A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials

Uttam Bandyopadhyay[1], Atanu Biswas[2] and Rahul Bhattacharya[3,*,†]

[1]*Department of Statistics, University of Calcutta, 35 Ballygunge Circular Road, Kolkata 700019, India*
[2]*Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India*
[3]*Department of Statistics, Asutosh College, 92 S.P. Mukherjee Road, Kolkata 700026, India*

SUMMARY

In the present work, we develop a two-stage allocation rule for binary response using the log-odds ratio within the Bayesian framework allowing the current allocation to depend on the covariate value of the current subject. We study, both numerically and theoretically, several exact and limiting properties of this design. The applicability of the proposed methodology is illustrated by using some data set. We compare this rule with some of the existing rules by computing various performance measures.

KEY WORDS: covariate; log-odds ratio; optimality; two-stage allocation

1. INTRODUCTION

Suppose patients, arriving sequentially for clinical trial, are to be allocated to one of the two competing treatments. Adaptive designs, utilizing earlier response history to determine the next assignment, are very popular in the recent years to skew the allocation proportion to the eventually best performing treatment. But this requires instantaneous responses, which perhaps is rarely met in any real clinical trial. Moreover, in any clinical trial, the resources are limited and the trials are generally time consuming. So it is not possible to accrue unlimited subjects. Therefore, a target sample size should be set in the planning stage. Two-stage designs are recommended for such purposes. The basic goal of such designs is to allocate a larger number of patients to the better-treatment arm and to get some ethical benefit out of that. In Coad [1], a total of N patients were allowed with the first stage as the initial randomization stage having m patients at each treatment arm and, in the second stage, the remaining $(N - 2m)$ patients were treated exclusively

by the treatment doing better in the primary stage of the experiment. Thus, a two-stage design is a particular adaptive design where the adaptation is carried out only once at the end of the first stage and all the remaining allocations are carried out accordingly. But the patients in the second stage are allocated without any randomization and hence results in selection bias. Since randomization is an important concern in clinical trials (see [2, 3]), the proposed design of Coad [1, 4] cannot be recommended for a real clinical trial from the clinician's point of view. Moreover, in any clinical trial, there are covariates of interest besides the treatment effect. Some covariates are known in advance influencing the outcome of a patient. For instance, a female patient may have a very different response to a treatment than a male patient. Therefore, it may be inappropriate to use data on the former patient's response to determine the latter patient's allocation when covariates are present. So a meaningful allocation design should take into account the current patient's covariate information. Unfortunately, the possibility of the presence of covariates is ignored in Coad [1]. The history of incorporating covariates in clinical trials is not too long. Some attempts with the presence of covariates can be found in Begg and Iglewicz [5], where optimum design theory is used to suggest a deterministic allocation criterion. Inclusion of covariates can also be found in the works of Atkinson [6] Bandyopadhyay and Biswas [7], Atkinson and Biswas [8, 9]. All these works (except Bandyopadhyay and Biswas [7]) provide a biased coin-type procedure utilizing optimum design theory. But these works involve a linear model with continuous responses, whereas many clinical trials deal with binary outcomes. Moreover, these works are adaptive in nature requiring instantaneous responses.

In the present work, we consider the case of two competing treatments with binary outcomes in the presence of prognostic factors. The proposed methodology is discussed in Section 2. The rest of the work is organized as follows. Section 3 provides the estimation procedure of true log-odds ratio with some related asymptotic result. Section 4 deals with some performance measures, exact and limiting, of the design. The exact properties include the overall proportion of allocation to the better treatment, its standard deviation, allocation proportion to the better treatment for a fixed covariate level along with the standard error whereas overall failure proportion is assessed asymptotically. Numerical computations, exploring the benefit of covariate information, are exhibited in Section 5. Choice of design parameters is discussed in Section 6. Section 7 includes a data study to illustrate the possible benefit of the proposed procedure. Finally, Section 8 ends up with a discussion of related issues.

2. PROPOSED ALLOCATION DESIGN

Patients arrive sequentially in a clinical trial and are to be allocated to one of the two competing treatment arms. For each incoming subject, a fixed number of covariate values are recorded. Let Y_{ki} denote the response of the ith subject for the treatment k, $i \geqslant 1$, $k = A, B$. We assume that

$$P(Y_{ki} = 1) = \pi_{ki} = G(\alpha_k + \boldsymbol{\beta}_k^{\mathrm{T}} \mathbf{x})$$

where G is some continuous symmetric distribution function, α_k is the effect corresponding to the kth treatment, and $\boldsymbol{\beta}_k$ is the effect of the covariate vector $\mathbf{x}$ for the kth treatment, $k = A, B$. In general, the treatment difference, for covariate $\mathbf{x}$, is denoted by

$$\Delta = \alpha_A - \alpha_B + (\boldsymbol{\beta}_A - \boldsymbol{\beta}_B)^{\mathrm{T}} \mathbf{x}$$

For simplicity, we consider $\boldsymbol{\beta}_A = \boldsymbol{\beta}_B$, and the treatment difference is $\Delta = \alpha_A - \alpha_B$. In the subsequent part of the paper we develop with the assumption of $\boldsymbol{\beta}_A = \boldsymbol{\beta}_B$. However, a similar analysis can be easily carried out for the case $\boldsymbol{\beta}_A \neq \boldsymbol{\beta}_B$. For practical purposes, one can either decide using prior idea, or a simple test of $\boldsymbol{\beta}_A = \boldsymbol{\beta}_B$ can be carried out based on the first-stage data.

Our main objective is to make a valid inference on $\Delta = \alpha_A - \alpha_B$. The allocation procedure is as follows: let N be a prefixed number of patients in the trial. The first stage involves $2m$ patients. These $2m$ patients are randomly allocated in a 50:50 way between the two competing treatments. The second stage involves $N - 2m$ patients. Any second-stage patient with covariate vector $\mathbf{x}$ is treated with probability $p = p(\mathbf{x})$, a function of the covariate of that patient, and also all the available data of the first-stage patients. As randomization is a basic requirement of a clinical trial (see [2, 3]), we, here, use an optimum design theory with some randomization device to determine p appropriately. Suppose δ_{ki} is the indicator of allocation with values 1 or 0 as the ith patient is treated by the kth treatment or not. From the allocation scheme, δ_{Ai} $(= 1 - \delta_{Bi})$'s are independent Bernoulli($\frac{1}{2}$) random variables, $i = 1, \ldots, 2m$, and conditionally, given the first-stage allocation-and-response-and-covariate data, δ_{Ai} follows Bernoulli($p(\mathbf{x}_i)$) distribution, where $\mathbf{x}_i$ is the covariate vector of the ith patient, and $i \geqslant 2m + 1$.

The likelihood function based on the $2m$ observations of the first stage and a single observation of the second stage (say the jth observation, $j \geqslant 2m + 1$) is proportional to

$$L_0 = \prod_{k=A,B} \prod_{i \in S_j} \{\pi_{ki}^{y_{ki}} (1 - \pi_{ki})^{1 - y_{ki}}\}^{\delta_{ki}}$$

where

$$S_j = \{1, 2, \ldots, 2m, j\}$$

Using $G(x) = e^x/(1 + e^x)$, the distribution function of a standard logistic distribution, we obtain the following log likelihood (apart from a constant)

$$l_0(\boldsymbol{\phi}) = \sum_{k=A,B} \sum_{i \in S_j} \delta_{ki} \{y_{ki}(\alpha_k + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i) - \log(1 + \exp(\alpha_k + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i))\} \tag{1}$$

where

$$\boldsymbol{\phi}^{\mathrm{T}} = (\alpha_A, \alpha_B, \boldsymbol{\beta}^{\mathrm{T}})$$

Now, to develop an elegant allocation rule, we compute Fisher's information matrix from (1) as

$$I_0 = \begin{pmatrix} \frac{1}{2}\mathbf{1}^{\mathrm{T}} \Lambda_A \mathbf{1} & 0 & \frac{1}{2}\mathbf{1}^{\mathrm{T}} \Lambda_A X^{\mathrm{T}} \\ & \frac{1}{2}\mathbf{1}^{\mathrm{T}} \Lambda_B \mathbf{1} & \frac{1}{2}\mathbf{1}^{\mathrm{T}} \Lambda_B X^{\mathrm{T}} \\ & & \frac{1}{2} X(\Lambda_A + \Lambda_B) X^{\mathrm{T}} \end{pmatrix}$$

$$+ \begin{pmatrix} p v_{Aj} & 0 & p v_{Aj} \mathbf{x}_j \\ & (1 - p) v_{Bj} & (1 - p) v_{Bj} \mathbf{x}_j \\ & & \{p v_{Aj} + (1 - p) v_{Bj}\} \mathbf{x}_j \mathbf{x}_j^{\mathrm{T}} \end{pmatrix}$$

where p is treated as a fixed quantity and $X = (\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_{2m})$ is an $r \times 2m$ matrix of covariates corresponding to the first-stage patients, $\Lambda_k = \text{Diag}(v_{k1}, v_{k2}, \ldots, v_{k2m})$ with $v_{kl} = \pi_{kl}(1 - \pi_{kl})$, $k = A, B$ and $l \geqslant 1$, and $\mathbf{1}$ is a $2m$ component vector with all elements unity. Since I_0 involves unknown parameters, it cannot be used to develop a meaningful allocation rule. So we replace I_0 by the expectation of I_0 with respect to some prior distribution ξ of $\boldsymbol{\phi}$. Denoting $E^{\xi}(v_{kj}) = \rho_{kj}$ and $\Phi_k = \text{Diag}(\rho_{k1}, \rho_{k2}, \ldots, \rho_{k2m})$, we observe that the structure of I_{ξ} is the same as that of I_0 except that, Λ_k is replaced by Φ_k and v_{kj} replaced by ρ_{kj}. We then suggest to determine the unknown p by maximizing the following utility function:

$$U(p) = \log |I_{\xi}| - \eta \left\{ p \log \left(\frac{p}{\pi_A} \right) + (1 - p) \log \left(\frac{1 - p}{1 - \pi_A} \right) \right\}$$

where π_A is obtained using the first-stage data such that

$$\pi_A = J \left(\frac{\hat{\alpha}_A - \hat{\alpha}_B}{T_0} \right)$$

with $\hat{\alpha}_A - \hat{\alpha}_B$ as the estimate of the treatment difference (see Section 3), J is a distribution function of a symmetric random variable such that $J(x) + J(-x) = 1$ for all x, and $T_0 \in (0, \infty)$ is a tuning constant. Here, η is the balancing factor between the optimal allocation and ethical allocation. For $\eta = 0$, optimum value of p is obtained by maximizing $\log |I_{\xi}|$, which gives the *piecewise D-optimal design* (see [10]). Again, $\eta \to \infty$ implies that the optimal value of p is obtained by maximizing

$$- \left\{ p \log \left(\frac{p}{\pi_A} \right) + (1 - p) \log \left(\frac{1 - p}{1 - \pi_A} \right) \right\}$$

with respect to p, which gives

$$p = \pi_A$$

the *ethical allocation* suggested by Bandyopadhyay and Biswas [7] for normally distributed responses. For any intermediate η, the optimal value of p for the jth patient, $j \geqslant 2m + 1$, with covariate vector $\mathbf{x}_j$, is obtained by maximizing $U(p)$ with respect to p. Let it be denoted by $p_m^*(\mathbf{x}_j)$ or simply p_{mj}^*. For any second-stage patient, we use this optimum choice of p by replacing $\mathbf{x}_j$ by the corresponding covariate vector.

Clearly, given the first-stage data and covariate vector corresponding to the jth patient, δ_{Aj} follows a Bernoulli(p_{mj}^*) distribution, $j = 2m+1, \ldots, N$, and is independent of $\delta_{A2m+1}, \ldots, \delta_{AN}$.

3. ESTIMATION PROCEDURE AND RELATED ASYMPTOTICS

Under the present formulation, we have

$$E(Y_{ki}) = \frac{\exp(\alpha_k + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i)}{1 + \exp(\alpha_k + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i)}, \quad k = A, B$$

with $\mathbf{x}_i$ as the covariate vector of the ith patient. Then the treatment difference can be measured by the quantity

$$\Delta = \alpha_A - \alpha_B$$

which is nothing but the log-odds ratio (see Newman [11]). For estimating Δ, we like to use the maximum-likelihood method. After implementing the proposed two-stage procedure, the likelihood function based on the first-stage sample is proportional to

$$L_1 = \prod_{k=A,B} \prod_{i=1}^{2m} \{\pi_{ki}^{y_{ki}} (1-\pi_{ki})^{1-y_{ki}}\}^{\delta_{ki}}$$

and, given the first-stage data, the likelihood based on the second-stage data is proportional to

$$L_{2|1} = \prod_{k=A,B} \prod_{j=2m+1}^{N} \{\pi_{kj}^{y_{kj}} (1-\pi_{kj})^{1-y_{kj}}\}^{\delta_{kj}}$$

We then use

$$L_W = L_1 \times L_{2|1}$$

as our working likelihood (see Cheung and Thall [12]). This gives, apart from a constant,

$$l(\boldsymbol{\phi}) = \log L_W = \sum_{k=A,B} \sum_{i=1}^{N} \delta_{ki} \{y_{ki}(\alpha_k + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x_i}) - \log(1 + e^{\alpha_k + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x_i}})\}$$

Then, $\widehat{\boldsymbol{\phi}}$, the maximum-likelihood estimate of $\boldsymbol{\phi}$, is obtained by maximizing $l(\boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$.

For the rest of the development we assume only a single covariate x with s categories $0, 1, \ldots, s-1$. Now, to establish some related asymptotics, we need the following conditions:

CI: For each N, there exists $m = m(N)$ such that, as $N \to \infty$

$$m \to \infty \quad \text{but } m/N \to \theta \in (0, 1/2)$$

CII: As $n \to \infty$, for any fixed $y = 0, 1, \ldots, s-1$

$$\frac{1}{n} \sum_{i=1}^{n} I_{[x_i = y]} \to w_y$$

with $\sum_{y=0}^{s-1} w_y = 1$.

Condition CII in turn implies, for any k, as $n \to \infty$

(a) $\frac{1}{n} \sum_{i=1}^{n} \pi_{ki} \to \sum_{j=0}^{s-1} w_j \frac{e^{\alpha_k + \beta_j}}{1 + e^{\alpha_k + \beta_j}}$,

(b) $\frac{1}{n} \sum_{i=1}^{n} v_{ki} \to \sum_{j=0}^{s-1} w_j \frac{e^{\alpha_k + \beta_j}}{(1 + e^{\alpha_k + \beta_j})^2}$.

Note that condition CII and the observation in (b) together with dominated convergence theorem imply

$$\frac{1}{n} \sum_{i=1}^{n} \rho_{ki} \to \rho_k$$

Now we have the following theorem.

Theorem 3.1
Let $p_m^*(x)$ be the optimum value of p corresponding to x, the covariate of a second stage patient. Then, with probability one,

$$\lim_{N\to\infty} p_m^*(x) = \pi_A^* = J\left(\frac{\alpha_A - \alpha_B}{T_0}\right)$$

The proof of this theorem relies on evaluating $|I_\xi|$ and use of conditions CI, CII and the related observations, together with the properties of the first-stage likelihood function.

If we ignore the covariate from the model, we have $E(Y_{Ai}) = e^{\alpha_A}/(1+e^{\alpha_A})$ and $E(Y_{Bi}) = e^{\alpha_B}/(1+e^{\alpha_B})$. A two-stage design can be carried out in a similar, but much simpler way. We compare the performance characteristics of this design with the design proposed in Section 2. The results are given in Section 5. It is observed that the introduction of covariate provides different results. The allocations depend on the way in which the covariates influence the responses.

4. SOME PERFORMANCE CHARACTERISTICS

We consider the following performance characteristics of the proposed procedure and the competitor.

4.1. Proportion of allocation

Proportion of allocation to treatment A, denoted by τ_{AN}, measures the ethical gain. Note that $E(\tau_{AN}) = \frac{1}{2}$ for equal allocation. We have the following theorem.

Theorem 4.1
As $N \to \infty$,

$$\tau_{AN} = \frac{n_{AN}}{N} \to \tau_A = \theta + (1-2\theta)\pi_A^*$$

almost surely.

Note: This limiting proportion of allocation is greater than $\frac{1}{2}$ if A is the better treatment.

4.2. Covariate imbalance

In any randomized study the objective should be to equalize the distribution of important covariates within each treatment regimen so as to minimize the biases due to covariate imbalance. Any imbalance in the numbers randomized to each treatment arm within each covariate strata may bias the results of the study. Therefore, it seems sensible to judge any allocation rule in the light of covariate imbalance to ensure equal distribution of the covariate within each treatment group.

To be specific, consider a randomized clinical trial with categorized covariate. Then the allocation proportion corresponding to a fixed covariate level can be used to judge any covariate imbalance. Then, with $I_{[.]}$ as the indicator function, $M_{kN}(\mathbf{y}) = \sum_{i=1}^{N} \delta_{ki} I_{[\mathbf{x_i}=\mathbf{y}]}$ denotes the number of patients assigned to treatment k for a given covariate level $\mathbf{y}$, and $M_N(\mathbf{y}) = \sum_{i=1}^{N} I_{[\mathbf{x_i}=\mathbf{y}]}$ indicates the total number of subjects with covariate level $\mathbf{y}$. Therefore, any covariate imbalance can be depicted from the conditional allocation proportion $\hat{\pi}_k(\mathbf{y}) = M_{kN}(\mathbf{y})/M_N(\mathbf{y})$.

The preceding development is no longer valid for continuous covariate vector. However, a suitable categorization of the covariates can help us to continue with the current formulation. But then $\widehat{\pi}_k(\mathbf{y})$ can only serve as a rough reference statistic. For a single categorical covariate we have the following theorem.

Theorem 4.2
If $w_j = 1/s$ for all j, then for a single covariate with s categories,

$$\widehat{\pi}_k(\mathbf{y}) \to \theta + (1-2\theta)\pi_k^*$$

almost surely as $N \to \infty$.

Proof
Writing

$$\widehat{\pi}_k(\mathbf{y}) = \frac{M_{kN}(\mathbf{y})}{N}\left\{\frac{M_N(\mathbf{y})}{N}\right\}^{-1}$$

the result follows from Theorem 3.1 together with the assumption of the theorem. □

Note: The procedure is covariate balanced in the limit. This result can also be extended in an obvious manner to incorporate more than one covariate.

4.3. Proportion of treatment failures

Another important consideration in any clinical trial is the minimization of expected treatment failures. This ensures the benefit from the better treatment for a comparatively larger number of subjects. Under the present formulation the quantity $F_N = (1/N)\sum_{k=A,B}\sum_{i=1}^{i=N}\delta_{ki}(1-y_{ki})$ can be used to have an idea about the overall failure proportion. Then we get the following theorem.

Theorem 4.3
Under the assumption of Theorem 4.2, for a single covariate with s categories $0, 1, \ldots, s-1$,

$$F_N \to \sum_{k=A,B}\{\theta + (1-2\theta)\pi_k^*\}(1-\pi_k(s))$$

almost surely as $N \to \infty$, where

$$\pi_k(s) = \frac{1}{s}\sum_{j=0}^{s-1}\frac{e^{\alpha_k+\beta j}}{1+e^{\alpha_k+\beta j}}$$

Proof
The proof follows exactly in the same direction as in the proof of Theorem 4.1 with the following additional observations:

(i) y_{ki} are independent *Bernoulli*(π_{ki}), $i \geqslant 1$.
(ii) For any fixed k, δ_{ki}'s are i.i.d. Bernoulli$(\frac{1}{2})$, $i = 1, \ldots, 2m$.
(iii) Given the first-stage data, δ_{Ai}'s are Bernoulli(p_{mi}^*), $i = 2m+1, \ldots, N$.
(iv) Given the first-stage data, the second-stage responses are independently distributed. □

5. NUMERICAL COMPUTATIONS

In this section we carry out a detailed simulation to illustrate the applicability of the proposed procedure. The simulation is carried out using 5000 repetitions in S-Plus with only one categorical covariate having four categories, namely, 0, 1, 2 and 3. Here, covariates are generated in such a way that, in the long run, we have the same frequency for all the four categories. We always consider treatment A to be better or equivalent to treatment B. So we take $\alpha_A \geqslant \alpha_B$, and set $\beta = 2$. The only difficulty is the appropriate choice of the prior distribution. We have considered a bivariate normal prior with correlation coefficient ρ for each of the pairs (α_A, β) and (α_B, β). We choose $\rho = 0.9$ and equal variances $\sigma^2 = 10^5$ with zero means in each case to make the prior distributions close to the *diffused prior*. In practice any prior distribution close to the non-informative one can be chosen. But, if the choice of appropriate prior is fraught with difficulties, one may safely use the first-stage estimates instead of taking the averages over the prior distribution. This will not affect greatly the inherent nature of the allocation procedure. We report the results at $N = 80$ and the first-stage sample size $m = 10, 15$, for brevity, although we computed for several other choices of m and N. The computation is done at $T_0 = 5$ and $\eta = 0, 0.01, 1.0, \infty$. Here, m is an arbitrary fixed positive integer. In practice a choice of m, at least three times larger than the number of parameters involved, is recommended. However, a large value of m outweighs the importance of the second stage. Therefore, m should be cautiously chosen. Allocation proportions together with their standard errors (in the parentheses) under different categories are presented in Tables I–IV. Expected proportion of allocation to treatment A (and its standard deviation)

Table I. Allocation proportions to treatment A and standard errors when $\eta = 0$.

m	Δ	$\hat{\pi}_A(0)$	$\hat{\pi}_A(1)$	$\hat{\pi}_A(2)$	$\hat{\pi}_A(3)$	τ_{AN}^W	τ_{AN}^{WO}
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.111)	(0.114)	(0.115)	(0.115)	(0.056)	(0.056)
	0.40	0.500	0.500	0.500	0.500	0.500	0.500
		(0.111)	(0.113)	(0.116)	(0.114)	(0.055)	(0.056)
	0.80	0.500	0.500	0.500	0.500	0.500	0.500
10		(0.049)	(0.050)	(0.050)	(0.049)	(0.050)	(0.055)
	1.20	0.500	0.500	0.500	0.500	0.500	0.500
		(0.112)	(0.112)	(0.115)	(0.115)	(0.056)	(0.056)
	1.60	0.500	0.500	0.500	0.500	0.500	0.500
		(0.111)	(0.115)	(0.113)	(0.112)	(0.055)	(0.055)
	2.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.111)	(0.115)	(0.114)	(0.113)	(0.055)	(0.056)
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.112)	(0.113)	(0.114)	(0.115)	(0.055)	(0.055)
	0.40	0.500	0.500	0.500	0.500	0.500	0.500
		(0.111)	(0.114)	(0.113)	(0.115)	(0.054)	(0.057)
	0.80	0.500	0.500	0.500	0.500	0.500	0.500
15		(0.112)	(0.112)	(0.115)	(0.112)	(0.054)	(0.055)
	1.20	0.500	0.500	0.500	0.500	0.500	0.500
		(0.112)	(0.114)	(0.113)	(0.112)	(0.054)	(0.056)
	1.60	0.500	0.500	0.500	0.500	0.500	0.500
		(0.116)	(0.113)	(0.112)	(0.112)	(0.055)	(0.056)
	2.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.112)	(0.111)	(0.115)	(0.114)	(0.055)	(0.056)

Table II. Allocation proportions to treatment A and standard errors when $\eta = 0.01$.

m	Δ	$\hat{\pi}_A(0)$	$\hat{\pi}_A(1)$	$\hat{\pi}_A(2)$	$\hat{\pi}_A(3)$	τ_{AN}^{W}	τ_{AN}^{WO}
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.274)	(0.275)	(0.280)	(0.286)	(0.266)	(0.128)
	0.40	0.545	0.545	0.548	0.550	0.545	0.534
		(0.277)	(0.274)	(0.282)	(0.282)	(0.267)	(0.156)
	0.80	0.579	0.579	0.578	0.580	0.580	0.539
10		(0.270)	(0.266)	(0.276)	(0.278)	(0.260)	(0.177)
	1.20	0.624	0.627	0.631	0.621	0.627	0.626
		(0.254)	(0.260)	(0.272)	(0.252)	(0.247)	(0.191)
	1.60	0.633	0.630	0.637	0.632	0.638	0.675
		(0.260)	(0.262)	(0.269)	(0.261)	(0.259)	(0.195)
	2.00	0.653	0.651	0.657	0.652	0.668	0.722
		(0.264)	(0.264)	(0.268)	(0.260)	(0.266)	(0.185)
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.207)	(0.205)	(0.214)	(0.213)	(0.189)	(0.080)
	0.40	0.539	0.538	0.542	0.541	0.541	0.525
		(0.212)	(0.211)	(0.219)	(0.216)	(0.195)	(0.089)
	0.80	0.573	0.574	0.576	0.574	0.575	0.552
15		(0.213)	(0.209)	(0.218)	(0.221)	(0.196)	(0.104)
	1.20	0.610	0.611	0.612	0.611	0.611	0.589
		(0.204)	(0.204)	(0.215)	(0.214)	(0.189)	(0.121)
	1.60	0.637	0.641	0.639	0.641	0.640	0.627
		(0.207)	(0.212)	(0.219)	(0.217)	(0.196)	(0.132)
	2.00	0.663	0.669	0.667	0.668	0.667	0.658
		(0.195)	(0.201)	(0.208)	(0.203)	(0.183)	(0.134)

is computed under both the scenarios—with (denoted as τ_{AN}^{W}) or without covariates (denoted as τ_{AN}^{WO}). We have simulated the proportions of treatment failures for various choices of parameters, but for brevity we provide only the asymptotic study for $\theta = 0.1$ and 0.2, when $\alpha_B = 1$, $\beta = 2.0$, $T_0 = 5$. These are shown in Figure 1. We observe that allocation proportions are always greater than 50% and, for a fixed configuration of the other parameters, the proportion gradually increases as the true treatment difference increases. We also report larger allocation proportions to the better treatment with smaller failure rate for the proposed design than the design ignoring covariate information. It is also observed that the proposed two-stage design assigns more subjects to the better treatment for a given covariate level. From the figures it is easily concluded that the consideration of covariate information amounts to experience much lower failure rate for the selected choices of first-stage sampling fractions.

6. CHOICE OF DESIGN PARAMETERS

The proposed design involves several design parameters. Although there can be different combinations of these parameters which are applicable in practice, some guideline in this context can ease the application.

Table III. Allocation proportions to treatment A and standard errors when $\eta = 1.0$.

m	Δ	$\hat{\pi}_A(0)$	$\hat{\pi}_A(1)$	$\hat{\pi}_A(2)$	$\hat{\pi}_A(3)$	τ_{AN}^{W}	τ_{AN}^{WO}
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.280)	(0.274)	(0.284)	(0.283)	(0.267)	(0.137)
	0.40	0.546	0.545	0.544	0.547	0.546	0.539
		(0.272)	(0.274)	(0.284)	(0.283)	(0.267)	(0.158)
	0.80	0.596	0.597	0.604	0.601	0.600	0.583
10		(0.273)	(0.270)	(0.284)	(0.282)	(0.265)	(0.178)
	1.20	0.625	0.626	0.630	0.629	0.628	0.638
		(0.266)	(0.264)	(0.271)	(0.275)	(0.257)	(0.191)
	1.60	0.641	0.647	0.647	0.645	0.646	0.683
		(0.271)	(0.274)	(0.279)	(0.282)	(0.266)	(0.191)
	2.00	0.665	0.669	0.671	0.667	0.668	0.723
		(0.269)	(0.278)	(0.281)	(0.284)	(0.267)	(0.187)
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.207)	(0.202)	(0.214)	(0.213)	(0.188)	(0.085)
	0.40	0.542	0.544	0.544	0.543	0.543	0.525
		(0.212)	(0.211)	(0.220)	(0.224)	(0.197)	(0.090)
	0.80	0.577	0.576	0.579	0.580	0.579	0.555
15		(0.216)	(0.210)	(0.223)	(0.225)	(0.200)	(0.105)
	1.20	0.615	0.608	0.618	0.617	0.616	0.591
		(0.209)	(0.205)	(0.213)	(0.218)	(0.192)	(0.124)
	1.60	0.646	0.653	0.653	0.645	0.650	0.627
		(0.216)	(0.223)	(0.225)	(0.227)	(0.206)	(0.128)
	2.00	0.674	0.684	0.680	0.683	0.681	0.665
		(0.209)	(0.217)	(0.219)	(0.219)	(0.199)	(0.131)

The parameter η is a tradeoff between intuitive complete ethical allocation and complete optimal allocation. As observed in Atkinson and Biswas [8], a typically small value of η (say 0.01 or 0.03 or 0.1) may well balance between the optimal and ethical criteria. Of course these values are not optimal, and may change depending on the problem, distribution, number and nature of the covariate(s).

As discussed in Bandyopadhyay and Biswas [7], Φ, the cumulative distribution function (c.d.f.) of the standard normal distribution, is the most obvious choice of J. Of course one may choose J in several ways depending on how much weight he/she wants to give in the past data for any allocation. The choice may be a heavy-tailed distribution like Cauchy, light-tailed distribution like normal or double exponential. A light-tailed distribution implies that the experimenter is relying more on the available data for adaptation. As suggested by Biswas and Angers [13], a sequence of t-distributions (starting from heavy-tailed to light-tailed) can also be employed, where with the accumulation of data one may choose lighter tailed distribution as the choice of J. Of course the choice of J should be restricted among the symmetric distributions.

The choice of T_0 largely depends on the choice of J. As discussed in Bandyopadhyay and Biswas [7], for standard normal c.d.f. as J, $T_0 = 2$ or 3 are reasonably good choices.

Table IV. Allocation proportions to treatment A and standard errors when $\eta = \infty$.

m	Δ	$\hat{\pi}_A(0)$	$\hat{\pi}_A(1)$	$\hat{\pi}_A(2)$	$\hat{\pi}_A(3)$	τ_{AN}^{W}	τ_{AN}^{WO}
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.279)	(0.279)	(0.287)	(0.288)	(0.271)	(0.139)
	0.40	0.544	0.544	0.544	0.545	0.545	0.540
		(0.284)	(0.283)	(0.293)	(0.293)	(0.276)	(0.160)
	0.80	0.587	0.589	0.594	0.591	0.591	0.586
10		(0.279)	(0.277)	(0.275)	(0.288)	(0.271)	(0.181)
	1.20	0.632	0.635	0.638	0.638	0.637	0.637
		(0.269)	(0.264)	(0.274)	(0.271)	(0.258)	(0.194)
	1.60	0.656	0.657	0.655	0.657	0.659	0.688
		(0.263)	(0.263)	(0.267)	(0.270)	(0.268)	(0.193)
	2.00	0.673	0.678	0.677	0.677	0.678	0.729
		(0.273)	(0.270)	(0.271)	(0.269)	(0.270)	(0.191)
	0.00	0.500	0.500	0.500	0.500	0.500	0.500
		(0.209)	(0.208)	(0.218)	(0.216)	(0.193)	(0.083)
	0.40	0.537	0.541	0.539	0.540	0.540	0.527
		(0.218)	(0.213)	(0.221)	(0.224)	(0.200)	(0.095)
	0.80	0.572	0.573	0.577	0.578	0.576	0.559
15		(0.218)	(0.215)	(0.225)	(0.2319)	(0.204)	(0.111)
	1.20	0.618	0.616	0.620	0.623	0.619	0.598
		(0.213)	(0.209)	(0.218)	(0.224)	(0.198)	(0.123)
	1.60	0.648	0.639	0.621	0.638	0.648	0.636
		(0.213)	(0.217)	(0.213)	(0.203)	(0.199)	(0.131)
	2.00	0.678	0.673	0.670	0.677	0.678	0.672
		(0.203)	(0.203)	(0.213)	(0.213)	(0.200)	(0.135)

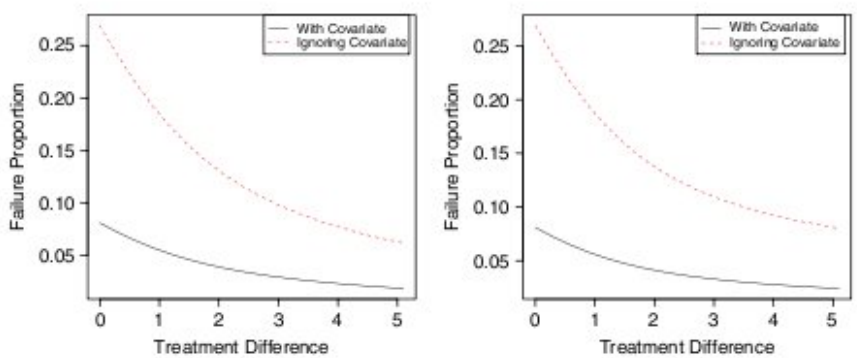

Figure 1. Asymptotic failure proportion for $\theta = 0.1, 0.2$ (left to right).

7. EXAMPLES

Example 1 (*Simulated example*)
We use a simulated data set from Whitehead [14, p. 297] on the response of 48 patients, randomized equally to two treatments A and B, A being the better treatment. The response in this example

Table V. Allocation probabilities to treatment A for a hypothetical second-stage patient.

η	$z=0$	$z=1$	z ignored
0.0000	0.500	0.500	0.500
0.0001	0.513	0.508	0.519
0.0010	0.579	0.557	0.594
0.0100	0.656	0.646	0.655
0.1000	0.673	0.671	0.666
∞	0.675	0.675	0.666

Table VI. Allocation probabilities to fluoxetine for a hypothetical second-stage patient for the fluoxetine trial.

η	$z=0$	$z=1$	z ignored
0.0000	0.500	0.500	0.500
0.0001	0.521	0.525	0.540
0.0010	0.588	0.604	0.623
0.0100	0.641	0.650	0.653
0.1000	0.650	0.657	0.657
∞	0.658	0.658	0.657

is binary and it is assumed that the responses are obtained shortly after the administration of the treatment. The available data are treated as the first-stage data with sex as the only covariate. We have calculated the allocation probability for a hypothetical second-stage patient with covariate z ($0=$male and $1=$female) for different values of η. Using $T_0=2$ and the same prior distribution as in the previous section, we provide Table V.

Table V shows the allocation probabilities corresponding to different values of z. As expected, the allocation probability increases with a small increase in η and females generally have a lower probability of allocation. The allocation probabilities are slightly higher when covariate information is ignored. But this is not surprising as the basic goal of the proposed rule is not to increase the allocation probability, but to set the assignment probability optimally when covariate information is available.

Example 2 (*Fluoxetine trial*)
As an example of the construction and applicability of the proposed design we use part of the data from Tamura *et al.* [15], which is a description of an Eli Lilly sponsored adaptive trial of the anti-depression drug *fluoxetine*. For formalization of our design, the two treatments used are control and fluoxetine. In this trial, the patients were classified according to their shortened rapid eye movement latency (REML), which is presumed to be a marker for endogenous depression. The indicator of shortened REML stratum is considered as covariate z. The final responses are taken as responses, and the misclassified subjects' data are not taken into account. The 83 correctly available data are considered as the first-stage data. The allocation probabilities of the possible second-stage patients for different η and z are given in Table VI.

As expected, the allocation probability increases with a small increase in η, as in Example 1, shortened REML results a higher probability of allocation. The allocation probabilities are slightly higher when covariate information is ignored, as in Example 1.

8. DISCUSSION

The methodologies of the present paper are applicable even if x_i's are stochastic. The theoretical developments can be carried out by assuming some probability distribution of x_i. We skip the details as the subsequent development is routine. Throughout the development of the proposed procedure, we have assumed a single categorical covariate but the proposed procedure is well applicable for more than one covariate variable with some being continuous. The method can also be extended where allocation probability for the second-stage patients can be updated using all available responses and current covariate value. Although the mathematics loses elegance, it is easily applicable in practice. Here, we have used some function of log-odds ratio to represent the ethical allocation, but we can also make the allocation probabilities proportional to odds ratio or variance to maintain the ethical norms. The development is also very similar and we ignore the details.

APPENDIX A

Proof of Theorem 4.1

From the sampling scheme, δ_{Ai}'s are i.i.d. Bernoulli($\frac{1}{2}$) for $i = 1, \ldots, 2m$, and conditionally given the first-stage data, δ_{Ai}'s are independently distributed Bernoulli(p^*_{mi}) random variables for $i \geqslant 2m+1$, where, as $N \to \infty$,

$$p^*_{mi} \to \pi^*_A = J\left(\frac{\alpha_A - \alpha_B}{T_0}\right)$$

almost surely. As n_{AN} is the total number of allocations to treatment A combining the two stages, we have

$$\frac{n_{AN}}{N} = \frac{1}{N}\sum_{i=1}^{2m} \delta_{Ai} + \frac{1}{N}\sum_{i=2m+1}^{N} \delta_{Ai}$$

Now, by SLLN, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{2m} \delta_{Ai} \to \theta$$

almost surely. Also

$$\frac{1}{N}\sum_{i=2m+1}^{N} \delta_{Ai} = \sum_{i=2m+1}^{N} \frac{(\delta_{Ai} - p^*_{mi})}{N} + \sum_{i=2m+1}^{N} p^*_{mi}/N$$

$$= T_N + \sum_{i=2m+1}^{N} p^*_{mi}/N \qquad \text{(A.1)}$$

Now, using Hajek–Renyi inequality (see Gnedenko [16, p. 244]), we have

$$
\begin{aligned}
P\left(\sup_{N\geqslant v}|T_N|\geqslant\varepsilon\right) &\leqslant \frac{1}{\varepsilon^2}E\left\{\sum_{i=2m+1}^{v}\frac{p_{mi}^*(1-p_{mi}^*)}{v^2}+\sum_{i=v+1}^{\infty}\frac{p_{mi}^*(1-p_{mi}^*)}{i^2}\right\}\\
&\leqslant \frac{1}{4\varepsilon^2}\left\{\frac{v-2m}{v^2}+\sum_{i=v+1}^{\infty}\frac{1}{i^2}\right\}\\
&\leqslant \frac{1}{4\varepsilon^2}\left\{\frac{1}{v}+\sum_{i=v+1}^{\infty}\frac{1}{i^2}\right\}\\
&\to 0 \quad \text{as } v\to\infty
\end{aligned}
$$

Hence, the first member of the right-hand side of (A.1) converges to zero almost surely. Also, as $N\to\infty$, second member of the right-hand side of (A.1) converges to $\pi_A^*(1-2\theta)$ almost surely. Hence, combining the result follows. □

ACKNOWLEDGEMENTS

The authors wish to thank two anonymous referees for their valuable suggestions which led to some improvement over an earlier version of the manuscript.

REFERENCES

1. Coad DS. Some results on estimation for two stage clinical trials. *Sequential Analysis* 1992; **11**:299–311.
2. Matthews JNS. *An Introduction to Randomized Controlled Clinical Trials*. Edward Arnold: London, 2000.
3. Rosenberger WF, Lachin JL. *Randomisation in Clinical Trials: Theory and Practice*. Wiley: New York, 2002.
4. Coad DS. Sequential estimation for two-stage and three-stage clinical trials. *Journal of Statistical Planning and Inference* 1994; **43**:343–351.
5. Begg CB, Iglewicz B. A treatment allocation procedure for sequential clinical trials. *Biometrics* 1980; **36**:81–90.
6. Atkinson AC. Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* 1984; **69**:61–67.
7. Bandyopadhyay U, Biswas A. Adaptive designs for normal responses with prognostic factors. *Biometrika* 2001; **88**:409–419.
8. Atkinson AC, Biswas A. Bayesian adaptive biased coin design for sequential clinical trials. *Biometrics* 2005; **61**:118–125.
9. Atkinson AC, Biswas A. Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal responses. *Statistics in Medicine* 2005; **24**:2477–2492.
10. Atkinson AC, Donev AN. *Optimum Experimental Designs*. Oxford University Press: Oxford, 1996.
11. Newman SC. *Biostatistical Methods in Epidemiology*. Wiley: New York, 2001.
12. Cheung YK, Thall PF. Monitoring the rates of composite events with censored data in phase I clinical trials. *Biometrics* 2002; **58**:89–97.
13. Biswas A, Angers J-F. A Bayesian adaptive design in clinical trials for continuous responses. *Statistica Neerlandica* 2002; **56**:400–414.
14. Whitehead J. *The Design and Analysis of Sequential Clinical Trials*. Wiley: New York, 1997.
15. Tamura RN, Faries DE, Andersen JS. Heiligenstein JH. A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association* 1994; **89**:768–776.
16. Gnedenko BV. *The Theory of Probability*. Chelsea: New York, 1962.