normalized within-stratum imbalances following minimization with independent covariates is proportional to the conditional

Statistica Sinica: Newly accepted Paper

covariance matrix of vector Y_z under Condition A, where the coefficient V depends on the bias p:

$$\text{Cov}(\boldsymbol{d}_N(z)) = V\, \text{Cov}(Y_z|A) \tag{3.4}$$

This is a purely empirical observation and while theoretical proof is not offered here, the available form of the covariance matrix might be helpful in eventually establishing a proof. Intuitively, the urn model with Condition A resembles minimization: with minimization, the covariates are drawn at random, while the treatment assignments are controlled only through minimizing the marginal imbalances (similar to condition A). The empirical evidence provided in this paper comes from simulations of minimization with 2 and 3 independent factors and unequal strata prevalence and simulations of equal prevalence minimization with 2 through 7 factors.

Since vector Y_z has an identity covariance matrix and the constraints are linear, conditioning is an orthogonal projection onto the subspace of arrays with zero one-way margins. The entries of the conditional covariance matrix $\text{Cov}(Y_z|A)$ can be explicitly expressed as follows. Let us denote $s(z) = \sum_{k=1}^{M} \frac{1}{w_{k,i_k}}$. For two strata $z_1 = (i_1, \ldots, i_M)$ and $z_2 = (j_1, \ldots, j_M)$, let us denote $g(z_1, z_2) = \sum_{k=1}^{M} \frac{I\{i_k = j_k\}}{w_{k,i_k}}$, where $I\{.\}$ is an indicator function.

Statistica Sinica: Newly accepted Paper

Then

$$\mathrm{Cov}(Y_{z_1}, Y_{z_2} \mid A) = I\{z_1 = z_2\} - \sqrt{w_{z_1} w_{z_2}}(g(z_1, z_2) - M + 1) \quad (3.5)$$

$$\mathrm{Var}(Y_{z_1} \mid A) = 1 - w_{z_1}(s(z_1) - M + 1) \quad (3.6)$$

$$\mathrm{Corr}(Y_{z_1}, Y_{z_2} \mid A) = \frac{I\{z_1 = z_2\} - \sqrt{w_{z_1} w_{z_2}}(g(z_1, z_2) - M + 1)}{\sqrt{(1 - w_{z_1}(s(z_1) - M + 1))(1 - w_{z_2}(s(z_2) - M + 1))}} \quad (3.7)$$

The derivation of (3.5) is provided in the Supplemenary Materials B1. The matrix $\mathrm{Cov}(Y_{z_1}, Y_{z_2}|A)$ can be expressed as a Kronecker (tensor) product of two matrices, each of which is a projection matrix. Consequently, its eigenvalues are products of the eigenvalues of these matrices, which are either 0 or 1.

Theorem 1. *The eigenvalues of* $\mathrm{Cov}(Y_{z_1}, Y_{z_2}|A)$ *are* 1 *(with multiplicity* $M_s - \sum_{k=1}^{M} n_k + M - 1$*) and* 0 *(with multiplicity* $\sum_{k=1}^{M} n_k - M + 1$*).*

Proof. The proof is provided in the Supplementary Materials B2.

It follows from Theorem 1 that as long as $V \leq 1$ in (3.4), the maximum eigenvalue of $\mathrm{Cov}(\boldsymbol{d}_N(z))$ does not exceed 1 and therefore, such minimization belongs to Type 2a allocation procedures.

Below we provide some empirical evidence that the covariance matrix of the vector of normalized imbalances $\mathrm{Cov}(\boldsymbol{d}_N(z))$ is proportional to $\mathrm{Cov}(Y_z|A)$ and show how V in (3.4) is related to the bias p.

Statistica Sinica: Newly accepted Paper

3.2 Covariance matrix of the normalized within-stratum imbalances with 2-factor minimization

For a particular case of $M = 2$, from (3.6), for a stratum $z = (l, i)$,

$$\mathrm{Var}(Y_{(l,i)})|A) = (1 - w_{1l})(1 - w_{2i}) \tag{3.8}$$

From (3.7), the conditional correlations are:

$$\begin{aligned}
&\mathrm{Corr}(Y_{(m,i)}, Y_{(m,j)} \mid A) = -\sqrt{\frac{w_{2i}w_{2j}}{(1 - w_{2i})(1 - w_{2j})}} && i \neq j \\
&\mathrm{Corr}(Y_{(l,i)}, Y_{(m,i)} \mid A) = -\sqrt{\frac{w_{1l}w_{1m}}{(1 - w_{1l})(1 - w_{1m})}} && l \neq m \\
&\mathrm{Corr}(Y_{(l,i)}, Y_{(m,j)} \mid A) = \sqrt{\frac{w_{2i}w_{2j}w_{1l}w_{1m}}{(1 - w_{2i})(1 - w_{2j})(1 - w_{1l})(1 - w_{1m})}} && l \neq m, i \neq j \\
&\mathrm{Corr}(Y_{(m,i)}, Y_{(m,i)} \mid A) = 1
\end{aligned} \tag{3.9}$$

Simulations of unequal prevalence minimization with 2 independent factors described below show that for all strata (l, i) the variance of the normalized within-stratum imbalances is proportional to the conditional variance (3.8):

$$\mathrm{Var}(l, i) = V(1 - w_{1l})(1 - w_{2i}) \tag{3.10}$$

Table 2 presents the estimates of the ratio

$$V = \frac{\mathrm{Var}(l, i)}{(1 - w_{1l})(1 - w_{2i})} \tag{3.11}$$

Statistica Sinica: Newly accepted Paper

within each stratum in the example of the multi-regional study simulations. The simulations mimic the study with 1:1 minimization that balances on two independent factors: region and risk group. Region has 5 levels with prevalence of 0.3, 0.2, 0.2, 0.15, and 0.15, respectively; the risk group has 4 levels with prevalence of 0.15, 0.2, 0.3, and 0.35, respectively. Three sets of 100,000 simulations with 12000 randomizations each were run for the values of bias of 0.9, 0.8, and 2/3.

As Table 2 shows, the ratios (3.11) are indeed very similar across the strata, supporting (3.10). The final estimate of V that uses the information from all strata was derived as the average of the estimates of V obtained within each stratum weighted proportionally to the strata prevalence. From Table 2, V increases as the bias decreases: it is below 1 for bias 0.9, is below 1 but very close to 1 with bias 0.8, and is slightly above 1 for bias of 2/3. This means that the maximum eigenvalue of the covariance matrix does not exceed 1 when bias is > 0.8 but is above 1 with bias 2/3. Similar conclusions follow from the results of the simulations of other examples of 2-factor minimization presented in Tables A1 – A4 in the Supplementary Materials.

Table 3 supports the formulae (3.9) for the correlations of the normalized

Statistica Sinica: Newly accepted Paper

within-stratum imbalances based on the same simulations of the minimization in the multi-regional trial example. The small maximum difference between the estimated correlation and the theoretical correlation of the normalized within-stratum imbalances across all pairs of strata agrees with the estimation error. The estimated and theoretical correlations for this and other examples of a 2-factor unequal strata prevalence minimization with independent factors provided in the Supplementary Materials (Tables A5–A10) also support formulae (3.9).

From (3.9) and (3.10), the covariances of the normalized within-stratum imbalances with the 2–factor minimization are:

$$
\begin{aligned}
\text{cov}((m,i),(m,j)) &= -V(1-w_{1m})\sqrt{w_{2i}w_{2j}}, \quad i \neq j \\
\text{cov}((l,i),(m,i)) &= -V(1-w_{2i})\sqrt{w_{1l}w_{im}}, \quad l \neq m \\
\text{cov}((l,i),(m,j)) &= V\sqrt{w_{2i}w_{2j}w_{1l}w_{1m}}, \quad l \neq m, i \neq j \\
\text{cov}((m,i),(m,i)) &= V(1-w_{1m})(1-w_{2i})
\end{aligned}
\tag{3.12}
$$

3.3 Covariance matrix of the normalized within-stratum imbalances with a 3-factor minimization

Table A11 in the Supplementary Materials summarizes the results of the simulations of 3-factor minimization that support relationship (3.4). The

Statistica Sinica: Newly accepted Paper

simulations used a bias of 0.9 and included 12 examples with 3 independent 3–level factors and unequal prevalence of the strata. These estimates in Table A11 were obtained using 10,000 simulations of minimization. Each simulation generated a randomization sequence for 42,000 subjects following three-factor minimization.

As Table A11 shows, the correlations provide a close match to the model correlations (3.7), with median absolute difference across the strata below 0.007 for all simulations. For each simulation, V was estimated as the weighted average of the stratum-specific ratios of the observed variance to the model variance (3.6) with weights proportional to the prevalence of the strata. For all simulations, the estimates of V were below 1, which means that the maximum eigenvalue of the covariance matrices in these examples is below 1.

3.4 Covariance matrix of the normalized within-stratum imbalances for minimization with equal prevalence of the strata

For equally prevalent strata, the conditional correlations (3.9) revert to formula (2.2). Tables A12–A14 present the estimates of V from (3.4) derived as the ratio of the estimated variances of the normalized within-stratum imbalances reported by Johnson, Gekhtman and Kuznetsova (2024) (Appendix

Statistica Sinica: Newly accepted Paper

1) for equal prevalence minimization with 2 to 7 factors and bias 0.9 over the conditional variance (3.4). These estimates are also presented based on the simulations with bias 0.8 (not reported in Johnson, Gekhtman and Kuznetsova (2024)), where the variance of the normalized within-stratum imbalances is estimated as the average across the strata of the stratum-specific variance estimates. Tables A12–A14 show that for $p = 0.9$ the estimate of V is very close to 1 and does not exceed 1. For a bias of 0.8, the estimate of V is mostly below 1, but in a few cases it exceeds 1 in the third digit (which may be attributed to simulation error).

The simulations with $p = 0.9$ and 0.8 across the range of the number of factors and factor levels show that as the number of strata increases, V (and thus, the maximum eigenvalue of $\mathrm{Cov}(\boldsymbol{d}_N(z))$) converges to 1 from below. That is observed in Johnson, Gekhtman and Kuznetsova (2024) (Appendix 1, tables A2–A4 with $p = 0.9$) and in Tables A12–A14 in the Supplementary Materials.

Statistica Sinica: Newly accepted Paper

4. Examining the properties of the log-rank test following minimization with unequal prevalence through simulations on the example of the multiregional clinical trial

In this section, the theoretical results are supplemented with the results of the simulations of the multi-regional study with 1:1 minimization that balances on two independent factors: region (with 5 levels of unequal prevalence) and risk group (with 4 levels of unequal prevalence), described in section 2. Here subject i has covariates $Z_i = (Z_{i1}, Z_{i2})$, $Z_{i1} = 1$ to 5, $Z_{i2} = 1$ to 4, $i = 1, \ldots, N$. Let $I_i = 0, 1$ be the treatment indicator for subject i; in this example, $I_i = 1$ is the experimental treatment indicator and $I_i = 0$ is the control treatment indicator.

The two treatment arms are compared in the overall survival (OS) at 0.025 Type I error level. The OS follows the proportional hazard model (Cox (1972, 1975)) with constant baseline hazard function and factors for region, risk group, and treatment. The OS for the i-th subject (in months)

Statistica Sinica: Newly accepted Paper

is exponentially distributed with the log hazard function

$$
\begin{aligned}
\ln(\lambda(t, Z_i, I_i)) = \ln(0.077016) &+ \ln(0.9)I(Z_{i1} = 2) + \ln(0.95)I(Z_{i1} = 3) \\
&+ \ln(0.8)I(Z_{i1} = 4) + \ln(0.85)I(Z_{i1} = 5) \\
&+ \ln(0.5)I(Z_{i2} = 2) + \ln(0.25)I(Z_{i2} = 3) \\
&+ \ln(0.125)I(Z_{i2} = 4) + \theta I_i
\end{aligned}
\tag{4.9}
$$

where $I(\cdot)$ is the indicator function. As we can see, region is a weak prognostic factor (the hazard ratio between the lowest risk level and the highest risk level is only 0.8, as might be seen in an oncology study), while the risk group is a very strong prognostic factor, with the hazard cut in half from level to level. The intercept $\ln(0.077016)$ in the log hazard function was chosen to provide a median OS of 9 months in the control arm for the highest risk stratum $z = (1, 1)$. Under the alternative hypothesis, the hazard ratio for treatment 1 vs. treatment 0 of 0.78, that would often be considered both realistic and clinically meaningful in an oncology trial, was assumed ($\theta = \ln(0.78)$). Under the null hypothesis $\theta = 0$. The study included 1,200 participants enrolled at a steady rate over 24 months and followed up until the 680 OS events are observed. With 680 events the study has approximately 90% power under the alternative with one-sided Type I error of 0.025 assuming proportional hazard.

Statistica Sinica: Newly accepted Paper

Six scenarios were investigated through simulations: for three values of the minimization bias: 0.9, 0.8, and 2/3, and for the null and alternative hypotheses. For each of the 6 scenarios, 5000 datasets with $N = 1200$ and independent region and risk group covariates were generated, after which subjects were randomized following minimization with the respective value of bias. The enrollment and survival times were generated for all subjects. The survival data followed (4.9) and were censored at the data cutoff time when 680 events were observed.

The survival data were then analyzed following the four approaches below:

1. The log-rank test stratified by region (5 levels) and risk group (4 levels) – the correct model.

2. The log-rank test with regions combined into 2 levels (first three levels with higher risk vs. the last 2 levels with lower risk), labeled "Region-pooled"

3. The log-rank test with region removed from stratification, labeled "Region-removed".

4. The unstratified log-rank test.

For each approach, the Type I error (under the null hypothesis) and the

Statistica Sinica: Newly accepted Paper

power under the alternative hypothesis were evaluated. The study design and the simulated data represent an example of a hypothetical oncology study with OS survival as a primary endpoint and a single efficacy analysis. The four approaches reflect typical choices the statistician faces when dealing with numerous strata (20 in this example). Shall all the strata be included in the analysis model or should the number of strata be reduced by pooling similar strata together or eliminating one of the factors (region, as the least prognostic factor) altogether? Finally, what will the impact of using the unstratified analysis be?

The results of the simulations are presented in Table 4. As expected per Ye and Shao (2020), when the correct analysis model, the log-rank test stratified by all strata used in the randomization, was used, the Type I error was preserved within the simulations error. Using the partially stratified log-rank test with the 5 region levels pooled into two ("Region-pooled" rows) as well as the partially stratified log-rank test with region removed from the analysis stratification ("Region-removed" rows), showed little if any impact on the Type I error. This is because of the weak prognostic value of the region and is consistent with the theory in Ye and Shao (2020). The unstratified analysis, however, is very conservative as a result of removal of a very strong prognostic factor (the risk group) from the analysis; this

Statistica Sinica: Newly accepted Paper

is also consistent with Ye and Shao (2020). The power numbers of the stratified and partially stratified analyses are similar, while the unstratified analysis has a substantially lower power – again, because of the removal of the highly prognostic factor from the analysis.

The sample size and the number of events in the simulated studies in Table 4 reflect the design of oncology trials in practice. However, with 20 strata and the expected size of the smallest stratum of 27 subjects, the setting might not be close enough to the asymptotic one. To explore the Type I error of the log-rank tests above in the near-asymptotic setting, the simulations were repeated for a hypothetical study where the number of subjects and the number of events at the data cutoff were increased 10-fold, to 12,000 subjects and 6800 events. In this setting, the expected size of the smallest stratum was 270 subjects. The enrollment duration was kept at 24 months and the log hazard was as described by (4.9) .

The Type I error results in Table 5 are similar to those in Table 4. Type I error was preserved within the simulations error for the fully stratified log-rank test, barely impacted by the pooling of the regions into two levels or removing region from the analysis stratification, and was very conservative for the unstratified log-rank test.

Statistica Sinica: Newly accepted Paper

5. Discussion

In this work the explicit form of the asymptotic covariance matrix for the normalized within-stratum imbalances for equal allocation two-arm minimization with independent covariates and unequal strata prevalence is provided based on empirical evidence. This matrix is important in evaluation of the validity of the log-rank and the robust score tests of survival data as follows from Ye and Shao (2020) and Johnson, Gekhtman and Kuznetsova (2024). It was conjectured that the covariance matrix is proportional with the coefficient V to the conditional covariance matrix of the normalized within-stratum imbalances from a simple urn model, where N balls that have M covariates and one of the two Treatments assigned independently, are drawn from an urn with a large number of balls. The condition is that all marginal imbalances are 0.

The simulations of minimization support that the coefficient V, and thus, the maximum eigenvalue of the asymptotic covariance matrix for minimization with independent factors does not exceed 1 for the values of bias of 0.8 or 0.9. From Johnson, Gekhtman and Kuznetsova (2024), this means that such minimization belongs to the Type 2a class of covariate-adaptive

Statistica Sinica: Newly accepted Paper

procedures (Table 1). Belonging to the Type 2a is a sufficient (but not a necessary) condition for the log-rank test and the robust score test to be valid or conservative for the misspecified model following such minimization. This was also illustrated in the simulations of the multi-regional clinical trial in overall survival with 2-factor minimization used to randomize the study subjects.

We also observed that for the bias of 2/3 the maximum eigenvalue of the asymptotic covariance matrix for the 2-factor minimization may exceed 1, which may lead to some inflation of the Type I error for certain misspecified models as explained in Johnson, Gekhtman and Kuznetsova (2024). Examples of the survival data when the 2-factor minimization with bias 2/3 leads to inflation of Type I error are provided by Zhao et al. (2022). Thus, for a study with time-to-event endpoints we would recommend to use bias at least as high as 0.8.

The explicit form of the asymptotic covariance matrix was provided only for independent factors. However, we would hypothesize that a similar urn model, where dependent covariates are assigned to the balls, will work for minimization with dependent factors. It is also possible, that the covariance matrix for the normalized within-stratum imbalances with minimization with equal allocation to $K > 2$ treatments can also be derived

Statistica Sinica: Newly accepted Paper

from a similar urn model. In this case, the asymptotic covariance matrix for the unequal allocation minimization obtained through mapping from equal allocation (Kuznetsova and Tymofyeyev (2012)) can also be derived. All these questions will be explored in future research. Additionally, while in our simulations we used the Freedman and White (1976) imbalance metric, we expect that the same covariance structure will be observed with other imbalance metrics.

Establishing the theoretical proof for the provided empirical form and explicit linking of the coefficient V to the bias p is a topic of future research. The theoretical path through solving Poisson equations associated with the Markov chain induced by minimization under random sampling from the specified population, outlined by Zhao et al. (2022) is challenging, but perhaps the closed form will provide helpful clues.

Johnson, Gekhtman and Kuznetsova (2024) generalized the robust stratified covariance estimator for the score test by Ye and Shao (2020) to Type 2a randomization procedures. They applied this robust estimator to minimization with equal prevalence of the factors, where the correlation matrix form was conjectured and the variance was estimated through Monte Carlo simulations. A similar robust estimator can be constructed in the unequal prevalence of strata case, using the estimated prevalence of the factors, the

Statistica Sinica: Newly accepted Paper

conjectured form of the asymptotic covariance matrix, and the estimate of coefficient V obtained through simulations. This approach that benefits from the knowledge of the form of the asymptotic covariance matrix may lead to a somewhat better performance of the test compared to the consistent estimator by Zhao et al. (2022), where the whole asymptotic covariance matrix is estimated. However, these two approaches will need to be compared through simulations. This can be explored further in future work .

Supplementary Materials

The Supplementary Materials document contains tables A1 – A14, the derivation of (3.5), and the proof of Theorem 1.

Acknowledgements

The authors are grateful to the anonymous reviewers and the Associate Editor for constructive recommendations that helped substantially improve the paper.

M.G.'s work was supported in part by the Merck & Co., Inc., Rahway, NJ, USA, BARDS grant "Demonstration that Type I error is preserved or reduced with the score test and the log-rank test in studies with Pocock

Statistica Sinica: Newly accepted Paper

and Simon covariate-adaptive randomization".

References

Baldi Antognini, A. and Giovagnoli, A. (2004). A new 'biased coin' design for the sequential allocation to two treatments. *Journal of the Royal Statistical Society* **53**(4), 651–664.

Barbachano, Y. and Coad D.S. (2013). Inference following designs which adjust for imbalances in prognostic factors. *Clinical Trials* **10**, 540–551.

Birkett, N .J. (1985). Adaptive allocation in randomized controlled trials. *Controlled Clinical Trials* **6**, 146–155.

Bugni,F.A., Canay I.A. and A. M. Shaikh. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association* **113**, 1784–1796.

Cox, D. R. (1975). Partial Likelihood. *Biometrika* **62**, 269–276.

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Ser. B* **34**, 187–220.

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

Forsythe, A. B. (1987). Validity and power of tests when groups have been balanced for prognostic factors. *Computational Statistics and Data Analysis* **5**, 193–200.

Freedman, L. S. and White S.J. (1976). On the Use of Pocock and Simon's Method for Balancing Treatment Numbers over Prognostic Factors in the Controlled Clinical Trial. *Biometrics* **32**(3), 691–694.

Statistica Sinica: Newly accepted Paper

Hu, F., Hu, Y., Ma, Z. and Rosenberger W.F. (2014). Adaptive randomization for balancing over covariates. *Wiley Interdisciplinary Reviews: Computational Statistics* **6**(4), 288–303.

Hu, F., Ye, X. and L. Zhang. (2023). Multi-arm covariate-adaptive randomization. *Science China Mathematics* **66**(1), 163–190.

Hu, Y. and Hu F. (2012). Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics* **40**, 1794–1815.

Johnson, V. P., Gekhtman, M. and Kuznetsova O.M. (2024). Validity of Tests for Time-to-Event Endpoints in Studies with the Pocock and Simon Covariate-Adaptive Randomization. *Statistics in Biopharmaceutical Research* **16**(4), 468–482.

Kahan, B. C. and Morris T.P. (2012). Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in Medicine* **31**, 328–340.

Kuznetsova, O. M. and Tymofyeyev Y. (2012). Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. *Statistics in Medicine* **31**, 701–723.

Lin, D. Y. and Wei L.J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**, 1074–1078.

Luo, X., Li, M., Xu, G. and Tu, D. (2016). Survival analysis following dynamic randomization. *Contemp Clinical Trials Communications* **3**, 39–47.

Ma, W., Hu, F. and Zhang, L. (2015). Testing hypotheses of covariate-adaptive randomized

Statistica Sinica: Newly accepted Paper

clinical trials. *Journal of the American Statistical Association* **110**(510), 669–680.

Liu, H., Tu, F. and Ma, W. (2022). Lasso-adjusted treatment effect estimation under covariate-adaptive randomization. *Biometrika* **asac036**. https://doi.org/10.1093/biomet/asac036

Ma, W., Tu, F. and Liu, H. (2022). Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective. *Statistics in Medicine* **41**, 5645–5661.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D.R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient: I. introduction and design. *British Journal of Cancer* **34**, 585–612.

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. *Statistical Science* **23**(3), 404–419.

Rosenberger, W. F. and Lachin, J.M. (2016). *Randomization in Clinical Trials: Theory and Practice* (2nd ed.). John Wiley & Sons, New York.

Scott, N. W., McPherson, G. C., Craig, R., Ramsay, C. R. and Campbell, M. K. (2002). The method of minimization for allocation to clinical trials: a review. *Controlled Clinical Trials* **23**, 662–674.

Shao, J. and Tu, D. (2012). *The jackknife and bootstrap*. Springer Science & Business Media.

Statistica Sinica: Newly accepted Paper

Shao, J. and Yu, X. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* **97**, 347–360.

Shao, J. and Yu, X. (2013). Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics* **69**, 960–969.

Shao, J. (2021). Inference after covariate-adaptive randomisation: aspects of methodology and theory. *Statistical Theory and Related Fields* **5**(3), 172–186.

Taves, D. (1974). Minimization: a new method of assigning subjects to treatment and control groups. *Clinical Pharmacology and Therapeutics* **15**, 443–453.

Taves, D. (2010). The use of minimization in clinical trials. *Contemporary Clinical Trials* **31**, 180–184.

Therneau, T. M. and Grambsch, P. M. (2000). *Modelling survival data: extending the Cox model*. Spring-Verlag, New York.

Therneau, T. M. (1973). How many stratification factors are 'too many' to use in a randomization plan. *Controlled Clinical Trials* **14**, 98–108.

Wei, L. J. (1978). An application of an urn model to the design of sequential controlled clinical trials. *Journal of American Statistical Association* **73**, 559–563.

Wei, L. J. (1978). The adaptive biased coin design for sequential experiments. *The Annals of Statistics* **6**, 92–100.

Weir, C. J. and Lees, K. R. (2003). Comparison of stratification and adaptive methods for

Statistica Sinica: Newly accepted Paper

treatment allocation in an acute stroke clinical trial. *Statistics in Medicine* **22**, 705–726.

Ye, T., Shao, J., Yi, Y and Zhao, Q. (2023). Toward Better Practice of Covariate Adjustment in Analyzing Randomized Clinical Trials. *Journal of the American Statistical Association* **118**(544), 2370–2382.

Ye, T., Shao, J. and Yi, Y. (2024). Covariate-adjusted log-rank test: guaranteed efficiency gain and universal applicability. *Biometrika* **111**(2), 691–705.

Ye, T. and Shao, J. (2020). Robust tests for treatment effect in survival analysis under covariate-adaptive randomization. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **82**(5), 1301–1323.

Ye, T. (2018). Testing hypotheses under covariate-adaptive randomisation and additive models. *Statistical Theory and Related Fields* **2**, 96–101.

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

Zhao, Z., Song, Y., Jiang, W and Tu, D. (2022). Consistent covariances estimation for stratum imbalances under marginal design for covariate-adaptive randomization. *Scandinavian Journal of Statistics* **51**(2), 861–890.

First author affiliation: Merck & Co., Inc., Rahway, NJ, USA

E-mail: olga_kuznetsova@merck.com

Statistica Sinica: Newly accepted Paper

Table 2: Estimates of the coefficient V in the multi-regional study example with the bias of 9/10, 8/10, and 2/3

Stratum	Bias = 9/10	Bias = 8/10	Bias = 2/3
11	0.986	1.002	1.010
12	0.983	0.995	1.016
13	0.977	1.003	1.009
14	0.979	1.002	1.021
21	0.986	0.996	1.020
22	0.976	0.992	1.008
23	0.972	0.998	1.015
24	0.973	0.989	1.020
31	0.975	0.992	1.017
32	0.985	0.998	1.019
33	0.972	0.991	1.022
34	0.974	0.991	1.023
41	0.993	0.997	1.013
42	0.990	0.994	1.023
43	0.987	0.996	1.014
44	0.980	1.000	1.018
51	0.991	0.995	1.015
52	0.984	0.996	1.014
53	0.979	0.997	1.017
54	0.981	0.999	1.011
Weighted estimate*	0.980	0.997	1.017

*Represents the weighted average estimate across all strata weighted proportionally to the strata prevalence.

Statistica Sinica: Newly accepted Paper

Table 3: Maximum difference between the estimated correlation and the theoretical correlation of the normalized within-stratum imbalances across all pairs of strata in the multi-regional study example with the bias of 9/10, 8/10, and 2/3

	Bias = 9/10	Bias = 8/10	Bias = 2/3
Maximum Difference	0.00645	0.00999	0.01072

Statistica Sinica: Newly accepted Paper

Table 4: Type I error estimate, power estimate, and the Monte Carlo standard errors of the estimates in the example of a multiregional trial with 1,200 participants randomized following a two-factor minimization with bias 9/10, 8/10, and 2/3

	Type I Error		Power	
	Estimate	Standard Error	Estimate	Standard Error
Bias $p = 9/10$				
Fully stratified	0.0236	0.0021	0.8924	0.0044
Region-pooled	0.0224	0.0021	0.8996	0.0043
Region-removed	0.0226	0.0021	0.9012	0.0042
Unstratified	0.0096	0.0014	0.7828	0.0058
Bias $p = 8/10$				
Fully stratified	0.0236	0.0021	0.8892	0.0044
Region-pooled	0.0236	0.0021	0.8958	0.0043
Region-removed	0.0230	0.0021	0.8990	0.0043
Unstratified	0.0120	0.0015	0.7790	0.0059
Bias $p = 2/3$				
Fully stratified	0.0246	0.0022	0.8748	0.0047
Region-pooled	0.0246	0.0022	0.8800	0.0046
Region-removed	0.0256	0.0022	0.8850	0.0045
Unstratified	0.0112	0.0015	0.7658	0.0060

Statistica Sinica: Newly accepted Paper

Table 5: Type I error estimate and the Monte Carlo standard error of the estimate under the near-asymptotic conditions in the example of a large multiregional trial with 12,000 participants randomized following a two-factor minimization with bias 9/10, 8/10, and 2/3

	Estimate	Standard Error
Bias $p = 9/10$		
Fully stratified	0.0254	0.0022
Region-pooled	0.0250	0.0022
Region-removed	0.0242	0.0022
Unstratified	0.0102	0.0014
Bias $p = 8/10$		
Fully stratified	0.0262	0.0023
Region-pooled	0.0260	0.0023
Region-removed	0.0260	0.0023
Unstratified	0.0132	0.0016
Bias $p = 2/3$		
Fully stratified	0.0270	0.0023
Region-pooled	0.0268	0.0023
Region-removed	0.0276	0.0023
Unstratified	0.0136	0.0016